OPTICAL NEURAL COMPUTING FOR ASSOCIATIVE MEMORIES

Thesis by
Ken Yuh Hsu

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

California Institute of Technology
Pasadena, California
1990
(Submitted July 5, 1989)

I dedicate this thesis to my parents, Jin-Shui and Yue-Lee, whose unconditioned love has made my dream come true. Without them there would be nothing for which to thank anyone else. I would also like to dedicate this thesis to my wife, Jen, and our children, George and Anne. Their unwavering devotion and understanding have made the most challenging period of my life also the most rewarding.

In a forest a fox bumps into a little rabbit, and says, "Hi, junior, what are you up to?"

"I'm writing a dissertation on how rabbits eat foxes," said the rabbit.

"Come now, friend rabbit, you know that's **impossible***!"*

"Well, follow me and I'll show you." They both go into the rabbit's dwelling and after a while the rabbit emerges with a satisfied expression on his face.

Comes along a wolf. "Hello, what are we doing these days?"

"I'm writing the second chapter of my thesis, on how rabbits devour wolves."

"Are you crazy? Where is your academic honesty?"

"Come with me and I'll show you." As before, the rabbit comes out with a satisfied look on his face and a diploma in his paw. Finally, the camera pans into the rabbit's cave and, as everybody should have guessed by now, we see a grand, huge lion sitting next to some bloody and furry remnants of the wolf and the fox. — Anonymous

The moral: "As a Caltech Ph.D candidate, you should always be curious." — Demetri Psaltis

ACKNOWLEDGEMENTS

Talking with my advisor, Professor Demetri Psaltis, is always inspiring, challenging, encouraging, and interesting. It has been a great privilege and a pleasure to be a member of his group. I am grateful to him for investing so much of his time and effort in me during my studies at Caltech. I appreciate especially the encouragement and friendship that he has extended to me, making my stay at Caltech an enjoyable experience.

I am deeply indebted to the past and present members of the Optical Information Group. I have benefited from them not only academically but also culturally. Special thanks go to Dr. Eung Gi Paek. The initial years of my research when I worked with Paek are memorable. His contribution to my research is immeasurable. I would like to thank David Brady whose friendship was well beyond my most optimistic hopes. The fifth chapter of my thesis is the result of a collaborative effort between Professor Psaltis, Dave, and myself. The collaboration was fruitful and the experience, enjoyable. Hsin Yu Li has spent his invaluable time in technical discussions and proofreading this thesis. The geometric method for analyzing the neural loops presented in this thesis was originally suggested by him. Mark Neifeld made helpful suggestions and proofread Chapter 1. I would also like to acknowledge the enormous help I received from Dr. Robert Snapp, Nabeel Riza, Cheol Hoon Park, Xiang Guang Gu, Alan Yamamura, Scott Hudson, Steve Lin, Chuanyi Ji, Charlie Stirk, Yong Qiao, Seiji Kobayashi, Dr. Jeff Yu, Dr. John Hong, Dr. Mok Fai, Dr. Kelvin Wagner, and Dr. Michael Haney. I am also grateful for the administrative help that I received from Mrs. Helen Carrier, Mrs. Su McKinley, and Mrs. Linda Dozsa.

For the privilege of having had the chance to join this group, I owe to the kindness of Dr. Hua Kuang Liu, who introduced me to Professor Psaltis five years ago. A Chinese proverb says, "When you drink water, think of its source ." I would like to thank Dr. Liu for his kindness and constant encouragement throughout my studies.

I gratefully acknowledge the financial support during the first two years at Caltech from a fellowship from the National Science Council of the Republic of

China. I am also very grateful to National Chiao Tung University in Taiwan for the leave for my studies. Professor Nan Hung Kuo, then President of the University, and Professor Yi Shung Guo, the Dean of the Science School, who committed themselves to encouraging young teachers to continue advanced studies. I am also grateful to Professor Ying Liang Chen at Chiao Tung, who offered me friendship and constant encouragement throughout my studies. There are also many friends and colleagues (too many to thank individually) who encouraged and cared in different ways. I thank you all.

ABSTRACT

Optical techniques for implementing neural computers are presented. In particular, holographic associative memories with feedback are investigated. Characteristics of optical neurons and optical interconnections are discussed. An LCLV is used for simulating a 2-D array of approximately 160,000 optical neurons. Thermoplastic plates are used for providing holographic interconnections among these neurons. The problem of degenerate readout in holographic interconnections and the method of sampling grids to solve this problem are presented.

Two optical neural networks for associative memories are implemented and demonstrated. The first one is an optical implementation of the Hopfield network. It performs the function of auto-association that recognizes 2-D images from a distorted or partially blocked input. The trade-off between distortion tolerance and discrimination capability against new images is discussed. The second optical loop is a 2-layer network with feedback. It performs the function of hetero-association, which locks the recognized input and its associated image as a stable state in the loop. In both optical loops, it is shown that the neural gain and the similarity between the input and the stored images are the main factors that determine the dynamics of the network.

Neural network models for the optical loops are presented. Equations of motion for describing the dynamical behavior of the systems are derived. The reciprocal vector basis corresponding to stored images is derived. A geometrical method is then introduced which allows us to inspect the convergence property of the system. It is also shown that the main factors that determine the system dynamics are the neural gain and the initial conditions.

Photorefractive holography for optical interconnections and sampling grids for volume holographic interconnections are presented. A periodic copying method for refreshening multiply exposured photorefractive holograms is presented, which allows the hologram to maintain the same diffraction efficiency as that when a single exposure scheme is used. This scheme provides us with the possibility of achieving maximum storage and maximum diffraction efficiency in holographic associative memories.

Contents

Chapter 1

Introduction

Optical neural computers are computing architectures based on the concept of neural networks and implemented with optical hardware. Interest in this type of computer has been motivated largely by the hope that by building a computer that shares some of the characteristics of biological systems, we will be able to address problems such as image recognition which animals do exceedingly well but current digital computers do not [1, 2, 3, 4]. Unfortunately, because of the lack of a detailed knowledge of the operation of biological neural networks or an adequate theoretical understanding of how to use an adaptive, massively parallel, densely connected computer architecture, the practical use of neural networks for the solution of difficult computational problems remains mostly a goal. Nevertheless, there has been a great deal of progress on the theoretical side to justify optimism about future applications, and this has focused attention on the hardware realization of neural architectures. In this respect it turns out that optical technology is particularly promising for constructing neural computers. Two questions must be answered before we start to build an optical neural computer: What are the characteristics of neural networks and why use optics?

Intuition tells us that the brain is a very complicated organ. Indeed, research in anatomy shows that the brain contains about 10^{12} neurons [5]. Each neuron receives information from hundreds or thousands of other neurons and in turn transmits information to hundreds or thousands of other neurons. Furthermore, the brain is organized in such a complicated way that the operation of large parts of it are still in question, not only in terms of how they work but also in terms of their biological purpose. Therefore, at the present time it is impossible to describe precisely how the

brain works. However, several models have being proposed [6, 7, 8, 9]. Though very much simplified from reality, each model successfully demonstrates certain aspects of brain function. Given this knowledge, we now define a neural computer to be characteristized by three properties: a) It consists of a large number of simple processing units, which we call "neurons." Each neuron performs a simple thresholding operation to its total input signal. b) Each neuron is connected to many others through weighted interconnections. Typically, the number of the interconnections range from several hundreds to several thousands. c) The network is programmed to respond appropriately to inputs by adjusting the weights between neurons during the learning phase.

Two features can be derived from the above definition, which distinguish neural computers from conventional computers. First, information is stored and distributed in the extensive interconnection links among the neurons. Each neuron does only the simple function of thresholding but influences thousands of others simultaneously through the massive communication links. It is this parallel and collective behavior that determines the dynamics of the network, which is also where the computational power of neural computers emerges from. This mode of computation is fundamentally different from that of conventional computers, which function sequentially, executing a step-by-step algorithm. Secondly, the interconnection pattern, or the information storage, is modifiable. It is formed by providing the network with enough "training" samples. With these two features in mind, we now proceed to find technologies that are appropriate for implementing a neural computer.

Electronics (analog, digital or hybrid) and optics are the two approaches under consideration for the hardware realization of neural computers. There are two basic components that need to be implemented: neurons and connections. The neurons are typically simple thresholding elements that can be implemented by a single switching device (e.g., transistors). The switching speed or the accuracy required

for the neurons is not beyond the capabilities of current electronic technology. A practical neural computer may require millions of neurons operating in parallel. This requirement by itself is also achievable in electronics [10, 11]. However, each of the neurons must be connected to several thousand other neurons and these connections must be modifiable so that learning can take place. This is relatively difficult to achieve electronically. The reason is that on a silicon chip, communication links are provided by electrical wires that must be kept separated by some minimum distance. Otherwise, the electrical signals carried by different channels interfere with one another. Therefore, the number of interconnections per electronic neuron is rather restricted and this limits the amount of communication.

On the other hand, optics dovetails nicely with the characteristics of neural computers. The first reason is that light beams can propagate in free space and remain unaltered even as they pass through each other. The second reason is that an optical system has its three-dimensional nature. Hence, if optical neurons are arranged in a planar architecture, then the neurons on different planes can communicate by emitting and receiving light beams through the free space between them. This provides optical systems a dense and parallel communication capability. There are several optical components available for the implementation of neural networks. The first optical component available is a two-dimensional array of optical switching elements. Each element of the array switches state depending on the total light input to this element; thus it simulates the function of a neuron. The array of optical neurons can be obtained by several approaches. Among them, two-dimensional spatial light modulators (2-D SLM), opto-electronic chips, and photorefractive crystals are most promising. One example of 2-D SLM is the liquid light crystal valve that we use in this thesis, which simulates 160,000 optical neurons in a one-inch aperture [12]. As we have already explained, the computational power of a neural computer comes from extensive connectivity among the neurons. For this purpose, a hologram is an

excellent component. A holographic grating provides a communication link between two optical neurons. A planar hologram of one-inch-square size can record as many as 10^8 independent gratings. This would allow $10,000$ neurons to talk to $10,000$ neurons simultaneously. If a volume hologram is used, then we can even obtain 10^{12} gratings in a cubic centimeter of photorefractive crystal [13]. This is extremely difficult to achieve with planar electronic technology.

An additional advantage of using holographic interconnections in a neural computer is that learning mechanisms can be realized. The reason for this is that the formation of holographic gratings simulates the Hebbian rule: The diffraction efficiency of the holographic grating depends on the modulation depth of the recording beams, which in turn depends on the light intensities from the two interconnecting neurons. By using adaptive recording materials such as photorefractive crystals to write dynamic holograms, learning can take place [14]. This feature again is unique in optical neural computers. We therefore are particularly interested in applying optics for the implementation of neural computers.

This thesis is an investigation into the holographic implementations of neural computers. Specifically, we consider the implementation of associative memories. The reason for choosing the associative memory is because it is the most interesting property emerging from the collective computational behavior of neural networks [15,16]. The features of fault tolerance, network dynamics, and learning in associative memories are quite general and can be extended to other types of neural computers. Two problems are addressed in this thesis. The first is the problem of how to design and implement a holographic neural computer for associative memories. The second is the dynamics of the associative loop with feedback. The remainder of this thesis is organized as follows: In Chapter 2 we discuss optical devices for the implementation of neural computers. Basic operational principles and the main characteristics of optical neurons and optical interconnections will be presented. In

Chapter 3 we present an associative feedback loop based on the Hopfield model [13]. The optical loop performs the function of image auto-association. The invariances and discrimination properties of the loop are demonstrated [17]. The neural net model of the loop and the dynamic equations are presented. A geometrical method for studying the convergence properties is presented and the parameters that are pertinent to the loop dynamics are discussed [18]. In Chapter 4 an optical loop that performs hetero-associations is described [19]. The method of using fractal sampling grids to avoid degeneracies in holographic reconstructions is presented, and algebraic formulas for designing the sampling grids are derived. The geometrical method is again used here to discuss the dynamics of the loop. In Chapter 5 we present photorefractive holograms for optical neural computers. The method of multiply exposed photorefractive holograms with maximum storage capacity and maximum diffraction efficiency is demonstrated [20]. The use of photorefractive crystals as dynamic holograms for increasing the storage capacity of optical associative memories is presented at the end of Chapter 5 [21].

1.1 References for Chapter 1

1. Y. Abu-Mostafa and D. Psaltis, "Optical Neural Computers,," *Scientific American*, Vol. **256**, No. **3**, 88-95, March (1987).

2. D. Psaltis, D. Brady, X. Gu, and K. Hsu, "Optical Implementation of Neural Computers," A chapter in *Optical Processing and Computing*, H. Arsenault ed., Academic Press, New York, 1989.

3. N. Farhat and D. Psaltis, "Optical Implementation of Associative Memory Based on Models of Neural Networks," A chapter in *Optical Signal Processing*, J. L. Horner ed., Academic Press, New York, 1987.

4. K. Hsu and D. Psaltis, "Optical Associative Memory," A chapter to appear in *Nonlinear Optics and Optical Computing*, S. Martellucci ed., Plenum Publishing, New York, 1988.

5. D. H. Hubel, *Eye, Brain, and Vision*, Scientific American Library series, # 22, W. H. Freeman and Company, New York, 1988.

6. W. S. McCulloch and W. H. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Math. Biophysics*, **5**, 115(1943).

7. D. Hebb, *Organization of Behavior*, Wiley, New York, 1949.

8. F. Rosenblatt, *Principles of Neurodynamics : Perceptrons and the theory of Brain Mechanisms*, Spartan Books, Washington, D. C., 1962.

9. K. Fukushima, S. Miyake, and Takayukiito, "Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol., SMC-13, No. **5**, 826-834, September (1983).

10. C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison- Wesley Publishing, Reading, Massachusetts, 1980.

11. C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley Publishing, Reading, Massachusetts, 1989.

12. K. Hsu, D. Brady, and D. Psaltis, "Experimental Demonstrations of Optical Neural Computers," *Neural Information Processing Systems*, D. Z. Anderson ed., American Institute of Physics, New York, 1987.

13. D. Psaltis, J. Yu, X. G. Gu, and H. Lee, "Optical Neural Nets Implemented with Volume Holograms," Paper **TuA3-1**, *Topical Meeting on Optical Computing*, Optical Society of America, March 16-18, 1987, Incline Village, Nevada.

14. D. Psaltis, D. Brady, and K. Wagner, "Adaptive Optical Networks using Photorefractive Crystals," *Appl. Optics*, Vol. **27**, 1752, May 1 (1988).

15. J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Natl. Acad. Sci. U. S. A.*, **79**, 2554 (1982).

16. J. J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons," *Proc. Natl. Acad. Sci. U. S. A.*, **81**, 3088 (1984).

17. K. Hsu and D. Psaltis, " Invariances and Discrimination Properties of the Optical Associative Loop," *IEEE Annual International Conference on Neural Networks*, San Diego, 1988.

18. S. Y. Li, "Analysis of Pinhole Array Associative Memory System Using the Method of Phase Planes," private communication.

19. D. Psaltis and K. Hsu, "The Hetero-Associative Loop using Planar holograms," to be submitted.

20. D. Brady, K. Hsu, and D. Psaltis, "Multiply Exposed Photorefractive Holograms with Maximal Diffraction Efficiency," Paper PDP2, Topical Meeting on Optical Computing, Optical Society of America, Salt Lake City, Feb. 27-March 1, 1989.

21. D. Brady, K. Hsu, and D. Psaltis, "Periodically Refreshed Multiply Exposed Photorefractive Holograms," submitted to *Opt. Lett.*

Chapter 2

Devices for Optical Neural Computers

2.1 Introduction

The generic structure of an optical neural computer is shown in Fig. 2.1. In the figure, the neurons are arranged in two-dimensional planes and they are interconnected through free-space communication links. There may be several layers in the network, with feedback paths in each layer. As was described in Chapter 1, the greatest advantage of this three-dimensional structure is to provide the neural computer with dense communication links among the neurons so that parallel and collective computations can emerge.

As shown in the figure, there are two basic components for optical neural computers: optical neurons and interconnections. This chapter presents the technologies for implementing these two devices and their characteristics. Section 2.2 discusses the devices for implementing optical neurons. The characteristics of the neurons such as the thresholding function, light-detection sensitivity, and temporal response are presented. In Section 2.3 the holographic method for interconnections in neural computers is presented. The operational principles of thermoplastic plates for planar holograms and the properties of planar holographic interconnections are presented.

2.2 Optical Neurons

The function of an optical neuron is two-fold. First, it integrates light beams emitted from all the neurons that its input is connected to. Second, the light-emitting or light-transmitting property of this neuron is modulated by the light level of the total input in a nonlinear manner; i.e., it performs a thresholding operation to the total

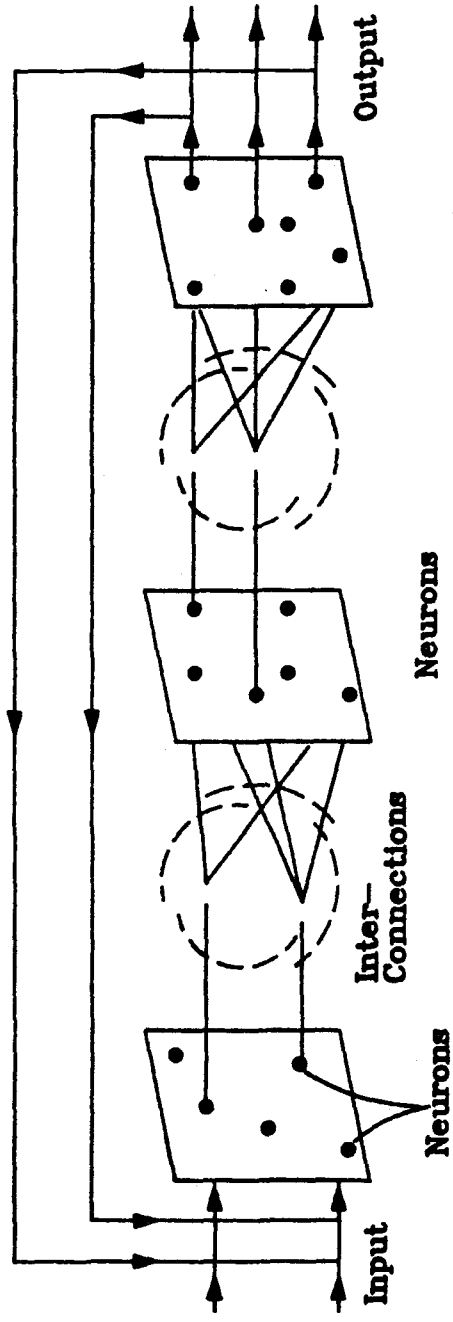


Figure 2.1: Optical Neural Computer

input. Several technologies are available for implementing optical neurons: two-dimensional spatial light modulators (2-D SLM) [1], integrated opto-electronic chips [2], arrays of nonlinear optical switches [3], and nonlinear photorefractive crystals [4].

Development of nonlinear optical devices with fast switching speed have been pursued for many years. In recent years considerable progress made in this area has led to numerous demonstrations of ultra-high speed (pico-second) optical switches. The major problem with these switches at present is the very high power required to switch each element. This makes large arrays impractical due to power dissipation limitation. For neural network simulation, speed is not critical. The response time of neurons in the brain is in the millisecond range. Instead, we need large arrays of neurons to perform collective computation. Thus, nonlinear optical switches are not considered here.

The opto-electronic approach to simulating optical neurons involves the integration of a two-dimensional array of light detectors, saturation amplifiers and light-emitters on a single chip. The amplifier controls the light-emitting level as a function of input light intensity. The combination of light detector, emitter, and the amplifier simulates an optical neuron. An opto-electronic chip with an array of 10×10 optical neurons has been demonstrated recently [2]. The arrays required for optical neurons are more regular and simpler than general circuits; hence larger arrays are possible. Furthermore, the technology required for making opto-electronic neurons is compatible with that currently being pursued for providing optical interconnections in VLSI chips. Therefore, the opto-electronic approach is most likely to provide practical and useful neurons in the future.

The photorefractive effect is a nonlinear phenomenon that has attracted much research effort in recent years [5]. The effect occurs in crystals that change their refractive index under light illumination. When a photorefractive crystal (PR) is

exposed to a light interference pattern, free charges are excited by the incoming photons. These photo-induced charges move inside the crystal and eventually are re-trapped. The pattern of re-distribution of the charges corresponds to that of the input light pattern. The presence of the spatially varying charge pattern then induces an internal electrical field, which in turn modulates the index of refraction through the linear electro-optic effect. Thus a phase hologram is recorded and can be reconstructed by illuminating the PR crystal with a light beam. The phase gratings can be used to provide interconnections between optical neurons. On the other hand, the simultaneous presence of two beams in a PR crystal provides a nonlinear coupling between the beams through the PR gratings. The amount of beam amplification and thresholding are functions of the two-beam modulation ratio and crystal parameters such as doping and orientation. The physical mechanisms and the use of the PR effect are still under current research.

The fourth technology for implementing optical neurons is two-dimensional spatial light modulators (2-D SLM). These devices have being investigated for many years, primarily for the purpose of optical image processing. Although thoroughly developed 2-D SLM's are still not available [6], there are two commercially available devices that can provide reasonably large arrays of optical neurons. This allows us to perform meaningful network experiments at this stage and in the foreseeable future. The two devices are the microchannel spatial light modulator (MSLM) and liquid crystal light valve (LCLV). Both these devices can operate in a high contrast mode such that the output light from each resolution pixel is a nonlinear function of the input light to that pixel. Thus, each pixel of the SLM simulates one optical neuron. For the MSLM that we have, the resolution is about 3 lines per millimeter in a 16×16 mm aperture. Thus, this device simulates 1,800 neurons. On the other hand, the resolution capability of the LCLV is 20 lines per millimeter in a 20×20 mm aperture. Consequently, this simulates 160,000 neurons. We use the LCLV in

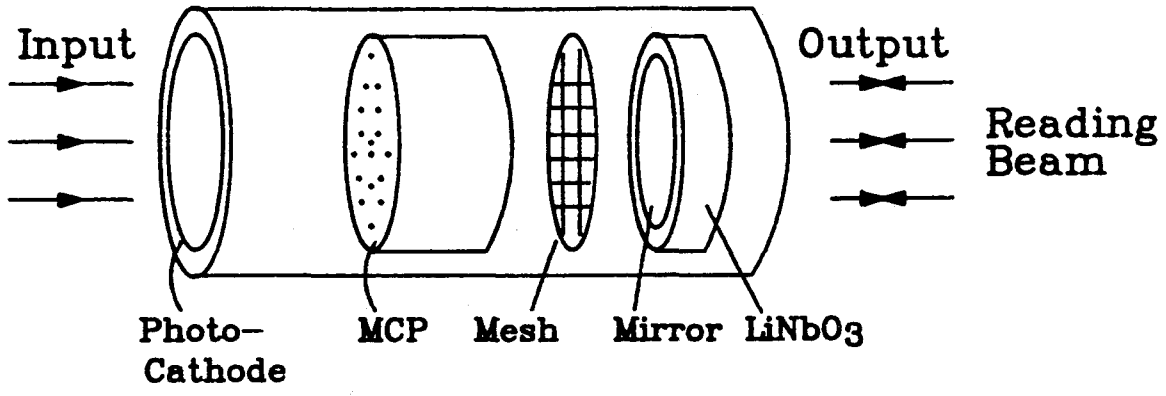


Figure 2.2: Microchannel Spatial Light Modulator

most of our experiments.

2.2.1 The MSLM as Optical Neurons

The MSLM is an optically addressed, electrically controlled 2-D SLM [7]. The basic structure is shown in Fig. 2.2. As shown in the figure, the device is a vacuum-sealed 2-D photo-transducer; its main components include the photocathode, microchannel plate (MCP), mesh electrode, and a $LiNbO_3$ single crystal. The photocathode is a multialkali type (S-20), which converts the optical input image into an electron image. The MCP is a thin $(0.5 - 1.0\ mm)$ semiconductor glass plate which consists of a 2-D array of 10 μm glass pores spaced 15 μm apart in a hexagonal pattern. Each element of the MCP functions as a continuous dynode electron amplifier. Under the application of 1 KV, the MCP can provide an electron gain of 10^4. The $LiNbO_3$ is a 300 μm thick single crystal with 55^0 cut. One side of the crystal which faces the MCP is a multilayer-coated dielectric mirror and the other side is coated with a transparent conducting electrode. Thus, a bias voltage can be applied on the

crystal to control the surface charge condition of the crystal. The reading light comes into the crystal from the rear side and is reflected by the mirror. There is a mesh electrode placed between the MCP and the crystal. The relative potential between the mesh electrode and the crystal surface determines whether the electrons emitting from the MCP are attracted to or repelled from the crystal.

The basic operational principle of the MSLM is described in the following. The 2-D optical pattern is first imaged on the photocathode of the MSLM, where it is converted to an electron image. The electron image is then amplified and re-emitted by the MCP to the mirror side of the $LiNbO_3$. This spatially varying charge pattern induces an internal electrical field inside the $LiNbO_3$, which in turn modulates the index of refraction through the linear electro-optic effect. Thus, the original optical pattern is recorded as a refractive index pattern in the $LiNbO_3$. A polarized reading beam, coming into the MSLM from the crystal side, is spatially phase-modulated by the refractive index pattern of the $LiNbO_3$. The phase modulation can be read out using a crossed polarizer. The light intensity of the read-out beam can be expressed as

$$I = A\Big[\sin^2(\Gamma/2)\Big] \tag{2.1}$$

where A is a constant, and Γ is the phase retardation induced by the $LiNbO_3$. The phase retardation is proportional to the voltage across the crystal, which is proportional to the charge density on the crystal surface. In the linear operation region, the charge density is proportional to the writing light intensity. Thus, the phase retardation can be written as

$$\begin{aligned} \Gamma &= \pi\left(\frac{V_x}{V_\pi}\right) \\ &= \pi\left(\frac{\sigma_x}{\sigma_\pi}\right) \end{aligned} \tag{2.2}$$

where V_x is the voltage across the crystal, V_π is the half-wave voltage of the crystal, σ_π is the charge density required to generate π phase retardation, and σ_x is the charge

density on the crystal surface, which is proportional to the total exposure. It is seen from the above equations that the intensity of the output image is proportional to the charge density accumulated on the crystal surface. The charge density is in turn proportional to the writing intensity, and is also controlled by the MCP gain. Fig. 2.3 shows the output intensity as functions of writing intensity and MCP gain. It is seen that the MSLM provides a nonlinear function of soft thresholding as well as optical gain. Suppose we deposit a uniform electron charge layer with density σ_π on the crystal surface before we write the input on the MSLM. Then the whole surface is biased by a half-wave voltage and the output is a bright background. Now we apply electrical voltages on the mesh and the crystal such that the mesh electrode has higher potential than the crystal surface. When an optical image is written on the MSLM, the corresponding electron image is generated by the photocathode and amplified by the MCP. The electrons re-emitted from the MCP are accelerated toward the crystal and bombard the deposited electrons out of the crystal surface. These secondary electrons are collected by the mesh electrode. Under appropriate voltage conditions the coefficient of the secondary emission is greater than one. Thus, the crystal surface loses electrons. The brighter the input, the more the loss. As a result, the brighter part produces less phase retardation. This reverses the contrast of the read-out image. Fig. 2.4(a) and (b) show the output images under normal operating conditions, and Fig. 2.4(c) and (d) show the contrast reversed version.

Although the MSLM can simulate neuron functions, its resolution is limited to 3 lines per millimeter with an aperture of 16×16 mm. The number of neurons that we can get from this device is approximately 1,800. This is too small for our image association purpose. Furthermore, the input image is written on the crystal surface as an electron charge image. Since the crystal is an insulator, unless an erase mechanism is applied to the MSLM, the electron image would stay on the crystal

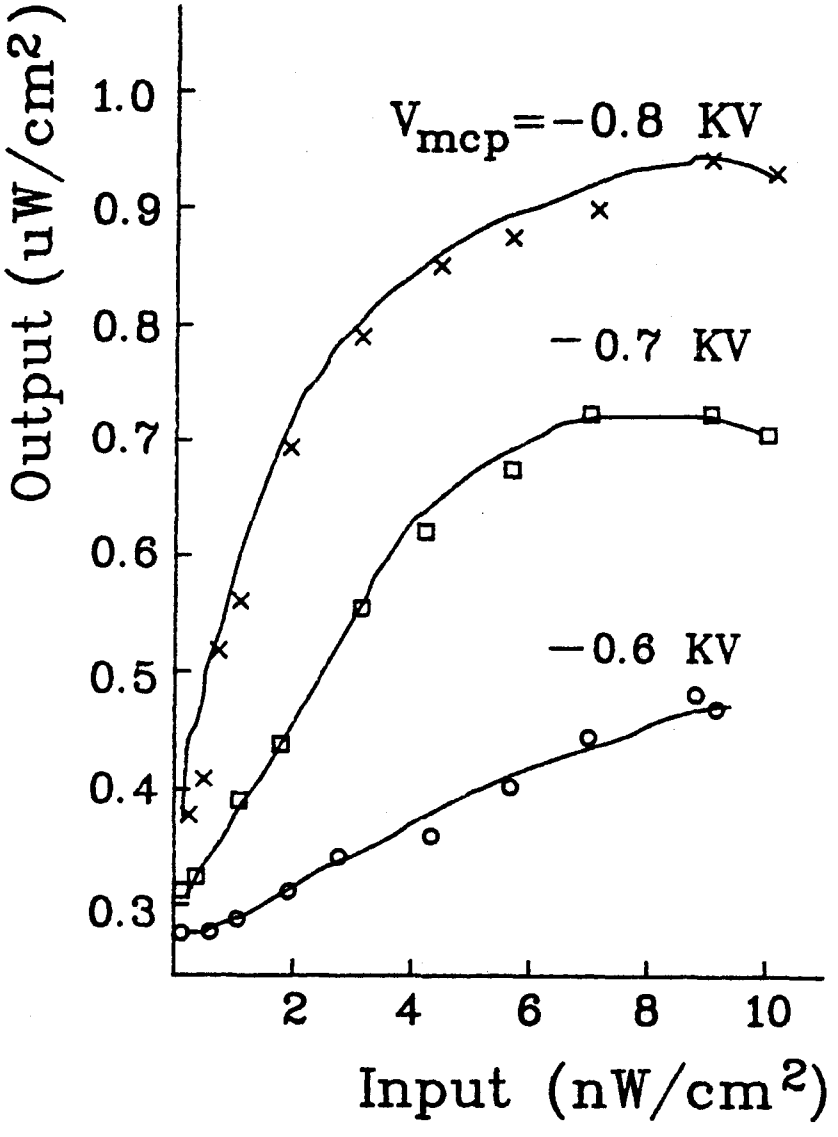
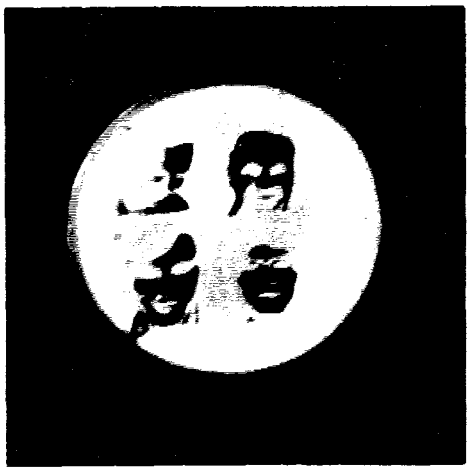


Figure 2.3: Characteristics of the MSLM

(a)

(b)



(c)

(d)

Figure 2.4: Readout Images of the MSLM: (a). Normal operation. (b). Contrast reversal.

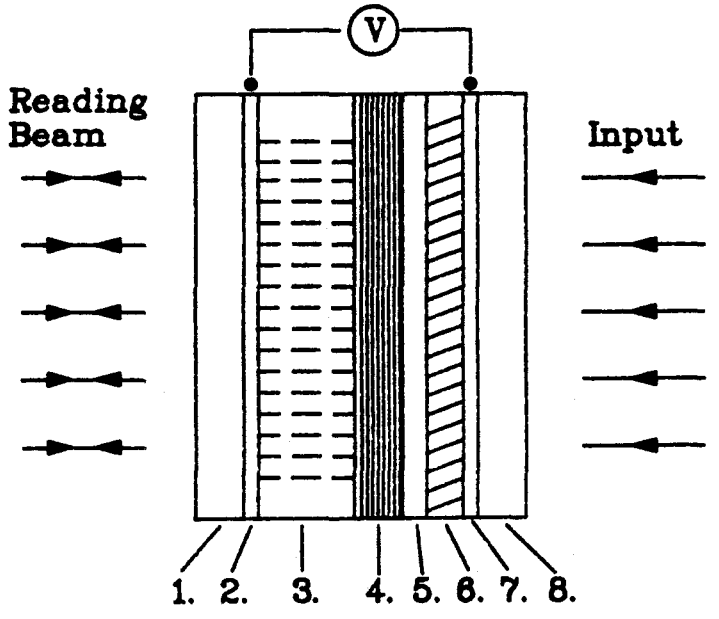


Figure 2.5: Liquid Crystal Light Valve

for a long time (longer than a week). This is a nice characteristic when the device is used as a buffer. If we want to use the MSLM for simulating the dynamics of neural networks, we need to apply a uniform erasing light on the MSLM periodically so that dynamics can occur.

2.2.2 The LCLV as Optical Neurons

The general structure of the liquid crystal light valve (LCLV) is shown in Fig. 2.5 [8]. As shown in the figure, ths basic structure of the LCLV is a multilayered device. On

one side it has the photosensor layer for writing the input image; on the other side it has the electro-optic layer for transferring the input information to the phase retardation of the reading beam. Similar to the MSLM, in the LCLV there is a dielectric mirror sandwiched between the light-sensing layer and the light-modulation layer for separating the reading beam from the writing beam. The differences between the two devices are the materials that are used. The photosensor of the LCLV is a CdS semiconductor layer, which has a peak photo-response at 515 nm in the green. We use an argon laser with wavelength 514 nm as the writing beam. The electro-optic material for the LCLV is a thin layer of nematic liquid crystal. The layer thickness is about a few microns. The liquid crystal (LC) molecules are aligned in a layered structure with their long axes parallel to each other. The axis of the aligned direction is equivalent to the optical axis of the LC layer. In the LCLV the LC molecules on the front surface of the reading side are aligned along some preferred direction. The molecules on the back surface of the LC layer are aligned along a direction that is 45^0 with respect to that of the front surface. As a result, the axes of the molecules in the bulk of the LC layer are rotated adiabatically from one surface to the other by 45^0. Thus, the optical axis is also rotated adiabatically by 45^0. This is called the twisted effect. When a reading beam comes into the LCLV with the polarization vector along the optical axis of the front surface, the polarization state will be rotated by 45^0 following the twisted angle of the liquid crystal molecules [9].

We now explain the operational principles of the LCLV. In operation, the optical image is input into the photosensor side and a uniform reading beam comes into the liquid crystal side of the LCLV. We apply an electrical bias voltage across the LC layer, dielectric mirror, and the photosensor layer. When there is no writing light, most of the voltages is across the high impedance photosensor layer; only a small portion of the voltage is across the LC layer. Thus, the orientation of the LC molecules remains in a twisted structure. As the reading beam passes through the

LC layer, its polarization state is rotated 45^0. However, the light beam is reflected by the mirror to make a second pass through the LC layer. Thus, the output polarization is rotated back to the incident state through the adiabatic following on the backward pass. Therefore, the output is zero after a crossed analyzer. On the other hand, when there is an input image writing on the LCLV, the impedance of the photosensor layer is reduced. The bias voltage is switched to the LC layer, in proportion to the input light intensity. This induced voltage produces an electric field in a direction perpendicular to the LC layer. If the voltage is above some threshold, the LC molecules begin to tilt toward the electric field. In this orientation of the molecules, between the parallel and perpendicular directions of the electrical field, the optical birefringence of the LC molecules as well as the twisted orientation both affect the polarization of the light. As a result, the output light from the LCLV after double pass through the LC layer is no longer linearly-polarized. The transmission through the crossed polarizer is proportional to the degree of tilt of the molecules, which is in turn proportional to the writing light intensity. Therefore, the intensity pattern of the input image is transferred to the output light amplitude. In the limit of applying a very high electric field across the LC layer, all the LC molecules are rotated 90^0 to the perpendicular direction. Then the optical axis of the LC layer is along the propagation of the reading beam. The LC layer is equivalent to an isotropic device. Thus, the polarization of the reading beam is unaffected, and the output is zero. Usually we keep the bias voltage under a certain level so that this does not happen.

Fig. 2.6 shows the measured results of the output vs. input intensity of the LCLV. Two observations can be seen from this figure. First, the device has the characteristics of soft thresholding; this simulates the nonlinear function of a neuron. Second, the device is sensitive to light intensities of 1 $\mu W/cm^2$ and produces output intensity of 100 $\mu W/cm^2$. This means that the LCLV provides an optical gain of

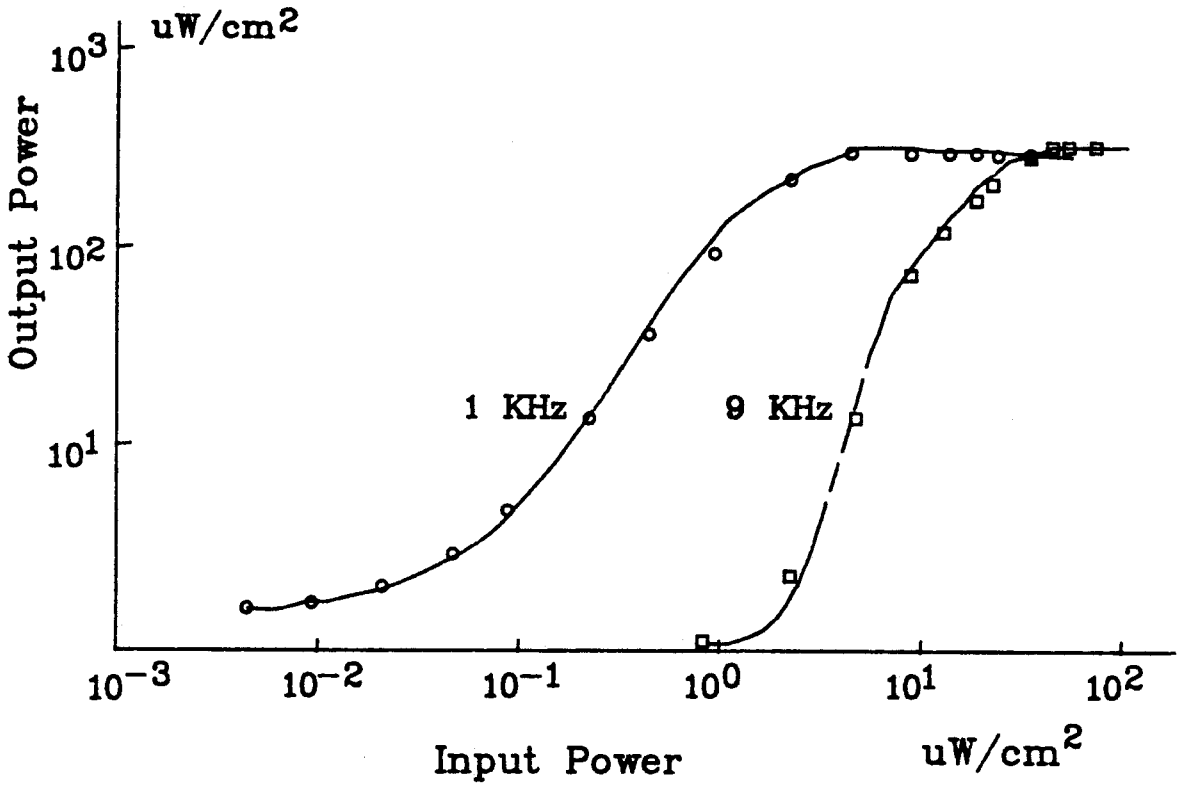


Figure 2.6: Output vs. Input Characteristics of the LCLV. Reading beam intensity @ 1 mW/cm^2.

100. The optical gain can be higher with higher intensity of the reading beam.

Besides the above characteristics, there is another important parameter for the implementation of neural computers: the number of neurons that the LCLV can simulate. This number is proportional to the number of the resolvable spots of the device. We build a Michelson interferometer to measure the resolution capability of the LCLV. The light beams of the two arms of the interferometer interfere to form a fringe pattern. The spatial frequency of the fringes is adjusted by rotating the mirror of one arm of the interferometer. The interference fringes are imaged on the photosensor of the LCLV. The read-out pattern of the LCLV is then Fourier-transformed onto a CCD camera and the first-order diffraction intensity is measured. The experimental result is shown in Fig. 2.7. It is seen that the LCLV has good resolution capability more than 20 lines per millimeter. The device has uniform response over a 20 mm aperture. Therefore, the device can simulate approximately 160,000 optical neurons.

In optical neural networks, the signal decays as it propagates through the networks. This is due to the losses from reflections and absorptions of the optical components as well as the from the finite diffraction efficiency of holograms. In some cases we need a high gain to compensate the losses. One way is to cascade a pre-amplifier to the LCLV. In our experiments we use an image intensifier to achieve this purpose. The image intensifier is a very sensitive 2-D device. The input side is a photocathode for converting an optical image into an electron image. The output side is a phosphorus screen for converting the electron image back into a visible image. Between them is a microchannel-plate (MCP) which can amplify the intensity of the electron image by up to 10^4. The output side of the image intensifier is a fiber optical plate. The writing side of the LCLV is also a fiber optical plate. Thus, the two devices can be attached directly to form a combination. The photosensor plate of the LCLV is driven by the output image displayed on the phosphorus screen of

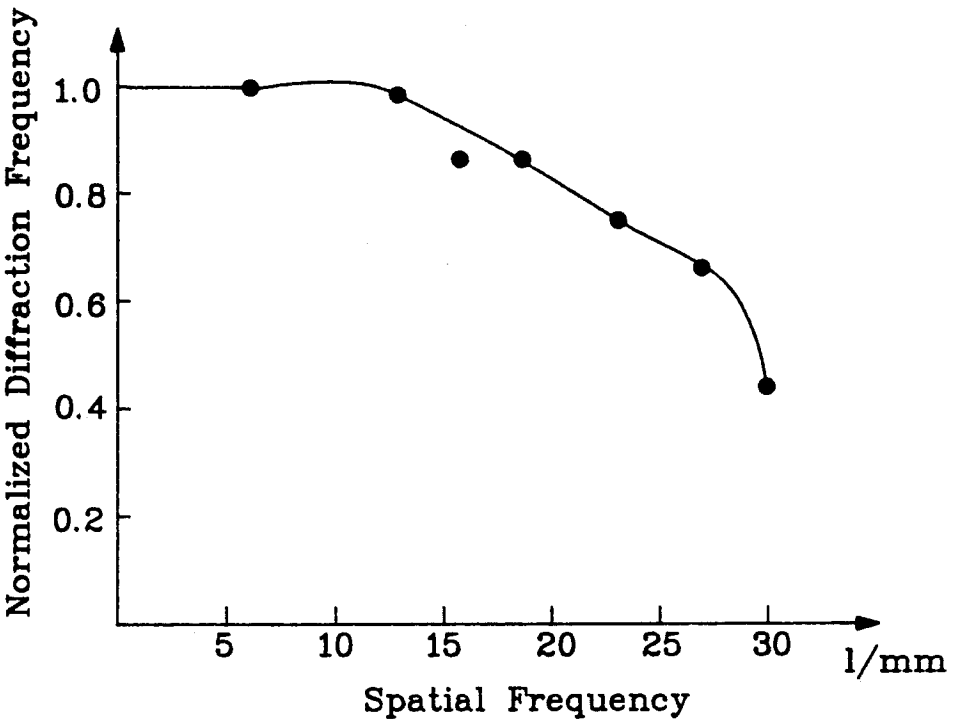


Figure 2.7: The Resolution Capability of the LCLV.

the image intensifier. This combination simulates a two-dimensional array of very sensitive optical neurons. Fig. 2.8 shows the output vs. input characteristics of this combination. It is seen that the combination has a soft thresholding characteristic. It is sensitive to the light intensity as low as nW/cm^2, and the output intensity is higher than 10^2 $\mu W/cm^2$. This corresponds to an optical gain of more than 10^5. We will see in Chapters 3 and 4 that the neuron gain is the key factor that determines the dynamic behavior of neural networks. The combination of the image intensifier and the LCLV provides a very versatile control on selecting proper neuron gains, because the the optical gain can be adjusted simply by changing the bias voltage of the microchannel-plate. Fig. 2.9 shows the optical gain of the neuron as a function of bias voltage for a specific reading beam intensity of 1 mW/cm^2.

In order to study the dynamic behavior of neural networks, we need to know the temporal response of the neurons. Fig. 2.10 shows the temporal response of the optical neuron under the writing intensity of 20 nW/cm^2 and the driving power supply at 1 KHz. In the figure, the upper curve represents the **ON** and **OFF** of the writing light. The lower trace is the response of the optical neuron. It is seen that both the rise and fall times of the device are in the range of 400 ms. The response time of the device is mainly determined by the RC time constant of the device. Since the impedance of the light valve is a function of the writing intensity as well as the driving frequency of the power supply, the response time will be different under different operating conditions. In our experiments the response times are in the range of a few hundred ms. We therefore take the response of an individual pixel x_i to its input y_i to mimic a nonlinear amplifier followed by an RC circuit as shown in Fig. 2.11. The equation of motion for each neuron can then be written as

$$\frac{dx_i}{dt} = -x_i + g(y_i) \tag{2.3}$$

where $g(y_i)$ is the nonlinear gain function performed by the neuron. Note that

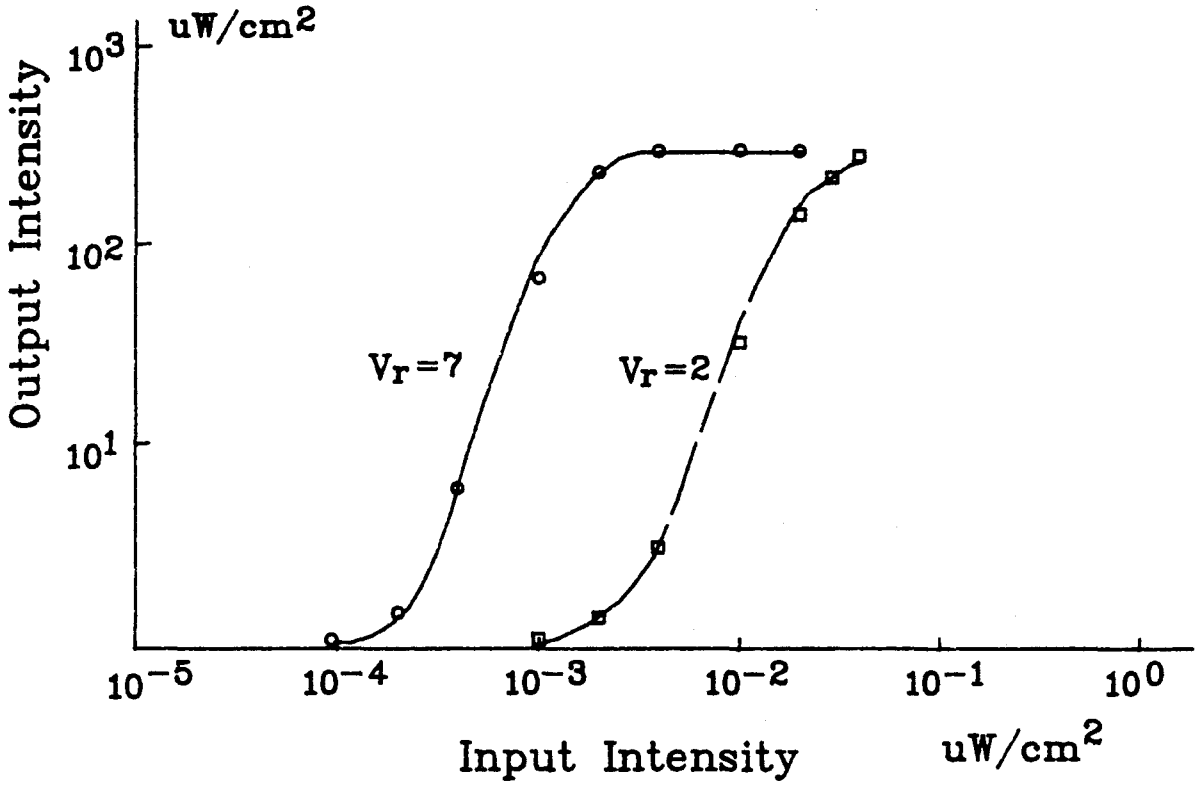


Figure 2.8: Operating Characteristics of the Optical Neuron. Reading beam intensity @ 1 mW/cm^2.

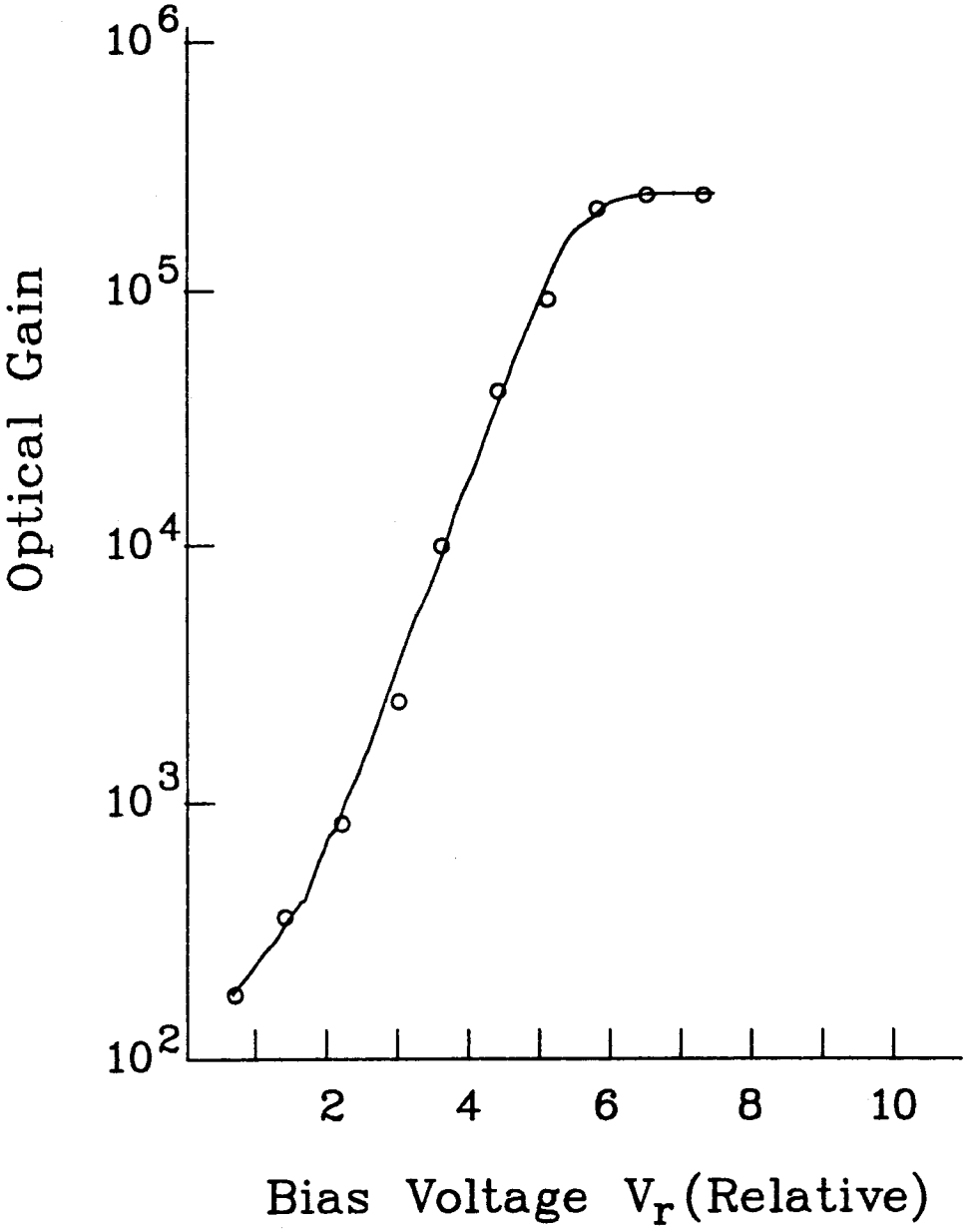


Figure 2.9: Control of Neural Gain by the Bias Voltage on the Image Intensifier.

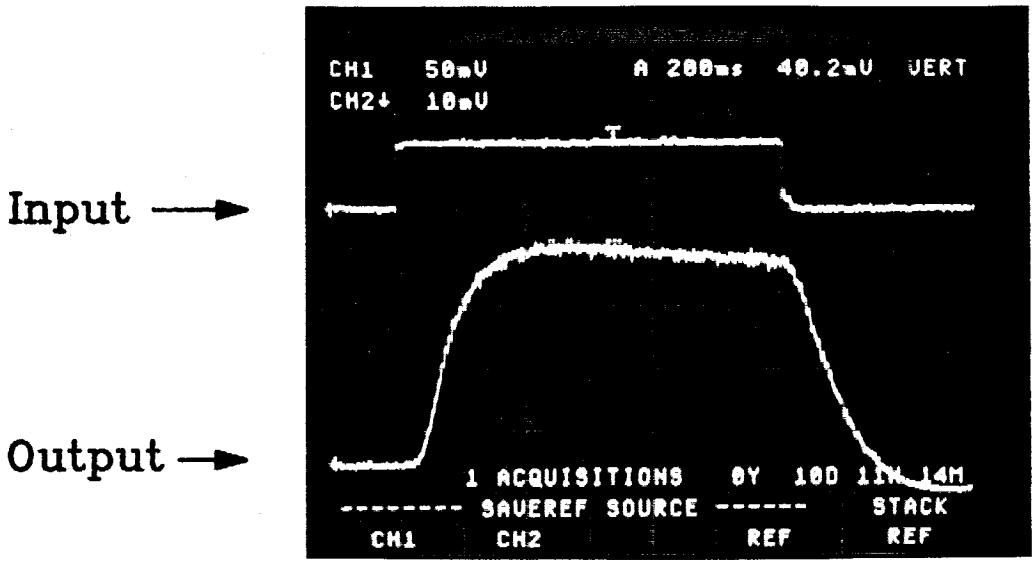


Figure 2.10: Temporal Response of the Optical Neuron.

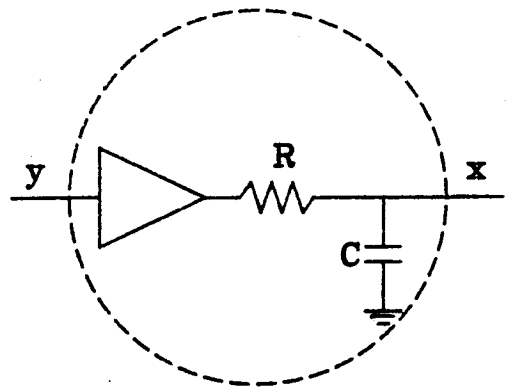


Figure 2.11: The Network Model of the Optical Neuron.

for simplicity we have scaled the time constant and gain function so that the time constant of the RC circuit is 1. One important comment about the gain function g should be made here. Our gain function is symmetric with respect to the input amplitude y_i. This is because the incoherent detection property of the LCLV and the image intensifier, which respond to light intensity and not to the amplitude of the input. As a result, the derivative of the gain function g is not always positive. This is different from the sigmoid-like function used in almost all other papers. Since the usual method for the convergence proof of the network uses the fact that the derivative of the gain function is always positive, that method does not apply to our system. We will develop a geometric method in Chapters 3 and 4 for studying the dynamics of neural networks. This method allows us to see not only how the system converges but also how the gain affects the dynamics. In our analysis, we do not need to specify a particular shape for the gain function, but need only assume that it approaches a saturation gain when the absolute value of the input exceeds a certain threshold, and that the slope at zero input is zero. The general form of the gain function is shown in Fig. 2.12.

2.3 Optical Interconnections

As we explained in Chapter 1 for the characteristics of neural networks, there are two crucial considerations for the interconnections in neural networks. First, the number of interconnections must be large enough for the collective computation to occur. Second, the interconnections must be dynamically modifiable so that the learning process can be realized. In the following, based on these two criteria, we discuss the technologies that are suitable for the implementation of optical interconnections.

For the networks we envision, we would like to have global interconnections among 10^4 neurons. This means 10^8 interconnections. In general, silver halide

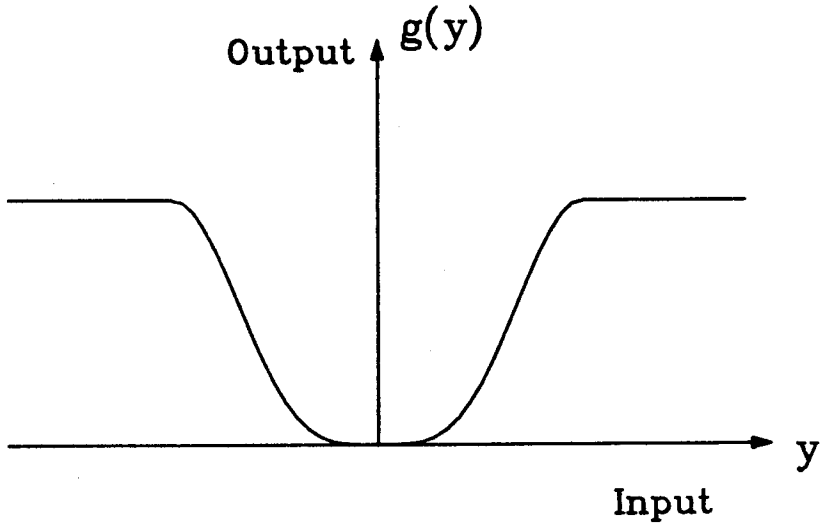


Figure 2.12: Nonlinear Characteristics of the Optical Neuron.

holographic films have a resolution capability of more than $1,000$ *lines/mm* over the size larger than one-inch square. Thus, it is sufficient for the required number of interconnections. However, once the interconnection are recorded and developed, they are fixed. Hence, these materials do not have the learning capability.

The second material that is promising for providing optical interconnections is erasable optical disks. Currently optical disks are used for data storage in electronic computers. We can have 10^{10} bits from a single disk. Thus, they are promising for recording a large number of interconnections. With current technology, information can be accessed randomly from any bit of the disk by mechanically scanning the disk and moving the reading laser head and then detecting the reflected beam. Each bit of the data is detected and processed sequentially; this limits the usefulness of the disks for optical neural computings. However, some optical architectures for parallel read-out of magneto-optic disks has being proposed and demonstrated [10, 11]. This suggests the potentiality of incorporating disks in optical neural networks.

Furthermore, magneto-optic disks can be erased and rewritten optically. Hence, the erasable optical disk is a promising technology for implementing optical neural computers with learning capabilities.

The third candidate is photorefractive materials.

As was explained in Section 2.2, photorefractive crystals are materials whose refractive indices can be modified by light illumination. Thus, the interconnections recorded in these materials are programmable. The second advantage for the photorefractive crystals is their volume storage capability. The storage capacity is limited by the number of resolvable spots in the crystal. In principle, this number is proportional to the volume of the crystal divided by the wavelength cubed. For a crystal of volume 1 cm^3 and with the laser wavelength of .5 μm, the storage capacity is approximately 10^{12}. The tremendous storage capacity and its modifiability make photorefractive crystals very attractive for applications in optical neural computers [12].

In the experiments of neural networks to be presented in Chapters 3 and 4, we are interested in studying the dynamics of associative memories. The information is recorded and fixed. We therefore use holographic plates for recording the interconnections. Thermoplastic plates are used for this purpose. In what follows we present the technique for recording optical interconnects in planar holograms and discuss its characteristics. The use of photorefractive crystals for neural networks will be presented in Chapter 5.

2.3.1 Holographic Gratings for Optical Interconnections

Historically, holograms were the first link between optics and neural networks. Both Van Heerden [13, 14] and Gabor [15] have discussed the analogy between holography and the way information may be stored in the brain. In holography the stored pattern is reconstructed by illuminating the hologram with the reference beam that

was used for recording. Thus holography is a form of associative memory. Recently, holograms have been considered as a means for providing interconnections in electronic computers [16, 17]. For our applications in neural networks, we use holograms for interconnecting neurons to form associative memories.

The way we make an interconnection between two neurons is by constructing a holographic grating between them. The schematic diagram of the optical system for the recording step is shown in Fig. 2.13. As shown in the figure, the architecture is similar to that of the Vander Lugt correlator [18]. Suppose the neurons a and b on different planes are to be associated. In order to develop the interconnection, we put a neuron b' at the imaging position of b in one of the input planes of the correlator. This input plane is called the training plane since it serves the purpose of training the network. We record a hologram using light emitted from the two training neurons a and b'. The recorded hologram is a Fourier-transform hologram of the two point sources. Suppose the light amplitudes coming from the neurons a and b' are A and B, respectively. Then the amplitude transmittance of the developed hologram is proportional to the exposure; i.e.,

$$\begin{aligned} t &\approx |Ae^{-j2\pi ua_1} + Be^{-j2\pi ua_2}|^2 \\ &= |A|^2 + |B|^2 + A^*Be^{j2\pi u(a_1-a_2)} + AB^*e^{-j2\pi u(a_1-a_2)} \\ &= I_0\Big[1 + me^{j2\pi u(a_1-a_2)} + me^{-j2\pi u(a_1-a_2)}\Big] \\ &= I_0\Big[1 + 2m\cos\left(2\pi u(a_1 - a_2)\right)\Big] \end{aligned} \tag{2.4}$$

where * denotes the complex conjugate, u represents the spatial frequency, a_1 is the position of the neuron a, a_2 is the position of the neuron b, and

$$\begin{aligned} I_0 &= |A|^2 + |B|^2 \\ &= I_a + I_b \end{aligned} \tag{2.5}$$

is the normalized intensity, $m = \frac{\sqrt{I_a I_b}}{I_a + I_b}$ is the modulation depth of the hologram. It

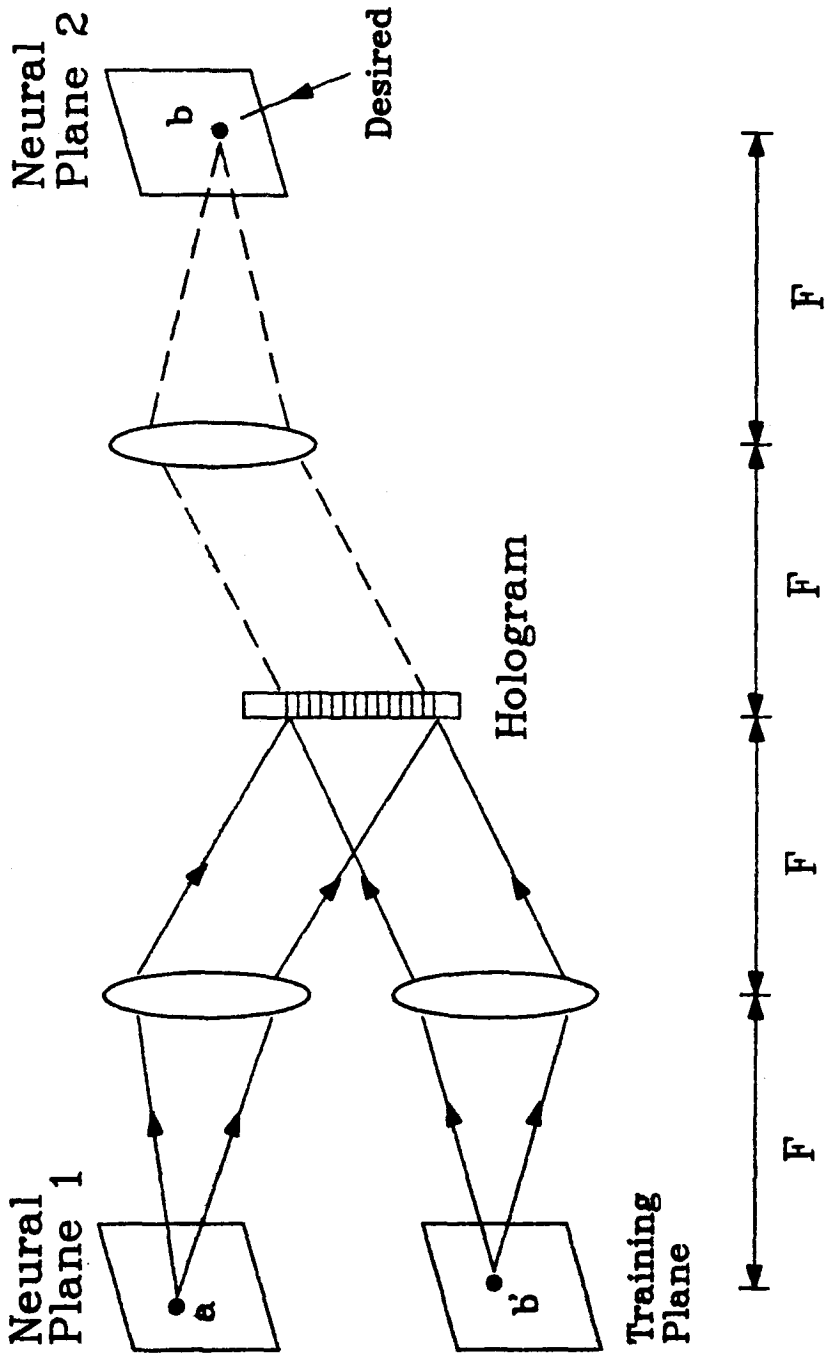


Figure 2.13: Recording the Interconnection for **a** and **b**.

is seen that a grating with spatial frequency of $\frac{1}{(a_1-a_2)}$ is recorded in the hologram. This grating works as the the interconnection between the neurons a and b. Suppose now that the neuron a is activated and emits light amplitude with A_1 . Then the reconstructed wave from the hologram is

$$\begin{aligned}&\left[A_1 e^{-j2\pi u a_1}\right] \cdot I_0\left[1 + m e^{j2\pi u(a_1-a_2)} + m e^{-j2\pi u(a_1-a_2)}\right] \\ &= A_1 I_0\left[1 + m e^{-j2\pi u a_2} + m^{-j2\pi(2a_1-a_2)}\right]. \end{aligned} \tag{2.6}$$

We see that the second term represents the light that is diffracted to the associated neuron b. This is the recalling process, and is shown in Fig. 2.14. Therefore, the interconnection between the associated neurons can be formed by recording the holographic grating with the neurons in the training plane.

Two points should be noted from the above result. First, the diffracted light amplitude is proportional to the modulation depth m of the hologram, which is in turn proportional to the product of the light amplitudes A and B that were used during recording. This implies that the interconnection strength is proportional to the product of the activities of two neurons. Thus holographic interconnection implements the Hebbian learning rule. Secondly, thin holographic gratings are shift-invariant. Hence, every neuron not only reads out its own gratings but also reads out other neurons' gratings and produces extra output. An example is shown in Fig. 2.15. In Figure 2.15(a), the neurons at the positions of the letter **A** are associated with a single neuron i. The holographic gratings are recorded. In the figure, w_{ij} represents the holographic interconnection between neurons j and i, and w_{ik} represents the holographic interconnection between neurons k and i. However, in the recalling process, neuron j is connected not only by w_{ij} to neuron i, but also by w_{ik} to neuron i_1. Similarly, neuron k also reads out w_{ik} and w_{ij}, thus activating an extra neuron i_2. Fig. 2.16 shows the experimental results. We see that in Figure (b) the reconstructed output is not the original point source. Therefore, if we want

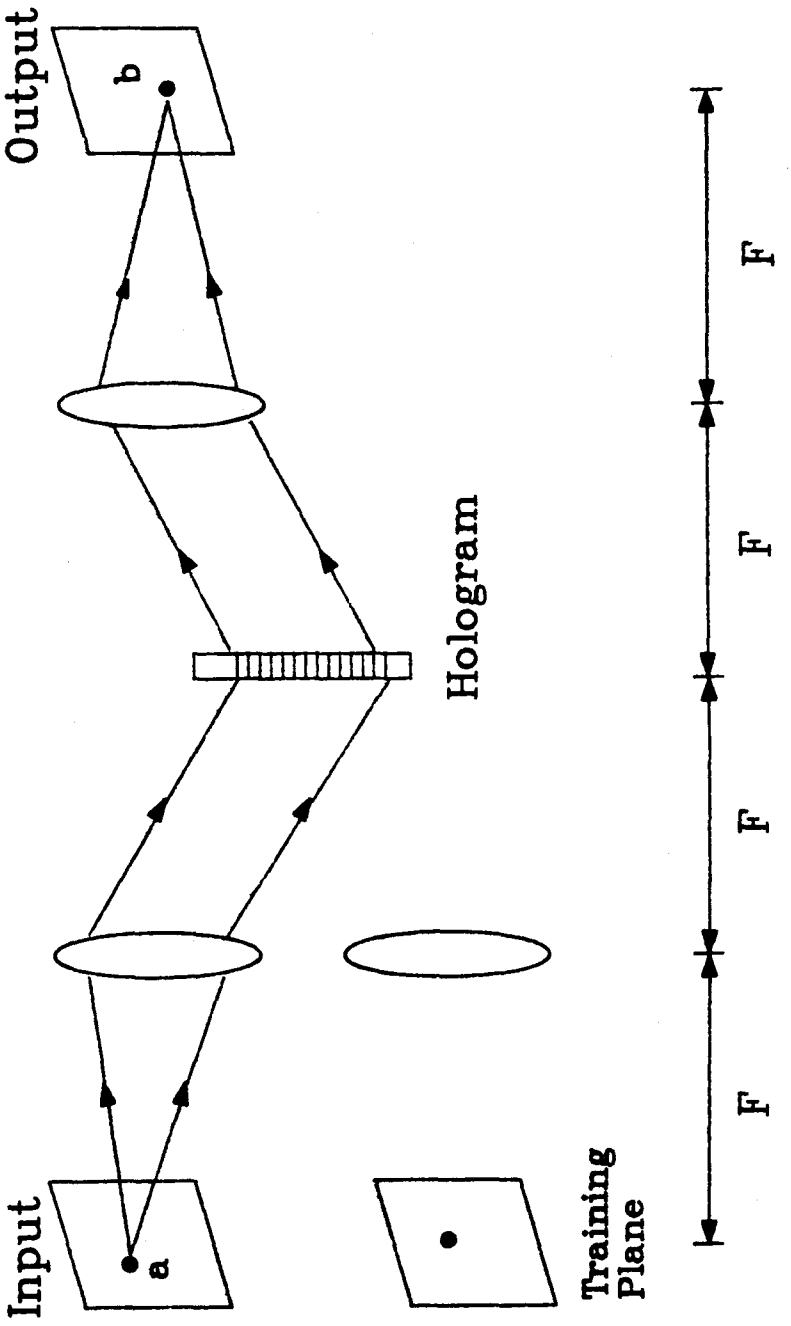
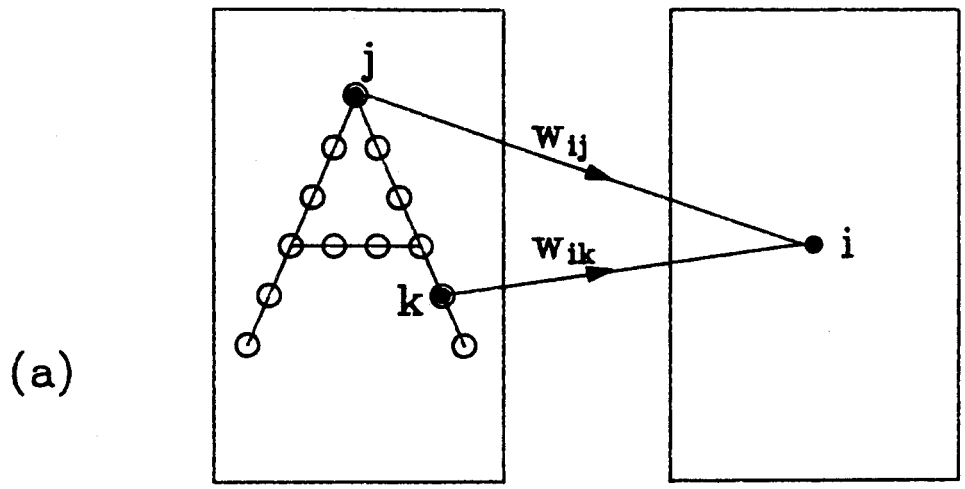


Figure 2.14: Recalling **b** from the Input **a**.



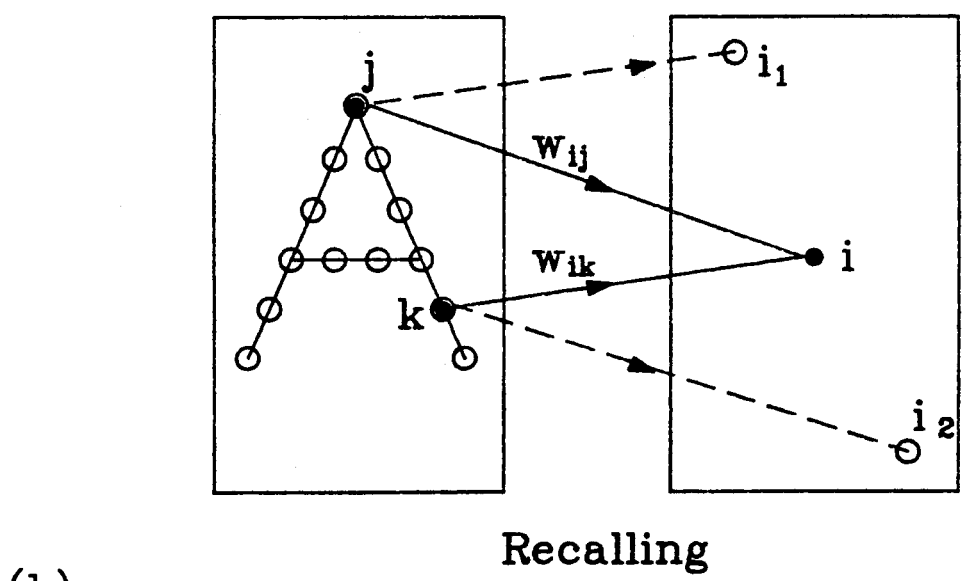


Figure 2.15: Association between the letter **A** and a point source. (a) Two gratings are recorded during the training phase. (b) The read-out with degeneracies.

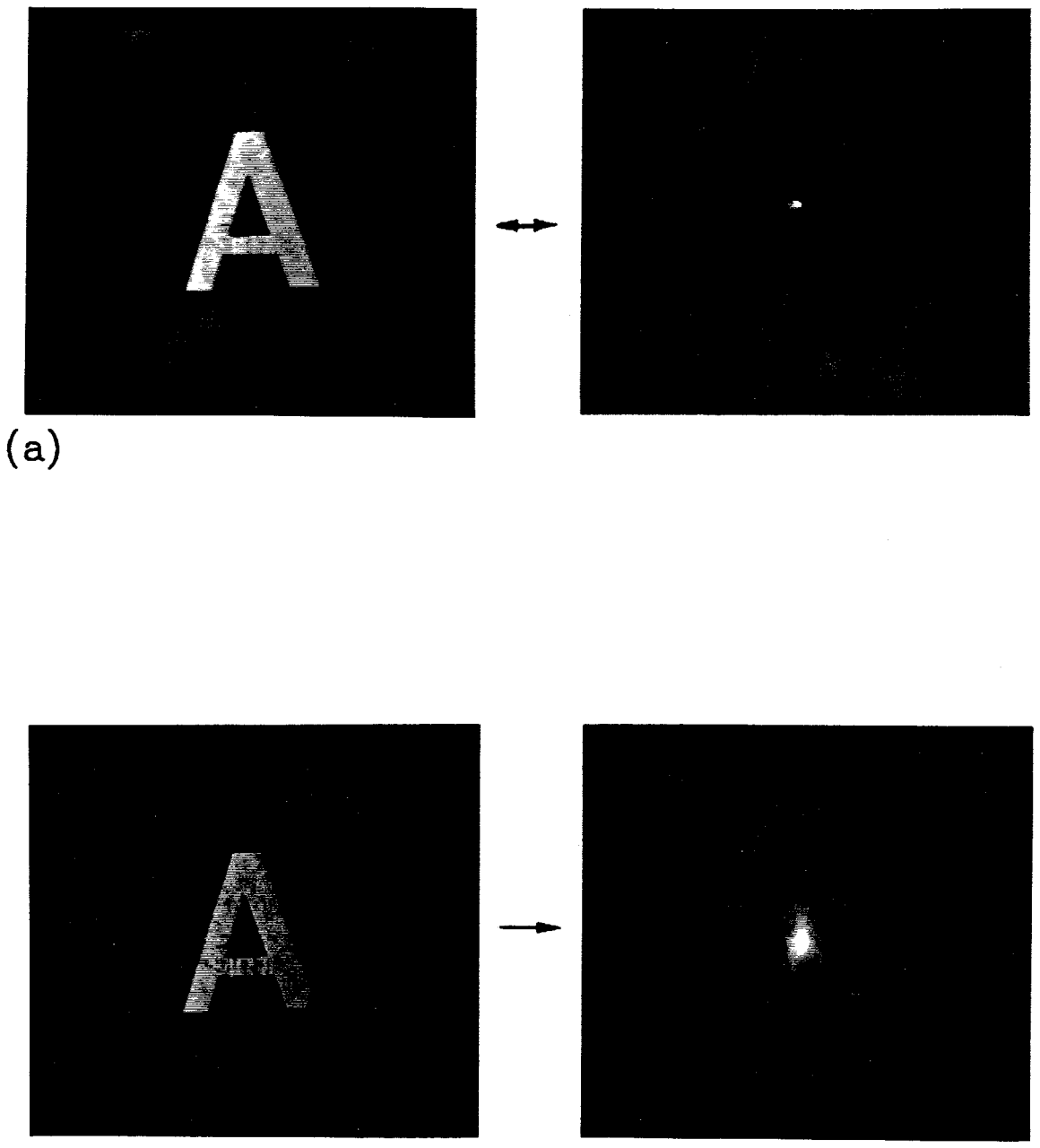


Figure 2.16: Association between the image of letter **A** and a point source.

to avoid reading out the degeneracies, the neurons i_1 and i_2 at the output plane should be prohibited. One way to achieve this is to use sampling masks to block out the neurons at the degenerate positions [19]. The technique of designing sampling masks will be presented in Chapter 4.

We now briefly describe the mechanisms of thermoplastic holograms. The thermoplastic plate consists of three layers coated on a quartz-plate substrate: an optically transparent conductor layer, a photoconductive organic polymer, and a thermoplastic layer. The plate is sensitive to the visible spectrum with an exposure sensitivity of 10 $\mu J/cm^2$. The optical interference fringes of the input beams are recorded as a surface relief pattern on the thermoplastic layer. The result is a thin phase hologram. The diffraction efficiency is typically 10%. The resolution capability of the plate is 800 $lines/mm$ with an aperture size of 30×30 mm square [20]. Thus, the plate can accommodate 10^8 gratings, allowing 10^4 neurons to talk to 10^4 others simultaneously. The thermoplastic holographic interconnections will be used in Chapters 3 and 4 for our associative loop experiments.

2.4 References for Chapter 2

1. A. Tanguay Jr., "Spatial Light Modulators for Real Time Optical Processing," in *Future Directions for Optical Information Processing*, Texas Tech Univ., Lubbock, Texas, 1981.

2. 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 Symposium on Lasers & Optics*, Paper 1043-07, Los Angeles, California, 15-20 Jan. (1989).

3. H. M. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic Press, Inc., Orlando, Florida, 1985.

4. L. S. Lee, H. M. Stoll, and M. C. Tackitt, "Continuous-time Optical Neural Associative Memory," *Opt. Lett.* **14**, 162 (1989).

5. *Photorefractive Materials, Effects, and Devices*, Technical Digest Series, **Vol. 17**, Optical Society of America, 1987.

6. J. N. Lee, "Devices for Optical Signal Processing," *Optics News*, October, 1985.

7. 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).

8. W. P. Bleha, L. T. Lipton, E. Wiener-Avnear, 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," *Optical Engineering*, Vol. **17**, No. 4, July/August, (1978).

9. A. Yariv and P. Yeh, *Optical Waves in Crystals*, John, Wiley and Sons, 1983, pp. 143 -147, pp. 266 -270.

10. D. Psaltis, A. A. Yamamura, M. A. Neifeld, and S. Kobayashi, "Parallel Read-out of Optical Disks," Paper ME3-1, *Topical Meeting on Optical Computing*, Optical Society of America, Salt Lake City, Utah, Feb. 27- March 1, (1989).

11. D. Psaltis, M. A. Neifeld, and A. Yamamura, "Optical Disk Based Correlation Architectures," Paper TuG3-1, *Topical Meeting on Optical Computing*, Optical Society of America, Salt Lake City, Utah, Feb. 27- March 1 (1989).

12. D. Psaltis, D. Brady, and K. Wagner, "Adaptive Optical Networks using Photorefractive Crystals," *Appl. Opt.* **27**, 1752, May (1988).

13 P. J. Van Heerden, "A New Optical Method of Storing and Retrieving Information," *Appl. Opt.*, **2**, 387 (1963).

14 P. J. Van Heerden, "Theory of Optical Information Storage in Solids," *Appl. Opt.*, **2**, 393 (1963).

15 D. Gabor, "Associative Holographic Memories," *IBM J. Res. Devel.*, **13**, 156, (1969).

16 J. Goodman, F. J. Leonberger, S. Y. Kung, and R. A. Athale, "Optical Interconnections for VLSI Systems," *Proc. IEEE*, Vol. **72**(1984).

17 A. A. Sawchuk and T. C. Strand, "Digital Optical Computing," *Proc. IEEE*, Vol. **72**(1984).

18 A. B. Vander Lugt, "Signal Detection by Complex Spatial Filtering," *IEEE Trans. Inform. Theory*, **IT-10:2**, 1964.

19. D. Psaltis, X. G. Gu, and D. Brady, "Fractal Sampling Grids for Holographic Interconnections," *Proceedings on Optical Computing 88*, SPIE Vol. 963 (1988).

20. *Operator's Manual- HC 300 Holographic Recording Device*, Newport Corporation.

Chapter 3

The Holographic Associative Memory Loop

3.1 Introduction

Associative memories are one of the most interesting applications of neural networks. In general, an associative memory stores a set of information, called memories. The information is stored in a format such that when an external stimulus is presented into the system, the system evolves to a stable state that is closest to the input data. We can view this process as a content-addressable memory since the stored memory is retrieved by the contents of the input and not by the specific address. In other words, the memory can recognize distorted inputs as long as the input provides sufficient information. Later in this chapter we will show the characteristics of the associative memory by presenting distorted versions of the stored images, e.g., rotated, scaled, shifted ones, etc. to the system and see how it converges.

In this chapter we present holographic implementations of Hopfield's model of neural network [1, 2]. This model has a very simple structure and is easy to implement. However, its principles and characteristics, e.g., neuron thresholding, global interconnections between the neurons, and feedback dynamics, are very flexible hence the results can be easily extended to other types of neural networks. In the following section, the basic principle of the Hopfield network and its one-dimensional implementation are reviewed. The optical implementation of the network for processing two- dimensional informations is presented in Section 3 and its experimental results are presented in Section 4. Special attention is focused on the dynamical phenomena of the feedback loop. Section 5 discusses the trade-off between distortion tolerance and image-recognition capability of the associative memory. Mathemati-

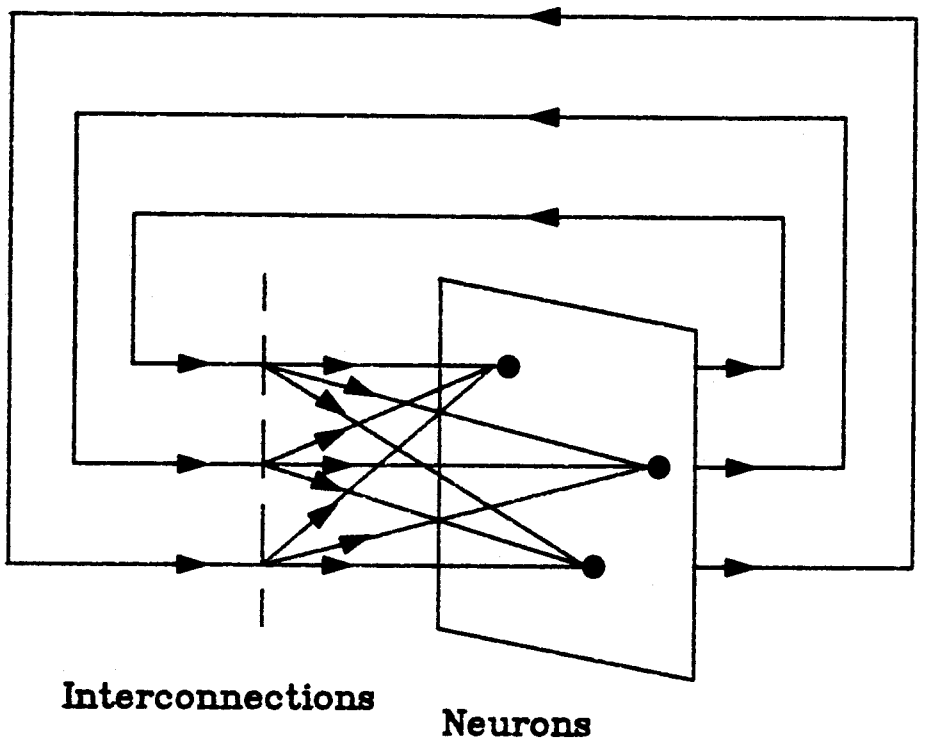


Figure 3.1: The 2-D Hopfield-type Neural Network

cal modeling of this particular optical loop and its convergence proof are presented in Section 6.

3.2 The Hopfield Model of Associative Memory

The basic structure of the Hopfield type neural network for two-dimensional (2-D) informations is shown in Fig. 3.1. It is a single-layer network with feedback. As shown in the figure there are two ingredients in the network: the neurons and the interconnection tensor. The neurons are distributed in a plane called the neural plane. The neurons receive input images, perform nonlinear thresholding on the received input, and then re-emit output patterns. The output of each neuron is

connected to the input of every other neurons to form a feedback network.

There are several ways to implement a 2-D neural plane [3]. Optoelectronic techniques integrate photo-detectors, electronic amplifiers, and light emitting sources to form a neural plane [4]. The thresholding characteristics of the neurons are controlled by adjusting the gains of the electronic amplifiers. Photorefractive $BaTiO_3$ crystal operating in a saturable, two-beam amplification mode provides another alternative [5] where the modulation depth between the two beams is an adjustable parameter. In our experiments we use the combination of an image intensifier and a liquid-crystal light valve (LCLV) to simulate the neural plane. The characteristics of this combination were already described in Chapter 2. For the case of 2-D neurons the interconnection pattern is a four-dimensional tensor. In our experiments the interconnection tensor is implemented by two Fourier-transform holograms. The holographic techniques of making this interconnection will be presented in the next section.

There are two phases involved in operating the Hopfield network, the learning phase and the recalling phase. In the learning phase, the information to be stored is recorded using an outer product scheme. This storage specifies the interconnection strengths between the neurons. In the recalling phase, an external input is presented to the system. The state of the system then evolves according to the correlation between the input and the stored data. Consider the learning phase in the 1-D case; we first store M N- bit binary words in a matrix $\omega_{i,j}$ according to

$$\omega_{i,j} = \begin{cases} \sum_{m=1}^{M} v_i^m v_j^m & \text{if } i \neq j; \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $v_i^m = \pm 1$, $i = 1, \cdots, N$, is the ith bit of the mth memory. This matrix can be calculated and recorded in a program for simulations or it can be plotted on a transparency for optical experiments. The recording of this synaptic matrix completes the learning phase. In the recalling phase, external data are fed into

the system and iterations are induced. For example, suppose $\mathbf{v}^{m0}$, the mth stored vector, is presented into the system. This vector is then multiplied by the stored matrix $\omega_{i,j}$, giving the output of the first iteration:

$$\begin{aligned} \hat{v}_i^{m0} &= sgn\left\{\sum_{j=1}^{N}\omega_{i,j}v_j^{m0}\right\} \\ &= sgn\left\{\sum_{j=1}^{N}\left[\sum_{m=1}^{M}v_i^m v_j^m\right]v_j^{m0}\right\} \qquad (3.2) \\ &= sgn\left\{(N-1)v_i^{m0}+\sum_{m\neq m0}^{M}\left[\sum_{j=1}^{N}v_j^m v_j^{m0}\right]v_i^m\right\}, \qquad (3.3) \end{aligned}$$

where $sgn\{\cdot\}$ means the thresholding function

$$sgn\{f(x)\} = \begin{cases} 1 & \text{if } f(x) \geq 0; \\ -1 & \text{if } f(x) < 0. \end{cases} \qquad (3.4)$$

The thresholded result of the first iteration re-enters the system for the next iteration so that the system evolves continuously. We see that in Eq. (3.3) two terms result from each iteration: The first term resembles the m^{th} memory which we call the signal; the second is the crosstalk between the m^{th} vector and other vectors, which we call the noise. We assume that the binary words are chosen randomly so that each bit is statistically independent; i.e., $P[v_i^m = 1] = P[v_i^m = -1] = 1/2$; then the signal-to-noise ratio (SNR) before thresholding is

$$\begin{aligned} SNR &= \frac{E[\hat{v}_i^{m0}]}{\sqrt{var[\hat{v}_i^{m0}]}} \\ &= \sqrt{(N-1)/(M-1)} \\ &\approx \sqrt{\frac{N}{M}}. \qquad (3.5) \end{aligned}$$

If N is sufficiently larger than M, then with high probability each element of $\hat{v}_i^{m0}$ equals that of v_i^{m0}. Therefore, the stored vector $\mathbf{v}^{m0}$ remains a stable state of the system in iterating operations. It was shown in [6,7] that if the stored vectors are

to be stable states, then the number of memories that can be stored in the system is limited by $M < N/(4 \ln N)$.

From Eq. (3.2) we see that we need three steps to implement the Hopfield network: vector-matrix multiplication, thresholding, and feedback. This was first done by Psaltis and Farhat using optoelectronic techniques [8,9]. They used a computer-generated transparency to provide the interconnection matrix. A 1-D array of 32 photodiodes followed by electronic thresholding plus a 1-D array of 32 LEDs was used to simulate 32 neurons. The arrays were used to detect and emit 32-bit data vectors. An optical vector-matrix multiplier was used to perform the multiplication between the the data vector and the transparency matrix. The multiplied result was detected by the photodiode array. The detected signal of each photodiode was electronically thresholded and fed back to the corresponding LED for further iterations. This system successfully demonstrated the dynamics and the capability of associative information recall of feedback neural networks. Another architecture using acousto-optics also demonstrated similar functions [10]. In this thesis we are interested in the processing of 2-dimensional images. The design and its optical implementation are presented in the following section.

3.3 Optical Implementation of the Associative Memory Loop

In this section we consider the implementation of the Hopfield-type neural network for 2-dimensional images. The interconnection pattern for 2-D images is described by the following equation:

$$\omega(x, y; \xi, \eta) = \sum_{m=1}^{M} f_m(x, y) f_m(\xi, \eta), \tag{3.6}$$

where $f_m(x, y)$ is the mth image to be stored, and M is the total number of images to be stored. Note that $\omega(x, y; \xi, \eta)$ is a four-dimensional tensor. It cannot be im-

plemented straightforward using a single transparency since a 2-D optical system has only two spatial coordinates. One could obtain additional variables by using wavelength multiplexing and time-domain processing. In [8], Psaltis and Farhat proposed a spatial-frequency multiplexing method in which a 2-D array of 2-D holograms, each separated by different spatial-frequencies, was used to perform the 4-D interconnection. Jang, et al. used a 2-D array of $N \times N$ diffused holograms to obtain the 4-D interconnection [11, 12].

We approach this problem from another point of view. Recall that in the recalling phase, the output of the system is described by the iterating equation

$$\hat{f}(x,y,t+1) = g\left\{\iint \omega(x,y;\xi,\eta) f(\xi,\eta,t) d\xi d\eta\right\}, \tag{3.7}$$

where $g\{\cdot\}$ represents the nonlinear thresholding of the neurons, $f(x,y,t)$ is the input to the system at time t, and $\hat{f}(x,y,t+1)$ is the output of the system. By inserting the expression of $\omega(x,y;\xi,\eta)$ into this equation and rearranging the order of integration and summation, we obtain

$$\begin{aligned}
\hat{f}(x,y,t+1) &= g\left\{\iint \left[\sum_{m=1}^{M} f_m(x,y) f_m(\xi,\eta)\right] f(\xi,\eta,t) d\xi d\eta\right\} \\
&= g\left\{\sum_{m=1}^{M} f_m(x,y) \left[\iint f_m(\xi,\eta) f(\xi,\eta,t) d\xi d\eta\right]\right\} \\
&= g\left\{\sum_{m=1}^{M} f_m(x,y) \left[\iint f_m(\xi,\eta) f(\xi-x,\eta-y,t) d\xi d\eta\right]_{x=0,y=0}\right\}. \qquad (3.8)
\end{aligned}$$

From Eq. (3.8) we can see that the implementation of the 2-D associative memory can be achieved in three steps [13]. The first step is to perform the inner product of the input image f with each of the memories f_m. Second, each inner product is multiplied by the associated memory. Third, these products are summed over all memories and thresholded by the neurons for iterations. The implementation of these steps matches holographic techniques exactly. The idea is shown in the block diagram of Figure 3.2. The first step described above can be realized by sampling

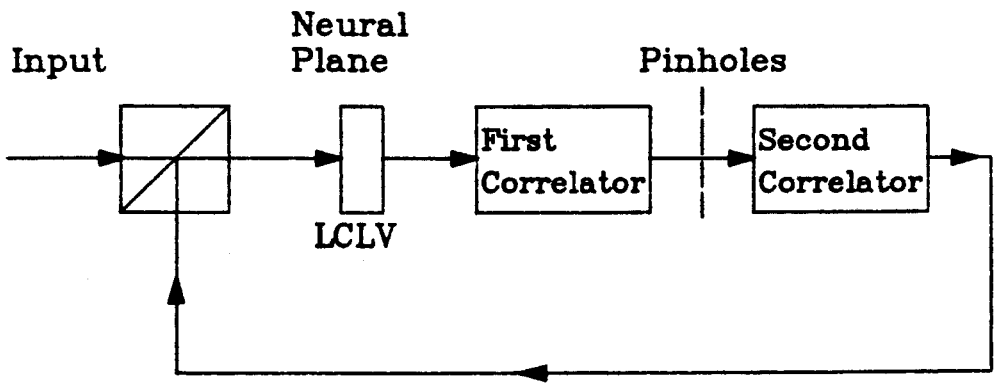


Figure 3.2: Block Diagram of the Optical Loop.

the correlation pattern at the origin in Vander Lugt system [14]. The first correlator and the pinholes carry out this operation. The signal passing through the pinholes reconstructs the images from the second correlator. This carries out the second step. In the final step, the reconstructed images are summed up at the input side of the neurons, in this case the writing side of the LCLV. The LCLV is used to simulate a plane of neurons to perform the thresholding and provides a signal for further iterations.

The operation of the associative loop can be explained with the aid of the diagram shown in Fig. 3.3. In this example four images are spatially separated and stored in the Fourier-transform holograms H_1 and H_2 as shown in the figure. When the input pattern **A** is presented on the plane $P1$ of the system, the first correlator produces the auto-correlation pattern along with three cross-correlations on plane P_2. The pinhole array on P_2 samples these correlation patterns at the center of each pattern where the inner products between the input and each of the stored images form. Each of the four beams that pass through the pinholes act as delta functions, reconstructing from the second correlator the four images stored in hologram H_2.

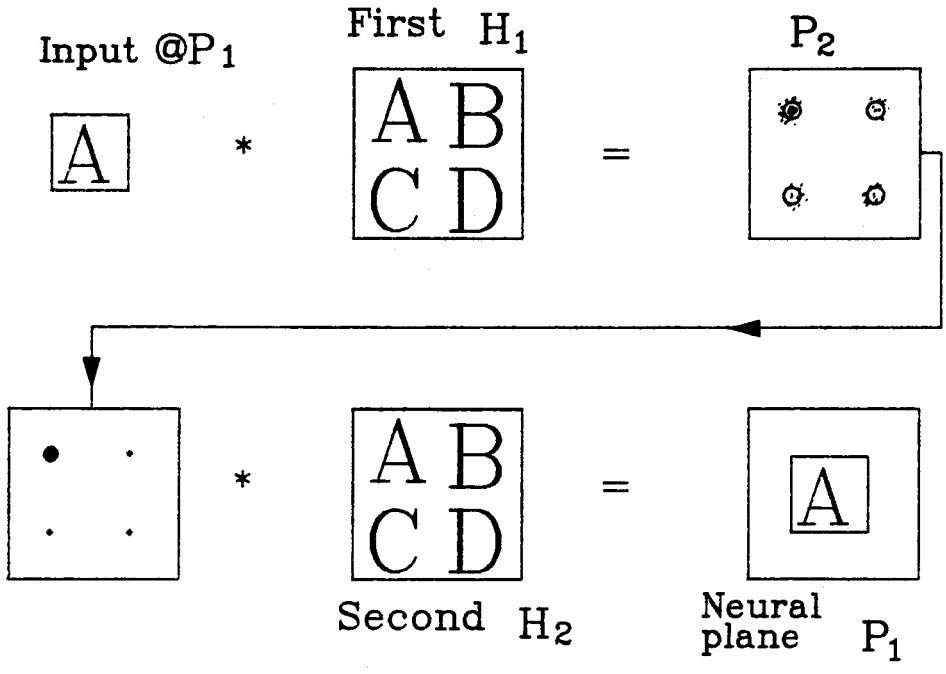


Figure 3.3: Operational Principle of the Optical Loop.

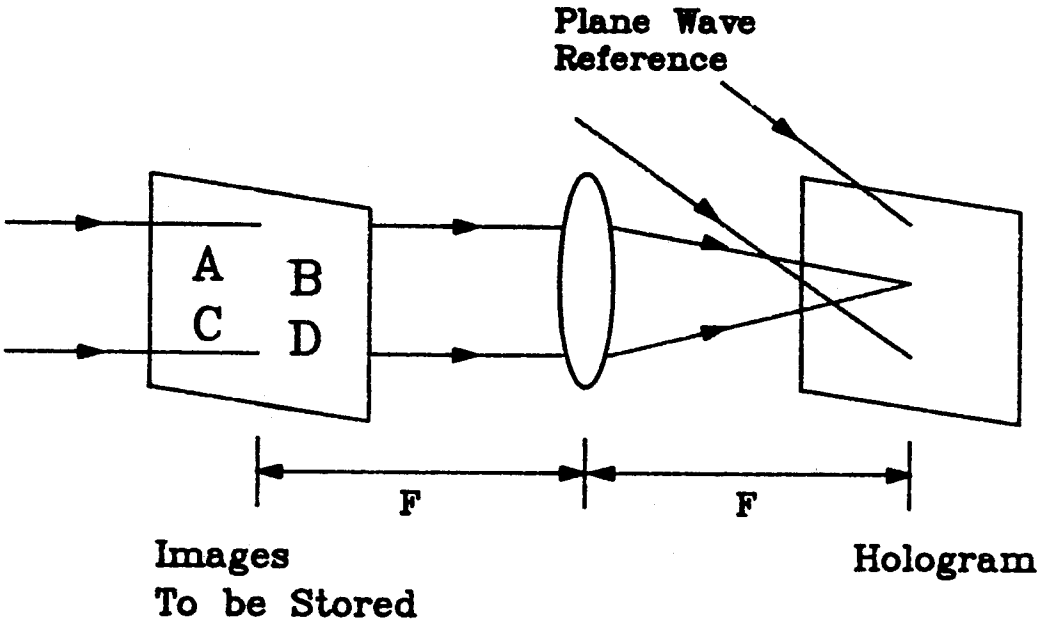


Figure 3.4: Recording of the Fourier Transform Hologram.

These reconstructed images are spatially translated according to the position of each pinhole and superimposed on plane P_1. At the center of the output plane of the second correlator we obtain the superposition of the four stored images. The stored image that is most similar to the input pattern gives the strongest correlation signal, hence the brightest reconstructed image. Here in Fig. 3.3 we show only the bright image that is reconstructed by the strongest auto-correlation peak. The weak read-out signal that is due to cross-correlations can be eliminated somehow by thresholding by the LCLV. The output of the LCLV becomes the new input image for the loop and forms a closed loop. The stable pattern that forms as a recirculating image in the loop is the stored image that is most similar to the original input. This image stays locked in the loop even when the external input in turned off.

In the system that we described in the previous paragraph, the images are recorded in a conventional Fourier-transform hologram, as shown in Fig. 3.4. The

four images, **A, B, C, D,** in this example are spatially separated at the input plane. A single plane wave is used as the reference. We record two identical holograms, $H1$ and $H2$, then put one in each of the two correlators of Fig. 3.2. A pinhole array with pinhole separations equal to that of the corresponding original images is placed at the correlation plane to sample the correlation signal.

The schematic diagram of the architecture of the above design is shown in Fig. 3.5; and a photograph of the experimental apparatus is shown in Fig. 3.6. In the system shown in Figure 3.5, the LCLV at plane P_1, the beam splitter cube BS_1, the lenses L_1, L_2, and the hologram H_1 form the first correlator. The part consisting of P_2, L_3, H_2, L_4, BS_3, and P_1 form the second correlator. The input pattern is imaged onto the LCLV by lens L_i and through beam splitter BS_3. A collimated argon laser beam illuminates the read-out side of the LCLV through beam splitters BS_2 and BS_1. A portion of the reflected light from the LCLV that propagates straight through BS_1, is diverted by BS_2, and it is imaged by lens L_0 onto a CCD television camera. This provides real time monitoring of the activity of the system. The portion of light that is reflected by BS_1 into the loop is Fourier-transformed by lens L_1 and illuminates hologram H_1. The correlation between the input image and each of the stored images is projected on plane P_2. The pinhole array at P_2 has spacing corresponding to the spatial separations of the stored images. The remainder of the optical system from P_2 back to the neural plane P_1 is essentially a replica of the first half, with the hologram H_2 storing the same set of images at H_1.

We now describe the operation of this system analytically. The amplitude transmittance of the transparency from which the holograms are recorded is

$$f(x,y) = \sum_{m=1}^{M} f_m(x - a_m, y - b_m), \tag{3.9}$$

where $f_m(x,y)$ is the amplitude transmittance of the mth image at the input plane in Fig. 3.5, (a_m, b_m) is the center of $f_m(x,y)$ on the (x,y) plane, and M is the total

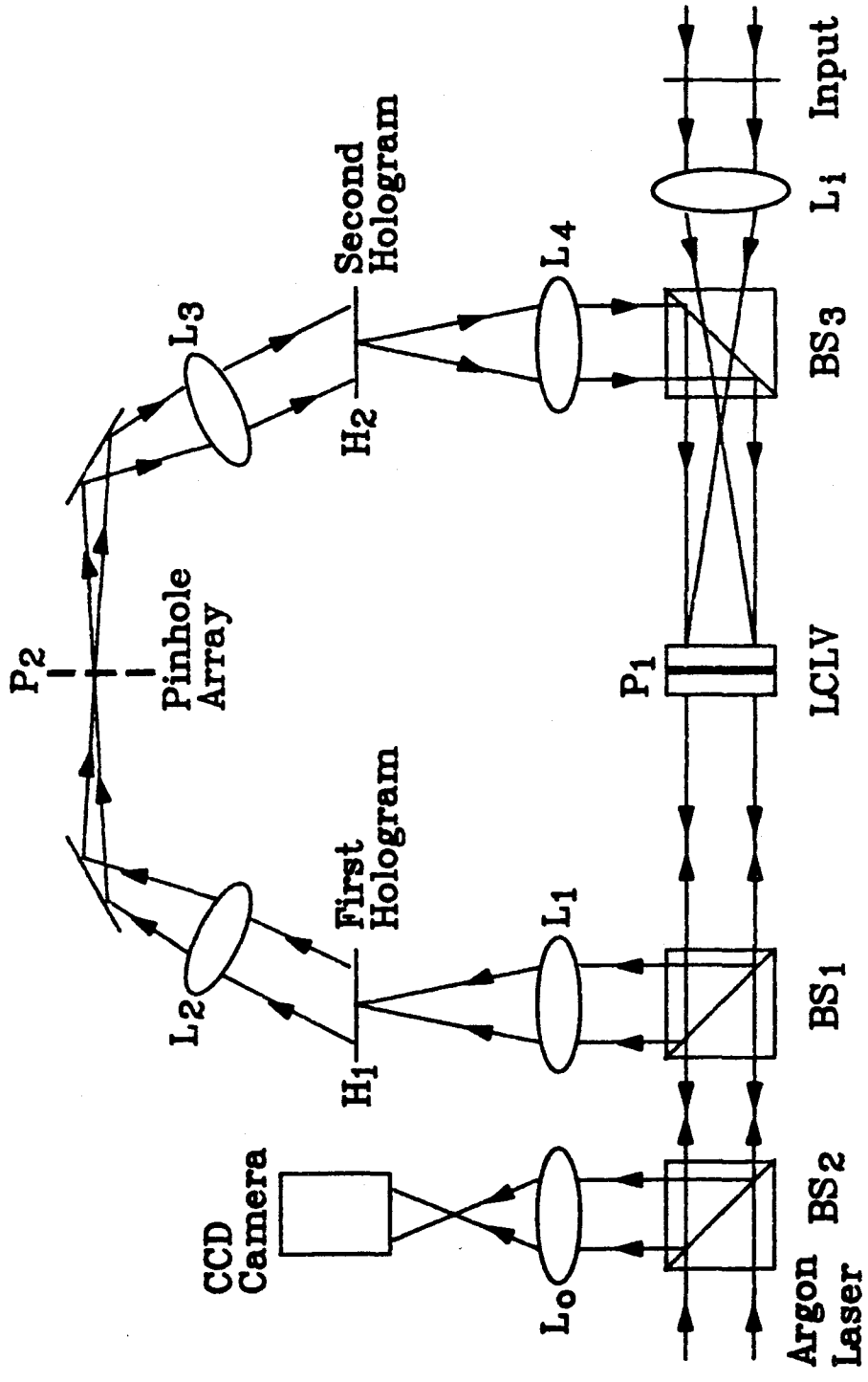


Figure 3.5: Schematic Diagram of the Optical Loop.

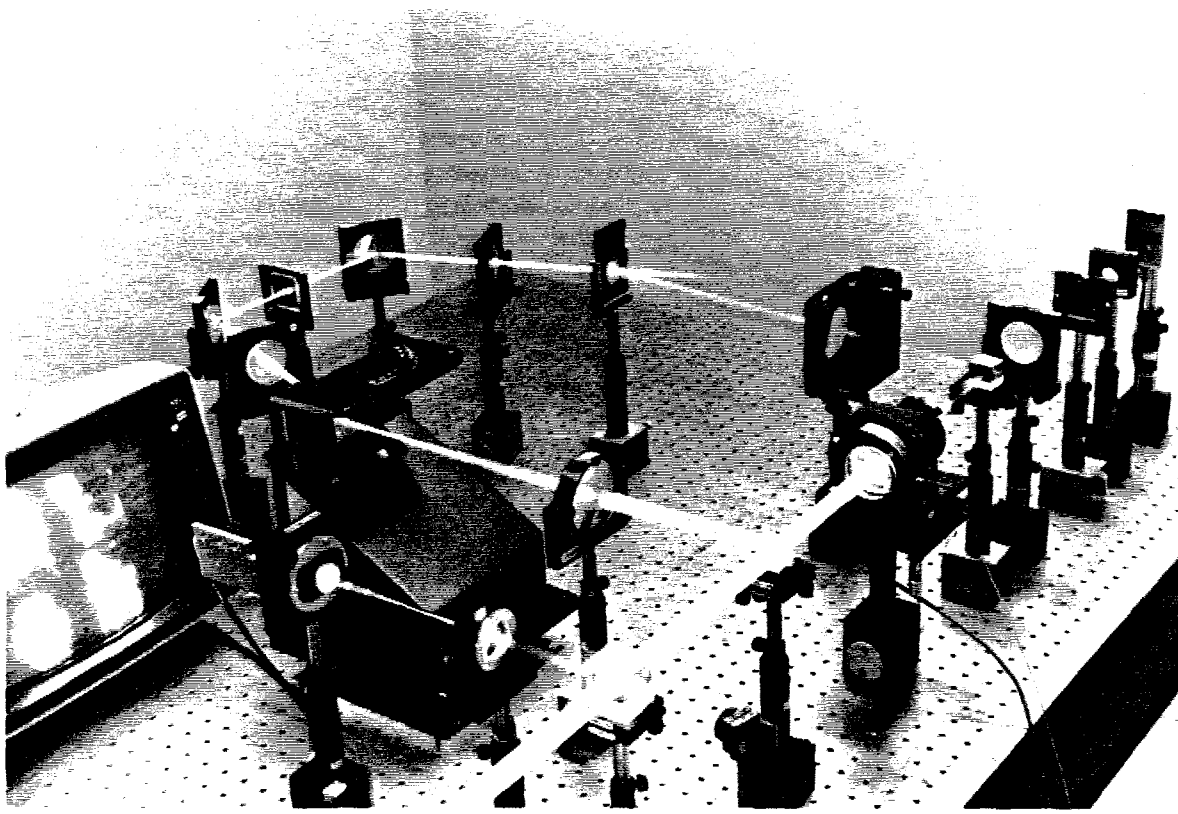


Figure 3.6: Photograph of the Optical Loop.

number of recorded images. In our experiments, the separations between different images are the same, i.e., $a_m = b_m = ma$, where a is a constant. The M images in Eq. 3.9 are Fourier-transformed to interfere with a reference wave $e^{-j2\pi ub}$ and recorded on a holographic plate. The amplitude transmittance of the developed hologram is

$$\begin{aligned} T(u,v) &= \left| \sum_m^M F_m(u,v) e^{-j2\pi(ua_m + vb_m)} + e^{-j2\pi ub} \right|^2 \\ &= \sum_m^M F_m^*(u,v) e^{j2\pi[u(a_m - b) + vb_m]} + Complex\ Conjugate + DC\ terms, \end{aligned} \quad (3.10)$$

where (u,v) is the coordinate in the Fourier plane. This hologram is placed in planes H_1 and H_2 of the system. In the above equation we are interested only in the first term since it is the part which the input image will correlate with. Suppose the amplitude of the input image at time t is $f(x,y,t)$ at plane P_1, then after hologram H_1 the light amplitude diffracted by the first term is

$$\sum_{m=1}^M F(u,v,t) F_m^*(u,v) e^{j2\pi[u(a_m - b) + vb_m]}. \quad (3.11)$$

At the correlation plane, P_2, the light amplitude is the Fourier transform of the above term:

$$\begin{aligned} &\sum_{m=1}^M f(-x',-y',t) \star f_m(-x',-y') * \delta(x' - (a_m - b), y' - b_m) \\ &\equiv \sum_m^M h_m(-x',-y',t) * \delta(x' - (a_m - b), y' - b_m), \end{aligned} \quad (3.12)$$

where $h_m(-x',-y',t)$ represents the correlation between f and the memory f_m, $\star$ means correlation and $*$ means convolution, and $(x\prime, y\prime)$ is the coordinate in correlation plane. The above equation is sampled by the array of pinholes having diameter W and locating at the positions $(a_m - b, b_m)$. In the limit $W \to 0$, the pinholes can be described as delta functions. Thus, the signal passing through the pinholes becomes

$$\sum_{m=1}^M \left[h_m(-x',-y',t) * \delta(x' - (a_m - b), y' - b_m) \right] rect\left(\frac{x' - (a_m - b)}{W} \right) rect\left(\frac{y' - b_m}{W} \right)$$

$$\approx \sum_{m=1}^{M} h_m(0,0,t)\delta(x'-(a_m-b),y'-b_m), \quad as\ W \to 0. \tag{3.13}$$

This signal reconstructs images from the hologram H_2, giving

$$\begin{aligned}&\sum_{m=1}^{M}\sum_{m'=1}^{M} h_m(0,0,t)e^{-j2\pi[u(a_m-b)+vb_m]}F^*_{m'}(u,v)e^{i2\pi[u(a_{m'}-b)+vb_{m'}]}\\ &=\sum_{m=1}^{M}\sum_{m'=1}^{M} h_m(0,0,t)F^*_{m'}(u,v)e^{-j2\pi[u(a_m-a_{m'})+v(b_m-b_{m'})]}.\end{aligned} \tag{3.14}$$

This signal is Fourier-transformed back to the input side of the neural plane P_1. The total field amplitude is

$$\hat{f}(x,y,t)=\sum_{m=1}^{M}\sum_{m\prime=1}^{M} h_m(0,0,t)f^*_{m\prime}(x+(a_m-a_{m\prime}),y+(b_m-b_{m\prime})). \tag{3.15}$$

There are $M \times M$ images reconstructed and imaged on the neural plane. Only the terms with $m = m'$ are on-axis and aligned with the original input $f(x,y,t)$. If we put a window centered at the optical axis, with the size equal to the size of each memory, we observe only the terms where $m = m'$:

$$\hat{f}(x,y,t)=\sum_{m=1}^{M} h_m(0,0,t)f^*_m(x,y). \tag{3.16}$$

The intensity of this light amplitude is detected by the photoconductor of the LCLV and gives an output light amplitude, to the first order, proportional to the detected intensity. Thus, the images are fed back into the loop. The images reconstructed by the auto-correlation peak become stronger and stronger until the LCLV saturates, whereas the images reconstructed by the cross-correlations become weaker and weaker until they die out.

The key elements in this optical loop are the holograms, the pinhole array, and the threshold device. The holograms in this system are thermoplastic plates. They have a resolution of 800 lines per millimeter over an area of one square inch. If we put a mirror [13] or a phase conjugate mirror [15, 16, 17] at the pinhole plane P_2 to reflect the correlation signal back through the system, then we only need one

hologram to form the associative memory. But then it losses feedback dynamics. The use of two holograms, however, not only provides dynamics but also improves system performance. We make the hologram at H_1 with a high-pass characteristic so that the input section of the loop has high spectral discrimination. On the other hand, we want the feedback images to have high fidelity with respect to the original images. Thus, the hologram at plane H_2 must have broad-band characteristics. We use a diffuser to achieve this when making H_2. Fig. 3.7(a) shows the four original images. Fig. 3.7(b) shows the images reconstructed from the first hologram $H1$, and Fig. 3.7(c) shows the images reconstructed from the second hologram $H2$. As expected, Fig. 3.7(b) is a high-pass version of the original image while Fig. 3.7(c) is broad band.

The pinhole array at plane P_2 samples the correlation signal between the image coming from the LCLV and the images stored in hologram H_1. The pinhole diameter used in these experiments range from 45 μm to 700 μm, depending on the images to be stored and the desired system performance. If the pinhole size is too small, then the light that passes through it to reconstruct the feedback image is too weak to be detected by the LCLV and no iterations can occur. On the other hand, large pinholes introduce excessive blurring and cross-talk in the feedback and make the reconstructed images unrecognizable. The pinhole size also affects the shift invariance property of the loop. In order to be recognized, the auto-correlation peak from an external image should stay within the pinhole. Larger pinholes allow more shift in the input image. The system performance under different selections of pinhole diameters will be discussed in the next section.

The purpose of the threshold device in this system is three-fold. The first is to bring into the system a coherent image from the external input (although either coherent or incoherent images can be used as the input). The second is to provide a thresholding operation to the feedback signal so that cross-correlation is reduced

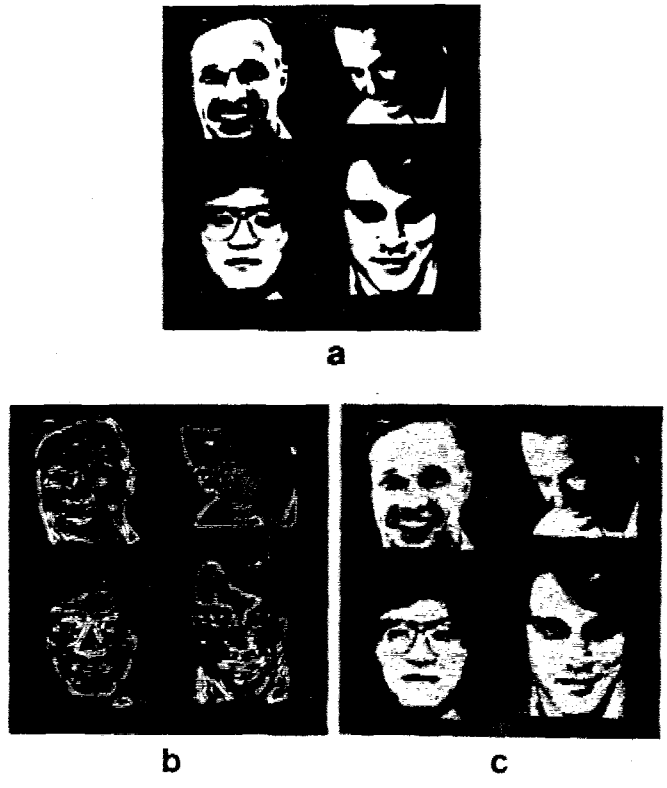


Figure 3.7: Images Stored in the Holograms. (a) The original images. (b) Images reconstructed from H_1. (c) Images reconstructed from H_2.

in successive iterations. The third is to provide gain to the feedback signal. The optical signal is attenuated in the loop because of the diffraction efficiencies of the Fourier transform holograms and the losses from pinholes as well as lenses and beam splitters. Therefore, we need to have optical gain to compensate this loss. In our system this is achieved by adding an image intensifier at the photoconductor side of the LCLV. As described in Chapter 2, the microchannel plate of the image intensifier is sensitive to a minimum incident intensity of approximately 1 nW/cm^2 and reproduces the input with an intensity 10^4 times brighter (10 $\mu W/cm^2$). This is bright enough to drive the LCLV. If we use a beam with intensity equal to 10 mW/cm^2 to read the LCLV, then the intensity of the output light is approximately 1 mW/cm^2. Thus, the combination of the image intensifier and the LCLV provides optical gain up to 10^6. This optical gain is similar to a sigmoid function and its slope can be adjusted by changing the bias voltage of the image intensifier. In Section 3.5 we will see that the setting of the gain is the key parameter that mediates the trade-off between distorsion invariance and the discrimination capability of the loop.

3.4 Experimental Results

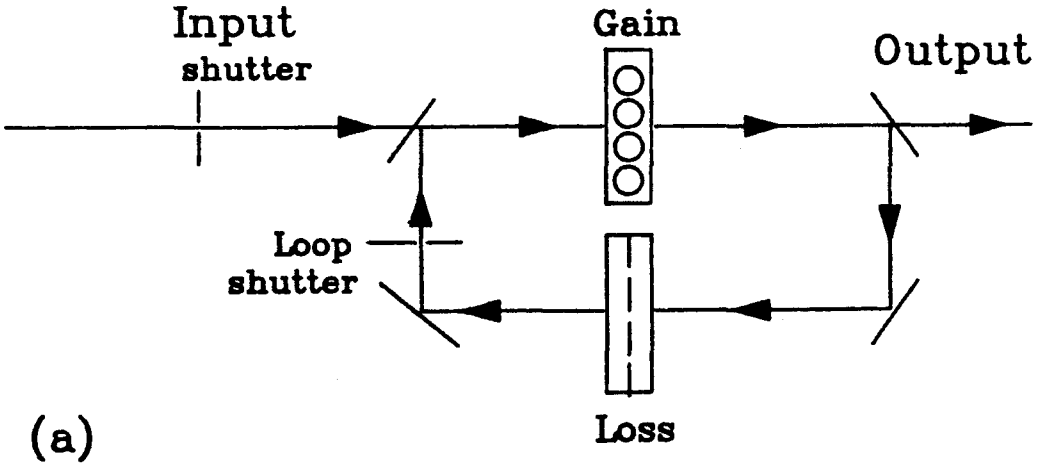
In this section we show, experimentally, many interesting properties of the optical associative loop. These include retrieval of complete images from partial inputs, recognition of the shifted, rotated and scaled images, error-correction capability, and neural network dynamics. The most interesting among these is the system dynamics. This is a unique property of a neural network with feedback and threshold. The state of the system evolves in time and the dynamics are determined by the operation parameters . The dynamics provide a trade-off for optimal system design. The dynamics of the associative memory is presented in the following section and the trade-off issue will be discussed in Section 3.5.

3.4.1 The Dynamics of the Associative Loop

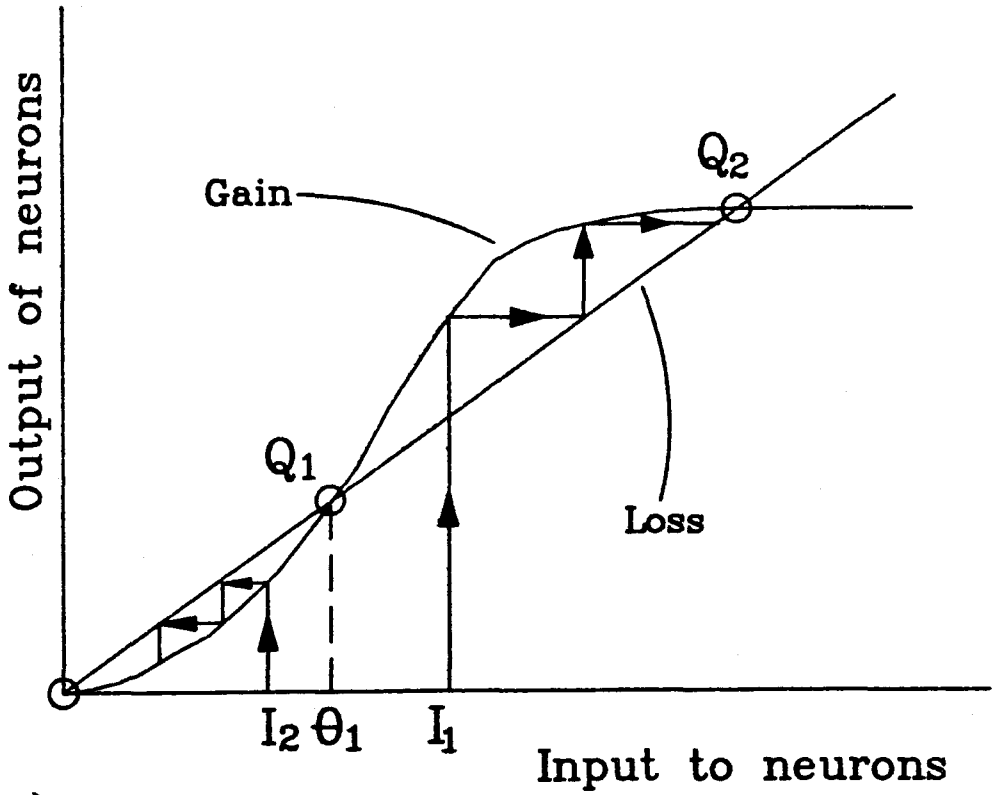
According to the characteristics of the optical components, i.e., active or passive, the optical associative loop of Fig. 3.5 can be lumped into a simplified diagram as shown in Fig. 3.8(a). Since the LCLV provides optical gain and thresholding to the signal, it is represented as the component **Gain** in Fig. 3.8(a). The other parts of the loop are all lossy components and are represented as the component **Loss** in Fig. 3.8(a). This includes the beam-splitter cubes, the lenses, the mirrors, the pinhole array, and the holograms. The circulation of the optical signal in the loop is also shown by the arrows in the figure. Note that in the loop the output of the neurons forms the input to the lossy part, and the output of the lossy part becomes the new input for the neurons in the next iteration.

Let us consider the loop dynamics. The dynamics of the recall process can be described by using an iteration map formed by the gain and loss curves as shown in Fig. 3.8(b). In the figure the sigmoid curve represents the input-output response of the neurons. The slope of the curve is proportional to gain of the neurons, whereas the slope of the straight line is proportional to loop loss because of the holograms and pinholes and is drawn on the same diagram as the input-output response of the neurons. The intersection point of this line with the neural gain curve at Q_1 determines the loop threshold level, and the intersection point Q_2 represents a stable point. If the initial condition of the neuron is above the threshold level θ_1, such as I_1 shown in the figure, the signal grows after each iteration until it arrives and latches at Q_2. On the other hand, if the initial condition is below θ_1, such as I_2 shown in the figure, the signal will decay to zero. The number of iterations depends on the distance of the initial condition from the threshold.

The loop dynamics was measured by controlling two shutters as shown in Fig. 3.8a. The temporal response of the loop to an input pattern is shown in Fig. 3.9. The

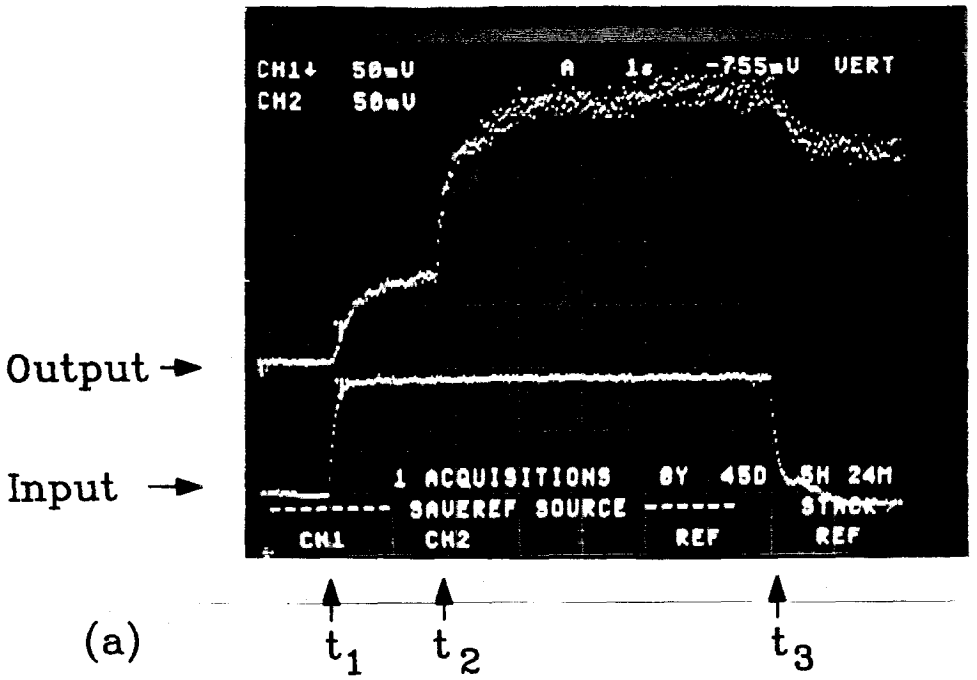


(a)



(b)

Figure 3.8: Iterations in the Optical Loop. (a) The gain and loss components in the loop. (b) The iterations in the loop.



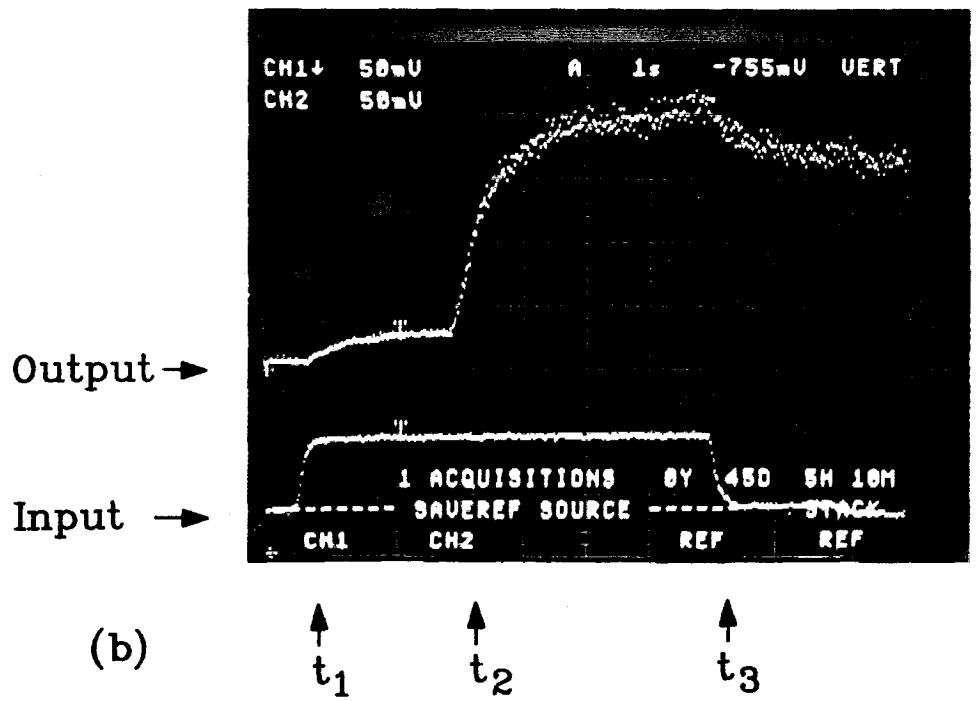


Figure 3.9: Temporal Response of the Loop. (a) Strong input. (b) Weak input. Timing: t_1 = Input ON, t_2 = Feedback ON, t_3 = Input OFF.

lower trace represents the intensity of the external input image and the upper trace represents the corresponding light intensity detected at the loop output. Before time t_1, both shutters are **OFF** and the responses are low. At time t_1 the input shutter is turned **ON**; hence the lower trace becomes high, but the loop shutter remains **OFF**. The upper trace shows the corresponding response of the neurons to the external input. The rise time of the LCLV is approximately one second in this experiment. At time t_2 the loop shutter is turned **ON** and the loop is closed. The feedback signal arrives at the neurons as an additional input and iteration occurs. From Fig. 3.9(a) we see that it takes about two seconds for the loop to reach a stable state. At time t_3 the input shutter is turned **OFF**; hence the lower trace becomes low. However, the loop remains latched to a stable state, which is one of the stored images. Fig. 3.9(b) shows the same experiment but with input intensity reduced to one-third of the first input. The first rise of the upper trace shows that the rise time of the neurons remains one second. And the second rise of the upper trace shows that it takes approximately four seconds for the loop to reach its stable state. However, after the input is turned off, the loop gives the same output intensity. This example shows that initial conditions affect the dynamics of the loop but do not affect the final state of the system. Fig. 3.10 shows the iteration map of this experiment where the initial input I_2 is lower than I_1. It shows that I_2 takes more iterations to reach the stable state, but the final state is the same as that of I_1.

Since the external input does not affect the shape of the final state, but rather selects which state is produced, we can build a degree of invariance in the system since a shifted, rotated or scaled version of a stored image can recall the stored image. The effect of such distortions of the input image is to decrease the level of the initial condition. As long as the initial condition is above the loop threshold (θ in Fig. 3.8(b)), the loop is always brought to the stable state that is most similar to the initial input. This means that the dynamics is determined by the relative

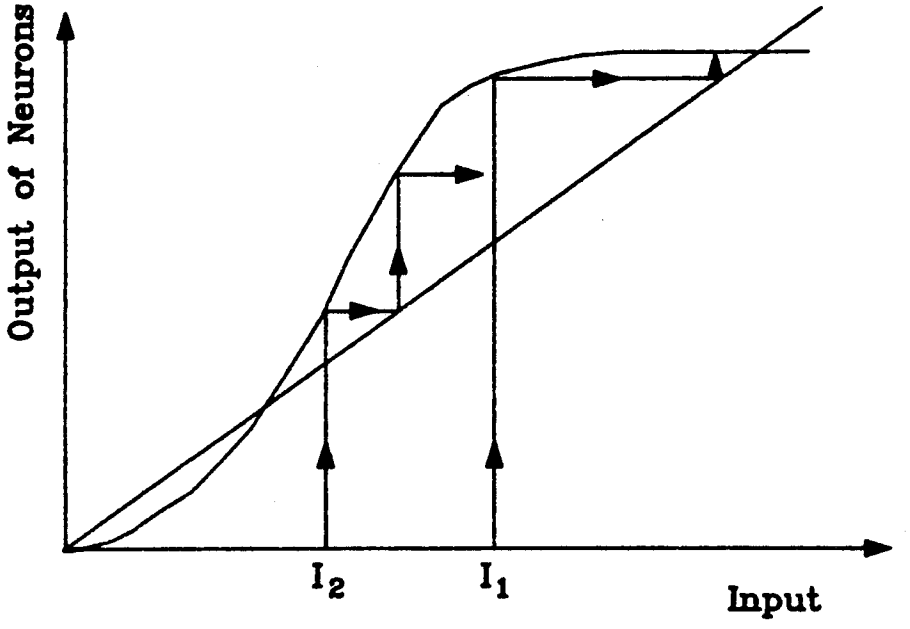


Figure 3.10: Convergence of the Loop.

position of the initial input with respect to the loop threshold. The initial condition is determined by the degree of distortion of the external image as well as its light intensity. On the other hand, the loop threshold is determined by system parameters such as the neural gain and loop loss. In the next section we give experimental results of system invariances against various distortions. The problem of how to select optimal system parameters to make the trade-off between distortion tolerance and capability of recognizing correct images will be presented in Section 3.5.

3.4.2 Retrieval of the Complete Image from a Distorted Input

In this section we show experimental results of the invariance property of the memory loop in recognizing a distorted image. The images stored in the loop are the four faces shown in Fig. 3.7(a). They are recorded as a Fourier-transform hologram as described in Section 3.3.

Fig. 3.11(a) shows the response of the memory when the half face of a recorded image is presented to the system with the loop shutter **OFF**. This sets the initial condition of the loop. We then turn the loop shutter **ON** to close the feedback loop. The signal then circulates in the loop and the state evolves. After many iterations the loop reaches the stable state and the complete face appears. The time for this process ranges from less than one second to several seconds, depending on the initial conditions and the system parameters. The complete image is locked in the loop even after the external input is turned **OFF**. Fig. 3.11(b) shows the system output at the moment the loop shutter is **ON**. We see that the feedback image is superimposed on the external input. Fig. 3.11(c) shows the output 0.4 seconds after the feedback loop is closed, and Fig. 3.11(d) shows the output 0.8 seconds after the feedback loop is closed. Fig. 3.11(e) shows the complete image recalled from the loop after 2 seconds. Fig. 3.11(f) shows that after we remove the external input the recalled image is latched in the loop. Fig. 3.12 shows the temporal sequence of the

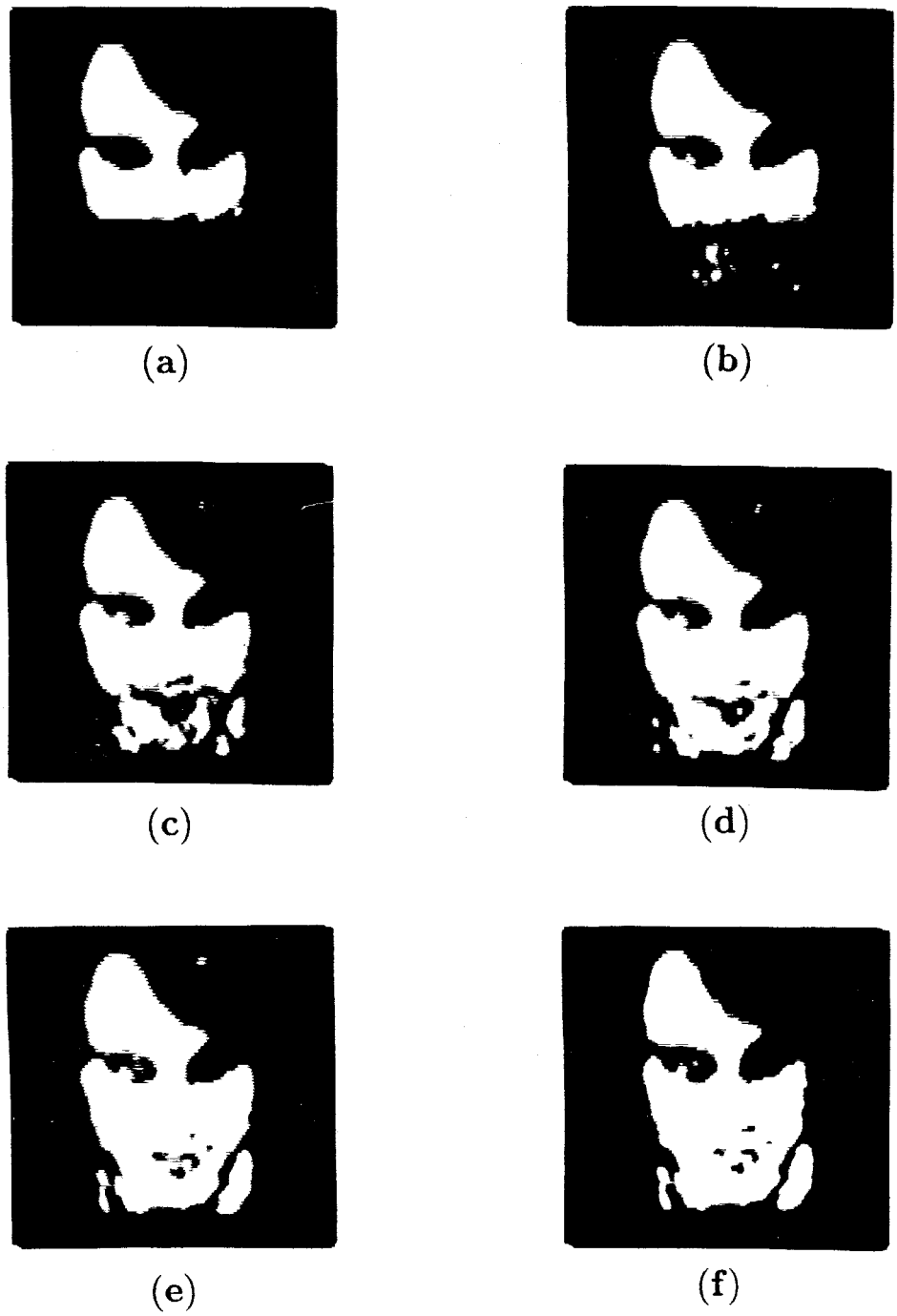


Figure 3.11: Retrieval of the complete Image from the Partial Input. (a) The partial input at $t = 0$. (b) $t = 0^+$ (Loop closed). (c) $t = 400\ ms$. (d) $t = 800\ ms$. (e) $t = 2$ sec. (f) Input OFF.

same experiment but with another half-face as the input. This experiment shows that the external input to the associative memory does not need to be exactly the same as the recorded images.

Now we use a rotated version of one of the stored images as the input to investigate the rotation invariance capability of the loop. Fig. 3.13(a) shows the response of the memory when a rotated version of a recorded image is presented into the system with the loop shutter **OFF**. The input is rotated by 6 degrees with respect to the original image. Fig. 3.13(b) shows the memory output at the moment the feedback loop is closed. The loop state then evolves to give the original image. The temporal sequence of this evolving is shown in Fig. 3.13(c) to Fig. 3.13(e). In this experiment the degree of rotation of the input image sets the initial condition of the loop. More rotation means more distortion; hence the initial condition is farther away from the stored memory. Thus, the loop needs more iterations to arrive at the stable state. Fig. 3.14(a) shows this result. In the figure the upper curve represents the stable state intensity of the output image and the lower curve represents the time for the loop to reach stable states, both as a function of rotation angle. It shows that as the input is rotated, it takes a longer time to be recognized. However, once the loop converges to the stable state, the output intensity is always the same regardless of initial rotation. The figure also shows that the output intensity drops to zero when the initial input is rotated over 8 degrees. This means that the initial condition is below the loop threshold and the rotated image is not recognizable. One way to increase the tolerance of rotation is to increase the neural gain so that it can detect weaker feedback signals from the distorted input. Fig. 3.14(b) shows the result of the same experiment but with the neural gain 10 times higher compared to that used in Fig. 3.14a. The rotation tolerance now increases to 16 degrees. One might infer that we can obtain more tolerance simply by increasing the neural gain. This is not true since high gain also enhances crosstalk and the loop may not converge

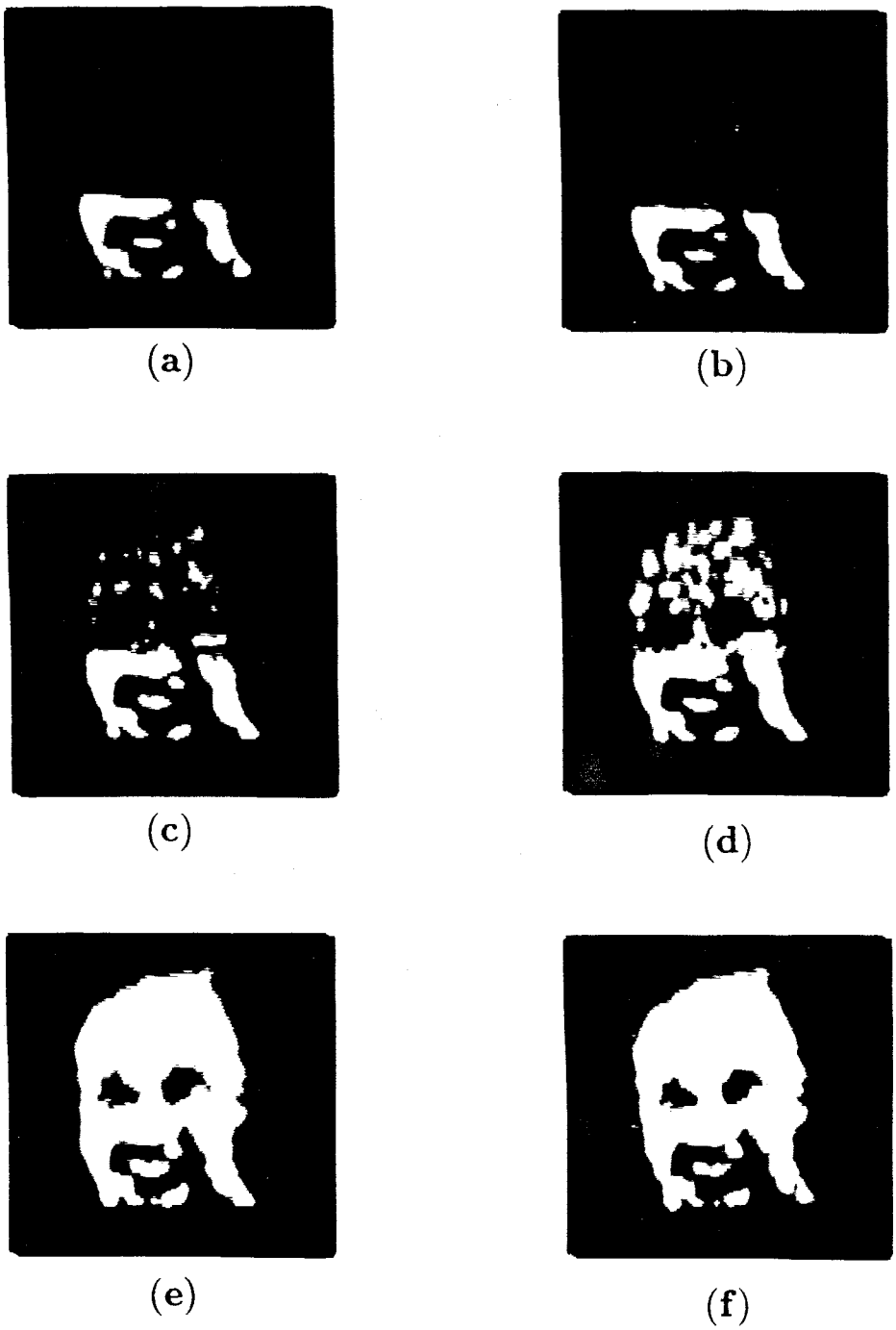


Figure 3.12: Retrieval of the complete Image from the Partial Input. (a) The partial input at $t = 0$. (b) $t = 0^+$ (Loop closed). (c) $t = 480\ ms$. (d) $t = 800\ ms$. (e) $t = 2$ sec. (f) Input OFF.

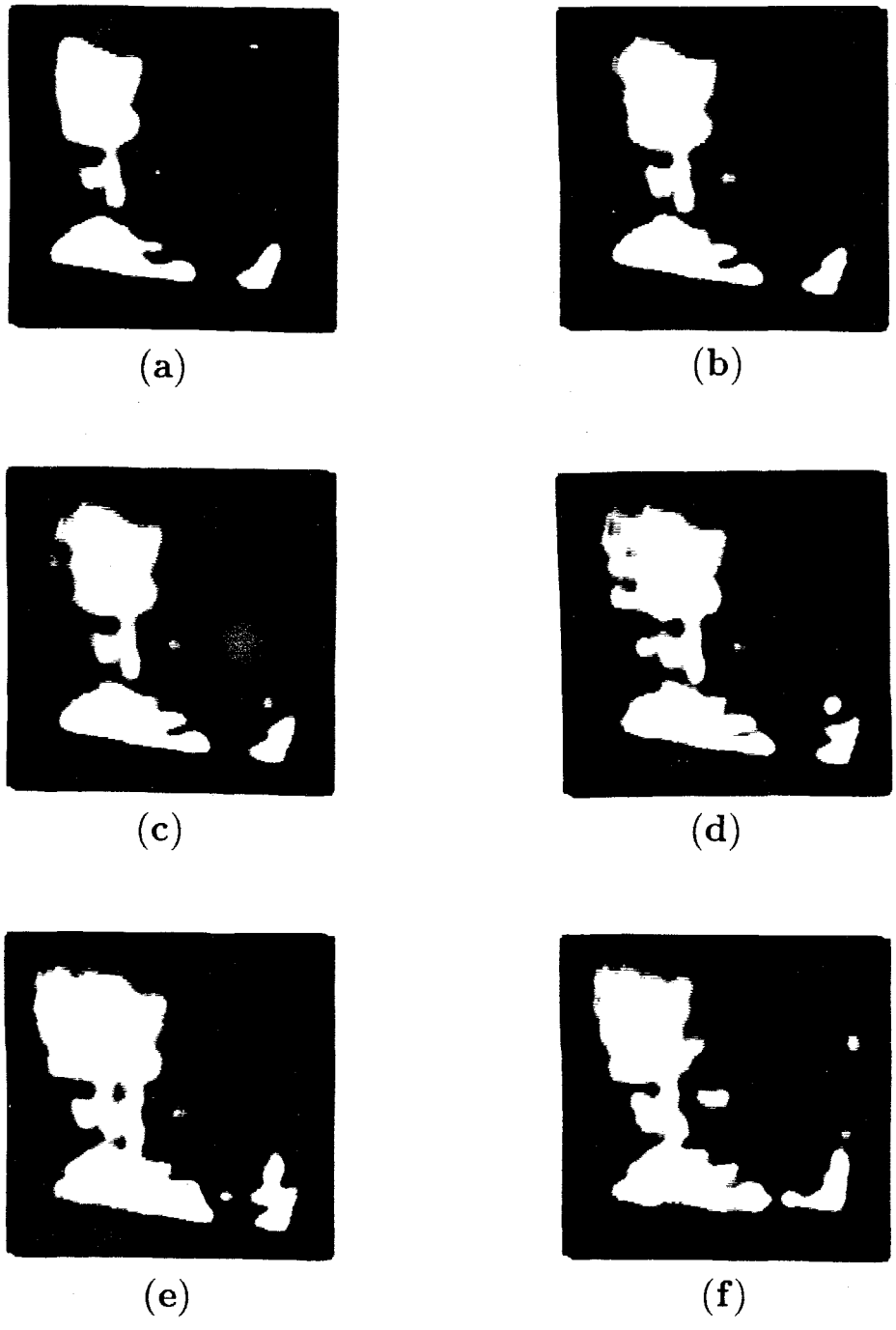
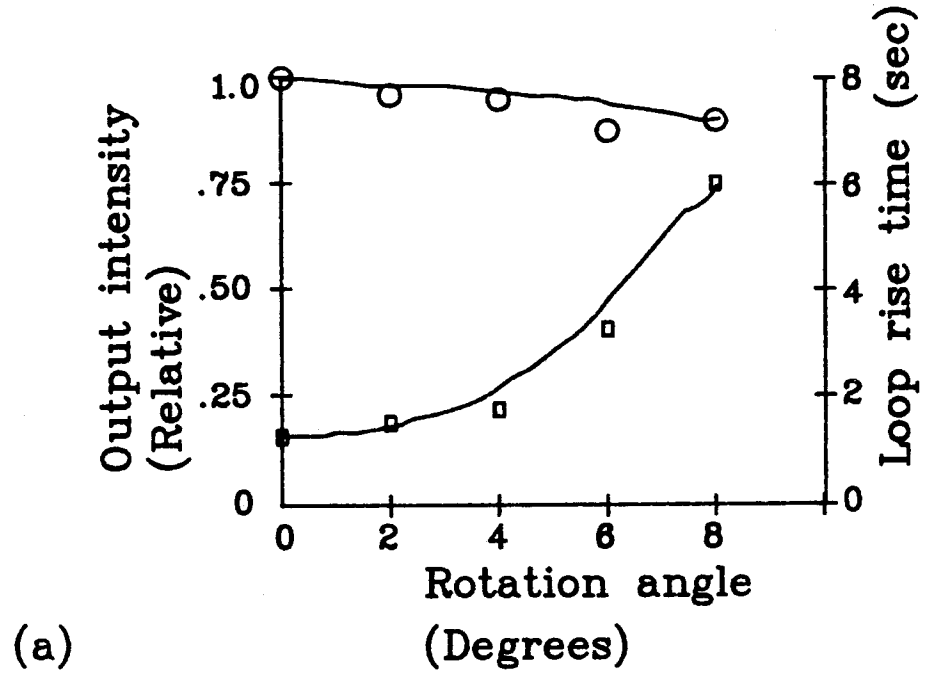


Figure 3.13: Retrieval of the complete Image from the Rotated Input. (a) The input at $t = 0$. (b) $t = 0^+$ (Loop closed). (c) $t = 1.8$ sec. (d) $t = 3.6$ sec. (e) $t = 4.8$ sec. (f) Input OFF.



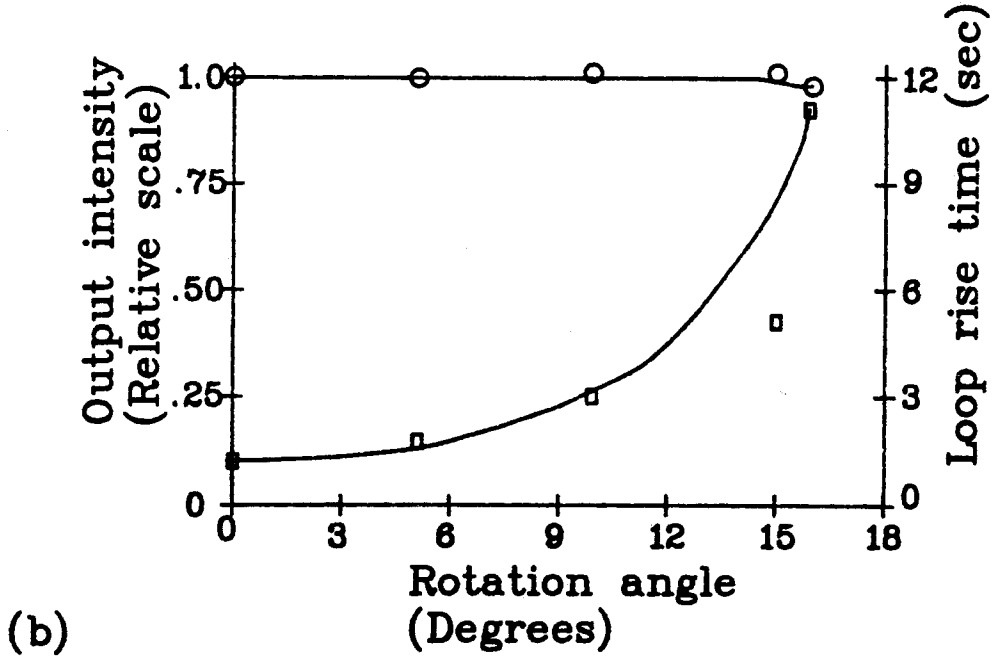


Figure 3.14: Rotation Invariance Property of the Optical Loop. (a) Optical gain = 10^4. (b) Optical gian = 10^5. (o: Output intensity. ◇: Loop rise time.)

to the correct image. This problem will be addressed in Section 3.5.

The third experiment on distortion tolerance is scale invariance. Fig. 3.15(a) shows the response of the memory when the scaled version of the face of a recorded image is presented to the system with the loop shutter **OFF**. This face is 85% of the size of the original stored image. After we turn **ON** the feedback shutter, the loop evolves to the stable state. The temporal sequence of this evolving behavior is shown in Fig. 3.15(b) to Fig. 3.15(e). We now remove the external input by turning **OFF** the input shutter. Fig. 3.15(f) shows that the loop remains latched to the original image. The convergence time in this case is 1.5 seconds. This is longer than what would be required if the input were the original image. However, the two cases give the same output intensity. When the input image is further scaled down to 70% of the original size we need to increase the neural gain from 10^4 to 10^5 for the loop to recognize the image. But this high gain results in low discrimination such that an input image that is not stored is also incorrectly recognized. These results are consistent with the dynamics and the invariance properties that we discussed above.

In principle, this loop is shift invariant since the images are stored in the Fourier-transform holograms. If one of the stored images appears at the input plane, there will be a bright spot at the correlation plane. If the input image shifts, the correlation peak will also shift to a corresponding position. However, if the peak shifts out of the pinhole position, then we lose the feedback signal and the image cannot be latched in the loop. Only when the shift equals an exact image spacing can the correlation peak pass through the pinhole to close the feedback path. But then the image that appeared in the window would be different from the previous one. Hence, this loop is shift-invariant only within the pinhole size. Without the pinholes the cross-correlation noise and the auto-correlation peak would be fed back to the loop together and the reconstructed images would not be recognizable. There is a

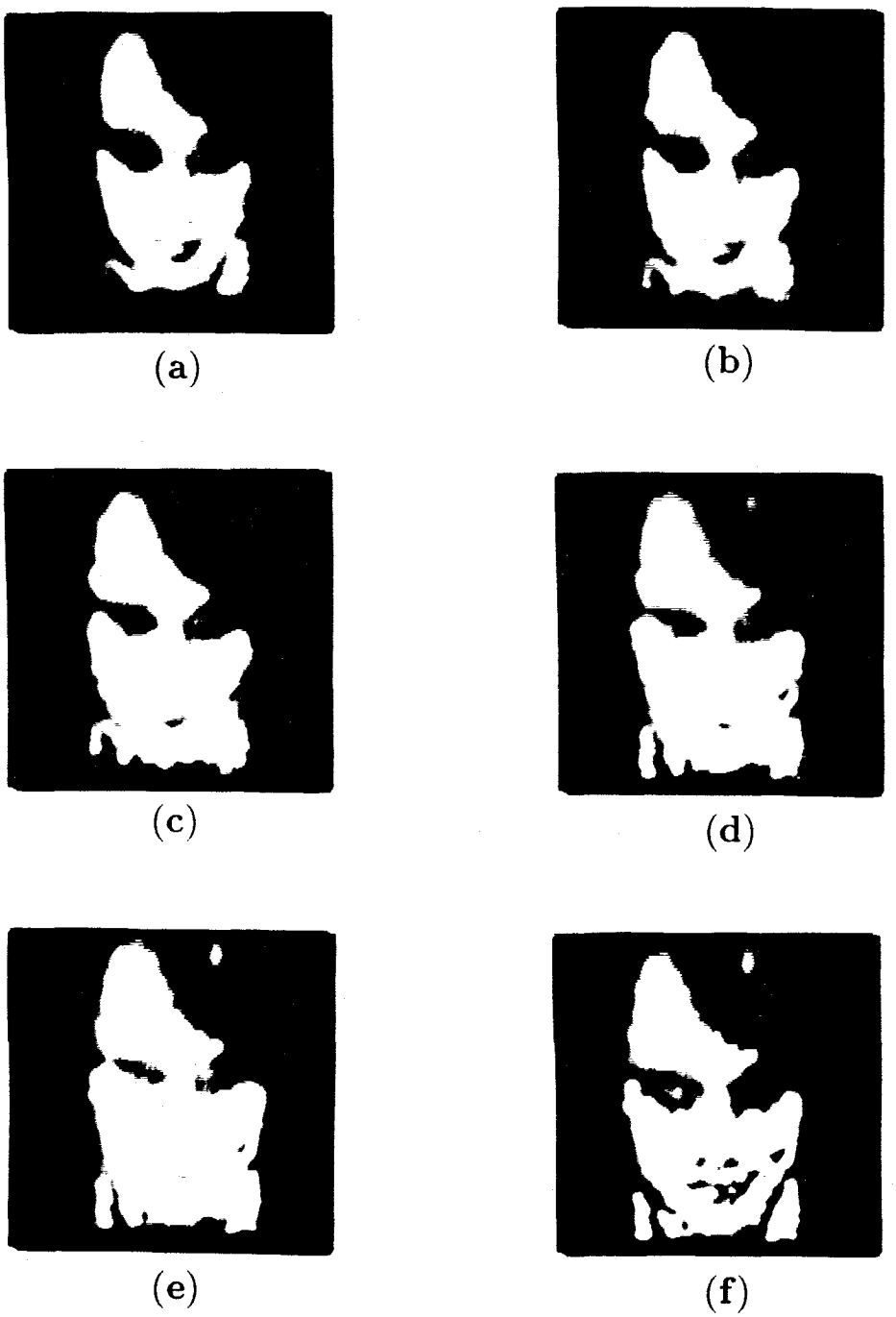


Figure 3.15: Retrieval of the Complete Image from the Scaled Input. (a) The input at $t = 0$. Image size = 85%. (b) $t = 0^+$ (Loop closed). (c) $t = 1.8$ sec. (d) $t = 3.0$ sec. (e) $t = 4.8$ sec. (f) Input OFF.

compromise between pinhole size and loop performance. Small pinholes allow good memory discrimination and sharp reconstructed images, but can also cut the signal to below the level that can be detected by the threshold device and reduce the shift tolerance of the system. The function of the pinhole array in this system might also be met by using a nonlinear spatial light modulator, in which case we can achieve full shift invariance [18, 19].

Fig. 3.16(a) shows the read-out image from the LCLV, which comes from an external input shifted away from its stored position. This shift moves its correlation peak so that it does not match the position of the pinhole. Thus, there is no feedback signal going through the loop. If we cut off the input image, the read-out image will die out with a characteristic time of about 50 to 400 ms, corresponding to the response time of the LCLV. Now we shift the input image around, trying to search for the correct position. Once the input image comes close enough to the correct position, the correlation peak passes through the right pinhole, giving a strong feedback signal superimposed with the external input on the neurons. The total signal then goes through the feedback loop and is amplified continuously until the neurons are saturated. Figs. 3.16(b) to (e) show the temporal sequence of this development from the moment that a sufficient portion of the correlation peak passes through the pinhole to the complete recall of the original image. Fig. 16(f) shows the image in the loop with the input turned **OFF** and the memory latched. In this experiment the pinhole diameters are 90 μm. As long as the correlation peak is close enough to the pinhole, the original image can be recalled. Depending on how far the input image is shifted from the original position and the optical gain of the neurons, the time required for the loop to reach a stable state is between two to several seconds. The lower curve of Fig. 3.17(a) shows the loop rise time as a function of shift when the neural gain is 10^4. The upper curve shows the output intensity as a function of shift. It shows that the shift tolerance for this particular image is 220 μm

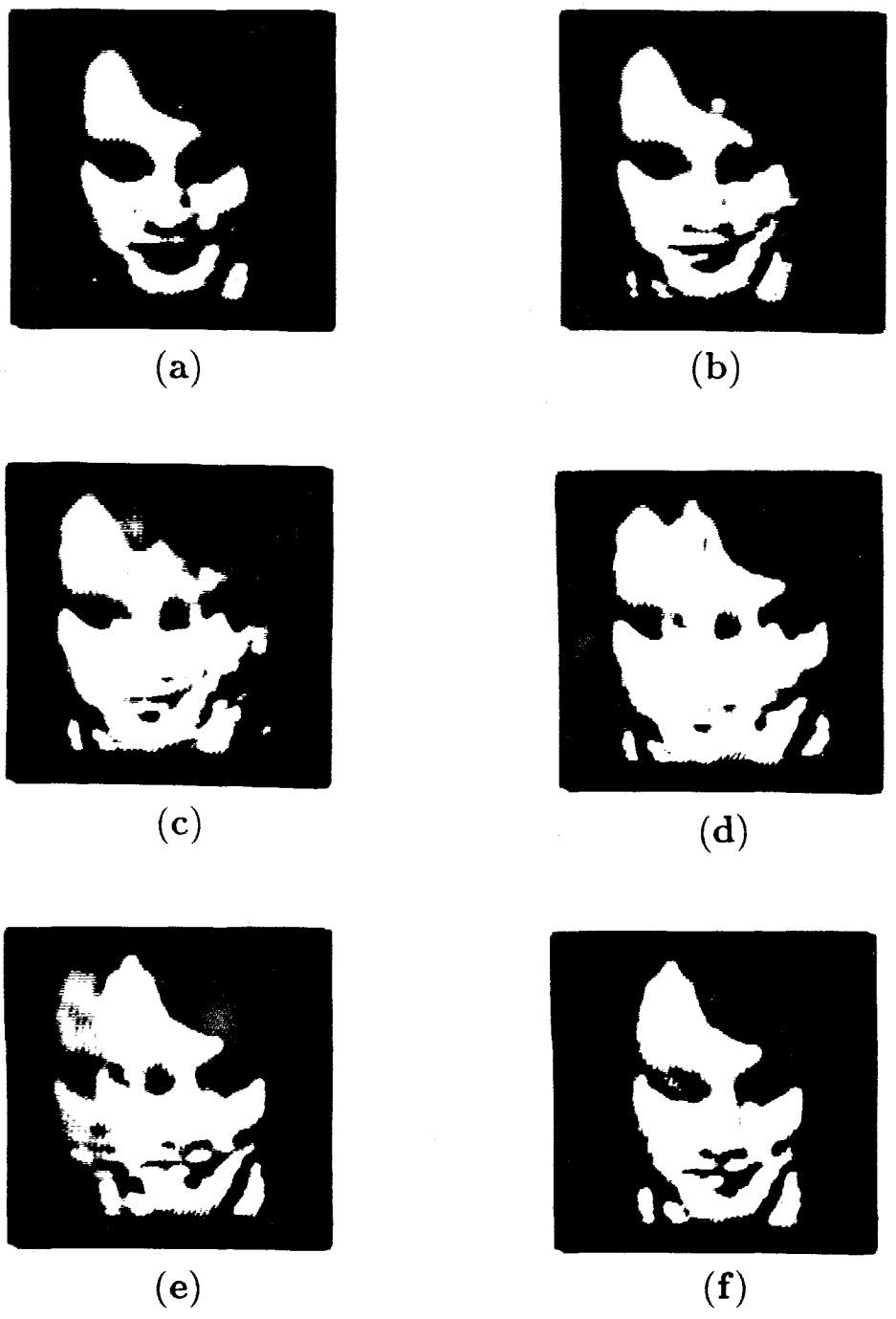


Figure 3.16: Retrieval of the Complete Image from the Shifted Input. (a) The input at $t = 0$. (b) $t = 0^+$ (Loop closed). (c) $t = 2.4$ sec. (d) $t = 3.0$ sec. (e) $t = 0.4$ sec after the input is OFF. (f) Stable state.

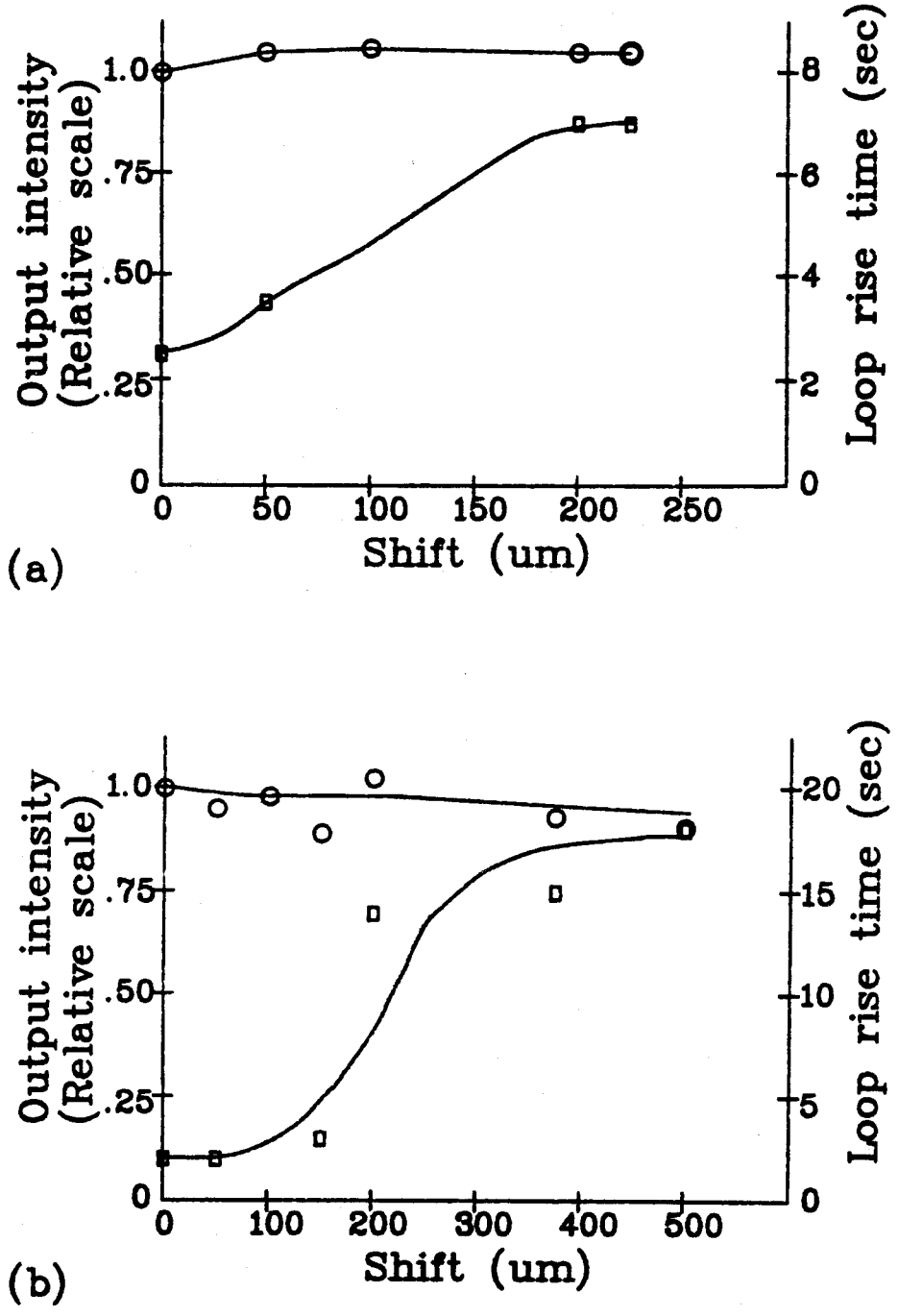


Figure 3.17: Shift Invariance Property of the Optical Loop. (a) Optical gain = 10^4. (b) Optical gian = 10^5. (o: Output intensity. $\diamond$: Loop rise time.)

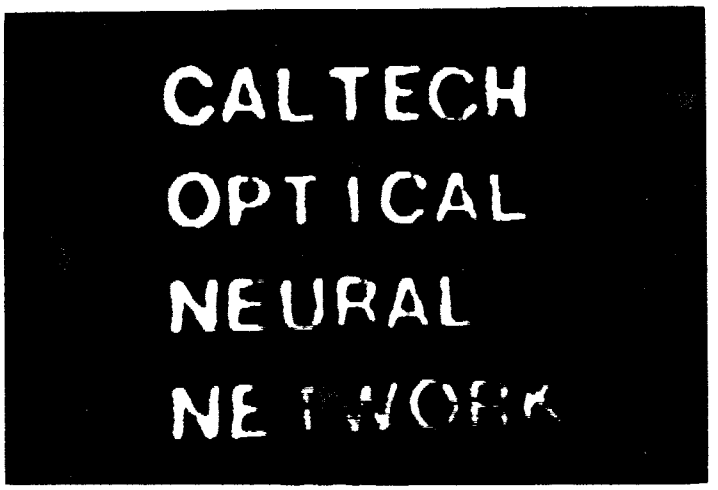
and that the loop converges to the image with the same intensity. If the shift is more than 220 μm, the feedback signal is too weak to trigger the loop and the output is zero. If we increase the neural gain by a factor of ten, then the shift invariance increases to about 500 μm. This is shown in Fig. 3.17(b). However, sidelobes are also amplified in the loop; hence the image may be uncorrectly recognized.

The dynamic and invariance properties of the associative loop shown above imply that the associative memory has error-correcting capability, since the input does not have to be the same as the original memory to be recognizable. We show this capability particularly by using four words as memory. Fig. 3.18(a) shows the four word s recorded in the Fourier-transform holograms of our system. Fig. 3.18(b) shows an input word that has three spelling errors. However, it has four correct letters, **O, P, I, A**, at the correct positions. Hence the correlation of (**OP4I&A**) with the memory gives sufficient feedback signal to trigger the loop iterations. Fig. 3.18(c) shows the word recalled from the loop, superimposed on the input. Fig. 3.18(d) shows the loop latched to the correct word after the input is turned **OFF**. If we increase the neural gain, the loop would be able to recognize more erroneous words. But then it will lose the discrimination capability because a completely different word would also induce enough feedback for the word to be recognized as one of the stored words.

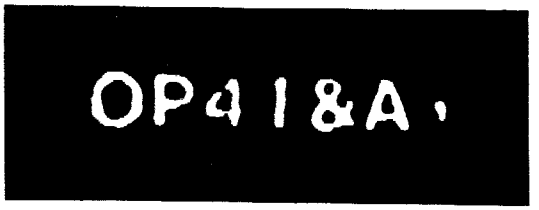
3.5 Trade-Off Between Distortion Tolerance and Discrimination Capability

The experimental results shown in the above subsection demonstrates the distortion-invariance capability of the associative loop. The input images do not have to match the memory exactly. Furthermore, by raising the neural gain, no matter how much we change the initial condition by rotating, shifting, and scaling the input image,

(a)



(b)



(c)

(d)

Figure 3.18: Error-Correction Capability of the Loop. (a) The stored images. (b) External input with errors. (c) Feedback image superimposes with the input image. (d) The stable state of the loop after the external is OFF.

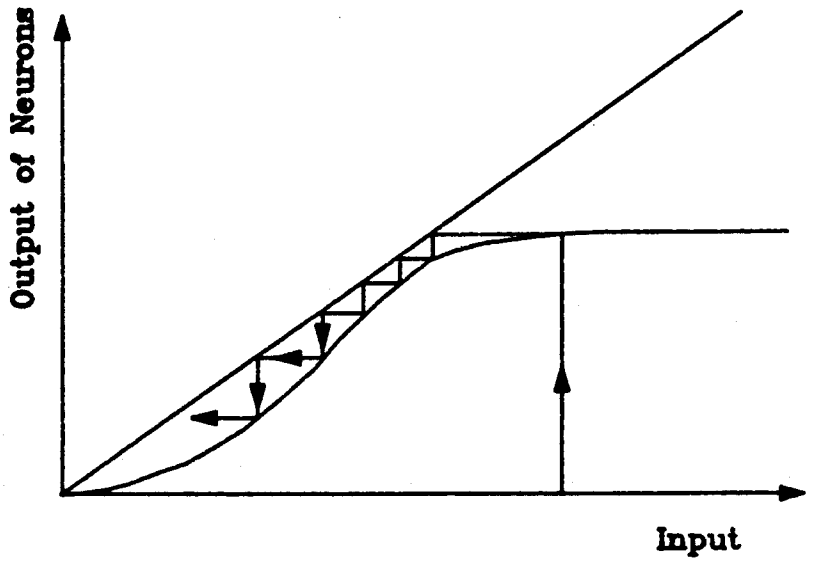


Figure 3.19: Iteration Map of Low Optical Gain.

the loop can always be made to produce an image as a stable state. But the ability to correctly recognize a stored image from a distorted input and the discrimination capability, i.e., the ability to distinguish images from one another, compromise each other. If there is too much gain, then just shining a flashlight at the input of the system causes it to lock on to one of its stable states. If the gain is set too low, then even an input that is a slightly distorted version of one of the stored images is not recognizable. In particular, there are two parameters under our control that can affect the gain in the loop: The gain of the neurons and the size of the pinholes.

The importance of selecting an optimal neural gain can be explained by the loop iteration map. Fig. 3.19 shows the iteration map of a low gain loop. In the figure there is no intersection between the gain curve and the loss line, and the gain curve is always below the loss line. Therefore, no matter how well the input image matches the stored images and no matter how bright the correlation peak is, the output of the neuron is always below the loss line. Hence, the loop signal becomes weaker and

weaker in successive iterations until finally it decays to zero. Fig. 3.20(a) shows the response of the loop to an input image, with the feedback shutter **OFF**. Fig. 3.20(b) and (c) shows the output after the loop is closed. Fig. 3.20(d) and (e) shows the loop output after the input image is **OFF**. The loop image decays because the neural gain is too low.

On the other hand, too much neural gain also causes problem. Fig. 3.21 shows the iteration map with high gain. In this figure the gain curve is higher than the loss line and they intersect at a low value. Hence, the loop has a low threshold. Therefore, a small correlation signal is enough to trigger the loop to evolve. As shown in the above subsection, this loop can tolerate more image distortions such as rotation, scaling, shift, etc., and can still recognize them. This means that the loop has a very large radius of attraction for each memory state. However, it also means that its discrimination capability to distinguish different images is poor. It is quite possible that the loop recognizes other images as one of the stored images. Fig. 3.22(a) shows an image that is not stored in the memory. Fig. 3.22(b) to (d) show that the unfamiliar image triggers the loop to evolve after the feedback shutter is turned **ON**. Fig. 3.22(e) shows that after the input image is turned **OFF**, the loop is locked to one of the stored images. This shows the loop making an incorrect recognition.

The gain required to sustain the iteration loop is determined by the loop loss. The factors in this system that decide the loss are the pinhole size, the hologram diffraction efficiency, and the reflections from optical components. Among them the pinhole size is the key factor since the other components are generally fixed. We use Fig. 3.3 as an example. Let $f_i(x, y), i = 1, 2, 3, 4$, represent the images of the letters **A, B, C, D**, respectively, and let the pinhole size be W. Then the reconstructed

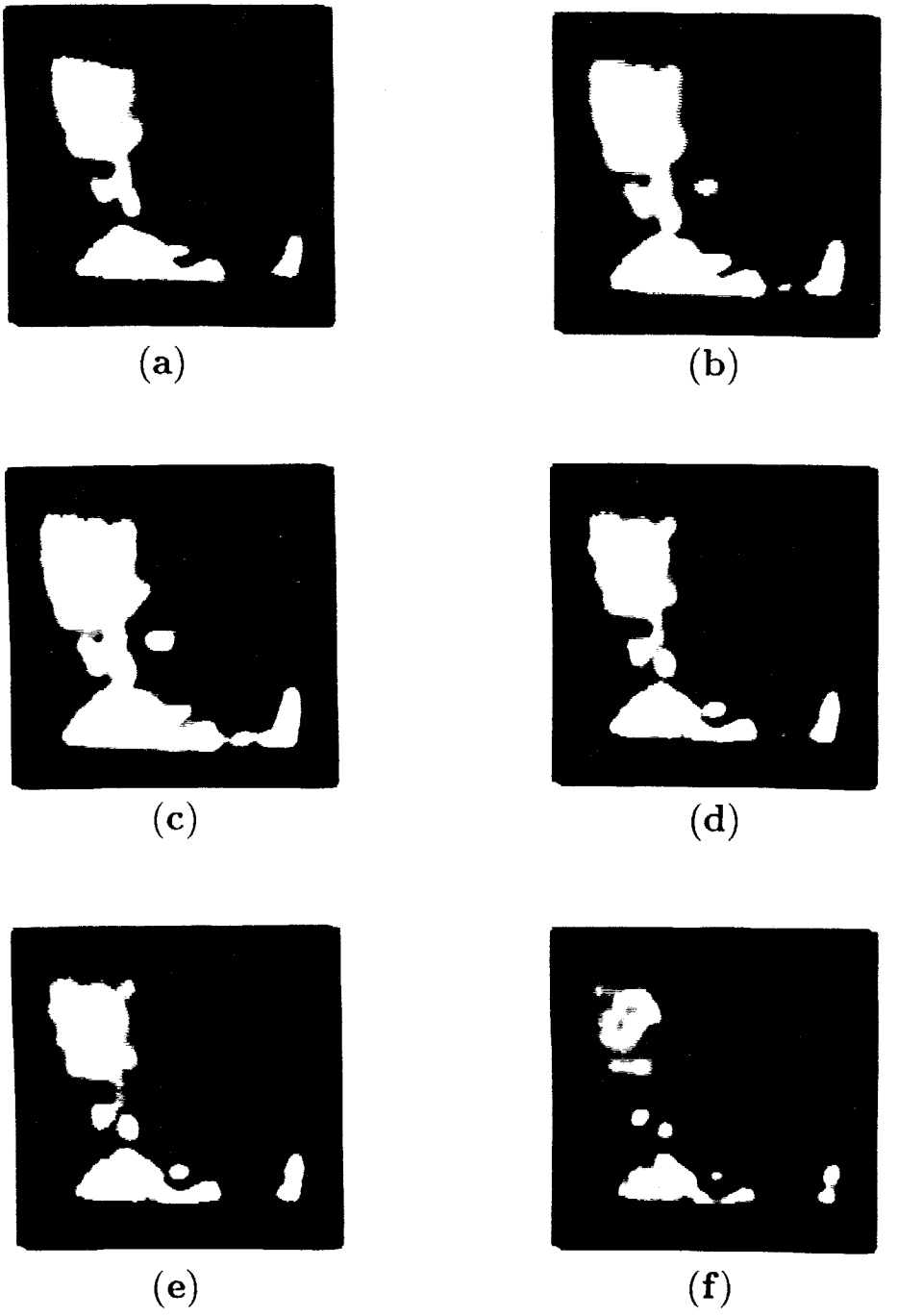


Figure 3.20: Loop dynamics of Low Optical Gain. (a) The input at $t = 0$. (b) $t = 0^+$ (Loop closed). (c) $t = 3$ sec. (d) Input OFF. (e) $t = 1.2$ sec after the input is OFF. (f) $t = 1.8$ sec after the input is OFF.

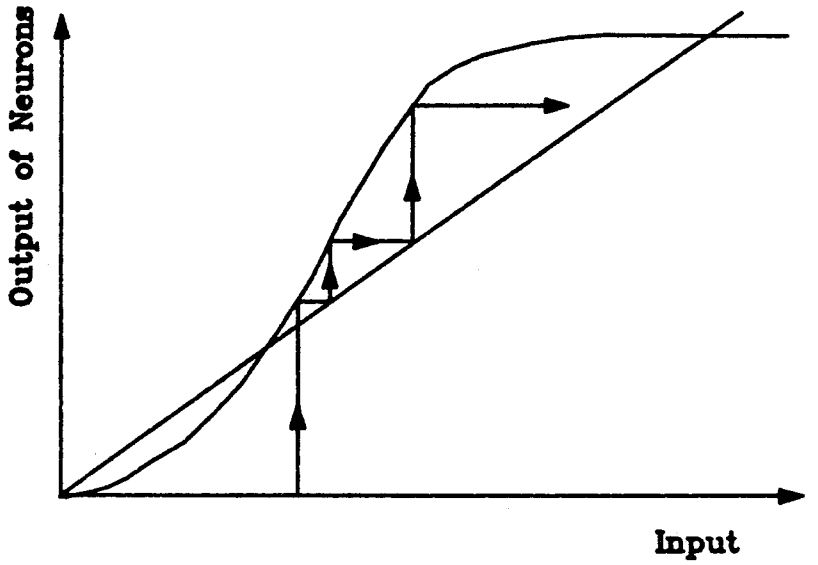


Figure 3.21: Iteration Map of High Optical Gain.

images in the window at P_1 can be shown to be

$$\sum_{i=1}^{4}[g_{1i}(x,y)\text{rect}(\frac{x}{W})\text{rect}(\frac{y}{W})] * f_i(x,y) \tag{3.17}$$

where $\text{rect}(\frac{x}{W})\text{rect}(\frac{y}{W})$ represents the finite size of the pinholes, * represents the convolution operation, $g_{11}(x,y)$ the auto-correlation of **A**, and $g_{1i}, i \neq 1$, the cross-correlations of **A** with **B, C, D**, respectively. We see that the images are blurred by the finite dimension of the pinholes. Decreasing W gives better image quality, but we need to increase the gain of the neurons to compensate for the loss caused by the small pinholes. At the other limit, if the pinhole size is increased, we do not need very high-gain neurons but the image quality deteriorates. Fig. 3.23 shows the effect of the pinhole size on the stable-state loop images. In the limit where W becomes infinitely large, the reconstructed image in the window at P_1 becomes a superposition of all the stored images, each approximately equally strong, and

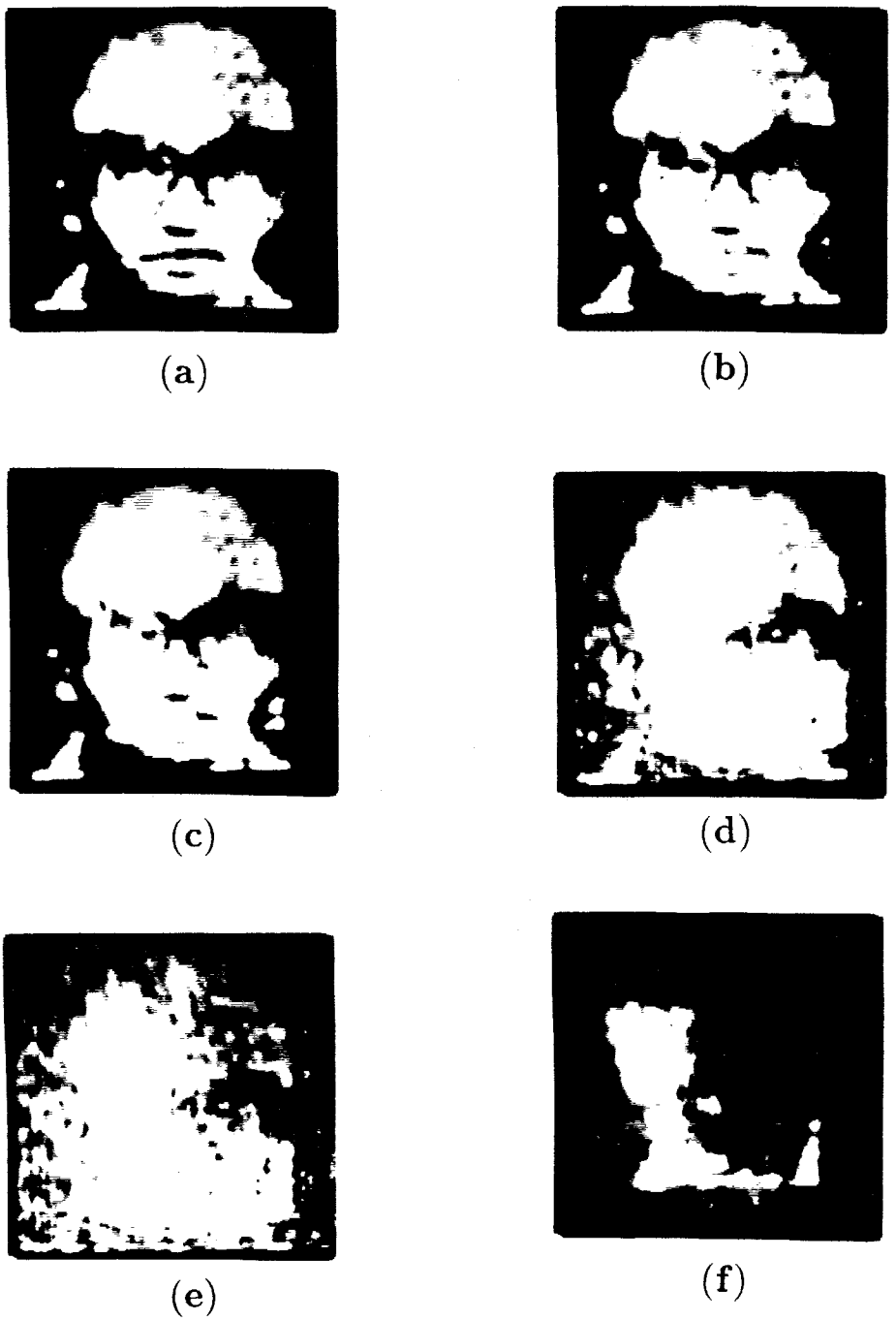


Figure 3.22: Loop dynamics of High Optical Gain. (a) The input at $t = 0$. (b) $t = 0^+$ (Loop closed). (c) $t = 1.2$ sec. (d) $t = 1.8$ sec. (e) The input is OFF. (f) Stable state.

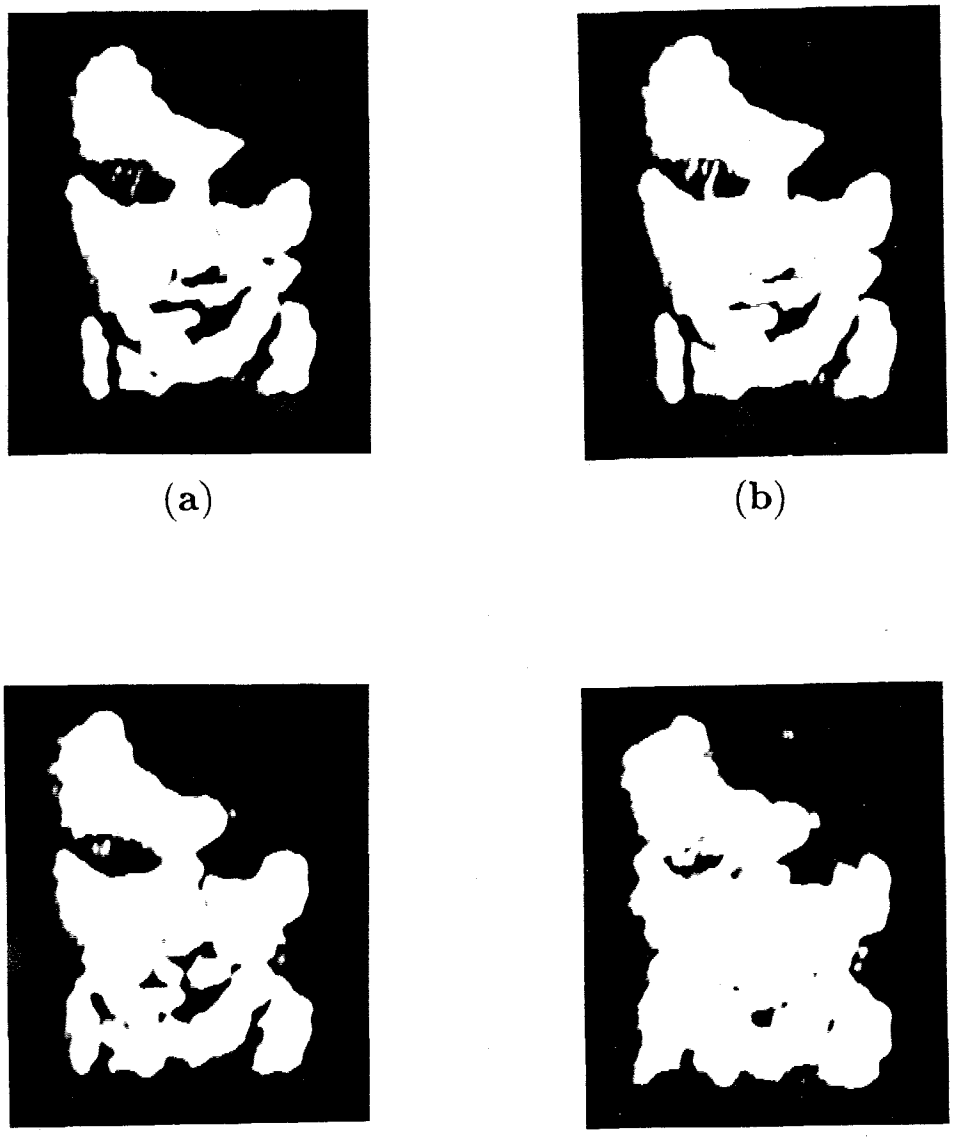


Figure 3.23: Output Image for Different Pinhole Sizes. (a) 40 μm. (b) 90 μm. (c) 180 μm. (d) 400 μm.

severely blurred. Fig. 3.24 shows the temporal sequence of the loop signal under the condition of infinite pinhole size. It is seen that final image in the loop is totally unrecognizable.

Thus, there is an optimum pinhole size and an optimum neural gain. Fig. 3.25 shows the experimental results of minimum gain required and maximum gain allowable for the loop to sustain a stable memory as a function of pinhole size. Below the minimum gain the loop can not recognize any image in the sense that once the external input is cut off, the loop activity decays to zero. Above the maximum gain the loop loses discrimination capability such that any input image, even a flashlight, will trigger the loop into a stable state. Note that the minimum gain increases when the pinhole size is increased to more than 250 μm. This is because the reconstructed images are blurred so much that the correlation peaks are weakened and the losses in the loop are increased. Fig. 3.25 shows that the optimum pinhole size in this system is in the range of 70 μm to 150 μm. We choose 90 μm for most of the experiments. Although this measurement is particular for the images we used, the above behavior is true in general.

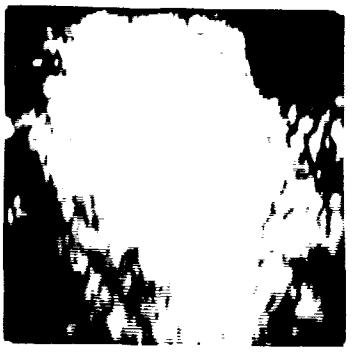
3.6 Neural Network Model for the Memory Loop

The architecture of the optical associative memory that we presented in the preceding sections is a Hopfield-type network. In the system, the neurons are simulated by the LCLV, and the interconnections between the neurons are achieved by holographic gratings. In the following, we present an analysis of the neural network model of the optical associative-memory loop. Based on the model, we will discuss the convergence properties of the system and the stability of the stored memories.

Recall from Eq. 3.16 that the feedback signal of each iteration in the optical loop

(a)

(b)



(c)

(d)

Figure 3.24: Loop Behavior without the Pinholes. (a) The input at $t = 0$. (b) $t = 0^+$. (Loop closed). (c) $t = 0.6$ sec. (d) $t = 2$ sec.

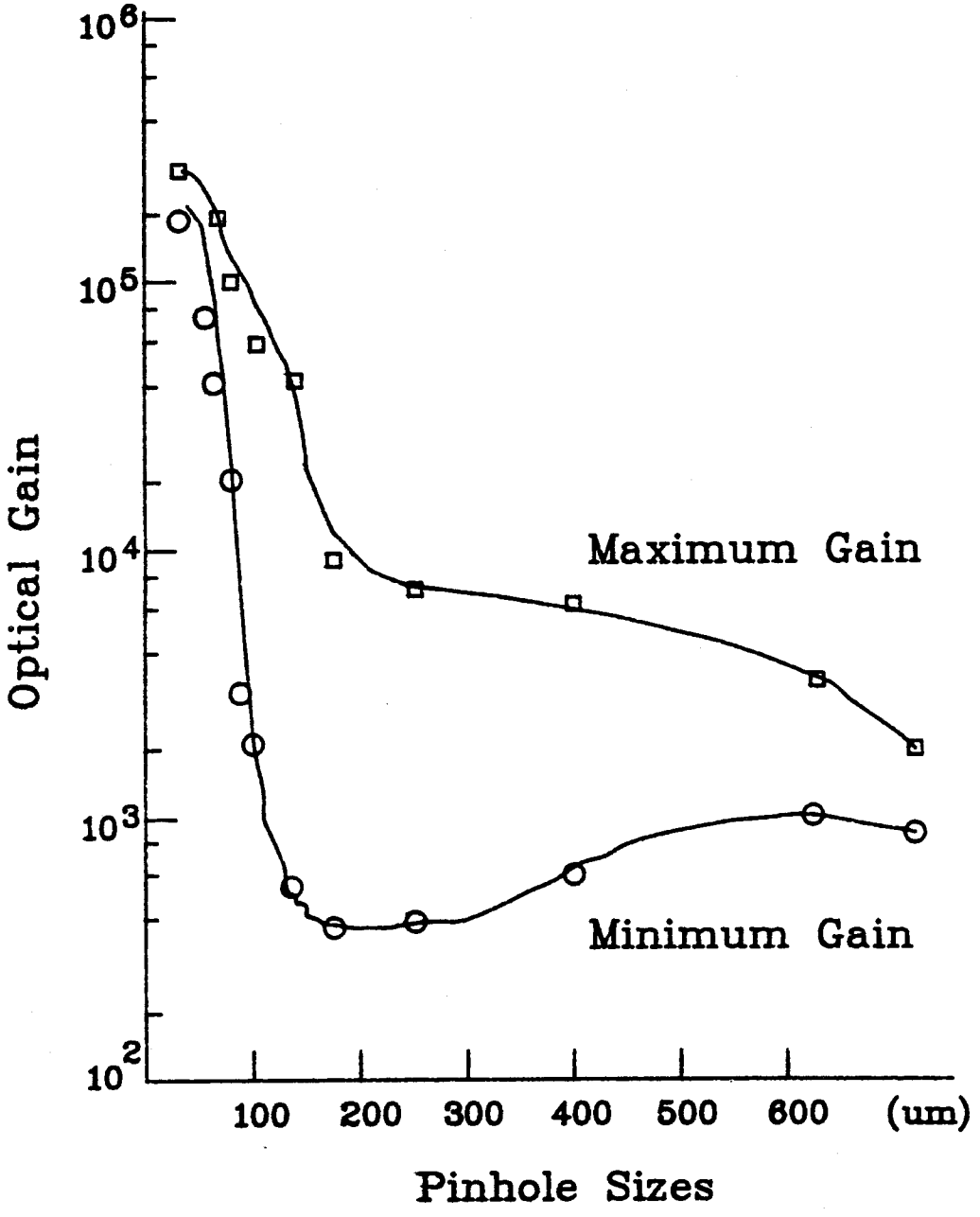


Figure 3.25: Optimum Pinhole Size and Optical Gain.

can be written as

$$f^o(x,y,t) = \sum_{m=1}^{M} (f \star f^m)(0,0,t)\ f^{m*}(x,y), \tag{3.18}$$

where $f^m(x,y)$ ($m = 1 \ldots M$) are the stored images, $f(x,y)$ is the input image to the LCLV, $\star$ is the cross-correlation operator, and f^{m*} is the complex conjugate of f^m. Eq. 3.18 shows that the feedback signal in the loop is the superposition of the reconstructed images from the second hologram, each being weighted by the cross-correlation of the input with the images stored in the first hologram. In the optical system, we make the first hologram H_1 with high-pass characteristics. The high-pass Fourier-transform hologram was obtained by adjusting the ratio of the recording intensities of the reference beam and the object beam such that the high-frequency part has good modulation depth while the low-frequency parts were overexposed. Thus, in reconstruction there is no diffraction from low-frequency gratings, and only high-frequency components reconstruct the image. The reconstructed image contains only the edges of the original image, as was shown in Fig. 3.7(b). The characteristics of the high-pass hologram can be described by subtracting the low-frequency portion from the original spectrum, which can be represented approximately by removing the dc signal. Using the high-pass hologram as H_1 in the optical loop, the cross- correlation term $f * f^m$ in Eq. 3.18 now should be replaced by the convolution of the input $f(x,y)$ with the stored images $g^m(x,y)$, i.e., $f \star g^m$, where $g(x,y)$ is the high-pass version of $f(x,y)$. Since $g^m(x,y)$ is the original image with the dc level removed, it can be described by

$$g^m(x,y) = f^m(x,y) - \iint f^m(\xi,\eta)\ d\xi\ d\eta, \tag{3.19}$$

where the integration is performed over the finite size of the images.

Instead of taking continuous functions $f^m(x,y)$, $f(x,y)$, etc., we will approximate by sampling discrete points (pixels), indexed by i, j, etc. (The number of the

sampled points N should be greater or equal to the SBP of the optical system.) We can approximate $(f \star g^m)(0,0)$ by summation of the inner product

$$(f \star g^m)(0,0) = \sum_{l=1}^{N}\sum_{k=1}^{N} f_{lk} g_{lk}^m \tag{3.20}$$

and write

$$g_{lk}^m = f_{lk}^m - \frac{1}{N^2}\sum_{i,j=1}^{N} f_{ij}^m. \tag{3.21}$$

The feedback signal can thus be written as

$$\begin{aligned} f_{ij}^{(o)} &= \sum_{m=1}^{M}\Big[\sum_{l=1}^{N}\sum_{k=1}^{N} f_{lk} g_{lk}^m\Big] f_{ij}^{m*} \\ &= \sum_{l=1}^{N}\sum_{k=1}^{N} w_{ij;lk}\, f_{lk}, \end{aligned} \tag{3.22}$$

where

$$\begin{aligned} w_{ij;lk} &= \sum_{m=1}^{M} g_{lk}^m f_{ij}^{m*} \\ &= \sum_{m=1}^{M}[f_{lk}^m - (\frac{1}{N^2}\sum_{l,k=1}^{N} f_{lk}^m)] f_{ij}^{m*}. \end{aligned} \tag{3.23}$$

Note that the matrix W is nonsymmetric and that its diagonal terms are not zero in this case.

For simplicity of discussion, we use one-dimensional signals in the following analysis. The extension to the two-dimensional case is straightforward. We also assume that the signals are real. In fact, this may not be true in the real system, because the nonuniformity and phase distortions of the optical components and the LCLV may cause signals to become complex. However, the analysis then becomes difficult if we try to take this into account. Furthermore, the intensity of the signal is detected in each iteration, and the phase information does not accumulate in the system; thus, our assumption is reasonable. Following these arguments, the w_{ij} for the 1-D case is modified as

$$w_{ij} = \sum_{m=1}^{M}(x_j^m - a_m)x_i^m, \tag{3.24}$$

where

$$a_m = \frac{1}{N}\sum_{j=1}^{N} x_j^m \tag{3.25}$$

is the average level of image m. The feedback signal shown in Eq. 3.22 can then be written as

$$y_i = \sum_{j=1}^{N} w_{ij} x_j. \tag{3.26}$$

Therefore, the feedback signal after one cycle and before thresholded by the neurons can be obtained by inserting Eq. 3.24 into 3.26, which gives

$$\begin{aligned} y_i &= \sum_{j=1}^{N}\Big(\sum_{m=1}^{M}(x_j^m - a_m)x_i^m\Big)x_j \\ &= \sum_{m=1}^{M}\Big[\Big(\sum_{j=1}^{N} x_j^m x_j\Big) - \Big(\sum_{j=1}^{N} x_j\Big)a_m\Big]x_i^m. \end{aligned} \tag{3.27}$$

Note that in this case the x_i^m and x_j are unipolar, i.e., 0 or 1; whereas y_i can be biploar because w_{ij} is bipolar. The above expression can be written in a matrix form,

$$\mathbf{y} = \sum_{m=1}^{M}\Big[\big(\mathbf{x}\cdot\mathbf{x}^m\big) - \Big(\sum_{j=1}^{N} x_j\Big)a_m\Big]\mathbf{x}_m. \tag{3.28}$$

The signal $\mathbf{y}$ is fed back at the neuron plane and is thresholded by the neurons to give the signal for the next iteration. Thus, the new signal for the next iteration is

$$\mathbf{x} = g(\mathbf{y}). \tag{3.29}$$

This process then repeats until the loop reaches an equilibrium state. Clearly, if we want to investigate the equilibrium states and the stability characteristics, we should solve the dynamic equations of the system. However, before we go to that step, we can qualitatively predict the performance of the loop simply by inspecting the physical meaning of each term of the feedback signal.

As is seen in Eq. 3.28, the total feedback signal is the weighted sum of the stored images $\mathbf{x}^m$. The weight of each $\mathbf{x}^m$ is determined by the two terms in the square bracket. The term $\mathbf{x}\cdot\mathbf{x}^m$ is the cross-correlation of the input $\mathbf{x}$ and the stored image

$\mathbf{x}^m$, while $\sum_{j=1}^{N} x_j$ is the dc level of the signal $\mathbf{x}$, and a_m is the dc level of the stored image $\mathbf{x}^m$. The subtraction of the product of the dc levels from the correlation signal gives the weight. Thus, only the pixels that have correlation with the stored $\mathbf{x}^m$ sufficiently larger than the dc levels contribute a significant component to $\mathbf{y}$. Therefore, only the strongest stored image component that appears in $\mathbf{x}$ will be enhanced in the feedback, while weaker components are suppressed by subtracting the average image. This argument suggests that the loop will work as an associative memory.

In the case of the optical loop, the neural gain is the main factor that determines whether the loop can recall a correct memory state. The modeled network shows the same behavior. To illustrate this point, we perform simulations with different gains. Fig. 3.26 shows an example of computer simulation. There are three patterns stored in the high-pass network; each pattern has a 30 pixels. In the figure, the height at each position represents the signal strength of that pixel. When a partial $\mathbf{x^2}$ is input into the system, it converges to a stable state that is closest to the $\mathbf{x^2}$ after 160 iterations. As the gain is reduced, the loop takes more iterations to reach the steady state. However, when the gain is reduced to below a certain level, the loop can no longer sustain the loop signal and the signal gradually decays to zero. This is illustrated in Fig. 3.27. On the other hand, as the gain is increased, the loop evolves to a steady state in fewer steps. But if the gain is too high, the loop converges to a mixed state of the stored images, which may be very distorted and unrecognizable. Fig. 3.28 shows simulations under high-gain conditions. The results of these simulations show that the modeled network has the same characteristics as the optical loop: It has the capability of recognizing distorted images, but there is a trade-off between distortion tolerance and discrimination capability, which is governed by the neural gains.

Note that in the simulation of Fig. 3.26 the steady-state image of the loop is

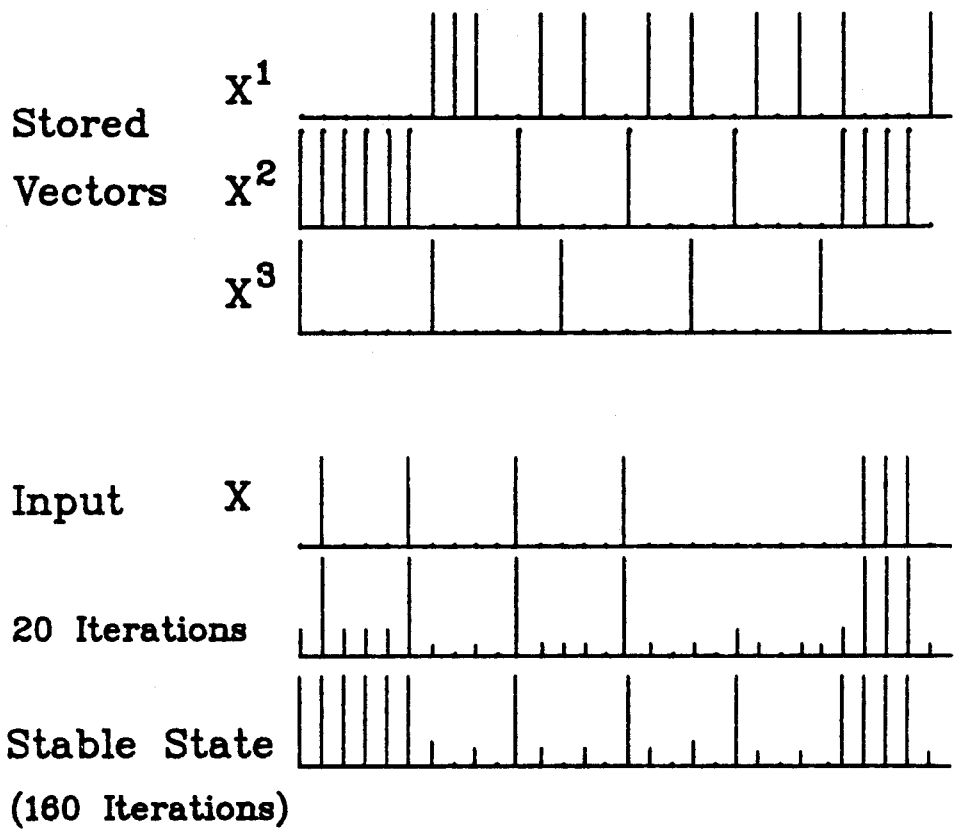


Figure 3.26: The dynamics of the high-pass loop with a good gain.

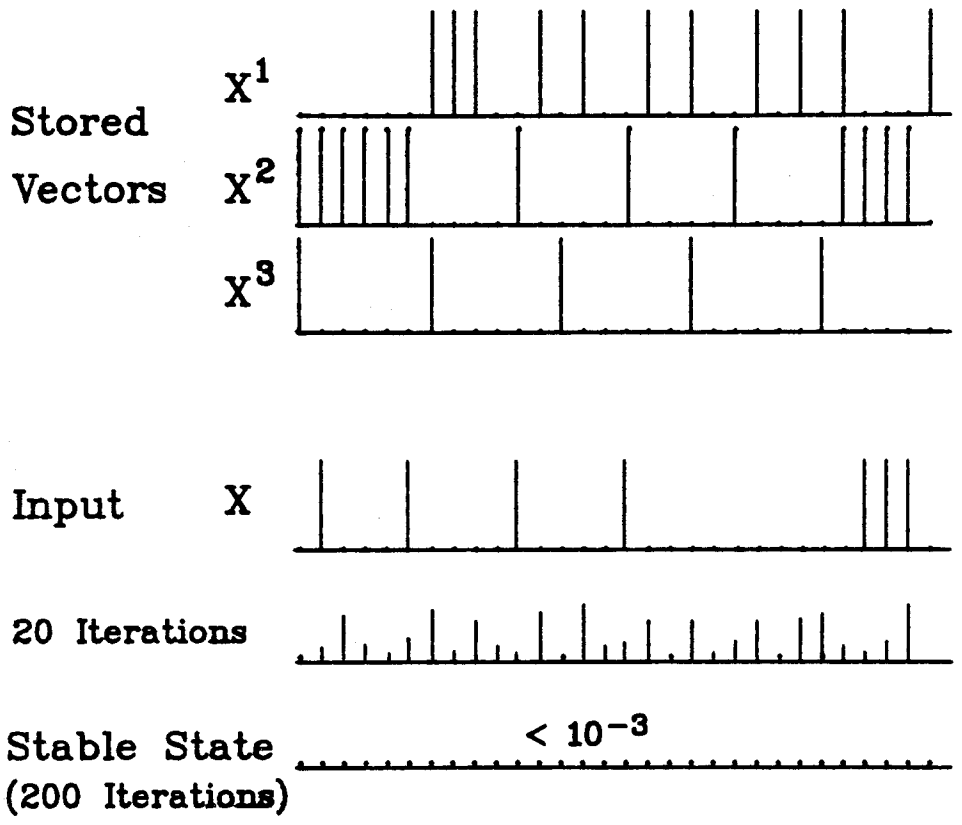


Figure 3.27: The dynamics of the high-pass loop with a low gain.

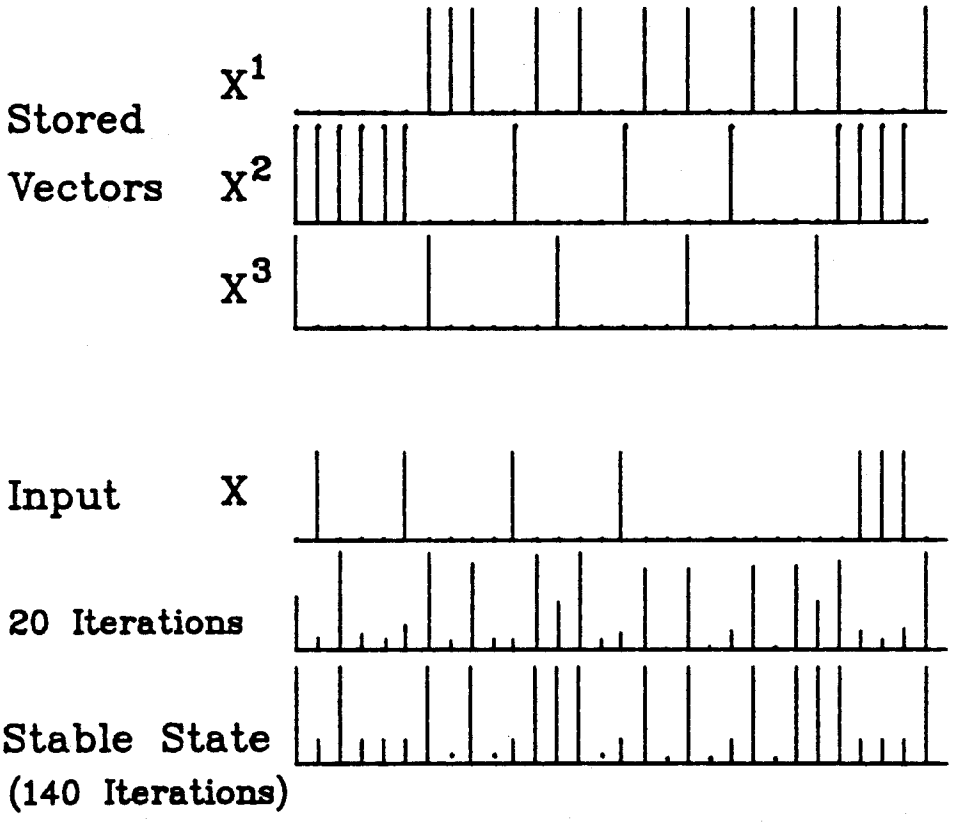


Figure 3.28: The dynamics of the high-pass loop with a high gain.

slightly different from the original images. But we still consider it a correct recall. Here the *correct* memory means that it has a *similar* shape as the original image, although there is a slight distortion and the signal level is weaker. The reason for that distortion is easily understandable by inspecting Eq. 3.27. We see that the feedback contains not only the autocorrelation signal, but also the cross-correlations with all other memories. Depending on the values of the cross-correlation and the gain function, some of the crosstalks will be detected by the neurons and survived in the loop; thus, the total output is distorted. The higher the neural gain the worse the distortions. We will return to this point later when we examine the steady-state solution of the dynamic equations of the system.

An interesting question that we want to address is whether the high-pass hologram is necessary and how it affects the system performance. Suppose the first hologram is not a high-pass version, then the interconnection strength can be obtained by setting $a_m = 0$ in Eq. 3.24. We get

$$w_{ij} = \sum_{m=1}^{M} x_i^m x_j^m. \tag{3.30}$$

Similarly, the feedback signal can be obtained by setting $a_m = 0$ in Eq. 3.28,

$$\mathbf{y} = \sum_{m=1}^{M} \left(\mathbf{x} \cdot \mathbf{x}^m\right) \mathbf{x}_m. \tag{3.31}$$

The feedback signal now is simply the superposition of the stored images each being weighted by the cross-correlation of the input image and the stored images. The strongest correlation gives the strongest feedback signal and it determines the state to which the loop will most possibly converge. Hence the loop still works as an associative memory. However, because there is no mechanism to suppress the crosstalks, noise is also easily picked up by the neurons. The discrimination capability of the loop will be poorer than the high-pass loop. Fig. 3.29 shows an example of the computer simulation. The images used in the high-pass simulations are used here.

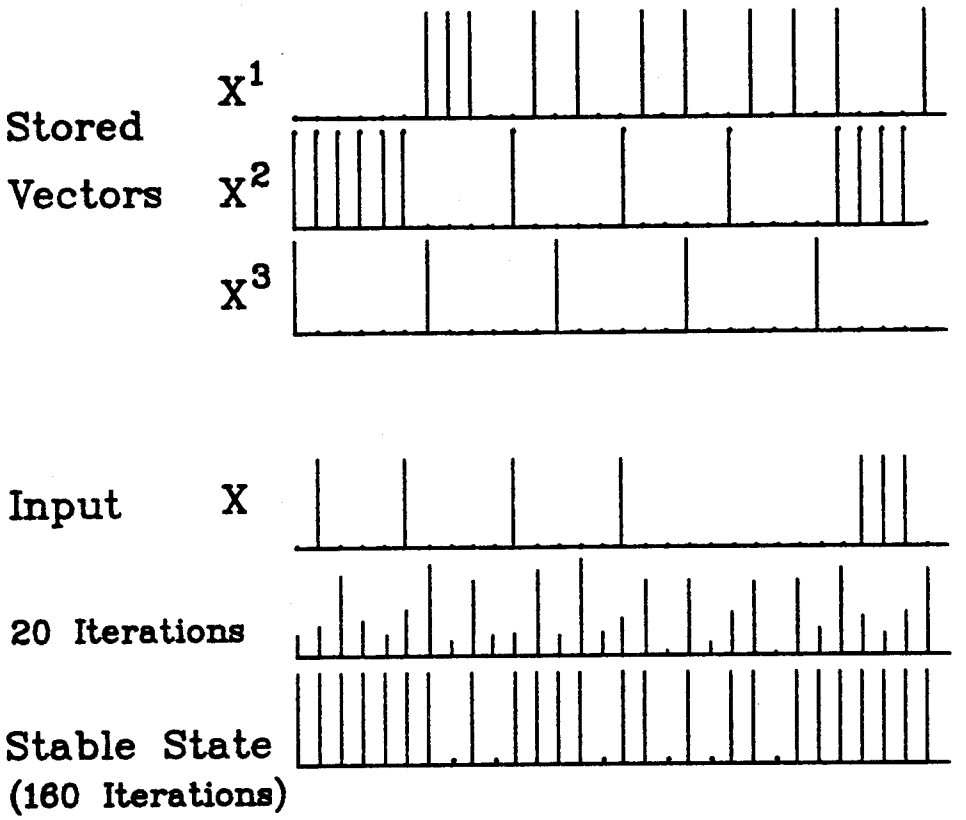


Figure 3.29: The dynamics of the loop without the high-pass filter.

Simulations show that as long as the stored images have crosstalks, then the low-pass loop always converges to a mixed state. There is only a very narrow range of the gain where the system gives the correct output. Therefore, the high-pass hologram is very crucial in the performance of the loop.

To this point, we have developed a neural network model of the optical associative loop, and we have investigated its characteristics with computer simulations. Results show that the model matches well with the optical loop. However, the above treatments are not very rigorous and do not show the dynamic behavior of the system. Therefore, it is not completely satisfactory. In general, only the solution of the dynamic equations can fulfill our purpose. As we have seen in Chapter 2, an optical neuron is simulated by one pixel of the LCLV, which gives response x_i to its input y_i. The dynamics of the neuron is described by Eq. 2.3. For the sake of convenience, we rewrite it in the following

$$\frac{dx_i}{dt} = -x_i + g(y_i), \qquad i = 1 \ldots N, \tag{3.32}$$

where $g(y_i)$ is a nonlinear function describing the neuron response. To complete the feedback loop, we substitute the expression of Eq. 3.26 into Eq. 3.32, we obtain

$$\frac{dx_i}{dt} = -x_i + g\Big(\sum_{j=1}^{N} w_{ij} x_j\Big), \qquad i = 1 \ldots N \tag{3.33}$$

where w_{ij} is the high-pass version of the interconnection strength given by Eq. 3.24. This is the equation describing the dynamics of the optical loop. In what follows, we will discuss the system's, dynamics using a geometrical method [20]. The main idea of the method is that corresponding to each set of stored memories we first define a new vector space; then the dynamic equations are transformed consequently into this space and the dynamics can be observed as a phase flow on the hyper-surface. We assume that the stored images $\mathbf{x}^1 \ldots \mathbf{x}^M$ are linearly independent. This is reasonable since the number of pixels in the image, N, is usually much larger than M, and we are

not considering the trivial case where one of the stored images is a linear combination of the others. Since the stored images are linearly independent, they span a vector space $\mathbf{V_1}$. Then, we can form a vector space $\mathbf{V_2}$ that is orthogonal to $\mathbf{V_1}$ such that the N-dimensional vector space $\mathbf{R}^N$ is the direct sum of the two sub-vector spaces; i.e.,

$$\mathbf{R}^N = \mathbf{V_1} \oplus \mathbf{V_2} \tag{3.34}$$

It is shown in the Appendix that we can derive a reciprocal basis $\beta_1 = \{\mathbf{y}^1 \ldots \mathbf{y}^M\}$ from $\{\mathbf{x}^1 \ldots \mathbf{x}^M\}$ for V_1, and an orthonormal basis $\beta_2 = \{\mathbf{y}^{M+1} \ldots \mathbf{y}^N\}$ for V_2 such that $\beta_1 \cup \beta_2 = \beta$ forms a basis for $\mathbf{R}^N$. Then any vector in $\mathbf{R}^N$ can be expressed as a linear combination of $\mathbf{y}^l, l = 1, \ldots N$.

It is shown in the appendix that the N-coupled differential equations of Eq. 3.33 can be represented in terms of the β coefficients as

$$\frac{dx_i}{dt} = -x_i + g\Big(\sum_{m=1}^{M} c_m x_i^m\Big) \qquad i = 1, \ldots, N. \tag{3.35}$$

Note that there are N-coupled differential equations, but there are only M parameters of c_m for all the equations. It is also derived in the appendix that Equation 3.35 in turn can be transformed into two sets of equations

$$\frac{dc_l}{dt} = -c_l + \sum_{i=1}^{N}(x_i{}^l - a_l)g\Big(\sum_{m=1}^{M} c_m x_i{}^m\Big) \qquad l = 1, \ldots, M, \tag{3.36}$$

$$\frac{dc_l}{dt} = -c_l + \sum_{i=1}^{N}(y_i{}^l - b_l)g\Big(\sum_{m=1}^{M} c_m x_i{}^m\Big) \qquad l = M+1, \ldots, N \tag{3.37}$$

where

$$b_l = \frac{1}{N}\sum_{j=1}^{N} y_j{}^l. \tag{3.38}$$

Several comments can be made on Eq. 3.36 and 3.37. First, we see that the components $c_1, \ldots, c_M$ are coupled together, but they are de-coupled from the components $c_{M+1}, \ldots, c_N$. On the other hand, the driving forces for $c_{M+1}, \ldots, c_N$ depend only on $c_1, \ldots, c_M$. This means that the dynamics of the system are completely determined

by $c_1, \ldots, c_M$, although $c_{M+1}, \ldots, c_N$ are not necessarily zero. Therefore, we only need to study the dynamics of these M components. The equilibrium states can be obtained from Eq. 3.35, by letting $d/dt = 0$, we get

$$x_i = g\Big(\sum_{m=1}^{M} c_m x_i^m\Big). \tag{3.39}$$

We see that the equilibrium states are indeed determined by $c_1, \ldots, c_M$. Each of these c_m represents the component of the high-pass version of the input vector on the reciprocal vectors of the m^th stored image, which in some sense is proportional to the correlation between $\mathbf{x}$ and $\mathbf{x^m}$. The stored images which have stronger correlation give stronger components to the equilibrium image. Note that if the neurons are sensitive enough, they will also pick up weak cross-correlation components. Therefore, the output image will not be exactly the same as the stored images. However, the summation of all components from the stored images is thresholded by the neurons. Thus, the nonlinear gain function provides an enhancement of the closest stored image. This effect helps the network converge to the state near the stored images.

In principle, the dynamics of the system and its convergence properties can be obtained by solving Eqs. 3.36 and 3.37. In practice, it is impossible to obtain explicit solutions. In what follows we present a geometrical method to illustrate how the system evolves to a stable state, and how it is influenced by the parameters such as gain and initial conditions. In order to illustrate the concept, we will consider the case where only two images, $\mathbf{x^1}$ and $\mathbf{x^2}$, are stored in the memory. As we shall see, the two-image case contains all the salient features of the dynamics. As discussed in the previous paragraph, we need only to solve two dynamic equations in the two-image case. By Eq. 3.36, we have

$$\frac{dc_1}{dt} = -c_1 + \sum_{i=1}^{N}(x_i^1 - a_1)g(c_1 x_i^1 + c_2 x_i^2) \tag{3.40}$$

$$\frac{dc_2}{dt} = -c_2 + \sum_{i=1}^{N}(x_i^2 - a_2)g(c_1x_i^1 + c_2x_i^2). \tag{3.41}$$

Recall that a_1 and a_2 are the average levels of the input images $\mathbf{x^1}$ and $\mathbf{x^2}$. Let $h_1(c_1, c_2)$ represent the summation term in Eq. 3.40, and $h_2(c_1, c_2)$ the summation term in Eq. 3.41. These two terms are the driving force for c_1 and c_2, respectively. We now discuss the dynamics of the two images. For simplicity, assume that $\mathbf{x^1}$ and $\mathbf{x^2}$ have no overlapping nonzero components. An example is shown in Fig. 3.30. It is seen that x_i^1 can be nonzero only when $x_i^2 = 0$, and vice versa. In this case, the driving forces can be written as

$$h_1(c_1, c_2) = \sum_{x_i^1 \neq 0}^{N}(x_i^1 - a_1)g(c_1x_i^1) - a_1 \sum_{x_i^2 \neq 0} g(c_2x_i^2) \tag{3.42}$$

$$h_2(c_1, c_2) = \sum_{x_i^2 \neq 0}^{N}(x_i^2 - a_2)g(c_2x_i^2) - a_2 \sum_{x_i^1 \neq 0} g(c_1x_i^1) \tag{3.43}$$

There are two terms in each of the driving forces. Consider $h_1(c_1, c_2)$. The first term comes from the correlation between the neuron state $g(c_1\mathbf{x}^1)$ and the stored image $(\mathbf{x}^1 - a_1)$, and the second term results from the coupling between c_1 and c_2 through the dc level a_1. Since a_1 and the gain function $g(x)$ are always positive, the second term gives a negative contribution to the driving force. This means that the coupling pulls the system away from $\mathbf{x}^1$. The same description also applies to c_2. We plot $h_1(c_1, c_2)$ against c_1 for $c_2 = 0$ and $c_2 \neq 0$ in Fig. 3.31(a). In the figure, the solid curve represents the case where $c_2 = 0$, and the dashed curve represents the case for $c_2 \neq 0$. We also plot the line $h(c_1) = c_1$ in the same figure. It is seen that there are three intersections, **P, Q**, and **R**, between the straight line and the solid curve. The plane is divided into four regions. In regions 1 and 3, c_1 is smaller than $h_1(c_1, c_2)$ and $\frac{dc_1}{dt} > 0$. Thus, in these regions the system state evolves in the direction of increasing c_1. This is represented by the arrow pointing to the right in the figure. On the other hand, in regions 2 and 4, $\frac{dc_1}{dt} < 0$; thus, the system evolves toward decreasing c_1. It can be seen that the points **P** and **R** are stable points, and **Q** is

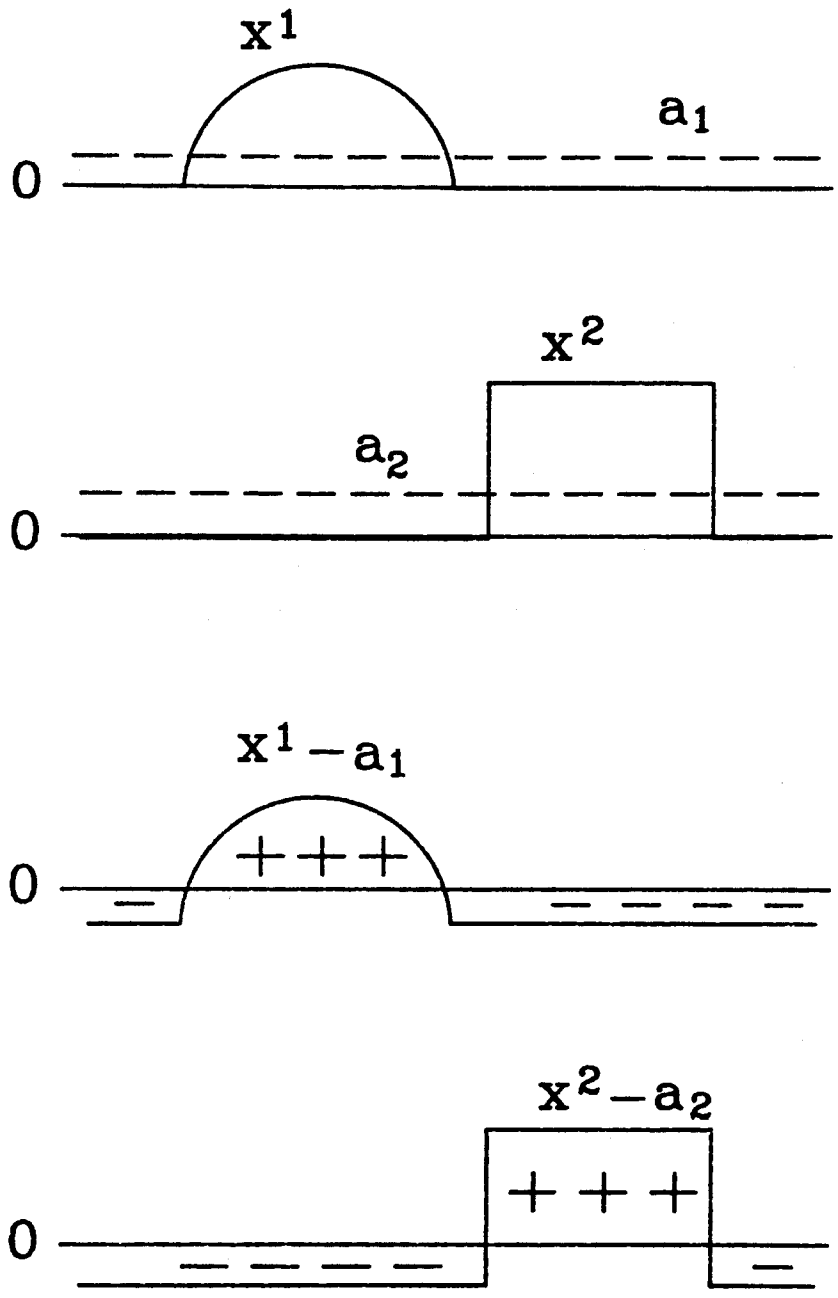


Figure 3.30: An example of two stored images without overlaps.

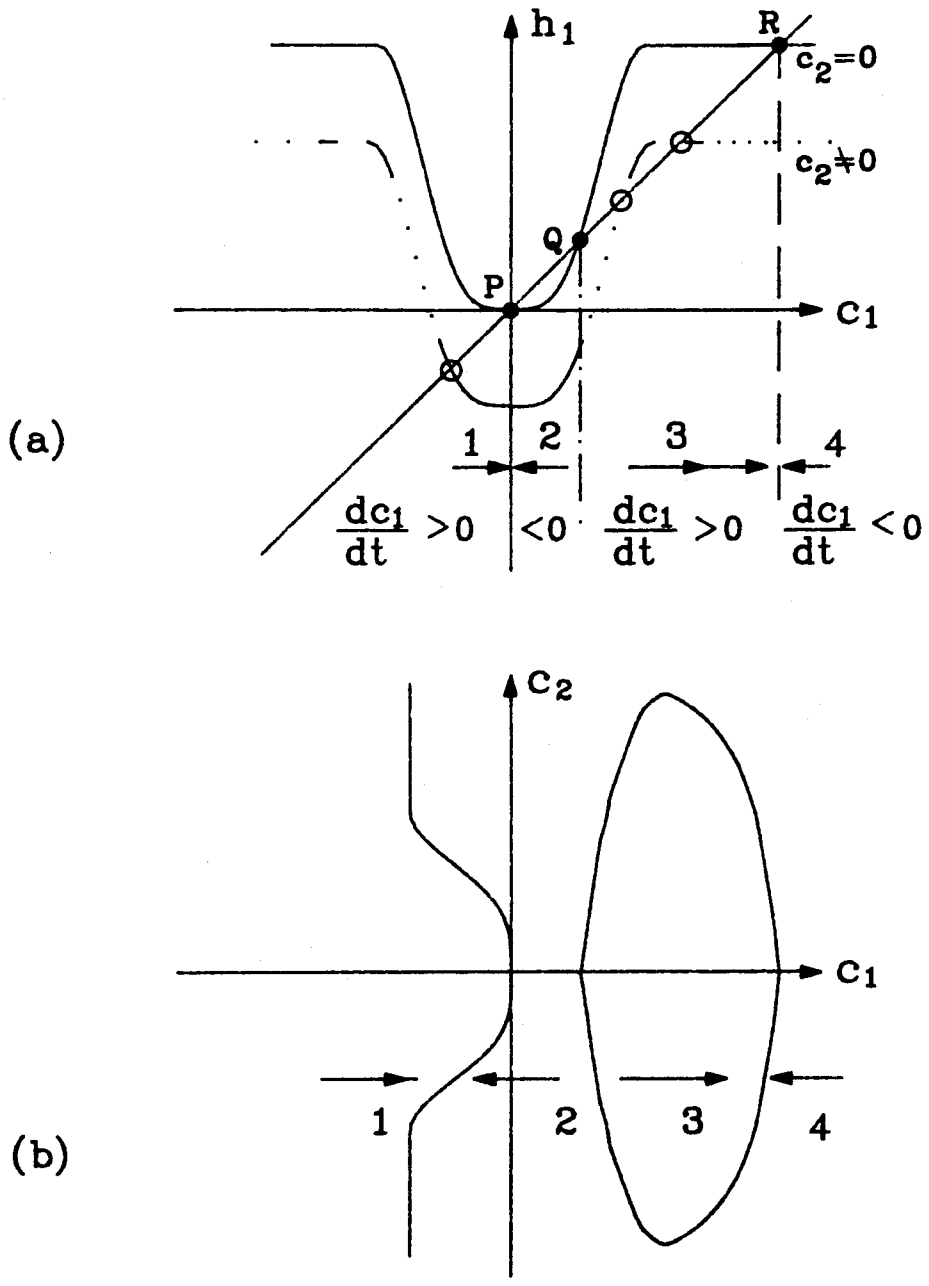


Figure 3.31: The Driving Force and the Dynamics of the Loop. (a) The driving force for the first stored image. (b) The trajectories of the equilibrium states of the first image.

a saddle point. Now suppose we increase c_2. Then $g(c_2 x_i^2)$ increases, and $h_1(c_1, c_2)$ decreases. It can be seen from Fig. 3.31(a) that points **Q** and **R** move towards each other. The two points merge into one point at large values of c_2. On the other hand, the point **P** moves downward as c_2 increases. If we plot the trajectories of **P**, **Q**, and **R** for the coefficient c_1 in the (c_1, c_2) plane, we obtain Fig. 3.31(b). Since the gain function $g(x)$ is positive and symmetric with respect to x, the trajectories are symmetric with respect to the c_2 axis. The above argument shows how the dynamical behavior of the system can be understood using a geometrical method.

By going through the same procedure, we can also obtain the trajectories leading to the equilibrium points of c_2. We plot the two groups of trajectories in the same (c_1, c_2) plane. Fig. 3.32 shows the result. We see that there are 7 equilibrium points: one source, three sinks, and three saddles. The three sinks represent the null state (no image) and the two stored images. Point 1 represents the stable state corresponding to stored image $\mathbf{x}^1$, since at that position c_1 is large and c_2 is small. On the other hand, at point 2 c_1 is small and c_2 is large. This represents the stable state corresponding to stored image $\mathbf{x_2}$. It can be seen from the figure that if we start from an initial state, which is close to one of the stored states, then the system will converge to that state. Otherwise, it will decay to zero.

From the geometrical diagram we see that the stable state is always a mixed state of the stored memories. The extent of mixture can be reduced by reducing the neural gain. However, if the gain is too small, then the system will not be able to sustain the stored memories. As shown in Fig. 3.33(a), when the gain is very low, there is only one intersection point **O** between the line $h(c) = c$ and the curve of the driving force. The corresponding trajectories leading to **O** are drawn in Fig. 3.33(b). It is seen that the only equilibrium state is the null state at **P**. No matter where the initial state is, the system always decays to zero.

On the other hand, suppose the neural gain is very high. The trajectories to **P**,

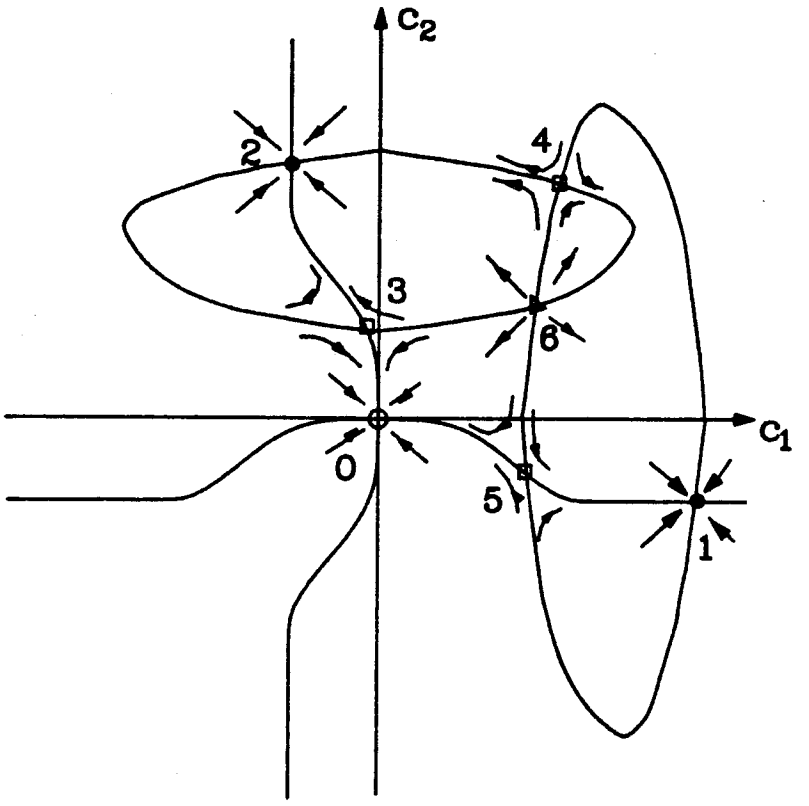


Figure 3.32: The phase flow of the two-image auto-associative memory. States O, 1, 2 are stable. States 3, 4, 5 are unstable (saddle points). State 6: Source state. (unstable)

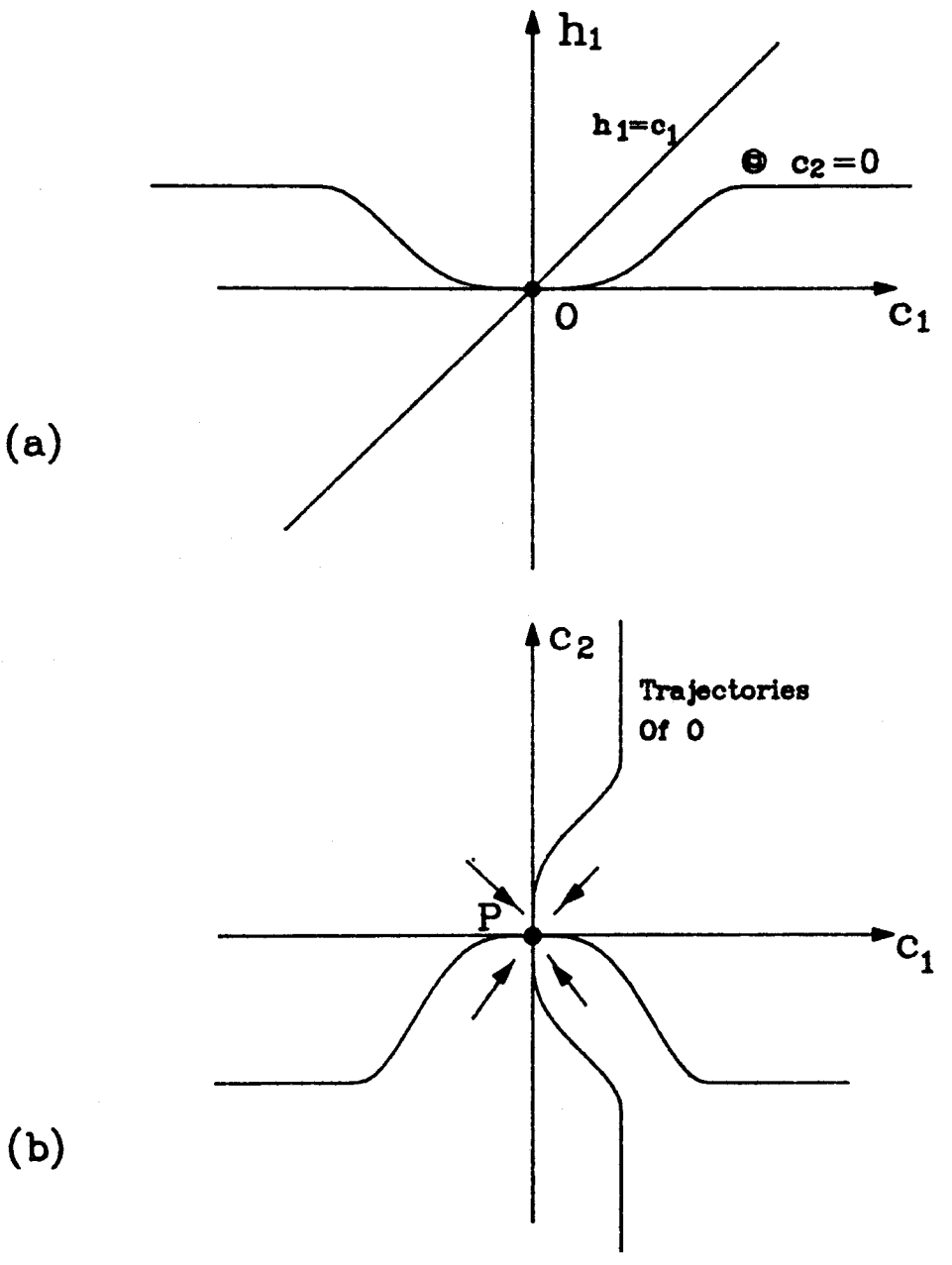


Figure 3.33: The Dynamics of the Loop at Low Gain. (a) The driving force for the first stored image. (b) Trajectories for the equilibrium points. Point P is the only stable state.

Q, and **R** are shown in Fig. 34. We see that there are two more equilibrium points than the case shown in Fig. 3.32: a stable point m and a saddle point s. The state m is a strongly mixed state of $\mathbf{x}^1$ and $\mathbf{x}^2$. We also see that m has a large region of attraction. Thus, we can not set too high a gain.

Next consider the case where the stored memories have some slight overlap. The shape and position of the trajectory lines will be altered somewhat. Since the neural function is continuous, so is the driving force. Thus, the general features of the system will be the same. However, as the overlapping between the stored states increases more and more, the trajectories of the equilibrium points will be more and more distorted. As a result, the stable points that give the stored memory states will have more "mixing" with the other stored state. Thus, we would not expect the system to work well in the case of strong overlapping.

We can now investigate the dynamics of the system for the all-pass hologram in the memory loop. If we do not make the dc-level subtraction in storing the memories, then Equations 3.40 and 3.41 become

$$\frac{dc_1}{dt} = -c_1 + \sum_{i=1}^{N} x_i^1 g(c_1 x_i^1 + c_2 x_i^2) \tag{3.44}$$

$$\frac{dc_2}{dt} = -c_2 + \sum_{i=1}^{N} x_i^2 g(c_1 x_i^1 + c_2 x_i^2). \tag{3.45}$$

By going through similar arguments, we can draw the trajectories leading to the equilibrium points of the system. Fig. 3.35 shows one example. It is seen that there are four stable states: two memory states, m_1 and m_2, one null state , and one mixed state **P**. If we decrease the neural gain, then the points m_1, m_2, **Q**, and **R** may disappear. However, the mixed state **P** always exists. Therefore, the high-pass hologram is crucial for good performance of the memory loop.

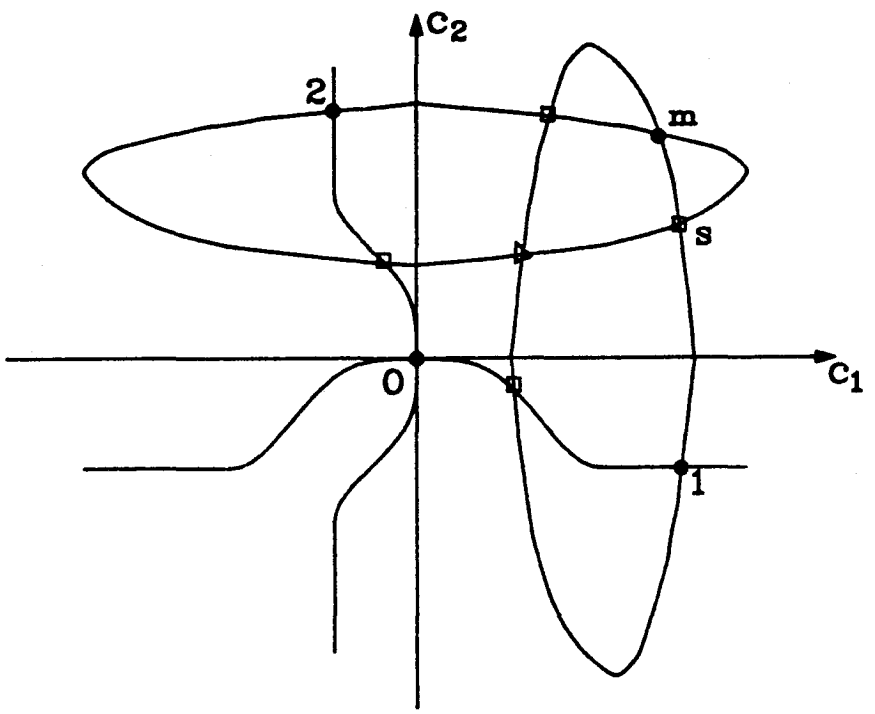


Figure 3.34: The Dynamics of the Loop at High Gain. Two new equilibrium states are generated: m is mixed and s is unstable state.

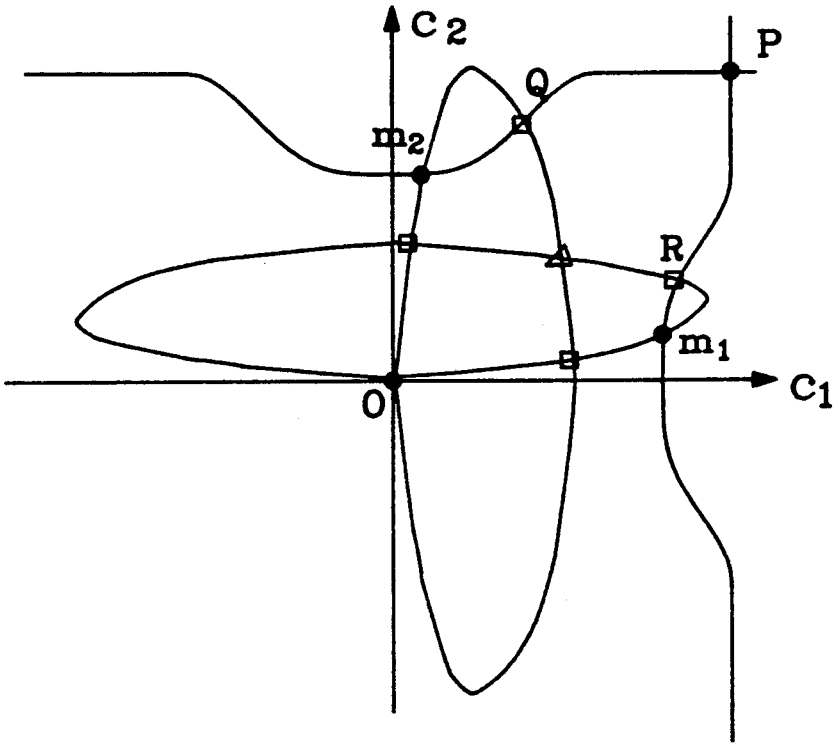


Figure 3.35: The Dynamics of the Loop without the High Pass Hologram. There are four stable states: m_1, m_2 are stored states, O is null state. P is mixed state. Other states are unstable.

3.7 Conclusion

In this chapter we have successfully implemented an optical loop of the Hopfield-type neural network. The loop consists of a single layer of neurons with feedback. The 2-D array of neurons are simulated by an LCLV. The interconnections between the neurons are implemented by using planar holograms.

Experimental results show that the loop performs the function of auto-association which recognizes 2-D images from a distorted input. The distortions may come from rotating, scaling, shifting, partial blocking, or erroneous input. As long as the input provides enough information, the loop always converges to a stable state corresponding to the correct stored image. The similarity between the input and the stored images determines how fast the loop converges.

The gain function is another factor that determines the dynamics of the loop. Too low gain cannot sustain a stable memory. High gain provides more tolerance to the system to recognize a distorted image. However, too high gain also induces mixed states such that the loop makes incorrect recognitions. The trade-off between distortion tolerance and discrimination capability is set by the gain function, and the value of the optimum gain depends on the specific images stored.

We have presented a neural network model for the optical loop. Dynamic equations for the network are formulated. Since the optical neurons respond to light intensities, the input vs. output relationship of the neurons is not a simple sigmoid function. Rather, it is a sigmoid function symmetrical for positive and negative inputs. Thus, the slope of the neuron function is not always positive, and the conventional proof of convergence does not apply here. Although we have not found an energy function for our loop, we have presented a geometrical method for inspecting dynamic behaviors of the loop. The flow of the state in the phase diagram helps us visualize the convergence properties of the loop. The diagram shows that the stable

state is always a mixed state of the stored images, the degree of mixture depending on the gain function.

3.8 Appendix

Dynamic Equations for the Optical Associative memory

This appendix will focus on the derivation of the dynamic equations for the optical associative memory loop described in this chapter. The schematic diagram of the optical loop is shown in Fig. 3.5. The main idea is that corresponding to the set of stored images, we can construct a vector space $\mathbf{R^N}$, where N is the number of neurons of the system. Then the N-coupled dynamic equations of the neurons can be expressed in terms of the basis coefficients in $\mathbf{R^N}$. This transformation simplifies the form of dynamic equations so that the dynamic properties of the system can be inspected using a geometrical method.

A. Vector Space $\mathbf{R^N}$ and the Basis

We assume that the stored images $\mathbf{x^1}, \mathbf{x^2}, \ldots, \mathbf{x^M}$ are linearly independent. We then decompose the vector space $\mathbf{R^N}$ into two subspaces $\mathbf{V_1}$ and $\mathbf{V_2}$, where $\mathbf{V_1}$ is the vector space spanned by the stored images and $\mathbf{V_2}$ is normal to $\mathbf{V_1}$; i.e.,

$$\mathbf{R^N} = \mathbf{V}_1 \oplus \mathbf{V}_2, \tag{3.46}$$

$$\mathbf{V_1} = span\{\mathbf{x^1}, \mathbf{x^2}, \ldots, \mathbf{x^M}\}, \tag{3.47}$$

$$\mathbf{V_2} = \{\mathbf{y} | \mathbf{x} \cdot \mathbf{y} = 0, \forall \mathbf{x} \in \mathbf{V_1}\}, \tag{3.48}$$

where $\oplus$ is the direct sum of vector spaces. We now define a reciprocal basis $\beta_1 = \{\mathbf{y^1}, \mathbf{y^2}, \ldots, \mathbf{y^M}\}$ for $\mathbf{V}_1$, such that

$$\mathbf{y^i} \cdot \mathbf{x^j} = \delta_{ij} \qquad i, j = 1, \ldots, M. \tag{3.49}$$

Next we select an orthonormal basis $\beta_2 = \{\mathbf{y^{M+1}}, \ldots, \mathbf{y^N}\}$ for $\mathbf{V_2}$. We then have

$$\begin{aligned} \beta &= \beta_1 \cup \beta_2 \\ &= \{\mathbf{y^1}, \cdots, \mathbf{y^N}\} \end{aligned} \tag{3.50}$$

which forms a basis for $\mathbf{R^N}$. Thus, any image in $\mathbf{R^N}$ can be expressed in terms of the basis β:

$$\mathbf{x} = \sum_{k=1}^{N} c_k \mathbf{y^k}. \tag{3.51}$$

B. Dynamic Equations for the Optical Loop

It was shown in Section 3.6 that the dynamic equation for the neurons in the optical loop is

$$\frac{dx_i}{dt} = -x_i + g\Big(\sum_{j=1}^{N} w_{ij} x_j\Big), \qquad i = 1 \ldots N \tag{3.52}$$

where w_{ij} is the high-pass version of the interconnection strength given by Eq. 3.24. By Eq. 3.24 and 3.25, we have

$$\begin{aligned}
\sum_{j=1}^{N} w_{ij} x_j &= \sum_{j=1}^{N} \Big(\sum_{m=1}^{M} (x_j^m - a_m) x_i^m \Big) x_j \\
&= \sum_{j=1}^{N} \Big(\sum_{m=1}^{M} x_i^m x_j^m \Big) x_j - \sum_{m=1}^{M} \sum_{j=1}^{N} a_m x_i^m x_j \\
&= \sum_{j=1}^{N} \Big(\sum_{m=1}^{M} x_i^m x_j^m \Big) x_j - \Big(\sum_{k=1}^{N} x_k \Big) \Big(\sum_{m=1}^{M} \sum_{j=1}^{N} \frac{1}{N} x_j^m x_i^m \Big) \\
&= \sum_{j=1}^{N} \Big(\sum_{m=1}^{M} x_i^m x_j^m \Big) \Big(x_j - \frac{1}{N} \sum_{k=1}^{N} x_k \Big).
\end{aligned} \tag{3.53}$$

Note that $\Big(x_j - \frac{1}{N}\sum_{k=1}^{N} x_k\Big)$ is the j^{th} component of the vector $\Big(\mathbf{x} - \frac{1}{N}\sum_{k=1}^{N} x_k\Big)$. Since $\Big(\mathbf{x} - \frac{1}{N}\sum_{k=1}^{N} x_k\Big)$ is also a vector in $\mathbf{R^N}$ and any vector in $\mathbf{R^N}$ can be expanded in the basis β, we let

$$\Big(\mathbf{x} - \frac{1}{N} \sum_{k=1}^{N} x_k\Big) = \sum_{k=1}^{N} c_k \mathbf{y^k}. \tag{3.54}$$

The feedback signal in the loop then becomes

$$\begin{aligned}
\sum_{j=1}^{N} w_{ij} x_j &= \sum_{j=1}^{N} \Big(\sum_{m=1}^{M} x_i^m x_j^m \Big) \Big(\sum_{k=1}^{N} c_k y_j^k \Big) \\
&= \sum_{m=1}^{M} \sum_{k=1}^{N} \Big(\sum_{j=1}^{N} x_j^m y_j^k \Big) c_k x_i^m \\
&= \sum_{m=1}^{M} \Big(\sum_{k=1}^{N} (\mathbf{x}^m \cdot \mathbf{y}^k) c_k \Big) x_i^m.
\end{aligned} \tag{3.55}$$

By applying the orthnormal property of the reciprocal vectors $\mathbf{x^m}$ and $\mathbf{y^m}$ in the above equation, we obtain

$$\sum_{j=1}^{N} w_{ij}x_j = \sum_{m=1}^{M} c_m x_i^m. \tag{3.56}$$

The dynamic equation then becomes

$$\frac{dx_i}{dt} = -x_i + g\Big(\sum_{m=1}^{M} c_m x_i^m\Big) \qquad i = 1 \ldots N. \tag{3.57}$$

Note that c_m is the projection of the high-pass version of the input on the stored vector $\mathbf{x^m}$. It is seen that the system dynamics is specified by the M coefficients of the reciprocal vectors of the stored images. In what follows we will derive the equation of motion for these coefficients.

Multiplying both sides of Eq. 3.57 by $1/N$ and summing over i, we get

$$\frac{1}{N}\sum_{i=1}^{N}\frac{dx_i}{dt} = -\frac{1}{N}\sum_{i=1}^{N} x_i + \frac{1}{N}\sum_{i=1}^{N} g\Big(\sum_{m=1}^{M} c_m x_i^m\Big) \qquad i = 1 \ldots N. \tag{3.58}$$

Subtracting Eq. 3.58 from Eq. 3.57 gives

$$\frac{d}{dt}[x_i - \frac{1}{N}\sum_{j=1}^{N} x_j] = -[x_i - \frac{1}{N}\sum_{j=1}^{N} x_j] + g\Big(\sum_{m=1}^{M} c_m x_i^m\Big) - \frac{1}{N}\sum_{i=1}^{N} g\Big(\sum_{m=1}^{M} c_m x_i^m\Big). \tag{3.59}$$

Note that the terms in the bracket can expanded in the basis β; thus, the equation of motion is given in terms of the β coefficients as

$$\sum_{k=1}^{N}\frac{dc_k}{dt}y_i^k = -\sum_{k=1}^{N} c_k y_i^k + g\Big(\sum_{m=1}^{M} c_m x_i^m\Big) - \frac{1}{N}\sum_{i=1}^{N} g\Big(\sum_{m=1}^{M} c_m x_i^m\Big). \tag{3.60}$$

Multiplying both sides of the above equation by x_i^l and then summing over i gives

$$\begin{aligned}\sum_{k=1}^{N}\frac{dc_k}{dt}(\mathbf{x^l}\cdot\mathbf{y^l}) \;=\; & -\sum_{k=1}^{N} c_k(\mathbf{x^l}\cdot\mathbf{y^l}) + \sum_{i=1}^{N} x_i^l g\Big(\sum_{m=1}^{M} c_m x_i^m\Big), \\ & -\Big(\frac{1}{N}\sum_{i=1}^{N} x_i^l\Big)\sum_{i=1}^{N} g\Big(\sum_{m=1}^{M} c_m x_i^m\Big)\end{aligned} \tag{3.61}$$

Applying the orthonormal relation between the stored images and the reciprocal basis, we obtain

$$\frac{dc_l}{dt} = -c_l + \sum_{i=1}^{N}(x_i^l - a_l)g\Big(\sum_{m=1}^{M} c_m x_i^m\Big) \qquad l = 1, \ldots, M. \tag{3.62}$$

Recall that a_l is the average level of the stored image $\mathbf{x}^l$. Similarly, the equation of motion for $c_{M+1}, \cdots, c_N$ can be obtained by multiplying both sides of Eq. 3.60 by y_i^l, and summing over i. We get

$$\frac{dc_l}{dt} = -c_l + \sum_{i=1}^{N}(x_i^l - b_l)g\Big(\sum_{m=1}^{M} c_m x_i^m\Big) \qquad l = M+1, \ldots, N, \tag{3.63}$$

where b_l is defined as

$$b_l = \frac{1}{N}\sum_{i=1}^{N} y_i^l. \tag{3.64}$$

The dynamic behavior of the optical loop is thus completely determined by equations 3.63 and 3.64. Instead of solving these equations, we have presented a geometrical method in Section 3.6 for investigating the convergence properties of the loop and the important parameters that affect the dynamics.

3.9 References for Chapter 3

1. J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Natl. Acad. Sci. USA*, Vol. **79** , pp.2554-2558, April (1982).

2. J. J. Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons," *Proc. Natl. Acad. Sci. USA*, Vol. **81**, pp. 3088-3092, May (1984).

3. Y. S. Abu-Mostafa and D. Psaltis, "Optical Neural Computers," Scientific American, Vol. **256**, No. 3, pp. 88-95, March (1987).

4. 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 Symposium on Lasers & Optics*, Paper 1043-07, Los Angeles, California, 15-20 Jan. (1989).

5 L. S. Lee, H. M. Stoll, and M. C. Tackitt, "Continuous-time Optical Neural Associative Memory," *Opt. Lett.* **14**, 162 (1989).

6. S. Venkatesh and D. Psaltis, "Linear and Logarithmic Capacities in Associative Neural Networks," submitted for publication in *IEEE Trans. Inform. Theory.*

7. R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. Venkatesh, "The Capacity of the Hopfield Associative Memory," *IEEE Trans. Inform. Theory*, Vol. **IT-33**, No. 4, 461-482, July (1987).

8. 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).

9. N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, "Optical Implementation of the Hopfield Model," *Appl. Opt.*, **24**, 1469-1475, (1985).

10. D. Psaltis, E. Paek, and J. Hong, "Acousto-Optic Implementation of the Hopfield Model," *JOSA*, **2**, No. 13, 48, December (1985).

11. J. S. Jang, S. W. Jung, S. Y. Lee, and S. Y. Shin, "Optical Implementation of the Hopfield Model for Two-Dimensional Associative Memory," *Opt. Lett.* **13**, 248 (1988).

12. 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).

13. E. G. Paek and D. Psaltis, "Optical Associative Memory using Fourier Transform Holograms," **26**, No. 5, 428-433, May (1987).

14. A. B. Vander Lugt, "Signal Detection by Complex Spatial Filtering," *IEEE Trans. Inform. Theory*, **IT-10**, No. 2, 139-145, (1964).

15. Y. Owechko, G. J. Dunning, E. Marom, and B. H. Soffer,"Holographic Associative Memory with Nonlinearities in the Correlation Domain," *Appl. Opt.*, **26**,(10), 1900-1910, May (1987).

16. D. Z. Anderson, "Coherent Optical Eigenstate Memory," *Opt. Lett.*, **11**, 56 (1986).

17. A. Yariv, S. K. Kwong, and K. Kyuma, "Optical Associative Memories based on Photorefractive Oscillations," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **613**, 1 (1986).

18. D. Psaltis and J. Hong, "Shift-Invariant Optical Associative Memories," *Opt. Eng.*, **26**, 10-15, January (1987).

19 G. J. Dunning, E. Maron, Y. Owechko, and B. H. Soffer, "All-Optical Associative Memory with Shift Invariance and Multiple-Image Recall," *Opt. Lett.* **12**, 346 (1987).

20. H. Y. Li, "Analysis of Pinhole Array Associative Memory System Using the Method of Phase Planes," private communication.

Chapter 4

The Hetero-Associative Memory Loop

4.1 Introduction

In the previous chapter we presented the holographic system with feedback, which performs the function of an auto-associative memory. The basic principle of that system is based on the correlation operation. When the input image has a strong correlation with one of the stored images $\vec{\mathbf{f}}_\mathbf{m}$, the system converges to the stable state that is most similar to the input image. The system contains two Vander Lugt correlators in which we used a plane wave as the reference for recording the Fourier-transform hologram of the stored images. The two holograms in the correlators store the same set of images $\vec{\mathbf{f}}_\mathbf{m}$. If, instead, we store a different set of images $\vec{\mathbf{g}}_\mathbf{m}$ in the second hologram, then the correlation of the input image with the images stored in the first hologram reconstructs images $\vec{\mathbf{g}}_\mathbf{m}$ from the second hologram. Thus, the system can perform the function of hetero-association. However, if we want to close the feedback loop, then we need a mechanism to map the images $\vec{\mathbf{g}}_\mathbf{m}$ back to the images $\vec{\mathbf{f}}_\mathbf{m}$ at the input plane. Actually, this can be achieved by using the images $\vec{\mathbf{g}}_\mathbf{m}$ as the reference wave to make the Fourier-transform hologram of the images $\vec{\mathbf{f}}_\mathbf{m}$ such that the two set of images are recorded associatively.

The basic idea of hetero-associative memory described in the above paragraph is easily understandable if we think of the principle of holography. A hologram is, in general, the recording of the intensity pattern of the interference between two waves. Normally, one of the two waves is called the *reference* and the other is called the *subject wave*. The distinction between the two is actually arbitrary. Let the complex amplitude of the *reference wave* be designated by $\mathbf{r} = r\ exp(j\phi_r)$, and that

of the *object wave* by $\mathbf{o} = o\ exp(j\phi_o)$. Suppose the hologram is recorded linearly; then the amplitude transmittance of the developed hologram is proportional to the exposure:

$$t \propto I = \mathbf{oo}^* + \mathbf{rr}^* + \mathbf{or}^* + \mathbf{o}^*\mathbf{r}, \tag{4.1}$$

where $*$ means the complex conjugate. Now if we input the *reference wave* $\mathbf{r}$ to reconstruct the hologram, the reconstructed wave would be :

$$w \propto \mathbf{r}t = \mathbf{roo}^* + \mathbf{rrr}^* + \mathbf{orr}^* + \mathbf{rro}^*. \tag{4.2}$$

Under suitable choice of $\mathbf{r}$ such that the intensity of $\mathbf{rr}^* = r^2$ is constant over the hologram plane, the third term of Eq. 4.2 becomes $r^2\mathbf{o}$, and therefore the original *subject wave* $\mathbf{o}$ is reconstructed. Similarly, if we use the *subject wave* $\mathbf{o}$ to reconstruct the hologram and make the arrangement such that $\mathbf{oo}^* = o^2$ is constant over the hologram plane, then we reconstruct the original *reference wave* as $o^2\mathbf{r}$. Therefore, the two waves $\mathbf{o}$ and $\mathbf{r}$ are stored in the hologram as an associative pair; whenever we input one of the two waves, we get the other as the output. We call the information stored in this form as hetero-associative memory. The idea of using holographic techniques for storing associative informations was first pointed out by Van Heerden [1,2] and also by Gabor [3].

In this chapter we present an optical feedback loop what performs the function of image hetero-association. In Section 4.2 we will describe how to make holographic interconnections between two image patterns. The problem of degenerate interconnections and the sampling techniques to solve this problem will be presented. The holographic interconnection techniques developed in this section will then be applied to the image-association loop in Section 4.3. The optical system of the hetero-associative memory and its experimental results will be presented in this section. Finally, the neural network model of the hetero-associative loop and its convergence properties will be presented in Section 4.4

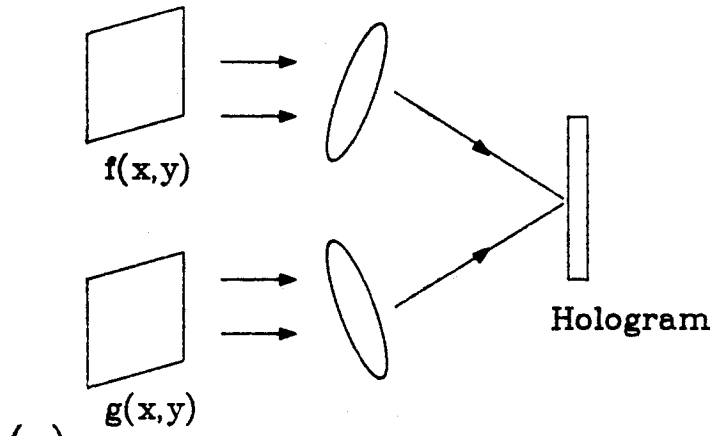
4.2 Sampling Grids for Planar Holographic Associative Memories

In this section we show how to use planar holograms to store hetero-association. A hetero-associative memory between two images is nothing but a transformation or an interconnection pattern between these two images such that one image is mapped to the other and vice versa. In Chapter 2 it was shown that a holographic grating can be used as an interconnection between two points. It was also shown that the Vander Lugt correlator performs the association between a pattern and a point. Given these results we will show how to extend the Vander Lugt correlator into a hetero-associative memory.

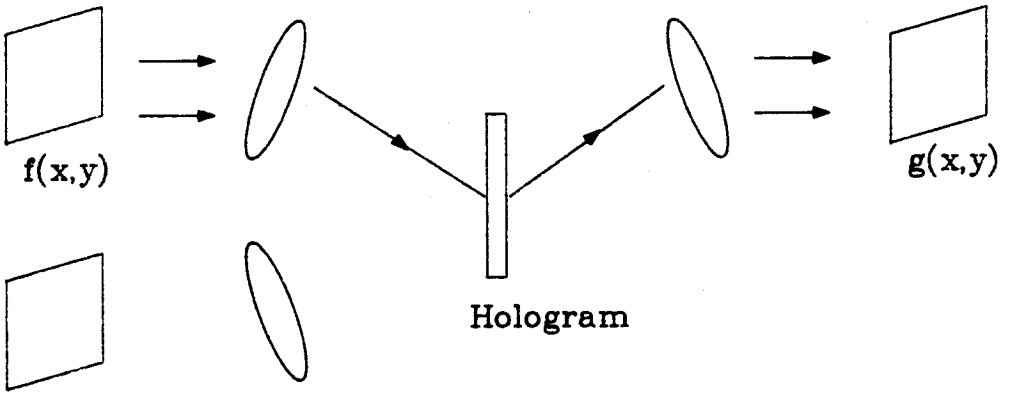
Consider the basic structure of the holographic hetero-associative memory shown in Fig. 4.1. The architecture shown in the figure is similar to that of the Vander Lugt correlator. However, instead of using a plane wave as the reference, we use images for inputs to both optical arms. The two patterns $f(x,y)$ and $g(x,y)$ to be associated are put at the front focal planes of the lenses, L_1 and L_2, respectively. A holographic plate is placed at the intersection of the back focal planes of lenses L_1 and L_2. The interference pattern of the Fourier transforms of $f(x,y)$ and $g(x,y)$ is recorded on the hologram. Suppose the hologram is linearly recorded and developed; then the amplitude transmittance of the hologram is proportional to the total exposure:

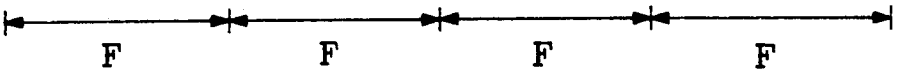
$$\begin{aligned} T &\propto tI(u,v) \\ &= t|F(u,v)+G(u,v)|^2 \\ &= t[|F(u,v)|^2+|G(u,v)|^2+F^*(u,v)G(u,v)+F(u,v)G^*(u,v)], \end{aligned} \tag{4.3}$$

where T is the amplitude transmittance of the developed hologram, t is the exposure time, $I(u,v)$ is the intensity distribution of the interference pattern, $F(u,v)$ and $G(u,v)$ are the Fourier-transforms of $f(x,y)$ and $g(x,y)$, respectively, u and v are



(a)





(b)

Figure 4.1: Hetero-Associative memory. (a) Recording. (b) Recalling.

spatial frequencies, and $*$ denotes the complex conjugate. It is seen from the third and fourth terms of Eq. 4.3 that $f(x,y)$ and $g(x,y)$ are associatively recorded in the Fourier-transform hologram. Now if the input to the system is $f(x,y)$, as shown in Fig. 4.1(b), then the transmitted wave from the hologram is

$$F(u,v)\left[|F(u,v)|^2+|G(u,v)|^2\right]+F(u,v)F^*(u,v)G(u,v)+F(u,v)F(u,v)G^*(u,v). \tag{4.4}$$

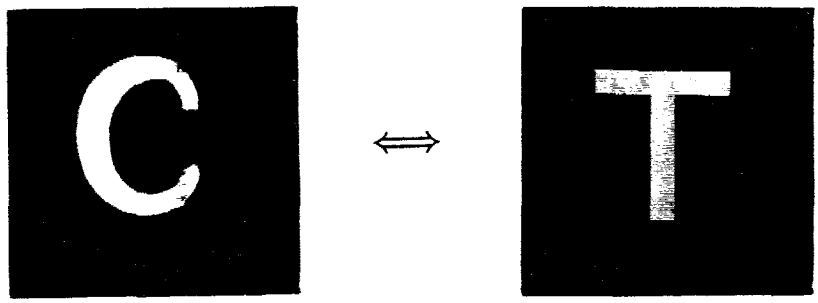
If the angular separation between the inputs $f(x,y)$ and $g(x,y)$ is large enough, then the four terms of the transmitted waves in Eq. 4.4 will also be angularly separated and will propagate along different directions. The third term is of particular interest to us because it gives a spectrum that is similar to that of $g(x,y)$. Fig. 4.1(b) shows that the diffracted wave from the third term is Fourier-transformed to the image plane. The reconstructed wave at the output plane is

$$\left[f(x,y)\star f(x,y)\right]*g(x,y), \tag{4.5}$$

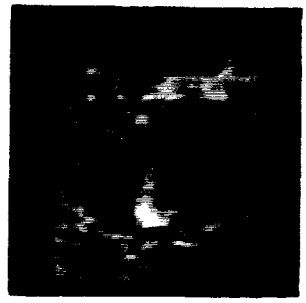
where $\star$ means auto-correlation and $*$ means convolution operators. We see that the input $f(x)$ produces its associated image $g(x)$. However, the read-out is blurred by the auto-correlation of $f(x,y)$. Fig. 4.2 shows an example of the optical experimental results. Fig. 4.2(a) shows the letters **C** and **T** that are associatively recorded using the above scheme. Fig. 4.2(b) shows the output image **T** when the input is **C**. The output image is blurred and is barely recognizable. Similarly, Fig. 4.2(c) shows the blurred output image **C** when the input is **T**.

In Eq. 4.5 if $f(x,y)\star f(x,y)=\delta(x,y)$, then the output image is the original $g(x,y)$. There are two cases when this condition is true: Either $f(x,y)$ is a point source or it is a random-phase diffuser. However, neither condition is of interest to us because the patterns to be associated are in general neither of the two cases. In the following we show how to use the Vander Lugt correlator to make interconnections between two patterns.

(a)



(b)



(c)

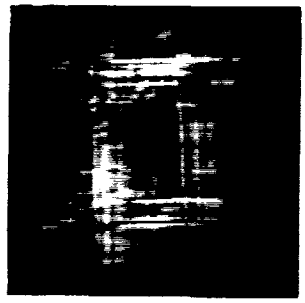


Figure 4.2: Experimental Results of Optical Hetero-Associative Memory. (a) Associated letters **C** and **T**. (b) Recalling **T** by **C**. (c) Recalling **C** by **T**.

The key characteristics of the Vander Lugt correlator-type planar holographic interconnections is shift invariance. Shift invariance means that when the input is shifted relative to the original position, the output response is shifted correspondingly, but the output pattern remains unchanged. First, we consider the very simple case when a point source $\delta(x)$ and a pattern $g(x)$ are associatively recorded in the Vander Lugt correlator architecture. According to Eq. 4.5, when the input is $\delta(x)$, the reconstructed output is

$$[\delta(x) \star \delta(x)] * g(x) = g(x). \tag{4.6}$$

Now if the input is shifted to $\delta(x-a)$, the output is shifted as

$$[\delta(x-a) \star \delta(x)] * g(x) = g(x-a). \tag{4.7}$$

This result demonstrates that an erroneous input point can also recall an associated stored image. Furthermore, if the input contains several point sources, then the total output will be the combination of the output coming from each source. The output can be written as

$$\Big[\sum_{i=1}^{M} \delta(x-a_i) \star \delta(x)\Big] * g(x) = \sum_{i=1}^{M} g(x-a_i), \tag{4.8}$$

where a_i is the position of each input source and M is the total number of inputs. Evidently, the output is very different from the original stored image $g(x)$. Fig. 4.3 shows an example. In the figure (a) shows that letter **A** is stored associatively with the point source, (b) shows the output **A** when the input is the original point source, and (c) shows the output when the input is five point-sources in which one is the original and four are its shifted versions. The output is a superposition of five shifted versions of the original memory.

The problem of the shift invariant property of holographic associative memories would become worse if both associative patterns consist of multiple points. Because

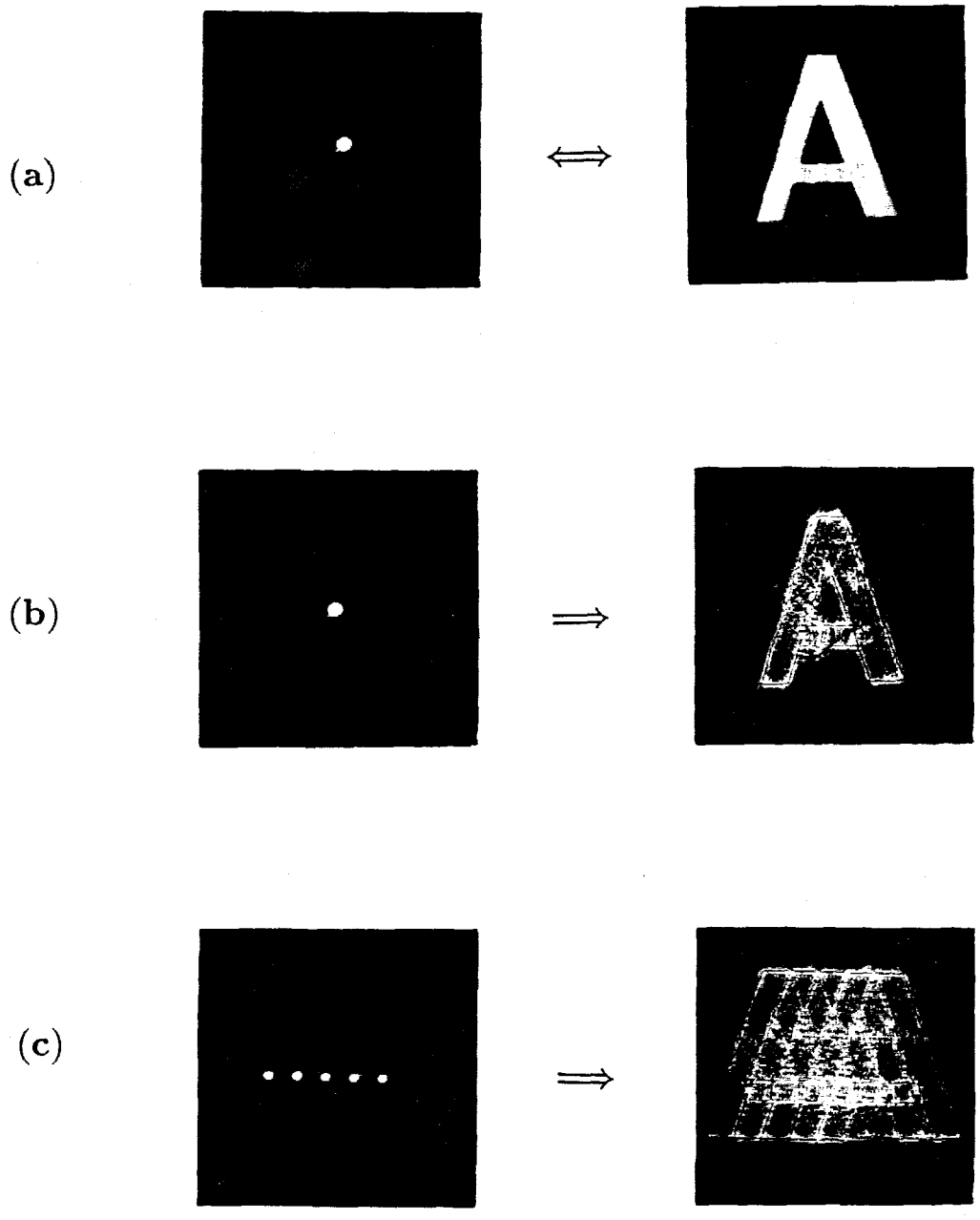


Figure 4.3: Shift Invariance of the Vander Lugt Correlator. (a) Two associated patterns. (b) Input one dot to recall the **A**. (c) Output when the input is five dots.

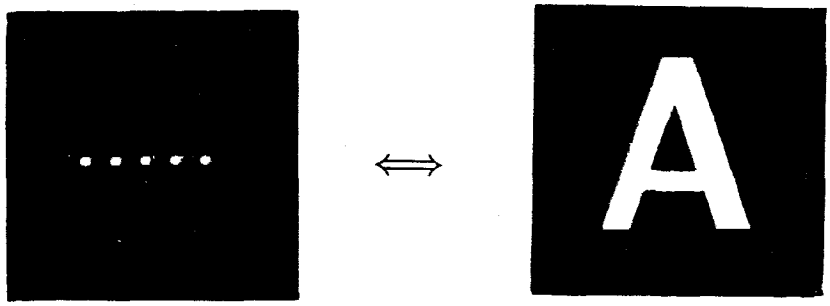
each point of one pattern is connected to every point of the other pattern, the light from one point reads out not only all of its own interconnections but also the shifted versions of interconnections of other points. The superposition of these read-outs will be mixed up and unrecognizable. Suppose the stored associative memories are image $g(x)$ and five point sources $\sum_{n=1}^{5}\delta(x-ma)$, then according to Eq. 4.5, the input $\sum_{m=1}^{5}\delta(x-ma)$ gives output

$$\begin{aligned}
&\Big[\sum_{m=1}^{5}\delta(x-ma)\star\sum_{n=1}^{5}\delta(x-na)\Big]*g(x)\\
=&\sum_{m=1}^{5}\sum_{n=1}^{5}\delta(x-(m-n)a)*g(x)\\
=&\sum_{m=1}^{5}\sum_{n=1}^{5}g(x-(m-n)a). \qquad (4.9)
\end{aligned}$$

Evidently the image is completely unrecognizable if the separation between the point source a is smaller then the size of $g(x)$. Fig. 4.4 shows the results of optical experiments of the association between the letter **A** and five point sources. In the figure (a) shows the original images that are associatively recorded, and (b) shows the output when the input is the original five points. It is seen that shifted versions of **A** are superimposed, as predicted by Eq. 4.9.

The main point of the above discussion is that the Vander Lugt correlator can be used as an associative memory, but it suffers from the problem of shift invariance. In order to achieve that purpose, we should find some ways of overcoming the problem of degenerate read-outs. This problem was first pointed out by Psaltis et al. [4, 5]. Based on the shift-invariance property of the Vander Lugt correlator, Psaltis derived a geometrical method to design the sampling grids for both planar and volume holographic interconnections. The condition derived for planar interconnections is that the diagram formed by connecting any two points of one pattern and any two points of the second pattern cannot be a parallelogram. In the following we present an alternate method of designing the sampling grids for planar interconnections. Sup-

(a)



(b)

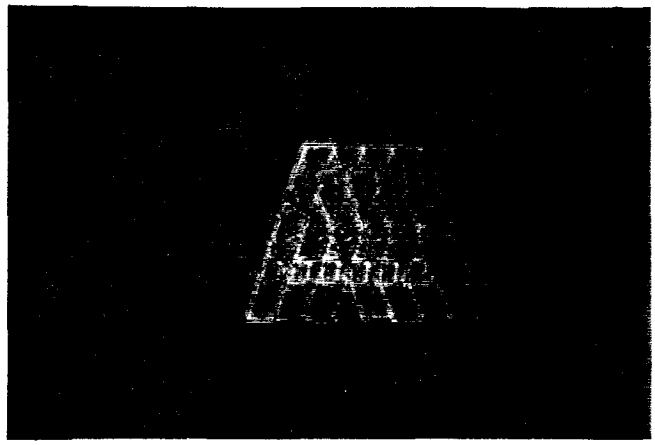


Figure 4.4: Hetero-Associative Memory. (a) Two associated patterns. (b) Input the five dot to recall the **A**.

pose two associative imagesthat are to be stored are sampled by the sampling grids $f(x)$ and $g(x)$, respectively. The sampling grids are represented by the following sampling functions

$$f(x) = \sum_{m=1}^{M} \delta(x - ma) \tag{4.10}$$

and

$$g(x) = \sum_{n=1}^{N} \delta(x - nb), \tag{4.11}$$

respectively, where a and b are the sampling periods, and M and N are the number of sampling points. For simplicity, we use 1-D signals in the analysis, but the results apply to the 2-D case as well. We also assume that the two grids are uniformly illuminated and that the light amplitude is normalized to one. The two patterns are recorded using the architecture of Fig. 4.1. After the hologram is developed, suppose the input is $f(x)$; then the output can be found by inserting Eq. 4.10 into Eq. 4.5 to obtain

$$\begin{aligned}
& \Big[\sum_{m=1}^{M} \delta(x - ma) \star \sum_{l=1}^{M} \delta(x - la)\Big] * \sum_{n=1}^{N} \delta(x - nb) \\
= & \Big[\sum_{m=1}^{M} \sum_{l=1}^{M} \delta(x - (m-l)a)\Big] * \sum_{n=1}^{N} \delta(x - nb) \\
= & \sum_{m=1}^{M} \sum_{l=1}^{M} \sum_{n=1}^{N} \delta(x - nb - (m-l)a) \\
= & \sum_{m=0}^{M-1} (M - |m|) g(x - ma).
\end{aligned} \tag{4.12}$$

We see that the read-out is a superposition of shifted versions of the original $g(x)$. Usually these images are mixed up. There are two cases when these images can be separated such that $g(x)$ is recalled correctly. The first case can be seen by inspecting Eq. 4.12. The output consists of the pattern $g(x)$ and its shifted versions, each separated by the period a. If the sampling period of $f(x)$ (viz. a) is larger than the image size of $g(x)$ and we pick up the output only from a window with the size of $g(x)$, then only one $g(x)$ goes through the window, whereas all other shifted

versions will be blocked out. Thus, the first condition for the sampling grids is

$$a > Nb. \tag{4.13}$$

An example of 2-D sampling functions $f(x)$ and $g(x)$ is shown in Fig. 4.5(a). In the figure the sampling spacing of $g(x)$ is set equal to one unit of the spacing between the resolution pixels of the optical devices; i.e., $b = 1$. There are two sampling pixels in each dimension of $g(x)$, i.e., $N = 2$. From Eq. 4.14, the sampling period of $f(x)$ should be equal to or larger than 2. We select $a = 2$. Fig. 4.5(b) illustrates the output of the memory for the input $f(x)$. It is evident that the shifted versions of $g(x)$ are blocked out and the original $g(x)$ is recalled correctly. Similarly, if the input is $g(x)$, then the output will also be recalled correctly. This is shown in Fig. 4.5(c).

The second case that produces a correct associative recall is obtained by inspecting Eq. 4.12. We see that for each input source at $\delta(x-ma)$, the output is the point $\delta(x-nb)$ plus $2(M-1)$ degenerate points that are designated by $\delta(x-nb-(m-l)a)$, where $m, l = 1, \cdots, M, m \neq l$. All the shifted points can be blocked out by using the sampling mask of $g(x)$such that

$$(m-l)a \neq kb \qquad m, l, = 1, \cdots, M, l \neq m;\ k = 1, \cdots, N. \tag{4.14}$$

This condition can be rearranged as

$$\frac{a}{b} \neq \frac{k}{m} \qquad k = 1, \cdots, N;\ m = 1, \cdots, (M-1). \tag{4.15}$$

This is the second sampling grid for planar interconnections. Under this condition, only one pair of associative patterns can be read-out. Fig. 4.6 shows an example. In the figure, $a = 2$, $N, M = 2$. According to Eq. 4.16, $b \neq 1, 2$. We select $b = 3$. Fig. 4.6(b) shows the output pattern associated to the input $f(x)$. It is seen that by using the sampling mask, $g(x)$ is correctly recalled. A similar result is illustrated in Fig. 4.6(c) when the input is the sampling pattern $g(x)$.

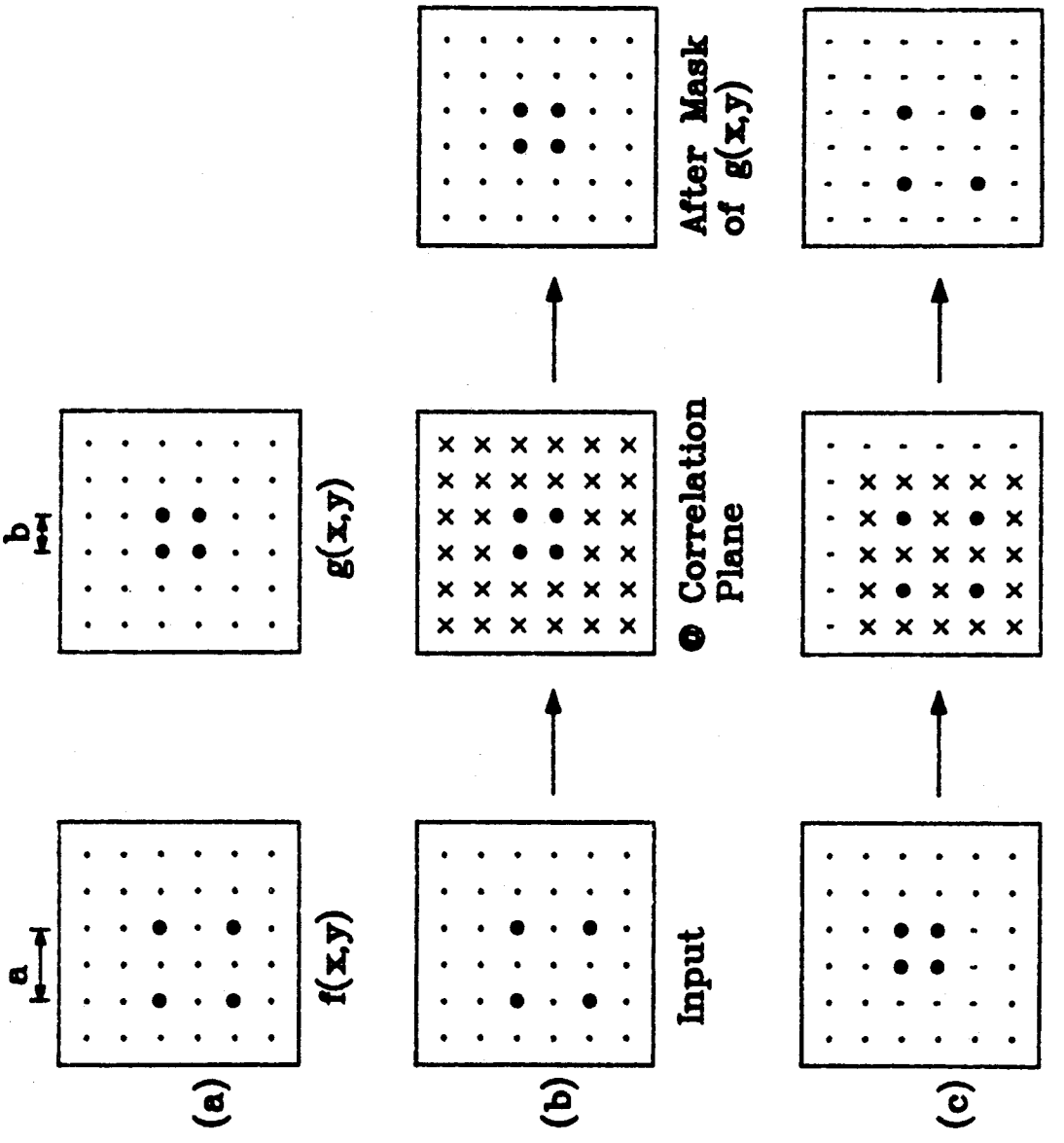


Figure 4.5: Sampling Patterns for Planar Holographic Interconnections. (a) Associated sampling grids f and g. (b) Input f to recall g. (c) Input g to recall f. ($\bullet$ represents the pixels selected for sampling grids. $\cdot$ represents the resolvable pixels. $\times$ represents the degenerate read-outs.)

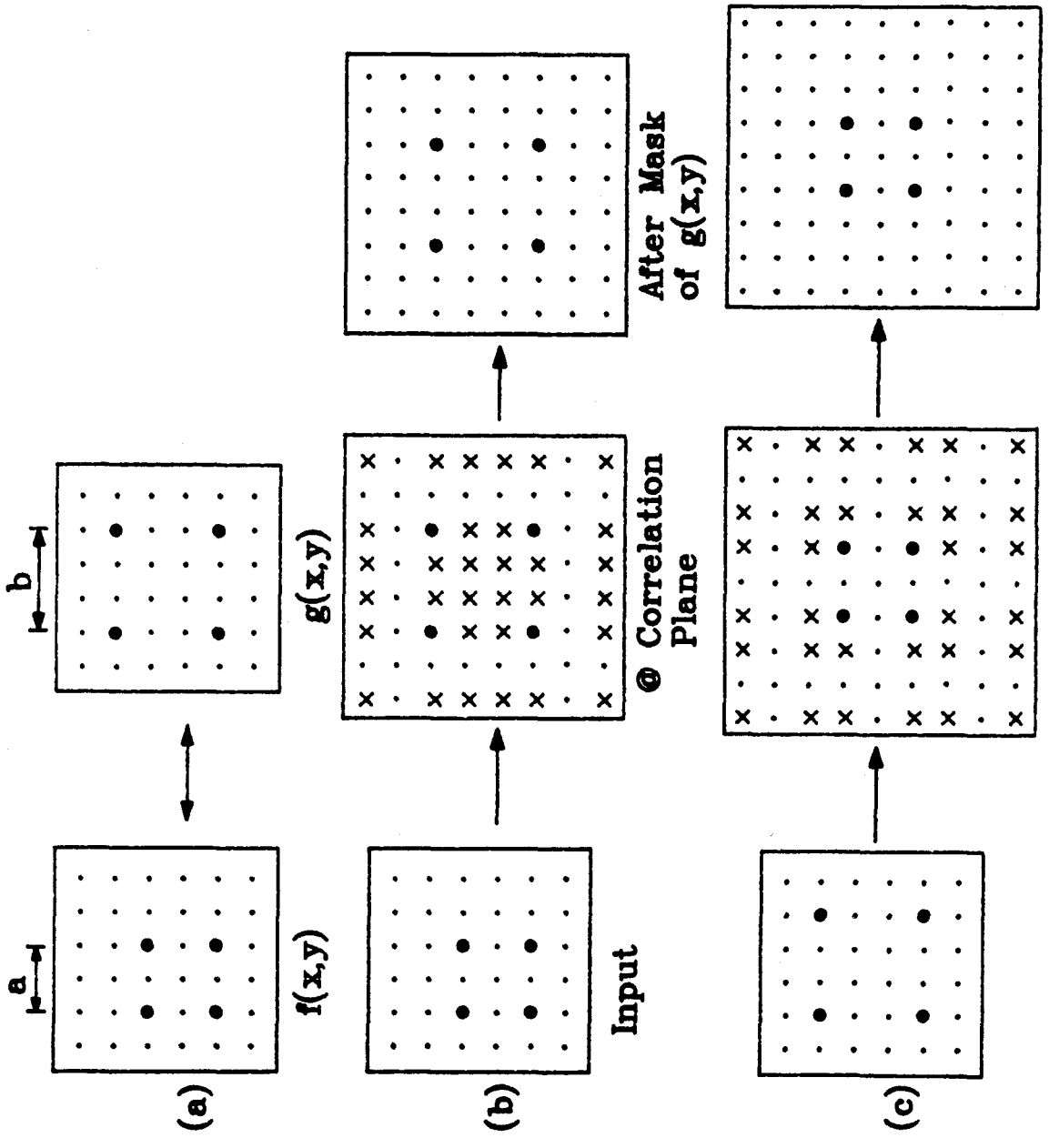


Figure 4.6: Another Sampling Pattern for Planar Holographic Interconnections. (a) Associated sampling grids f and g. (b) Input f to recall g. (c) Input g to recall f.

Thus, if the sampling grids are properly designed, the Vander Lugt correlator can be used as a hetero-associative memory. The next issue that we want to address is how to choose the number of sampling points M and N and how to specify the sampling spacings a and b. The design of these parameters depends on the application and system architecture. For example, for the purpose of image storage, we want to have pixel density as high as possible to obtain high-fidelity sampling of the original images. We choose the sampling grids of the type shown in Fig. 4.5 for this purpose. But the density of pixels that can be used is usually limited by the resolution capability of the image input devices such as spatial light modulators. For some applications, one of the patterns is small, in which case we also choose dense sampling grids such as the type for $g(x)$ shown in Fig. 4.5(a). On the other hand, if both patterns are uniform and of the same size, then we choose the uniform type such as the patterns shown in Fig. 4.6(a). Once a and b are determined, then M and N are determined by the sizes of the image input devices. Furthermore, since the interconnection patterns are stored in holographic gratings, M and N are also limited by the space bandwidth product of the holographic recording material. Suppose M and N are the total number of sampling points of the two patterns, then the total number of independent gratings needed for interconnection is $M \times N$. If the total number of resolution points of a hologram is N_H, then

$$M \times N \leq N_H. \tag{4.16}$$

In summary, we obtain the rules of designing sampling grids for associative memories. The specific design depends on the application and is usually not unique. As an example, in the following experiments we use English alphabets as the associated patterns. Since the patterns have similar sizes, we choose uniform gratings. This means that the grids have similar sampling periods; i.e., $a \approx b$ or $\frac{a}{b} \approx 1$. For a good sampling of an alphabet, we choose ten sampling pixels in each dimension;

i.e., $M = N = 10$. By Eq. 4.16 the prohibited value of $\frac{a}{b}$ which is closest to 1 is $\frac{10}{9}$. We choose $a = 10$ and $b = 11$. Therefore, one of the sampling grids has a sampling period of 10 pixels and the other has 11 pixels. The actual size of the pixel depends on the optical system and imaging device used. For example, the resolution limit of the LCLV described in the next section is about 50 μm; thus, we choose the pixel size to be 100 μm for all of our experiments. The magnified sampling grids based on this design are illustrated in Fig. 4.7. The sampling grids are used in the experiments in this chapter. Fig. 4.8 shows the associative memory experiment using the sampled patterns of **C** and **T**. In the figure (a) shows the two sampled patterns. The two patterns are recorded using the architecture of Fig. 4.1(a). After the recording, when the input is **C**, the read-out pattern is shown in Fig. 4.8(b). It is seen that the output is the mixture of all shifted patterns of **T**. If the sampling grid of **T** is superimposed on this pattern, we obtain a clean read-out as shown in the final picture. This example illustrates how to design sampling grids for a specific application. The results will be used in the next section for the associative memory loop.

4.3 The Hetero-Associative Memory Loop

As discussed in the previous section, holographic associative memories can be implemented by using suitable sampling grids in the Vander Lugt correlator. The results are used in this section for the holographic hetero-associative memory feedback loop. The memory loop has the capability of recalling the complete associative image from a partial input. Similar to the auto-associative loop described in the previous chapter, the most interesting property of this loop is its dynamics. In this section we present the experimental results of the effects of optical gain and loss on the growth and decay of the loop signal. The neural network model of the system

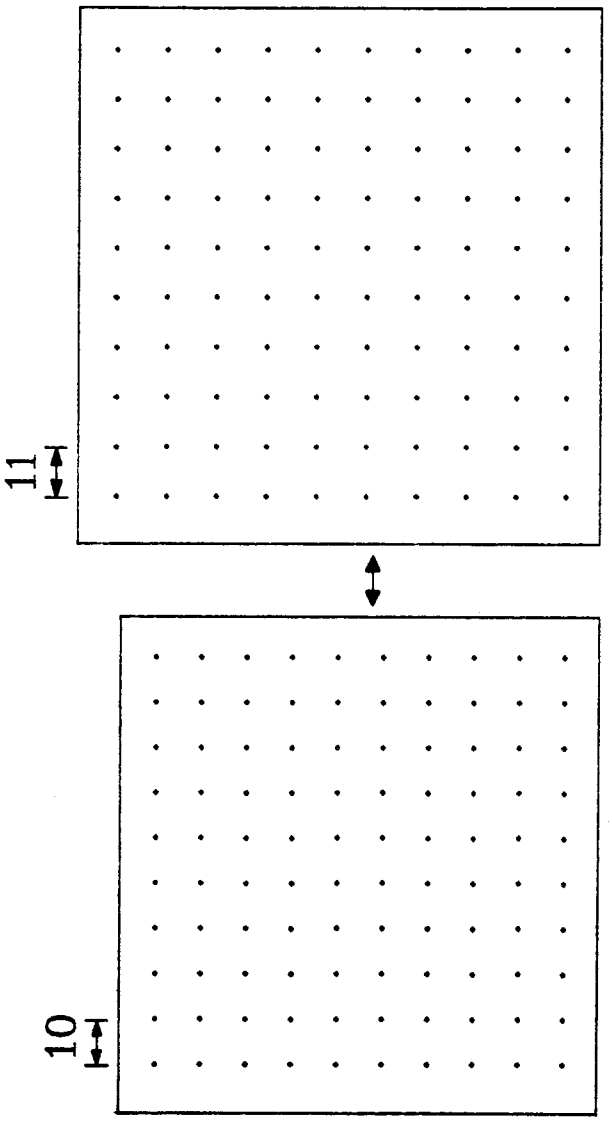


Figure 4.7: Sampling Grids for the Optical Experiments in this Chapter.

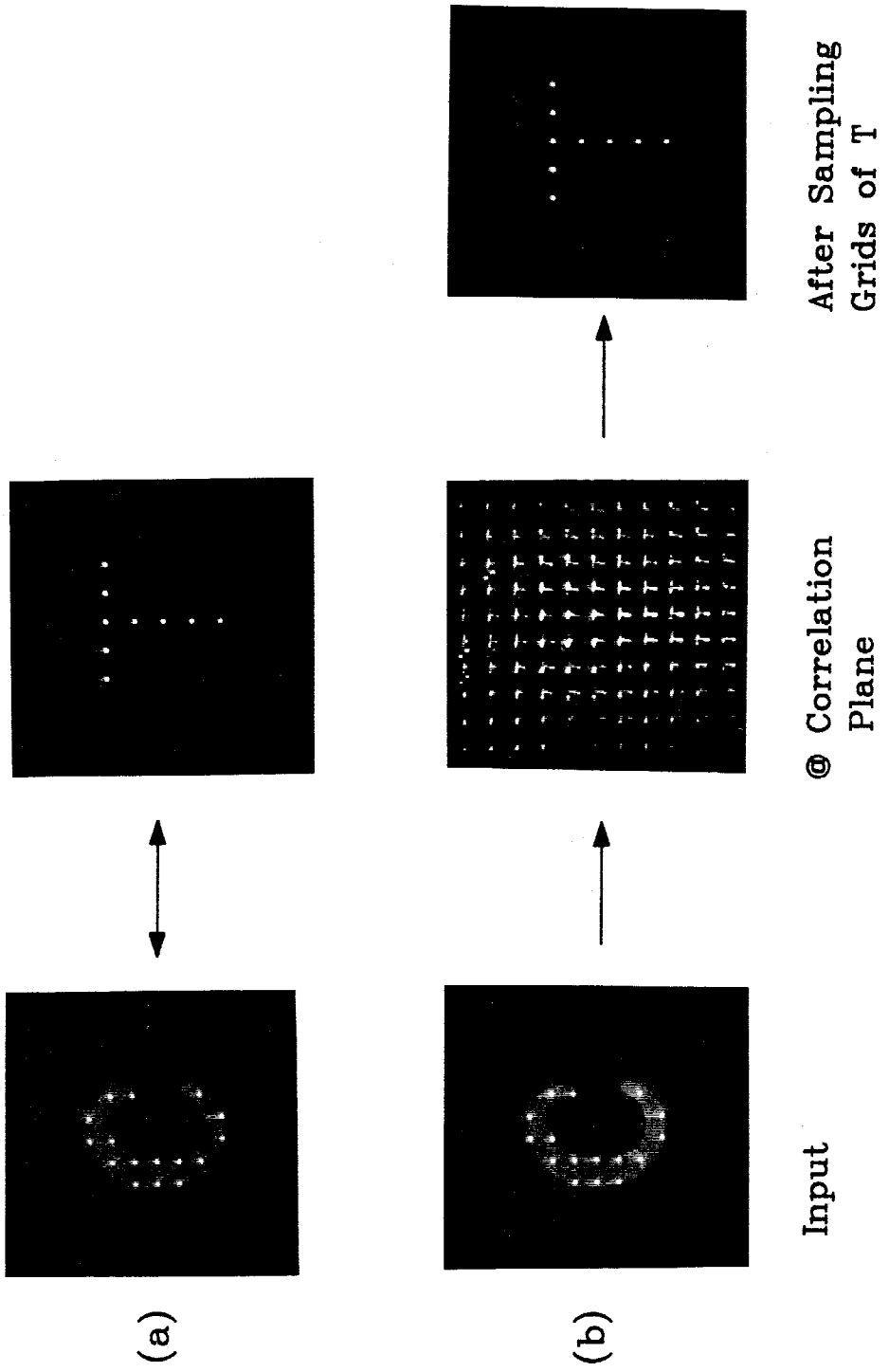


Figure 4.8: Hetero-Associative Experiment. (a) Two associated patterns. (b) Recalling the correct pattern using the sampling grids.

and its dynamics will be discused in the next section.

Fig. 4.9 shows the schematic diagram of the holographic loop, and a photograph of the experimental apparatus is shown in Fig. 4.10. In the system shown in Fig. 4.10, the LCLV at plane P_1, the beam splitter cube $BS1$, the lenses L_2, L_3, and the hologram H form one arm of the correlator. An identical input I_2 and the lens L_5 form the second arm of the correlator. A collimated argon laser beam illuminates the read-out side of the LCLV through beam splitters BS_4 and BS_1. The reflected light from the LCLV is reflected by BS_1 into the loop and is Fourier-transformed by lens L_2 to illuminate hologram H. A portion of the loop light is reflected by BS_2 and is imaged by lens L_5 onto a CCD television camera. This provides real-time monitoring of the activity of the system. A portion of the illuminating argon laser is reflected by BS_4 to illuminate the input I_2. The input I_2 is Fourier-transformed by lens L_4 and illuminates H. The lens L_3 gives the inverse Fourier-transform of H and thus images the input I_2 onto the writing side of the LCLV. This image is superimposed with the image from I_1 and forms a closed feedback loop.

We now describe the operation of the loop. There are five shutters, S_1, S_2, S_3, S_4, and S_5 in the system for controlling the light beams. The two patterns in each associative pair are arranged side by side as shown in the figure. In the recording process, the association from f to g is first recorded on the hologram H. During this process, only shutters S_1 and S_4 are open. The recorded hologram interconnects the f pattern of the reading side of the LCLV to the g pattern of the writing side of the LCLV. The optical path is indicated by a dashed line in the figure. We then make a second hologram to interconnect the g pattern of the reading side of the LCLV to the f pattern of the writing side of the LCLV. During this process, shutters S_2 and S_3 are open and the others are closed. This optical path is indicated by the solid line in the figure. Thus, the complete loop for one pair of associations is formed. The associations for other pairs of patterns are also formed in a similar manner and

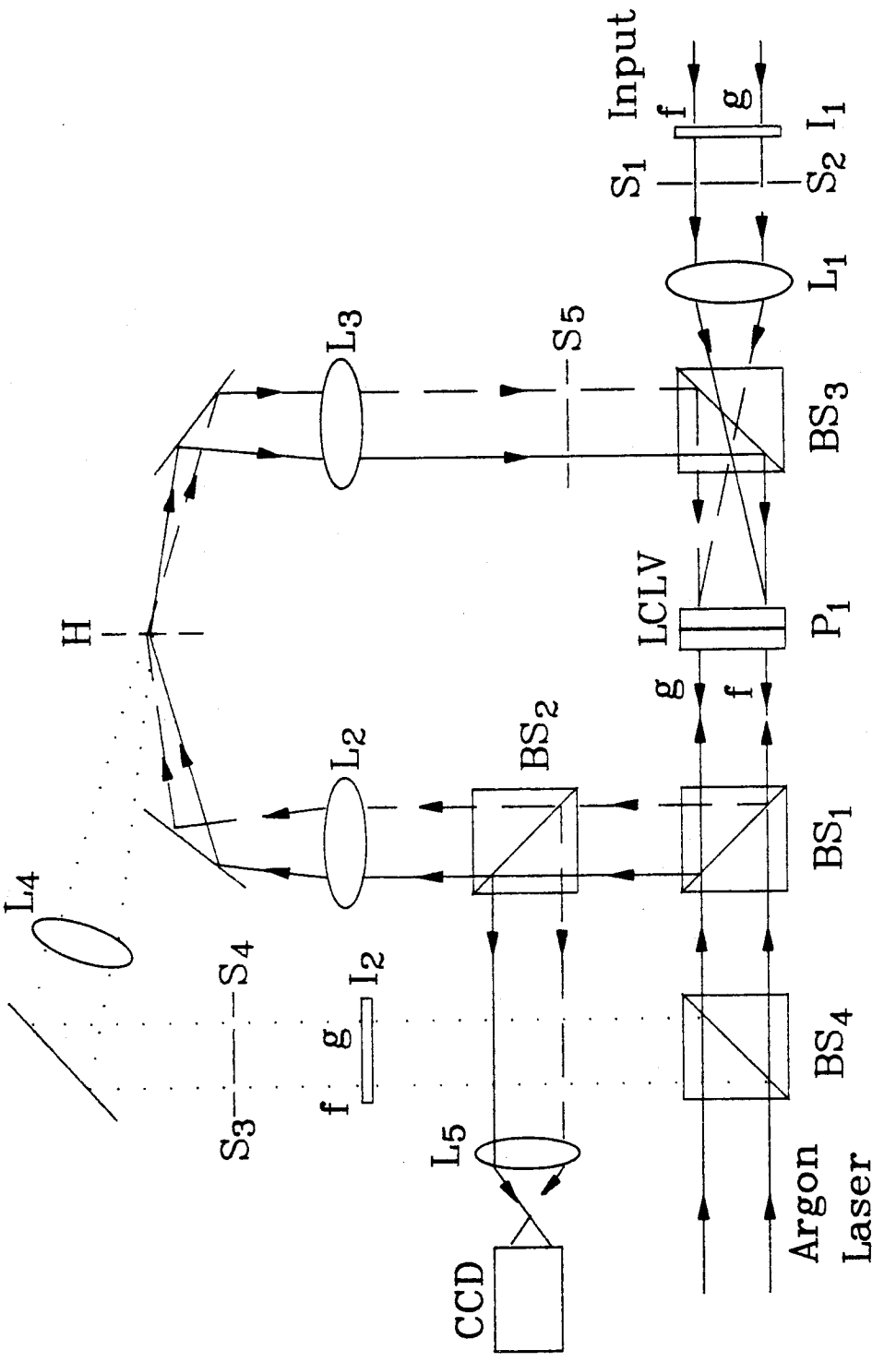


Figure 4.9: Schematic Diagram of the Hetero-Associative Loop.

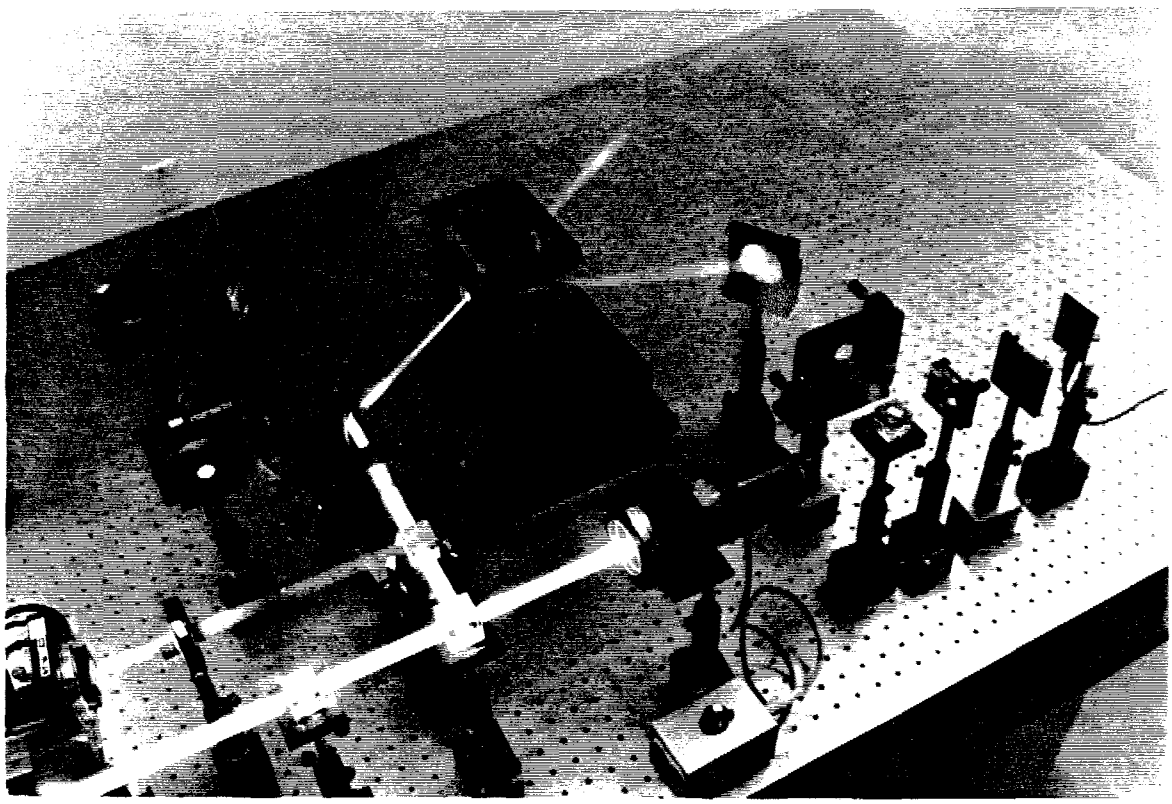


Figure 4.10: Photograph of the Hetero-Associative Loop.

recorded on H.

In the recalling process, shutters S_3 and S_4 are kept closed. The input is fed into the system from I_1 through the shutter S_1 or S_2. The input from I_1 is correlated with the patterns stored in the hologram and reconstructs its associative pattern I_2. The reconstructed pattern is imaged on the writing side of the LCLV as a feedback signal. The feedback loop is closed by turning **ON** the shutter S_5. The loop state then starts to evolve until it reaches a steady state. We then turn **OFF** the input shutter. If the input signal has strong correlation with the stored patterns and if the optical gain of the LCLV is large enough, then the loop will be locked in the steady state. Otherwise, the loop signal dies out gradually. Note that this loop is equivalent to a two-layer network with feedback, because there are two optical paths circulating in the complete loop and each path has a thresholding operation. However, as we described in the previous paragraph, the optical system is designed such that two patterns in an associative pair are spatially separated and are imaged on the LCLV side by side. Therefore, we need only one LCLV in the loop. During operation, suppose the the shutter S_1 is opened to present the input f from I_1 into the system; then the reconstructed output from the hologram is g and is fed back to the LCLV as the second input to the system. Note that g is spatially separated from f; thus, it reconstructs f as the second feedback signal. Therefore, the operation sequence is $f \rightarrow g \rightarrow f \rightarrow \cdots$ and forms a complete feedback loop.

Our first experiment of the loop is the hetero-association between the sampled patterns of **C** and **T**. The sampling grids shown in Fig. 4.7 are used. The interconnections between the two patterns are recorded on H. Now we block I_2 and input a partial image of **C** from I_1. When the feedback loop is **ON**, the associated **T** is recalled and starts to grow. **T** then enters the loop and recalls **C** from the hologram as the second feedback signal. This completes one loop and the next iteration starts. Thus, the two associated patterns reinforce each other in the sequence

$\mathbf{C} \rightarrow \mathbf{T} \rightarrow \mathbf{C} \rightarrow \cdots$ until the loop reaches steady state. If we now turn **OFF** the external input, the pattern stays latched in the loop. The loop output of this operation is shown in the sequence of pictures in Fig. 4.11.

The second experiment of the loop concerns the shift invariance property of the associative memory. The result is shown in Fig. 4.12. As shown in the figure, a partial image of **T** is input from I_1 but is shifted away from the original position of recording. There is no response in the feedback loop, as shown in Fig. 4.12(a). We move the input around until it matches the original position; only at this moment will the signal start to grow. The associated patterns then iterate in the loop until they reach the steady state. Fig. 4.12(b) to (d) shows the loop output of the sequence from the operation. This result is consistent with our prediction. As we explained in the previous section, the operation of the holographic hetero-associative loop is based on the principle of using sampling grids to remove the shift-invariance property of the Vander Lugt correlator. Hence all the shifted signals reconstructed from the hologram are blocked out by the sampling grids. This destroys shift invariance.

As signals circulate in the loop, they decay because of the loss due to diffraction efficiency of the holograms and other optical components. This loss can be compensated by appropriately adjusting the optical gain of the loop as we discussed in the previous chapter. An image intensifier is attached on the writing side of the LCLV so that the gain adjustment can be made by adjusting the bias voltage of the image intensifier.

In fact, gain and loss are the main factors that determine the dynamics of the feedback loop. In what follows, we examine the dynamic behavior of the loop. Two pairs of associated patterns, $\mathbf{C} \leftrightarrow \mathbf{T}$ and $\mathbf{A} \leftrightarrow \mathbf{O}$, are stored in the hologram H. The sampling grids designed in the previous section are used here. The sampled patterns of the two associated pairs are shown in Fig. 4.13. Note that there is a four pixel overlap between **A** and **C**, and a one-pixel overlap between **O** and **T**.

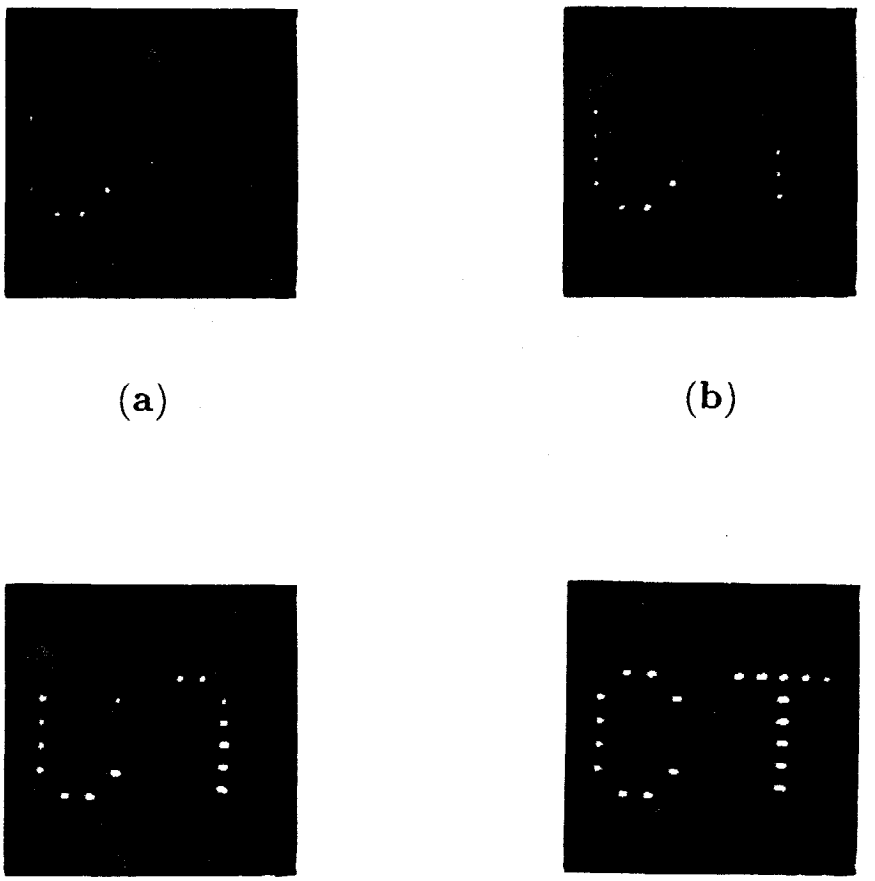


Figure 4.11: Dynamics of the Hetero-Associative Loop. (a) Partial input of **C** at $t = 0$. (b) $t = 450\ ms$. (c) $t = 750\ ms$. (d) $t = 4$ sec.

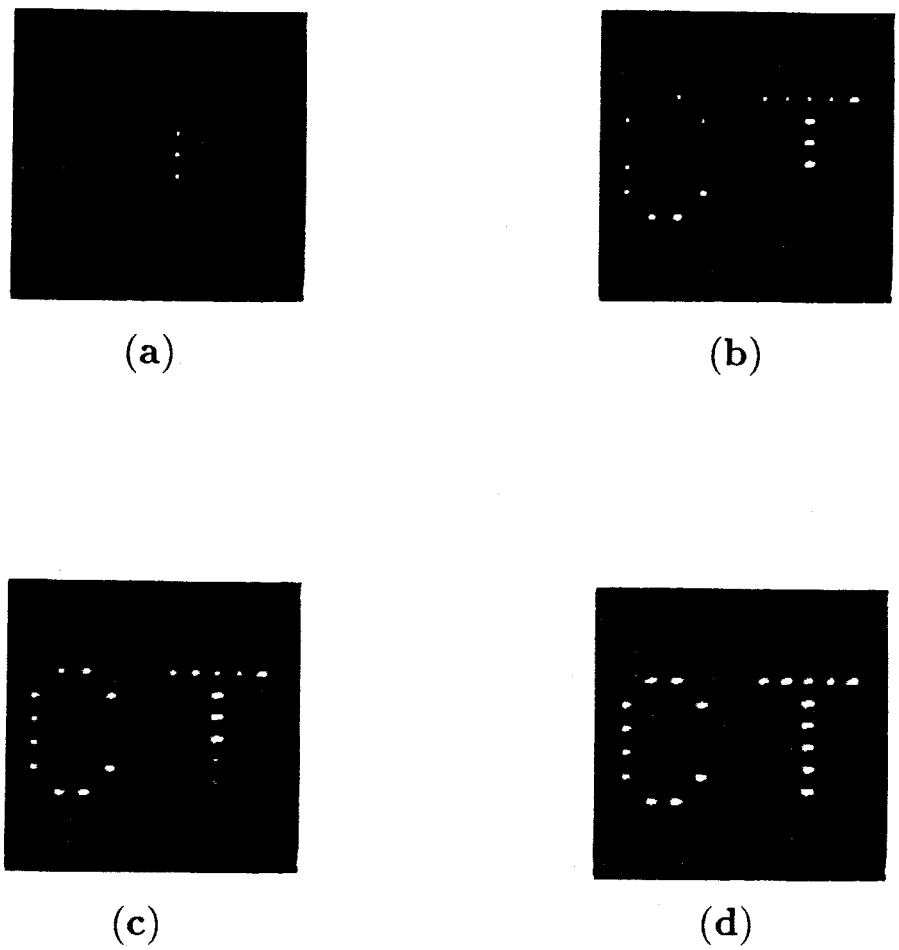
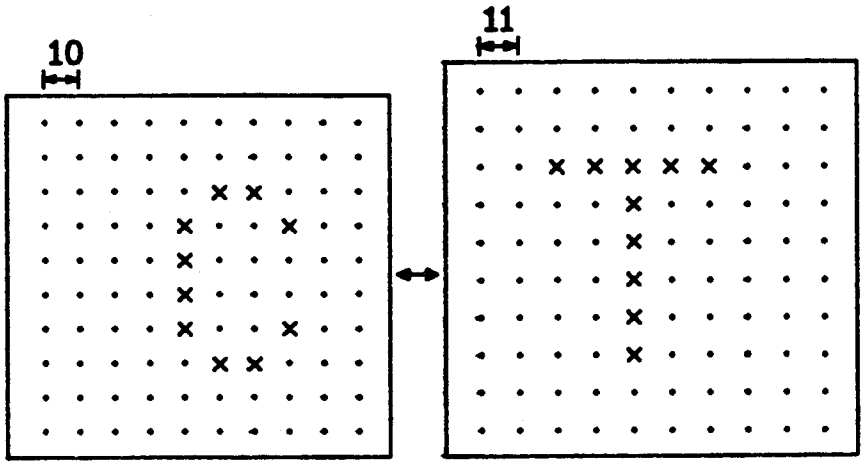
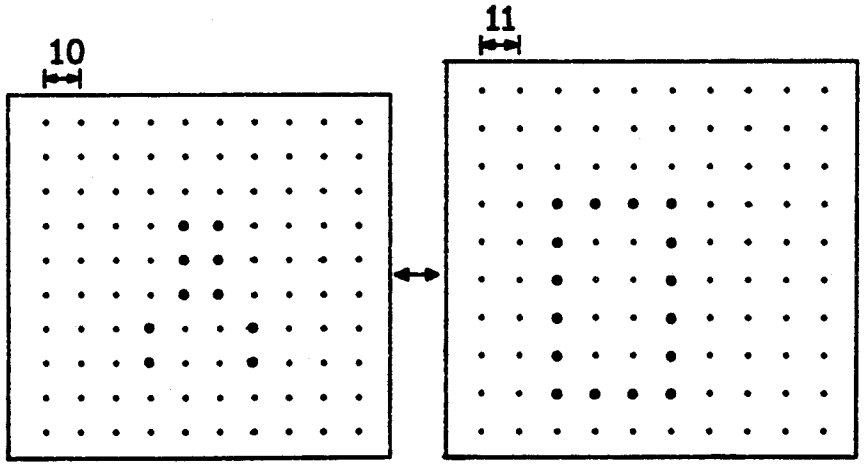


Figure 4.12: Dynamics of the Hetero-Associative Loop. (a) Partial input of **T** at $t = 0$. (b) $t = 450\ ms$. (c) $t = 750\ ms$. (d) $t = 4$ sec.



(a)



(b)

Figure 4.13: The Sampled Patterns Stored in the Optical Loop. (a) The **C** $\leftrightarrow$ **T** pair. ($\times$'s represent the sampled points.) (b) The **A** $\leftrightarrow$ **O** pair. ($\bullet$'s represent the sampled points.)

In recording, we made the Fourier-transform holograms for the **A** $\leftrightarrow$ **O** pair with diffraction efficiency higher than that of the holograms for the **C** $\leftrightarrow$ **T** pair. This means that the optical gain needed to support the loop of **A** $\leftrightarrow$ **O** is lower than that of the **C** $\leftrightarrow$ **T** pair. Measurements in our experiments show that the minimum gain required for the **A** $\leftrightarrow$ **O** pair is 5×10^4, while that of the **C** $\leftrightarrow$ **T** pair is 1.5×10^5.

In operation, we first set the optical gain at 6×10^4. When **A** is input into the system and the feedback loop is closed, the system evolves to the stable state where the association **A** $\leftrightarrow$ **O** is recalled and latched. The external input is then turned **OFF** but the patterns stays locked in the loop. The output of this sequence is shown in the photographs in Fig. 4.14(a) to (d). Note that the output gives not only the patterns **A** $\leftrightarrow$ **O** but also has a slight mixture of the **C** $\leftrightarrow$ **T** pair. This is expected. It is due to the crosstalk between the original patterns shown in Fig. 4.13. Therefore, when **A** reconstructs its associated pattern which is **O**, at the same time the four crosstalk signal contributes to reconstruct **T**, which is the associated pattern of **C**. Similarly, the crosstalk between **O** and **T** contribute to the reconstruction of **C**. However, there is only one pixel overlap between these patterns; therefore, the reconstructed **C** that is due to this crosstalk is very weak. Since the optical gain is not enough to sustain the **C** $\leftrightarrow$ **T** pair, the crosstalk signals do not grow. If we now increase the optical gain from 6×10^4 to 1.5×10^5, the intensities of **C** and **T** also increase until the pair is latched. The mixed state of the two associated pairs co-exist in the loop. The output of this state is shown in Fig. 4.14(e). If we decrease the optical gain back to 6×10^4, the loop returns to the previous **A** $\leftrightarrow$ **O** state, as shown in Fig. 4.14(f). This experiment shows that optical gain is the key factor that determines the loop state.

We consider next the loop behavior when the initial optical gain is high. We set the gain at 1.5×10^5. When **C** is input into the system, the loop evolves to a stable state where both **C** $\leftrightarrow$ **T** and **A** $\leftrightarrow$ **O** are locked in the loop. Fig. 4.15(a) to (d)

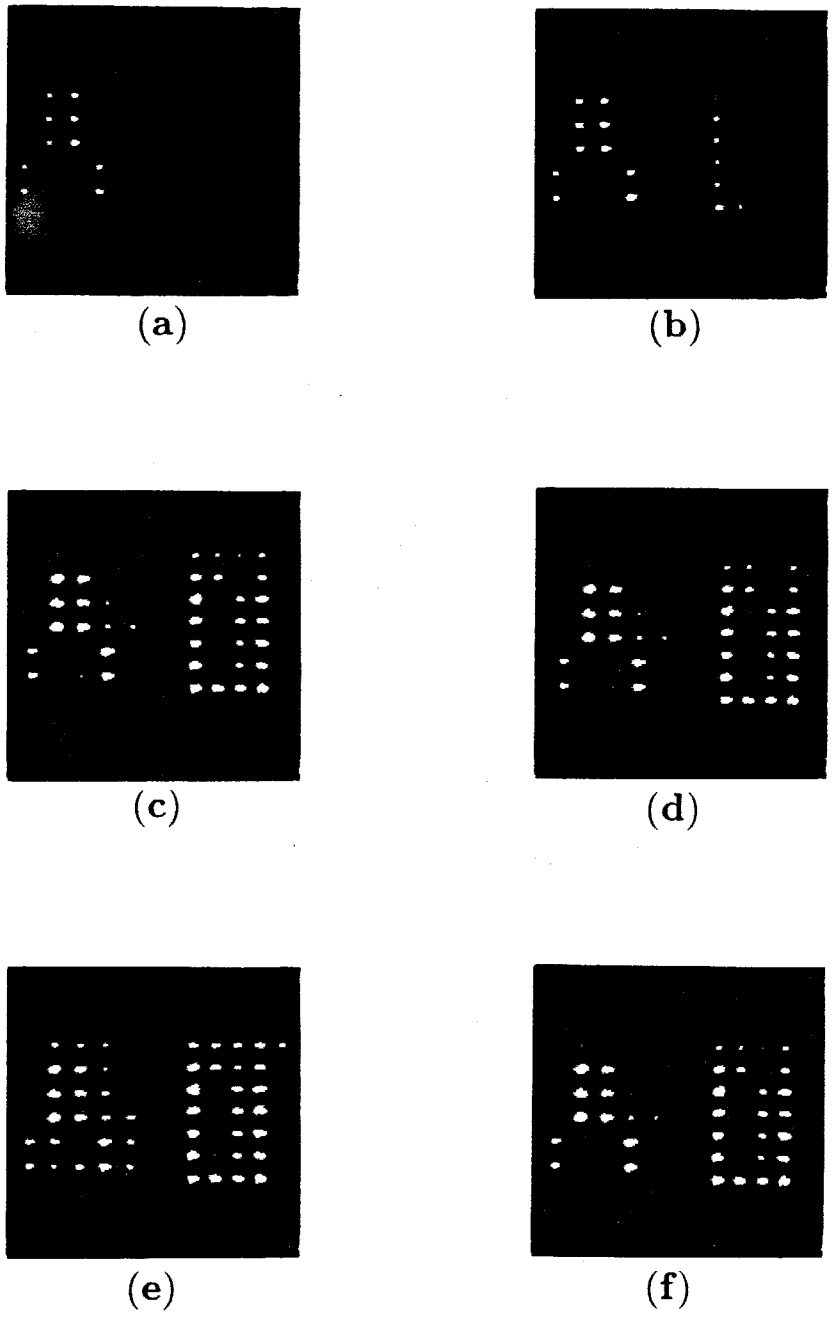


Figure 4.14: The Dynamics of the Loop at Gain = 6×10^4. (a) Input **A** at $t = 0$. (b) $t = 600$ *ms*. (c) $t = 1$ sec. (d) Loop with the input OFF. (e) Gain increased to 1.5×10^5. (f) Gain is decreased to= 6×10^4.

shows the output of this evolution. We see that the stable state is similar to that of Fig. 4.14(d), which has the same gain. The **A** $\leftrightarrow$ **O** pair in this case is recalled through the crosstalk between the pairs. Since the **A** $\leftrightarrow$ **O** pair needs lower gain to sustain the loop, we expect that as the gain is reduced it will remain latched, while the **C** $\leftrightarrow$ **T** pair will decay. Fig. 4.15(f) shows the output when the gain is reduced to 6×10^4. It is seen that indeed only **A** $\leftrightarrow$ **O** survives in the loop. Again, this experiment shows that gain determines the dynamics of the associative loop.

The preceding experiments show the problem of crosstalks in associative memories. As we have seen, as long as the original patterns have overlap, recalling one associative pair may also induce extra associative patterns from the other pairs. The extent depends on the amount of overlap in the original patterns and also on the optical gain of the system. The appearance of mixed states can be reduced by careful control of the gain, but in general, the stable state of the loop is a mixed state of patterns that have crosstalks.

Finally, let us consider one more characteristic of the interconnections of the memory loop. Since each point of one pattern is connected to every point of its associative pattern through independent interconnections, as long as one point has input, its associated pattern will be recalled by this point and is independent of the state of all other points. This causes a problem in the memory. Suppose the system is illuminated by a flashlight. Although this is not a correct input, every point of the input is excited and all the memory pairs are recalled. The memory will then converge to a state depending on the specific gain. Fig. 4.16(a) to (d) shows the sequence of the loop operation. The gain is 6×10^4. The initial state is set by a flashlight; the loop then evolves to the state of **A** $\leftrightarrow$ **O**. This result shows a fundamental limit of the hetero-associative memory. It is a restriction that we must take into account when considering the applications of the memory loop.

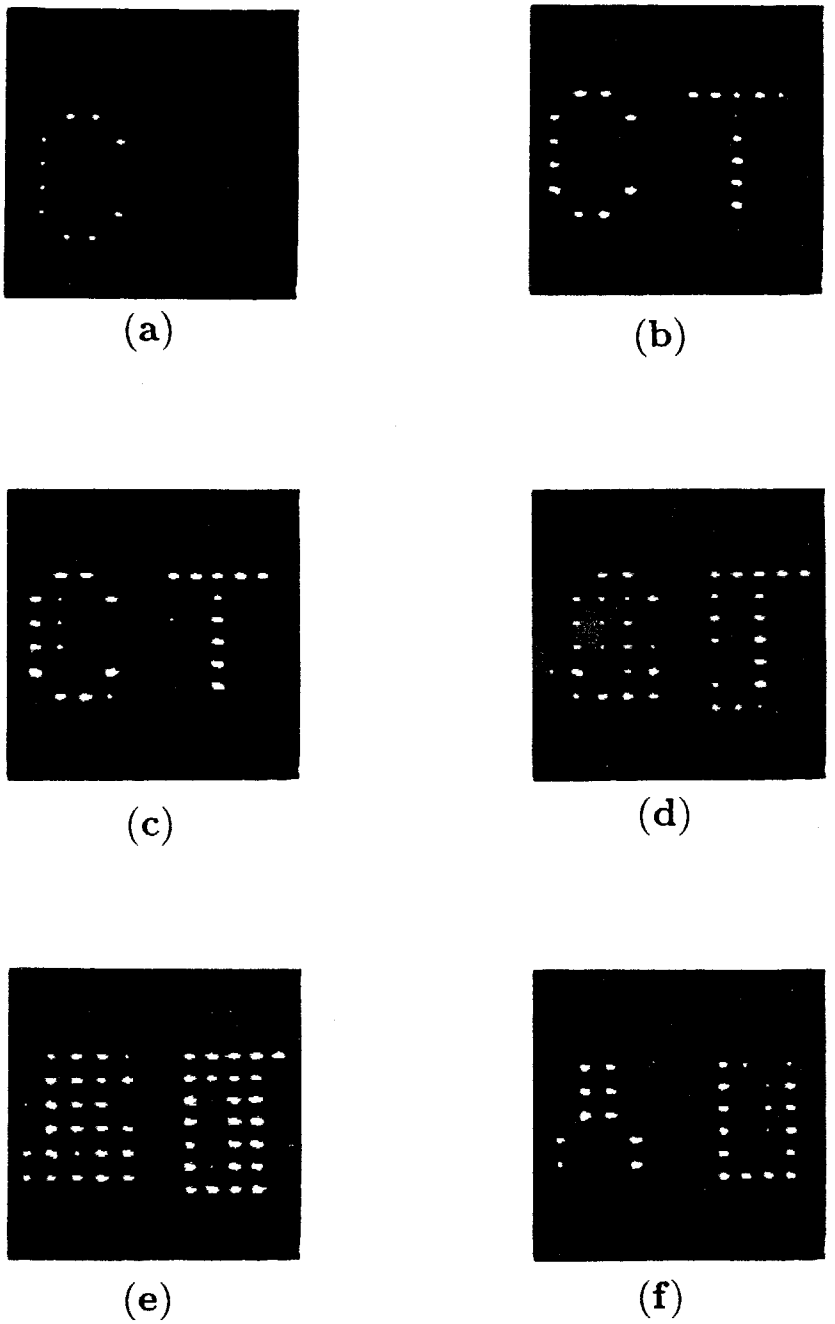


Figure 4.15: The Dynamics of the Loop at Gain = 1.5×10^5. (a) Input **C** at $t = 0$. (b) $t = 750$ *ms*. (c) $t = 1$ sec. (d) $t = 1.5$ sec (Input OFF). (e) Gain increased to 3×10^5. (f) Gain decreased to= 6×10^4.

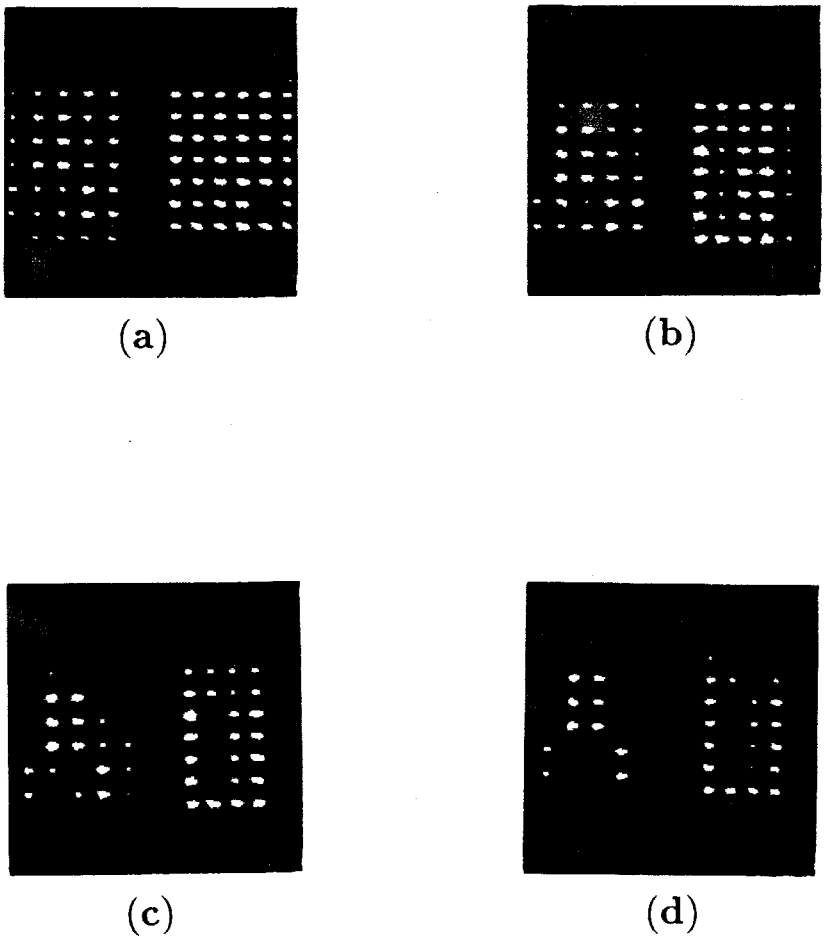


Figure 4.16: The Dynamics of the Loop at Gain $= 6 \times 10^4$. (a) Input at $t = 0$. (b) $t = 600\ ms$. (c) $t = 1.5$ sec. (d) $t = 2$ sec.

4.4 Neural Network Model for the Hetero-Associative Memory

The dynamic behavior of the holographic hetero-associative memory loop demonstrated in the preceding section can be analyzed using a neural network model. The analysis will follow the same line of argument we used in the previous chapter for the auto-associative memory loop, because basically both memory loops are Hopfield-type networks. The main difference is in their feedback architectures, i.e. the interconnection matrices. In the material follows, we will first present a neural network model of the memory loop. Then we derive the dynamic equations of the system. We will construct a vector space corresponding to the stored patterns to investigate the convergence properties of the loop and the stability of the stored memories. A graphic method will be presented which will allow us to visualize the loop dynamics.

To begin, referring to the optical loop shown in Fig. 4.10, we see that the complete loop of the hetero-associative memory is equivalent to a two-layer network, since it takes two steps for the network to finish one iteration, e.g., $\mathbf{C} \rightarrow \mathbf{T} \rightarrow \mathbf{C}$, and since each step passes through a nonlinear operation by the LCLV. The neural network model of the optical loop is illustrated in Fig. 4.17. In the figure, neurons represent the LCLV pixels and holograms represent the interconnections between the associated patterns. The associated patterns $\mathbf{X} \leftrightarrow \mathbf{X}'$ are imaged on the LCLV side by side. $\mathbf{W}$ is the hologram for connecting $\mathbf{X} \rightarrow \mathbf{X}'$, and $\mathbf{W}'$ is the hologram for connecting $\mathbf{X}' \rightarrow \mathbf{X}$. Thus, each neuron of pattern $\mathbf{X}$ receives signals from every neuron of pattern $\mathbf{X}'$, and vice versa. This network is similar to the bidirectional associative memory of Kosko and Guest [6]; Kosko [7]. However, there is a fundamental difference in our optical loop: The neurons respond to light intensity instead of to amplitude. Thus, the slope of the neuron function with respect to the input

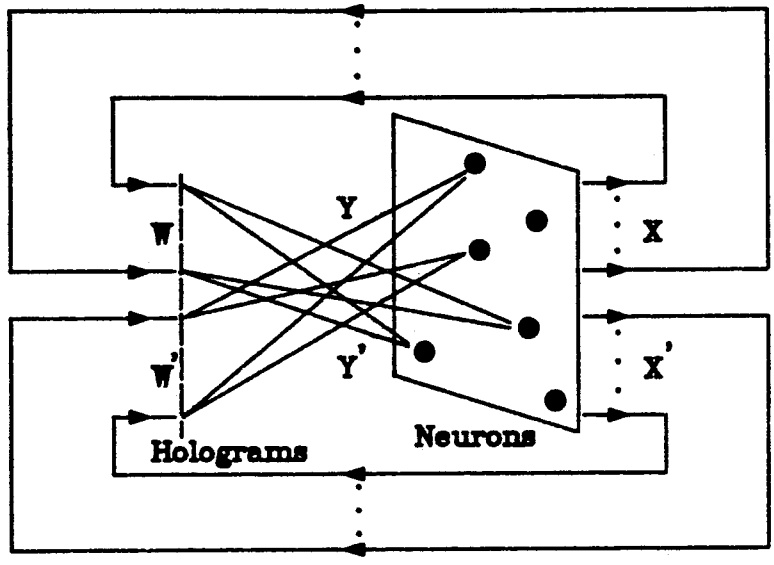


Figure 4.17: Neural Network Model for the Hetero-Associative Loop.

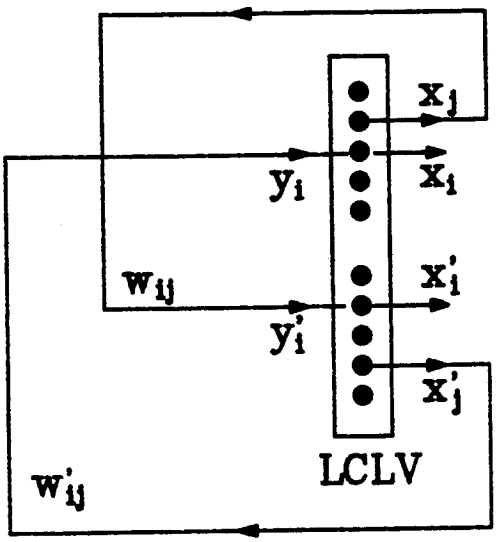


Figure 4.18: Simplified Model for the Hetero-Associative Loop.

signal is not always positive and hence the usual convergence proof does not apply here. Furthermore, the convergence proof does not give the loop dynamics, such as the behavior of the stable states. In what follows we will present a graphic way of inspecting the dynamic properties of the network [8].

For convenience, we simplify the neural network model of Fig. 4.17 to the schematic diagram shown in Fig. 4.18. Suppose the input to the neuron i of the pattern $\mathbf{X}$ is y_i and its output activity is x_i. The dynamic behavior of this neuron can be described by the equation

$$\frac{dx_i}{dt} = -x_i + g(y_i), \qquad i = 1 \ldots N, \tag{4.17}$$

where $g(y_i)$ represents the neural nonlinear function. Since the input y_i comes from outputs of $\mathbf{X}'$ through the holographic interconnections $\mathbf{W}'$, we have

$$y_i = \sum_{i=1}^{N} w'_{ij} x'_j, \tag{4.18}$$

where x'_j is the output of the neuron j of $\mathbf{X}'$, and w'_{ij} is the holographic grating that connects x'_j to y_i. Thus, the dynamic equation of the neuron i becomes

$$\frac{dx_i}{dt} = -x_i + g\Big(\sum_{j=1}^{N} w'_{ij} x'_j\Big), \qquad i = 1 \ldots N. \tag{4.19}$$

Similarly, we obtain the dynamic equation of the neuron i of $\mathbf{X}'$ as

$$\frac{dx'_i}{dt} = -x'_i + g\Big(\sum_{j=1}^{N} w_{ij} x_j\Big), \qquad i = 1 \ldots N. \tag{4.20}$$

We see that the dynamics of the system is determined by the neuron functions $g(x)$ and the interconnection matrices $\mathbf{W}$ and $\mathbf{W}'$. In general, it is very unlikely, although not impossible, to solve the $2N$-coupled nonlinear differential equations analytically. However, following the same line of arguments as in Chapter 3, we can construct a vector space corresponding to the stored patterns to inspect the dynamic properties of the network.

Suppose the associated patterns to be stored are two pairs, $\mathbf{A^1} \leftrightarrow \mathbf{B^1}$ and $\mathbf{A^2} \leftrightarrow \mathbf{B^2}$. The patterns $\mathbf{A^1}$ and $\mathbf{A^2}$ are connected to the layer $\mathbf{X}$, and the patterns $\mathbf{B^1}$ and $\mathbf{B^2}$ are connected to the layer $\mathbf{X}'$. The interconnection holograms are recorded as the outer products of the associative patterns; i.e.,

$$\mathbf{W} = \mathbf{B^1 A^{1T}} + \mathbf{B^2 A^{2T}}, \tag{4.21}$$

and

$$\mathbf{W}' = \mathbf{A^1 B^{1T}} + \mathbf{A^2 B^{2T}}. \tag{4.22}$$

In what follows, we will first define a vector space corresponding to the patterns $\mathbf{A^1}$, $\mathbf{A^2}$, $\mathbf{B^1}$, and $\mathbf{B^2}$. Assume that there are N neurons in each layer; then every pattern

in $\mathbf{X}$ or $\mathbf{X}'$ can be represented as an N-dimensional vector. Since the neurons detect light intensity, we further assume that the signal from each pixel is positive. Here the phase nonuniformity caused by the LCLV and optical components is neglected. Thus, we are considering the patterns in an $\mathbf{R^N}$ vector space.

Consider the nontrivial case of the hetero-associative memory where the stored patterns are linearly independent. We can decompose $\mathbf{R^N}$ into two subspaces $\mathbf{V_1}$ and $\mathbf{V_2}$, where $\mathbf{V_1}$ is the vector space spanned by the stored patterns and $\mathbf{V_2}$ is normal to $\mathbf{V_1}$, i.e.

$$\mathbf{R^N} = \mathbf{V}_1 \oplus \mathbf{V}_2, \tag{4.23}$$

$$\mathbf{V_1} = span\{\mathbf{A^1}, \mathbf{A^2}, \mathbf{B^1}, \mathbf{B^2}\}, \tag{4.24}$$

$$\mathbf{V_2} = \{\mathbf{y} | \mathbf{x} \cdot \mathbf{y} = 0, \forall \mathbf{x} \in \mathbf{V_1}\}, \tag{4.25}$$

where $\oplus$ means direct sum of the vector spaces. We now define a reciprocal basis $\beta_1 = \{\mathbf{a^1}, \mathbf{a^2}, \mathbf{b^1}, \mathbf{b^2}\}$ for $\mathbf{V}_1$, such that

$$\mathbf{a^i} \cdot \mathbf{A^j} = \delta_{ij}, \tag{4.26}$$

$$\mathbf{b^i} \cdot \mathbf{B^j} = \delta_{ij}, \tag{4.27}$$

$$\mathbf{a^i} \cdot \mathbf{B^j} = \mathbf{b^i} \cdot \mathbf{A^j} \tag{4.28}$$

$$= 0, \qquad i, j = 1, 2. \tag{4.29}$$

Let $\beta_2 = \{\mathbf{e^5}, \mathbf{e^6}, \cdots, \mathbf{e^N}\} \subset \mathbf{V_2}$ be an orthonormal basis for $\mathbf{V_2}$; then

$$\begin{aligned} \beta &= \beta_1 \cup \beta_2 \\ &= \{\mathbf{a^1}, \mathbf{a^2}, \mathbf{b^1}, \mathbf{b^2}, \mathbf{e^5}, \cdots, \mathbf{e^N}\} \end{aligned} \tag{4.30}$$

forms a basis for space $\mathbf{R^N}$. Thus, any pattern in $\mathbf{X}$ and $\mathbf{X}'$ can be expressed in terms of the basis β. Let

$$\mathbf{X} = \alpha_1 \mathbf{a^1} + \alpha_2 \mathbf{b^1} + \alpha_3 \mathbf{a^2} + \alpha_4 \mathbf{b^2} + \alpha_5 \mathbf{e^5} + \cdots + \alpha_N \mathbf{e^N} \tag{4.31}$$

and

$$\mathbf{X}' = \beta_1\mathbf{a^1} + \beta_2\mathbf{b^1} + \beta_3\mathbf{a^2} + \beta_4\mathbf{b^2} + \beta_5\mathbf{e^5} + \cdots + \beta_N\mathbf{e^N}. \tag{4.32}$$

It is shown in the appendix that the dynamic equations (Eq. 4.20 to 21) for $\mathbf{X}$ and $\mathbf{X}'$ can be written in the vector space $\mathbf{R^N}$ as

$$\frac{dx_i}{dt} = -x_i + g(\beta_2 A_i^1 + \beta_4 A_i^2), \tag{4.33}$$

and

$$\frac{dx_i'}{dt} = -x_i' + g(\alpha_1 B_i^1 + \alpha_3 B_i^2). \tag{4.34}$$

It is also shown in the Appendix that, by using Eq. 4.32 and 4.33, the dynamic equations can be further transformed into a set of 2N equations in terms of the coefficients of the β basis. However, among these 2N equations, only the four equations of $\alpha_1, \alpha_3, \beta_2$, and β_4 are coupled together, while the equations for all other coefficients are decoupled, and they are functions of these four variables. Therefore, the dynamic behaviors of the system are completely governed by $\alpha_1, \alpha_3, \beta_2$, and β_4. We need only inspect the dynamic equations of these four variables. The four equations are written in the following

$$\frac{d\alpha_1}{dt} = -\alpha_1 + \sum_i^N A_i^1 g(\beta_2 A_i^1 + \beta_4 A_i^2), \tag{4.35}$$

$$\frac{d\alpha_3}{dt} = -\alpha_3 + \sum_i^N A_i^2 g(\beta_2 A_i^1 + \beta_4 A_i^2), \tag{4.36}$$

$$\frac{d\beta_2}{dt} = -\beta_2 + \sum_i^N B_i^1 g(\alpha_1 B_i^1 + \alpha_3 B_i^2), \tag{4.37}$$

$$\frac{d\beta_4}{dt} = -\beta_2 + \sum_i^N B_i^2 g(\alpha_1 B_i^1 + \alpha_3 B_i^2), \tag{4.38}$$

where A_i^1 is the light amplitude of the ith pixel of pattern $\mathbf{A_1}$, A_i^2 is the light amplitude of the ith pixel of pattern $\mathbf{A_2}$, etc. It can be seen in Eq. 4.36 that α_1 is affected by β_2 and β_4, while β_2 and β_4 are in turn affected by α_1 in Eq. 4.38 and 4.39. Similarly, α_3 and β_2, β_4 also affect each other in the same way. Thus,

mathematically, the four variables form a feedback loop. Eq. 4.36 states that the driving force for α_1 is proportional to the product of each pixel of the stored pattern A_i^1 and the corresponding feedback signal after thresholding $g(\beta_2 A_i^1 + \beta_4 A_i^2)$, and summed over all pixels. The first term of the arguments of the thresholding function $g(x)$ represents the contribution by the pixel A_i and its associated pattern B_1 (since β_2 is proportional to the projection of $\mathbf{X}$ along $\mathbf{B_1}$). If $\mathbf{X}$ has a large projection along $\mathbf{B_1}$ (i.e., β_2 is large), then this term is large and it drives the system toward the state of increasing α_1. The second term in the argument represents the contribution by the associated pair $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ through the overlaps between $\mathbf{A_1}$ and $\mathbf{A_2}$. If state $\mathbf{X}$ has a large projection along $\mathbf{B_2}$ (i.e. β_4 is large), and the stored pattern $\mathbf{A_2}$ overlaps with $\mathbf{A_1}$ at the pixel i (i.e., both A_i^1 and A_i^2 are not zero), then this term represents a force that also drives the system toward increasing α_1. The overall effect of these two terms is to drive the system toward the state $\mathbf{A_1}$. In reality, if the interconnection strength between the associated patterns is strong and if the input has strong correlation with the associated pattern, then the system indeed converges to that stored state. Therefore, the mathematical model matches well with the optical system.

Consider first the simple case where the patterns $\mathbf{A^1} \leftrightarrow \mathbf{A^2}$ have no overlaps and $\mathbf{B^1} \leftrightarrow \mathbf{B^2}$ have no overlaps. Consider the dynamic equation for α_1. Each term of the summation has value only when $A_i^1 \neq 0$ and $g(x) \neq 0$. But since $\mathbf{A^1}$ and $\mathbf{A^2}$ have no overlaps, $A_i^2 = 0$ at the pixel positions where $A_i^1 \neq 0$; i.e., A_i^2 has no contributions to the summation term. Thus, Equation 4.36 can be simplified to

$$\frac{d\alpha_1}{dt} = -\alpha_1 + \sum_i^N A_i^1 g(\beta_2 A_i^1). \tag{4.39}$$

Similarly, the other three equations can be simplified to

$$\frac{d\beta_2}{dt} = -\beta_2 + \sum_i^N B_i^1 g(\alpha_1 B_i^1), \tag{4.40}$$

$$\frac{d\alpha_3}{dt} = -\alpha_3 + \sum_i^N A_i^2 g(\beta_4 A_i^2), \tag{4.41}$$

$$\frac{d\beta_4}{dt} = -\beta_2 + \sum_i^N B_i^2 g(\alpha_3 B_i^2). \tag{4.42}$$

We see that the dynamics are governed by two set of equations, where the first set (Eq. 4.40 and 4.41) describes the dynamics of the $\mathbf{A^1} \leftrightarrow \mathbf{B^1}$ pair, and the other set decribes the dynamics of the $\mathbf{A^2} \leftrightarrow \mathbf{B^2}$ pair. The two sets are un-coupled. Thus, the dynamics of the two pairs are independent of each other. For this reason, it is enough to investigate the dynamics of one pair, e.g. $\mathbf{A^1} \leftrightarrow \mathbf{B^1}$. We define the new variables $\alpha_1 = u$, $\beta_2 = v$, and let

$$\sum_i^N A_i^1 g(\beta_2 A_i^1) = f(v), \tag{4.43}$$

$$\sum_i^N B_i^1 g(\alpha_1 B_i^1) = h(u). \tag{4.44}$$

Then, the dynamic equations for the $\mathbf{A^1} \leftrightarrow \mathbf{B^1}$ pair become

$$\frac{du}{dt} = -u + f(v), \tag{4.45}$$

$$\frac{dv}{dt} = -v + h(u). \tag{4.46}$$

We now use the phase diagram in (u, v) space to illustrate the dynamics of the loop. By definition, $h(u)$ is a linear combination of neural functions under different magnifications. Since $g(x)$ is positive and symmetric with respect to the x-axis, hence $h(u)$ is always positive and symmetric with respect to the u-axis. Fig. 4.19(a) shows one $h(u)$ curve in (u, v) phase space. In the figure, the dashed curves represent components of $h(u)$, $B_i^1 g(\alpha_1 B_i^1)$, as defined in Eq. 4.45. The solid curve represents the total value of $h(u)$. It is seen that the region above the curve $h(u)$ is where $v > h(u)$. By Eq. 4.46 , $\frac{dv}{dt} < 0$. Hence the loop in this region tends to move down toward the $h(u)$ curve. On the other hand, the region below $h(u)$ is where $v < h(u)$; thus $\frac{dv}{dt} > 0$, and thus the state in this region tends to move up toward $h(u)$. The

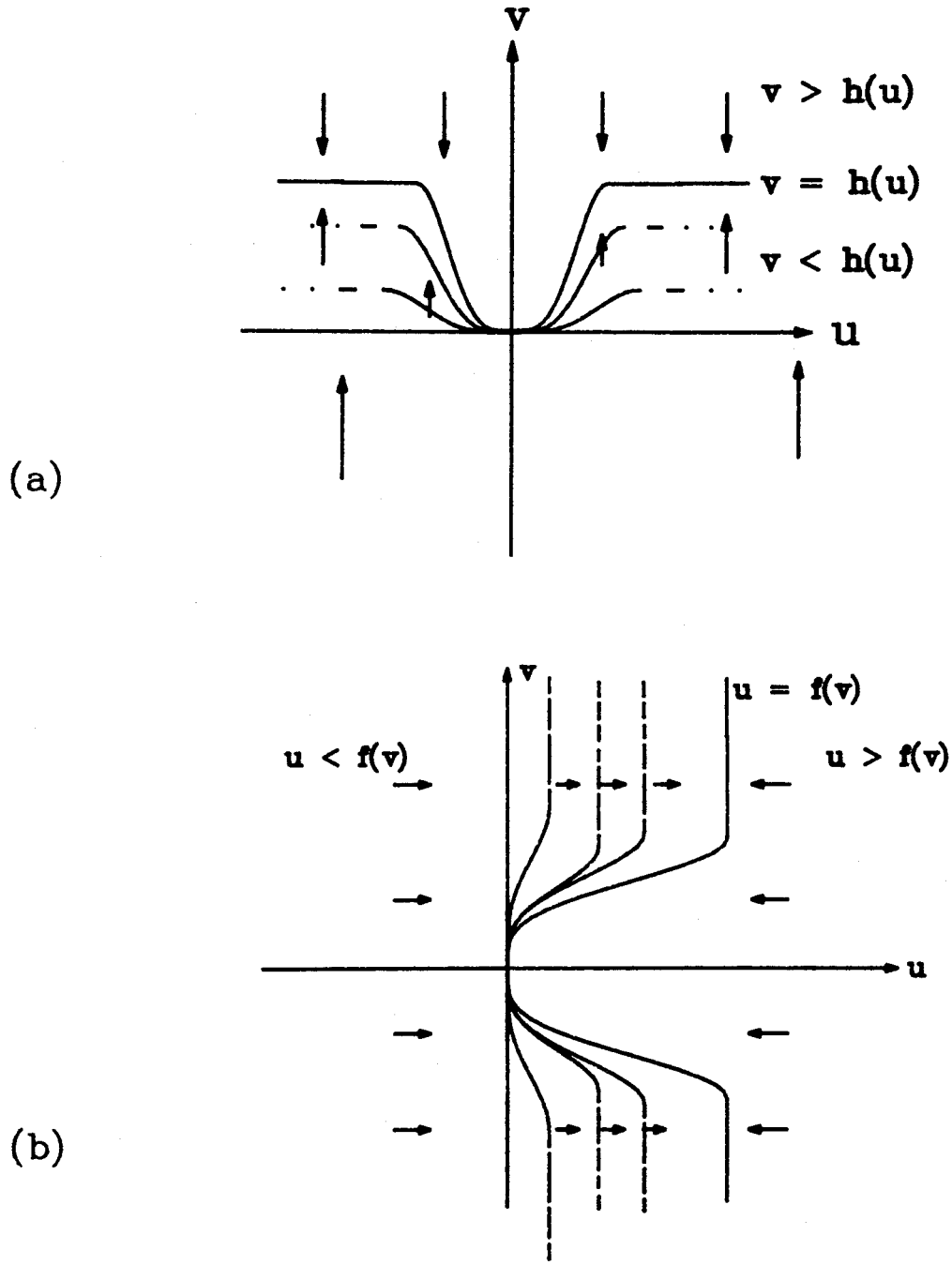


Figure 4.19: The Driving Forces for the Stored Images. (a) The driving force for the pattern $\mathbf{B}_1$. (b) The driving force for the pattern $\mathbf{A}_1$.

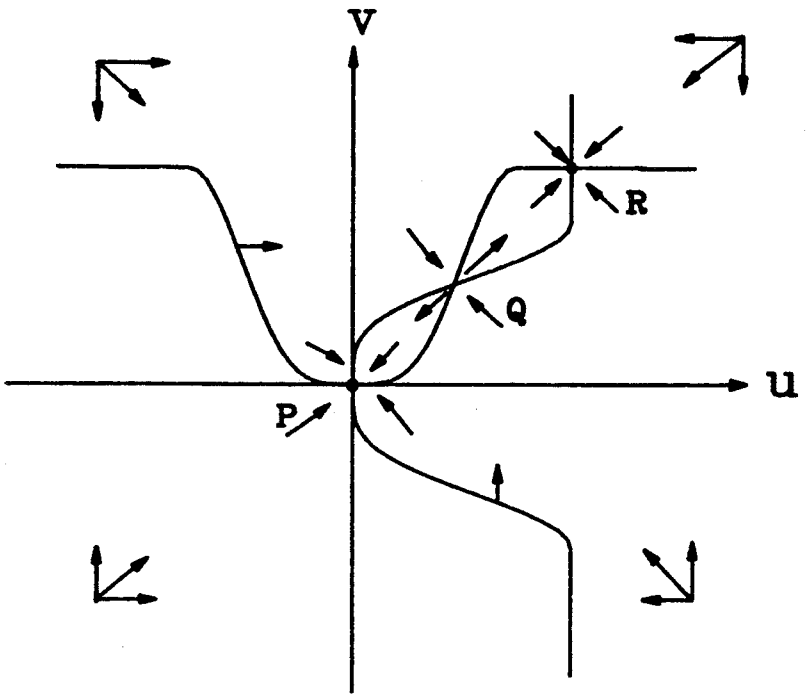


Figure 4.20: The Phase Diagram of the Stored Pair $\mathbf{A}^1 \leftrightarrow \mathbf{B}^1$.

flows of the states are shown by the arrows in the figure. Similarly, the dynamics of Eq. 4.47 is shown in Fig. 4.19(b). The two diagrams are combined in the same phase space of (u, v) as shown in Fig. 4.20.

We see that there are three intersections: P, Q, and R. The point Q is a saddle point where the loop always moves away from this position. The points P and R are stable positions where the loop converges to either one of them. The initial conditions determine which point the loop will converge to. By Eq. 4.32, the initial values of u (i.e., α_1) are set by the projection of the input $\mathbf{X}$ on $\mathbf{a}^1$, which in turn is proportional to the correlation of the input with the stored pattern $\mathbf{A}^1$. Similarly, the initial values of v (i.e., β_2) are set by the correlation of the input $\mathbf{X}'$ with the

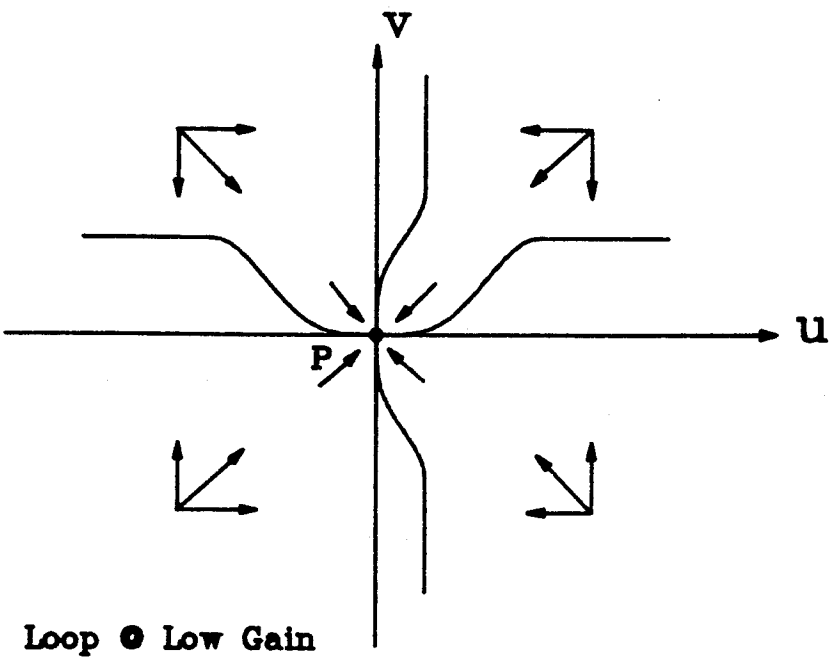


Figure 4.21: Ths Phase Diagram at the Low Gain Condition

stored pattern $\mathbf{B^1}$. The higher the initial values of u and v, the more likely it is that the loop converges to R. On the other hand, if the initial values of u and v are too small, then the loop will converge to P, where it dies out.

When either the neuron gain or the diffraction efficiency of the interconnection gratings is too low, the two curves $f(v)$ and $h(u)$ have only one intersection point at the origin P. This is shown in Fig. 4.21. It is seen that in this case the loop always dies out.

To find the steady state of the loop, we apply Eq. 4.34 and 4.35. Thus, setting $\frac{d}{dt} = 0$, we obtain

$$x_i = g(\beta_2 A_i^1), \tag{4.47}$$

$$x'_i = g(\alpha_1 B_i^1). \tag{4.48}$$

The above equations show that the stable state output of $\mathbf{X}$ is a pure function of $\mathbf{A^1}$, and the stable output of $\mathbf{X}'$ is a pure state of $\mathbf{B^1}$. It also shows that only positions corresponding to the original stored pixels give output; this means that the output pattern has the same shape as the stored pattern. Furthermore, the output patterns in $\mathbf{X}$ and $\mathbf{X}'$ are spatially separated. There is no mixing between the two patterns. Thus, the associated pair $\mathbf{A^1} \leftrightarrow \mathbf{B^1}$ is correctly retrieved. Similarly, the associated pair $\mathbf{A^2} \leftrightarrow \mathbf{B^2}$ can also be correctly retrieved by setting an appropriate initial condition.

Up to this point , we have neglected possible overlap between the stored patterns. Consider next the more general case where $\mathbf{A^1} \leftrightarrow \mathbf{A^2}$ and/or $\mathbf{B^1} \leftrightarrow \mathbf{B^2}$ overlap. The dynamic behavior of the system is described by the set of four coupled equations, 4.36 to 4.39. The dynamic behaviors of the system can be understood by using similar technique as that used for the nonoverlapping case. By changing variables in Equations 4.36 to 4.39, we obtain

$$\frac{du_1}{dt} = -u_1 + f_1(v_1, v_2), \tag{4.49}$$

$$\frac{du_2}{dt} = -u_2 + f_2(v_1, v_2), \tag{4.50}$$

$$\frac{dv_1}{dt} = -v_1 + h_1(u_1, u_2), \tag{4.51}$$

$$\frac{dv_2}{dt} = -v_2 + h_2(u_1, u_2), \tag{4.52}$$

where $u_1 = \alpha_1$, $u_2 = \alpha_3$, $v_1 = \beta_2$, and $v_2 = \beta_4$, and

$$f_1(v_1, v_2) = \sum_i^N A_i^1 g(v_1 A_i^1 + v_2 A_i^2), \tag{4.53}$$

$$f_2(v_1, v_2) = \sum_i^N A_i^2 g(v_1 A_i^1 + v_2 A_i^2), \tag{4.54}$$

$$h_1(u_1, u_2) = \sum_i^N B_i^1 g(u_1 B_i^1 + u_2 B_i^2), \tag{4.55}$$

$$h_2(u_1, u_2) = \sum_i^N B_i^2 g(u_1 B_i^1 + u_2 B_i^2). \tag{4.56}$$

In general, we need 4-dimensional phase diagrams in (u_1, v_1, u_2, v_2) space to describe the system dynamics. However, the system operation can be visualized by inspecting the individual 2-D phase flow of each associated pattern. In the case of two pairs of associative patterns, we use two 2-D phase diagrams with $f_1(v_1, v_2) \leftrightarrow h_1(u_1, u_2)$ for the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair and $f_2(v_1, v_2) \leftrightarrow h_2(u_1, u_2)$ for the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair. Let us first inspect the effects of overlaps on the driving force functions f and h. By Equation 4.56, we see that the function $h_1(u_1, u_2)$ can be de-composed into two terms. One has contributions from B_i^1 pixels that have no overlap with B_i^2; the other one has contributions from the pixels that B_i^1 and B_i^2 are overlapped; i.e.,

$$h_1(u_1, u_2) = \sum_{B_i^2=0}^N B_i^1 g(u_1 B_i^1) + \sum_{B_i^2 \neq 0}^N B_i^1 g(u_1 B_i^1 + u_2 B_i^2). \tag{4.57}$$

The second term of the above equation shows that the effect of the overlapped pixels B_i^2 to $h_1(u_1, u_2)$ is to change its slope. $h_1(u_1, u_2)$ is no longer a symmetrical function of u_1. Furthermore, $h_1(u_1, u_2)$ is not zero at the origin of the u_1-axis; i.e.,

$$\begin{aligned} h_1(u_1, u_2)|_{u_1=0} &= \sum_{B_i^2 \neq 0}^N B_i^1 g(u_2 B_i^2), \\ &\neq 0. \end{aligned} \tag{4.58}$$

The physical meaning of this equation is that even when the input pattern has no components in $\mathbf{B_1}$, there is still a driving force toward $\mathbf{B_1}$. The driving force comes from the overlap between $\mathbf{B_1}$ and $\mathbf{B_2}$ as well as the component of the input on $\mathbf{B_2}$. We also noted that the additional contributions from the overlapped pixels change the shape of the driving force. However, two things remain unchanged. First, the driving force h is always positive. Hence, h remains positive everywhere. Second, the asymptotic value of h remains the same as the nonoverlapping case. This is because no matter how the overlap changes the shape of $g(x)$, it has the

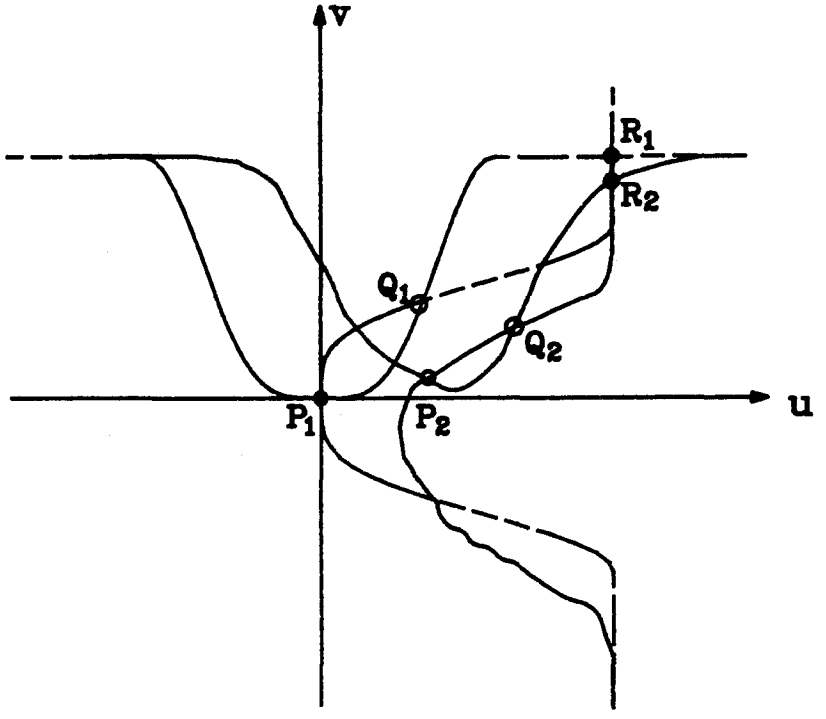


Figure 4.22: Phase Diagram for One Pair of Patterns When the Two Stored Pairs Have Overlaps.

same saturation value for large values of $|x|$. Consider the case where the stored patterns have small overlap. As we start from the nonoverlapped case and increase continuously the portion of overlapping, so will the neural functions g, and hence the driving function h. However, we would expect these changes to be continuous. Therefore, the overall effect of overlaps would change only the shape of the driving function near the origin.

Fig. 4.22 shows the phase diagram for one associative pair. The dashed line represents the nonoverlapping case, while the solid line represents the overlapped case. The figure shows that the minimal points of the functions $h(u_1, u_2)$ and $f(v_1, v_2)$

deviate from the origin. The functions are no longer symmetric with respect to the u_1 and u_2 axes. The intersection points move from P_1, Q_1, and R_1, to P_2, Q_2, and R_2, respectively. We should note that the two curves in the figure correspond to some specific values of u_2 and v_2. The driving forces f_1, f_2, h_1, and h_2 define 3-D hyper-surfaces in the 4-D hyper-space. In general, the intersections for these four hyper-surfaces are isolated points unless one of the surfaces is tangent to the others at the intersection. However, as discussed in the previous paragraph, the distortion of the phase diagram from the nonoverlapped case is continuous and small for small overlaps. Hence, the total number of stable points remains four. The stable states are near to those of the nonoverlapping case. Note that, by Equations 4.50 to 4.53, the origin is still one of the stable states. The actual state where the system converges depends on the neural gain and initial conditions. In what follows, we will draw the phase diagrams for different system conditions to show their stable states and how they are reached.

Consider the case where the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair has strong interconnections (i.e., high diffraction efficiencies in holograms) and the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair has week interconnection strength. Suppose the loop gain is enough to sustain only one pair. Since the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair has stronger interconnections, we expect that under the low-gain condition the loop can sustain only this pair, and the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair will die out. The phase diagram corresponding to this condition is shown in Fig. 4.23. In the figure, the stable state of the system is shown on separated 2-D cross sections of the phase diagrams. Figure (a) shows the phase flow for the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair and (b) is for the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair. The effect of weak interconnections of the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair is that f_2 and h_2 have low values. The result is that there is only one stable point in the phase diagram, P_2, which is very close to the origin. This means that only a weak mixture of $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ exists. On the other hand, stronger interconnection strength of the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair means high values of f_1 and h_1. The result is that there

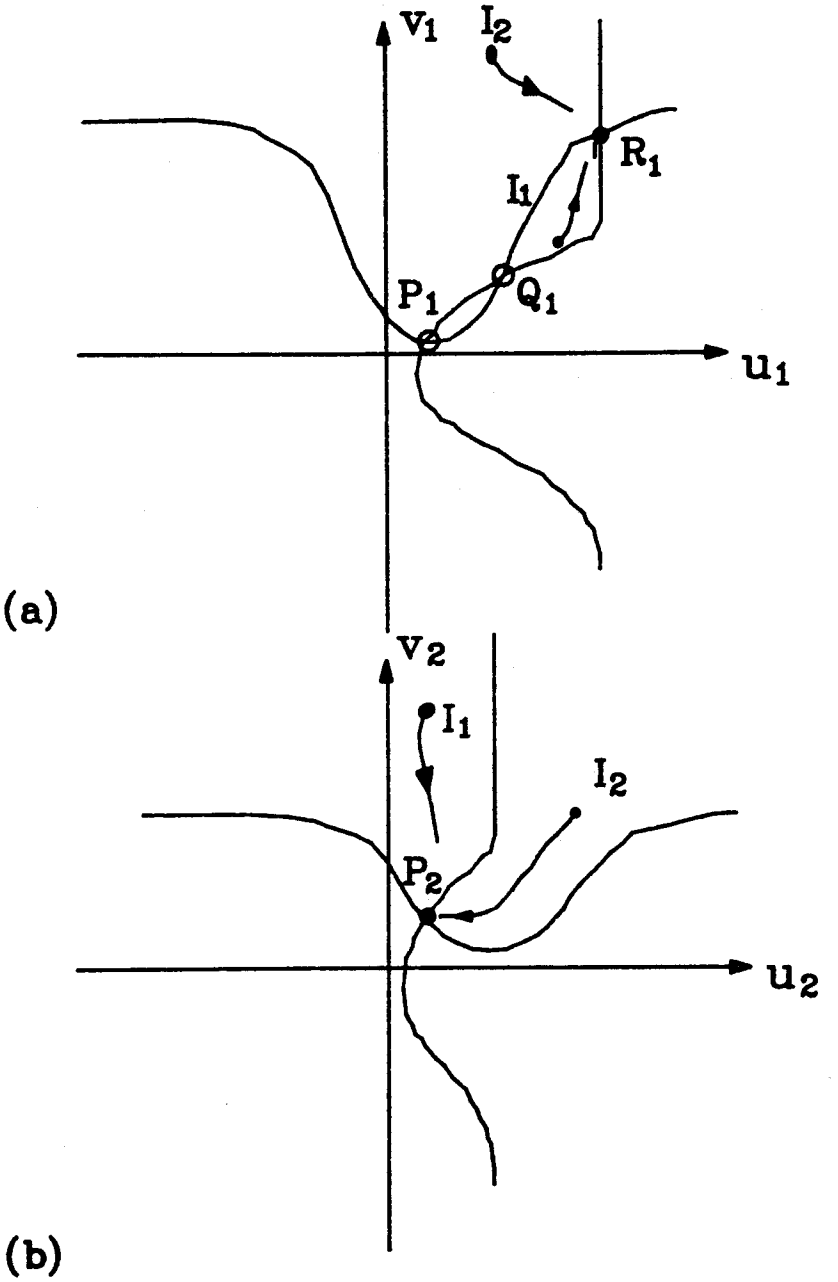


Figure 4.23: Phase Diagram of the Two Associated Pairs. (a) Phase flow of $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ (with strong interconnections). (b) Phase flow of $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ (with weak interconnections).

are three intersection points P_1, Q_1, and R_1, with P_1 and R_1 as the stable points. It is seen that the stable point P_2 for the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair always exists as long as $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ stays at R_1. This means that the total output consists of the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair plus a weak mixing from the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair. As shown in Fig. 4.14(d), the optical experiment confirms this prediction. Note that from the phase diagram in order to reach the point R_1, we should start from initial conditions that have large u_1 or v_1, or both. On the other hand, the initial values of u_2 and v_2 do not affect the system convergence, because it always decays to the point P_2. Note that the small mixing output P_2 comes from the couplings through overlaps between $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ and $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$. The less the overlap the smaller the mixing. The mixing can also be reduced by reducing the neural gain such that P_2 is closer to the origin. However, as the gain is reduced, so are f_1 and h_1. As a result, the intersection point R_1 disappears and the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair decays toward P_1. Then the values of f_2 and h_2 are reducing through the coupling between the two pairs. This further reduces f_1 and h_1. This effect is fedback in the loop until finally both pairs decay to zero.

Next, we consider the case where the neural gain is high enough such that the loop can also sustain the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair. An optical experiment corresponding to this condition was shown in Fig. 4.15. It was shown that the stable output for $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ contains a weak mixture of $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$. As we increase the gain, $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ becomes stronger. When the gain is high enough, the two pairs co-exist in the loop. Then if we reduce the gain, the loop signals become weaker. When the gain is reduced to below the minimal loop gain for $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$, $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ dies out. But $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ stays latched. Thus, by changing the gain we can jump from one memory to the other. This behavior is easily described by using our neural network model. Consider the phase diagrams shown in Fig. 4.24. The figure shows that under high gain conditions the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair also has three intersection points in the phase space, and the loop is driven to the stable point (P_1, R_2) by a suitable initial

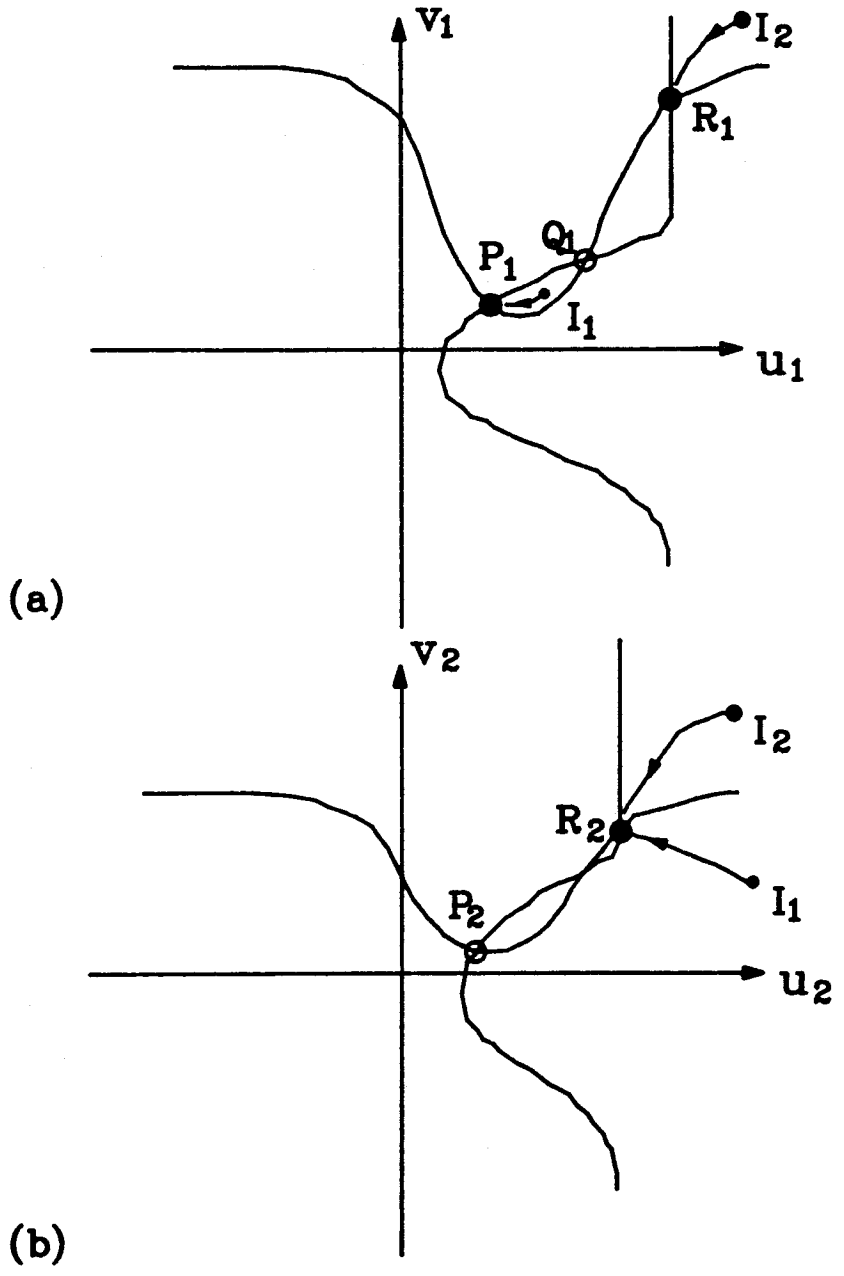


Figure 4.24: Phase Diagram for the Loop with a High Gain.

condition. The loop output is a mixed state of strong $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ and weak $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$. Note the importance of the initial conditions in determining the dynamics of the loop. If, instead of starting from the point I_1, we set the initial condition at I_2, then the loop will converge to the point (R_1, R_2). The output is a combination of $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ and $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ with almost equal brightness. In reality, this occurs if we use a flashlight as the input.

Considering the system at the stable state (P_1, R_2) as is shown in Fig. 4.24. As we increase the gain, the values of f_1, h_1, f_2, and h_2 are also increased, and the intersection points move away from the origin. When the gain reaches a certain high level such that the point P_1 now becomes P and passes over the saddle point Q_1 of Fig. 4.24, the loop is stable at the point (P, R_2). However, the state P now is so bright that the complete pattern of $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ appears in the loop. Thus, both pairs are locked in the loop. This condition is shown in Fig. 4.25. It is seen that the point P is higher than the saddle point Q of the low gain loop. The output of the optical loop corresponding to this condition was shown in Fig. 4.15(e), which indeed confirms our prediction. Now if we decrease the gain, then the values of f_1, h_1, f_2, and h_2 as well as the loop signal decrease. When the gain is reduced to below the minimal gain for the $\mathbf{A_2} \leftrightarrow \mathbf{B_2}$ pair, this pair dies out. However, since the state P was higher than the saddle-point Q_1, instead of decaying, the $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ pair evolves to the stable state R_1. Thus $\mathbf{A_1} \leftrightarrow \mathbf{B_1}$ stays locked in the loop. This condition is illustrated in Fig. 4.26. The corresponding optical loop was shown in Fig. 4.15(f). Again, the neural network model provides a reasonable description of dynamics of the optical loop.

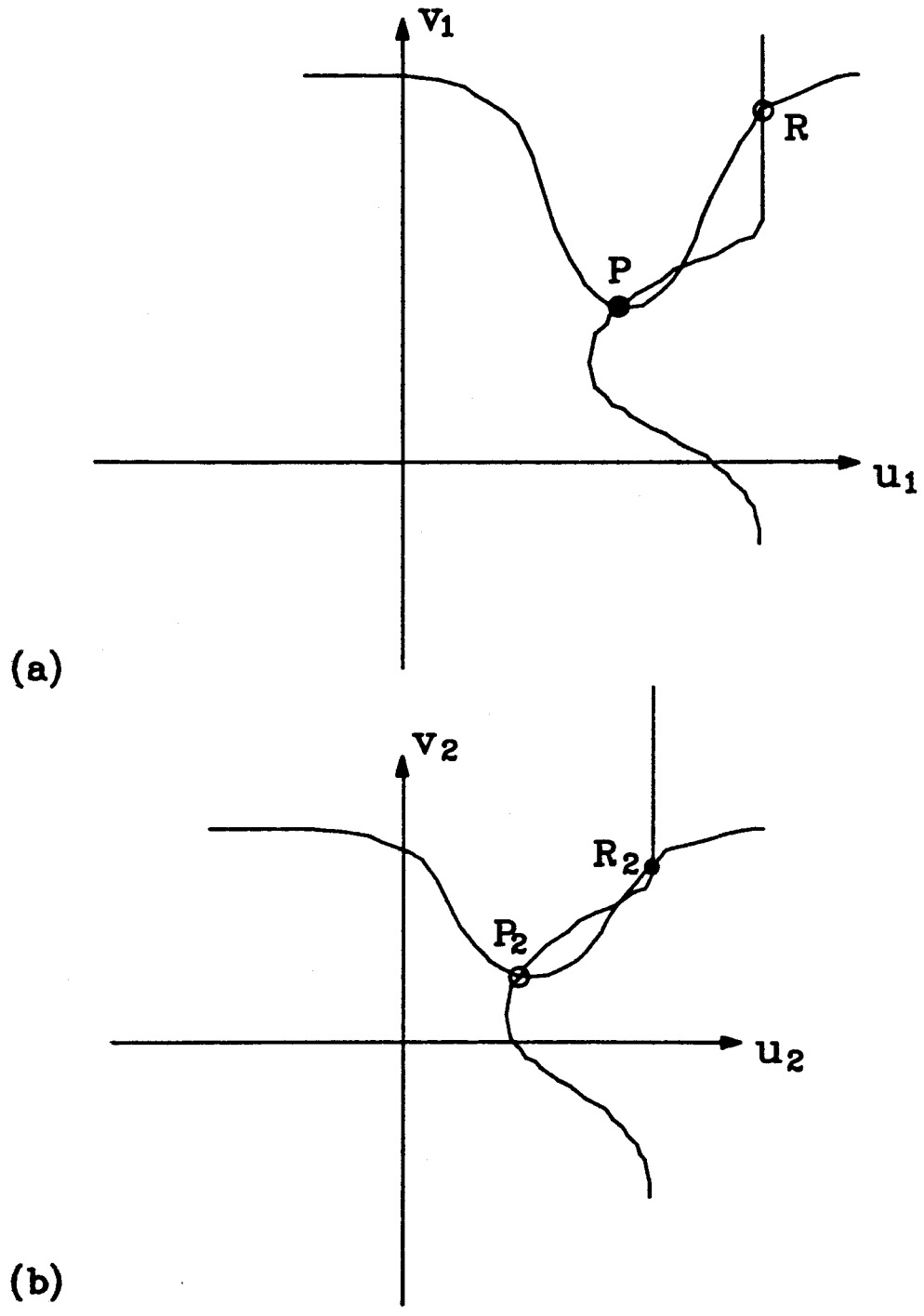


Figure 4.25: Phase Diagram for the Loop with a Very High Gain (Increased from that of Fig. 4.24).

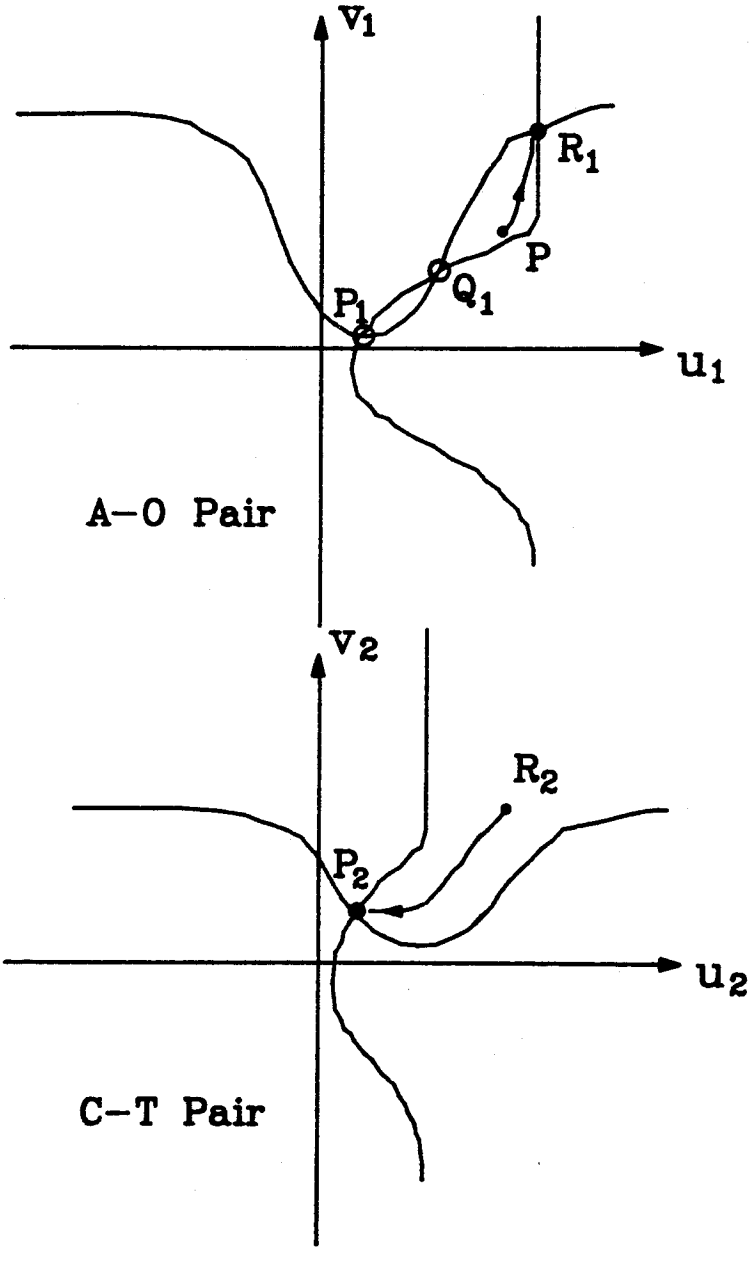


Figure 4.26: Phase Diagram for the Loop when the Gain is Decreased to that of Fig. 4.23).

4.5 Conclusion

In this chapter we have successfully implemented an optical loop that performs the function of hetero-associative memory. We have also developed a neural network model for the optical loop to describe its dynamics and the convergence properties.

The Vander Lugt correlator is modified to provide the means for storing associative memories. Planar Fourier-transform holograms are used for interconnecting the associative patterns. The holograms are recorded by using one of the patterns as the reference wave for recording the Fourier-transform pattern of the other pattern. In operation, input of one pattern retrieves the other as the output. The problem of degenerate reconstructions that are due to the shift invariant property of the Vander Lugt correlator has been analyzed. Given this analysis we have derived two sets of rules for designing sampling grids for associative memories. Under the sampling conditions, all shifted degeneracies in the reconstructed patterns are blocked out by the sampling masks; hence the shift-invariant property of the Vander Lugt correlator is removed. Thus, the grating interconnections between each pair of the associative pixels are unique, and the original patterns can be reconstructed correctly. Optical experiments of associative memories for the sampled patterns have been performed to illustrate the design principle.

We have used the sampled patterns in the optical associative-memory system with feedback. The most interesting property of the optical loop is its dynamic behaviors. The dynamic property of the loop is governed by the gain function and initial conditions of the system. The gain required to sustain a memory in the loop is in turn determined by the interconnection strengths between the associative patterns, which is the diffraction efficiency of the holographic gratings. Once the interconnection holograms are made, the convergence behavior can be controlled only by setting suitable gain and initial external inputs.

In the case of storing nonoverlapping patterns, we can always retrieve the original patterns by presenting enough input of the stored patterns. The recalled patterns are exactly the same as the stored patterns, although the light-intensity distributions may be different. In the case of storing overlapping patterns, the output is always a mixing pattern combining the retrieved patterns and the patterns with which it has overlap. The way to reduce crosstalks is to reduce the gain to as small as possible such that only one pair of associated patterns is locked in the loop. However, the mixed output is not totally avoidable as long as the stored patterns have crosstalks. This characteristics set the storage limits of the optical loop.

We have developed a neural network model to analyze the convergence property of the optical loop. In general, the dynamics of the system can be described by the set of equations that describes the equation of motion of each neurons. In the case of the 2N-neuron optical system, we have 2N nonlinear equations which are coupled together through the interconnections. This is difficult to solve analytically, because the number N is usually large (at least 100 for one image). However, we have shown that we can construct an N-dimensional vector space using each of the stored patterns as one of the coordinates in the vector space. Any input pattern (or the current state of the system) can then be represented using the basis of the vector space, and the 2N dynamic equations are transformed into M equations, where M is the number of stored patterns. The dynamic property of the system then can be visualized by inspecting the phase flow in M-dimensional hyper-space using a graphic method. The method enables us to see how the system converges and where it converges to. If the stored patterns have no overlaps, the original patterns can always be retrieved correctly. If there are overlaps, then the output is always a mixed state of the stored patterns. The model shows that the gain function, interconnection strengths, crosstalks between the stored patterns, and the initial conditions are the factors that determine the dynamic behavior of the system.

4.6 Appendix

Dynamic Equations for the Hetero-Associative Memory

This appendix will focus on the derivation of the dynamic equations for the optical hetero-associative memory loop that has been implemented in this chapter. The neural network model of the optical loop is shown in Fig. 4.17. For simplicity, in what follows we assume that there are two pairs of associative patterns stored in the system. However, the method and the results are applicable to more general cases. Suppose that there are N neurons in each neural plane of $\mathbf{X}$ and $\mathbf{X}'$ layers. The dynamic equations for each neuron were derived in Section 4.4. For convenience, we re-write the equations here:

$$\frac{dx_i}{dt} = -x_i + g\Big(\sum_{j=1}^{N} w'_{ij}x'_j\Big), \qquad i = 1 \ldots N, \tag{4.59}$$

$$\frac{dx'_i}{dt} = -x'_i + g\Big(\sum_{j=1}^{N} w_{ij}x_j\Big), \qquad i = 1 \ldots N, \tag{4.60}$$

where x_i and x'_i are the output activities of the ith neuron in the $\mathbf{X}$ and $\mathbf{X}'$ layers, respectively. w'_{ij} is the holographic grating for connecting x'_j to the input of the ith neuron in $\mathbf{X}$, and w_{ij} is the holographic grating for connecting x_j to the input of the ith neuron in $\mathbf{X}'$. The interconnections provide feedback signals for the loop. It can be written in a matrix form; i.e.,

$$\Big(\sum_{j=1}^{N} w'_{ij}x'_j\Big) = \Big(\mathbf{W}'\mathbf{X}'\Big)_i. \tag{4.61}$$

By Equation 4.23, this signal is determined by the correlation between the current state and the stored patterns, and can be written as

$$\begin{aligned}\Big(\mathbf{W}'\mathbf{X}'\Big)_i &= \Big[(\mathbf{A^1B^{1T}} + \mathbf{A^2B^{2T}})\mathbf{X}'\Big]_i \\ &= \Big[\mathbf{A^1}(\mathbf{B^1}\cdot\mathbf{X}') + \mathbf{A^2}(\mathbf{B^2}\cdot\mathbf{X}')\Big]_i.\end{aligned} \tag{4.62}$$

By Equation 4.33, any pattern in $\mathbf{X}'$ can be expressed in terms of the β basis for the $\mathbf{R^N}$ vector space. Then, by using the orthonormal property (Eqs. 4.27 to 4.30),

we can further simplify the above equation. We obtain

$$\left(\sum_{j=1}^{N} w'_{ij}x'_j\right) = \left(\beta_2\mathbf{A^1} + \beta_4\mathbf{A^2}\right)_i. \tag{4.63}$$

Hence, the dynamic equations for neurons in layer $\mathbf{X}$ now become

$$\frac{dx_i}{dt} = -x_i + g\left(\beta_2 A_i^1 + \beta_4 A_i^2\right), \qquad i = 1 \ldots N. \tag{4.64}$$

Similarly, the dynamic equations for layer $\mathbf{X}'$ can be written as

$$\frac{dx'_i}{dt} = -x'_i + g\left(\alpha_1 B_i^1 + \alpha_3 B_i^2\right), \qquad i = 1 \ldots N. \tag{4.65}$$

Now, since $\mathbf{X}$ can be expanded using the β basis, we may write

$$\begin{aligned} x_i &= \left(\alpha_1\mathbf{e^1} + \alpha_2\mathbf{e^2} + \cdots + \alpha_N\mathbf{e^N}\right)_i \\ &= \left(\sum_j^N \alpha_j e_i^j\right), \end{aligned} \tag{4.66}$$

where $\mathbf{e^1} = \mathbf{a^1}$, $\mathbf{e^2} = \mathbf{b^1}$, $\mathbf{e^3} = \mathbf{a^2}$, and $\mathbf{e^4} = \mathbf{b^2}$. Therefore, the dynamic equation for x_i can be expressed in terms of coefficients of the β basis. We obtain

$$\sum_j^N \frac{d\alpha_j}{dt} e_i^j = -\sum_j^N \alpha_j e_i^j + g\left(\beta_2 A_i^1 + \beta_4 A_i^2\right), \qquad i = 1 \ldots N. \tag{4.67}$$

Multiplying both sides by A_i^1 and sum over i, the above equation becomes

$$\sum_i^N A_i^1\left(\sum_j^N \frac{d\alpha_j}{dt} e_i^j\right) = -\sum_i^N A_i^1\left(\sum_j^N \alpha_j e_i^j\right) + \sum_i^N A_i^1 g\left(\beta_2 A_i^1 + \beta_4 A_i^2\right), \quad i = 1 \ldots N. \tag{4.68}$$

Re-arranging the order of summations, we obtain

$$\sum_j^N \frac{d\alpha_j}{dt}\left(\mathbf{A^1} \cdot \mathbf{e^j}\right) = -\sum_j^N \alpha_j\left(\mathbf{A^1} \cdot \mathbf{e^j}\right) + \sum_i^N A_i^1 g\left(\beta_2 A_i^1 + \beta_4 A_i^2\right). \tag{4.69}$$

By applying the orthonormal relation of $\mathbf{A^1} \cdot \mathbf{e^j}$ (Eq. 4.27), we obtain final the form of the dynamic equation as

$$\frac{d\alpha_1}{dt} = -\alpha_1 + \sum_i^N A_i^1 g\left(\beta_2 A_i^1 + \beta_4 A_i^2\right). \tag{4.70}$$

Similarly, by multiplying Equation 4.68 by A_i^2, B_i^1, B_i^2, and e_i^l, respectively, and going through the same process, we obtain the dynamic equations for other components as well:

$$\frac{d\alpha_3}{dt} = -\alpha_3 + \sum_i^N A_i^2 g\Big(\beta_2 A_i^1 + \beta_4 A_i^2\Big), \tag{4.71}$$

$$\frac{d\alpha_2}{dt} = -\alpha_2 + \sum_i^N B_i^1 g\Big(\beta_2 A_i^1 + \beta_4 A_i^2\Big), \tag{4.72}$$

$$\frac{d\alpha_4}{dt} = -\alpha_4 + \sum_i^N B_i^2 g\Big(\beta_2 A_i^1 + \beta_4 A_i^2\Big), \tag{4.73}$$

$$\frac{d\alpha_l}{dt} = -\alpha_l + \sum_i^N e_i^l g\Big(\beta_2 A_i^1 + \beta_4 A_i^2\Big), \qquad l = 5, 6, \ldots, N. \tag{4.74}$$

Next, starting from Equation 4.66, and going through similar procedures for neurons in $\mathbf{X}'$, we obtain dynamic equations as

$$\frac{d\beta_1}{dt} = -\beta_1 + \sum_i^N A_i^1 g\Big(\alpha_1 B_i^1 + \alpha_3 B_i^2\Big), \tag{4.75}$$

$$\frac{d\beta_2}{dt} = -\beta_2 + \sum_i^N B_i^1 g\Big(\alpha_1 B_i^1 + \alpha_3 B_i^2\Big), \tag{4.76}$$

$$\frac{d\beta_3}{dt} = -\beta_3 + \sum_i^N A_i^2 g\Big(\alpha_1 B_i^1 + \alpha_3 B_i^2\Big), \tag{4.77}$$

$$\frac{d\beta_4}{dt} = -\beta_4 + \sum_i^N B_i^2 g\Big(\alpha_1 B_i^1 + \alpha_3 B_i^2\Big), \tag{4.78}$$

$$\frac{d\beta_l}{dt} = -\beta_l + \sum_i^N e_i^l g\Big(\alpha_1 B_i^1 + \alpha_3 B_i^2\Big), \qquad l = 5, 6, \ldots, N. \tag{4.79}$$

The above 2N equations determine the dynamic behavior of the system. However, from Equations 4.71 and 4.72 it is seen that the driving forces for α_1 and α_3 are only functions of β_2 and β_4. On the other hand, from Equations 4.77 and 4.79 it is seen that the driving forces for β_2 and β_4 are functions of only α_1 and α_3. These four coefficients are coupled together. Since these coefficients are proportional to the projections of the system state on the stored patterns, the physical meaning of the coupling is that the the system dynamics is determined by all the stored patterns through the interconnections and feedback. We note that in the remaining $(2N-4)$

equations, all other coefficients are de-coupled. These coefficients are independent of each other and the driving forces for each of them are functions of only α_1 and α_3, or of β_2 and β_4. Physically, this means that they are in a open loop, and are determined by the four coefficients. In conclusion, the loop dynamics is completely governed by α_1, α_3, β_2, and β_4. As long as we solve the four coefficients, we obtain a system state. The four equations were re-written as Equations 4.36 to 4.39. And as shown in Section 4.4, they are the starting point for studying the convergence property of the loop.

4.7 References for Chapter 4

1. P. J. Van Heerden, "A New Optical Method of Storing and Retrieving Information," *Appl. Opt.*, **2**, 387 (1963).

2. P. J. Van Heerden, "Theory of Optical Information Storage in Solids," *Appl. Opt.*, **2**, 393 (1963).

3. D. Gabor, "Associative Holographic Memories," *IBM J. Res. Devel.*, **13**, 156 (1969).

4. D. Psaltis, J. Yu, X. G. Gu, and H. Lee, "Optical Neural Nets Implemented with Volume Holograms," Paper **TuA3-1**, *Technical Digest of Topical Meeting on Optical Computing*, Optical Society of America, 1987.

5. D. Psaltis, X. G. Gu, and D. Brady, "Fractal Sampling Grids for Holographic Interconnections," *Proceedings on Optical Computing 88*, SPIE Vol. 963, pp. 468-474 (1988).

6. B. Kosko and C. Guest, "Optical Bidirectional Associative Memories," *SPIE Proceedings*, SPIE Vol. 758, Jan. (1987).

7. B. Kosko, "Competitive Adaptive Bidirectional Associative Memories," *Proceedings of the International Conference on Neural Networks*, San Diego, June (1987).

8. H. Y. Li, private communications.

Chapter 5

Photorefractive Holography for Associative Memories

5.1 Introduction

The essential characteristics of the optical neural computers for associative memories have being discussed in Chapters 3 and 4. These include the capability of retrieving stored information from a partial and/or distorted input, the discrimination capability to reject unfamiliar images, the stability of the memory states, and the parameters that determine the dynamics of the system. However, before they can be used in practical systems, there are three more properties that should be addressed: the storage capacity of the memory, the size of the optical system, and the adaptability of the memory. In this chapter we consider these three issues in photorefractive holographic associative memories.

Based on the consideration of signal-to-noise ratio and the statistical property of the Hopfield model, the total number of the memory states that can be stored in the associative memory is shown to be $M < N/(4lnN)$ [1, 2], where N is the number of neurons and M is the number of stored memories. As was described in Chapter 3, $N = 400 \times 400 = 160,000$ in our optical system. Thus, theoretically, the number of images that can be stored in the system is approximately 3,300. Unfortunately, this requires a large array of images at the input plane. As was described in Section 3.3, the images are stored on Fourier-transform holograms. All the images are spatially separated and arranged in a two-dimensional array for writing the holograms. Thus 3300 images require a 60×60 image array for the input. This is very difficult to obtain directly. One way to do this is to reduce photographically the image array

and to record them on a high-resolution plate using the photolithographic method, then to use the plate as the input for recording the Fourier-transform holograms. However, we have to add an image magnification/demagnification scheme into the optical loop so that the resolution requirement at the neural plane does not exceed the capability of the LCLV while the image magnification of the optical loop is kept to unity. The alignment process of this optical system will become very tricky and makes the system impractical. A better way of doing it is to use a multiple exposure scheme so that the LCLV does not move during both the recording and recalling steps. However, the diffraction efficiency of reconstructed images falls off quickly as the number of multiple exposures increases. Therefore, multiple exposure can not be applied directly. We present in this chapter a new method for recording multiply exposed photorefractive volume holograms with improved diffraction efficiency to overcome this problem.

The use of volume holograms for optical neural computers provides another advantage over planar holographic interconnections. Suppose we want to interconnect N neurons in one neural plane to N neurons in the other plane. The total number of interconnections required is $N \times N = N^2$. This requires an interconnection medium having N^2 resolvable spots. If we use planar holograms, then we need a holographic plate that has N resolving spots in one dimension. Suppose the linear size of each resolvable spot is δ; then the linear size of the plate is $N\delta$ and the area is $(N\delta)^2$. Assume that we use a 4-F optical system with f-number equals to 1, then the volume of the system is proportional to $4N\delta \times (N\delta)^2 = 4(N\delta)^3$. Now, if instead, we use volume holograms to record the N^2 interconnections, then the number of resolving spots in one dimension of the holographic medium requires only $(N^2)^{1/3}$ pixels. This requires a linear size of $N^{2/3}\delta$. We also assume that the f-number of the optical system is with the order of 1; then the volume of the system becomes $4(\delta)^3 N^2$. This volume is N times smaller than that of the planar holographic inter-

connection architecture [3]. Usually N is a large number, e.g., 10,000; hence using volume holographic interconnections can reduce the system volume by a huge factor.

The use of photorefractive crystals for holographic interconnections in optical neural networks not only permits the storage of a very large number of interconnections per unit volume [4, 5,], but also makes feasible the implementation of learning algorithms [6]. Based on the dynamic response characteristics of photorefractive materials, the holograms previously recorded are reinforced or erased depending on the new patterns they are exposed to. Thus, the most distinctive features of a set of images will be extracted and recorded in the photorefractive holograms simply by repeatedly exposing the crystal to the training samples. Furthermore, by applying appropriate updating algorithms and exposure schemes, the photorefractive interconnections can be modified continuously so that the network responds to correct input-output mappings [8]. In this chapter we consider applying the adaptability of photorefractive crystals to increase the storage capacity of optical associative memories. Basic characteristics of the photorefractive effect and volume holographic interconnections will be presented in the next section. The optimum scheme for obtaining multiply exposed photorefractive holograms with increased storage and improced diffraction efficiency will be presented in Section 5.3.

5.2 The Photorefractive Effect and Volume Holographic Interconnections

The photorefractive effect is a phenomenon in which a photo-induced space charge field and a corresponding refractive index change can be formed [9]. This effect has been found in a variety of materials. In this chapter we consider the use of $SBN:Ce$ photorefractive crystals for holographic interconnections [4, 5, 6, 7].

The basic mechanism of the photorefractive effect is shown in Fig. 5.1.

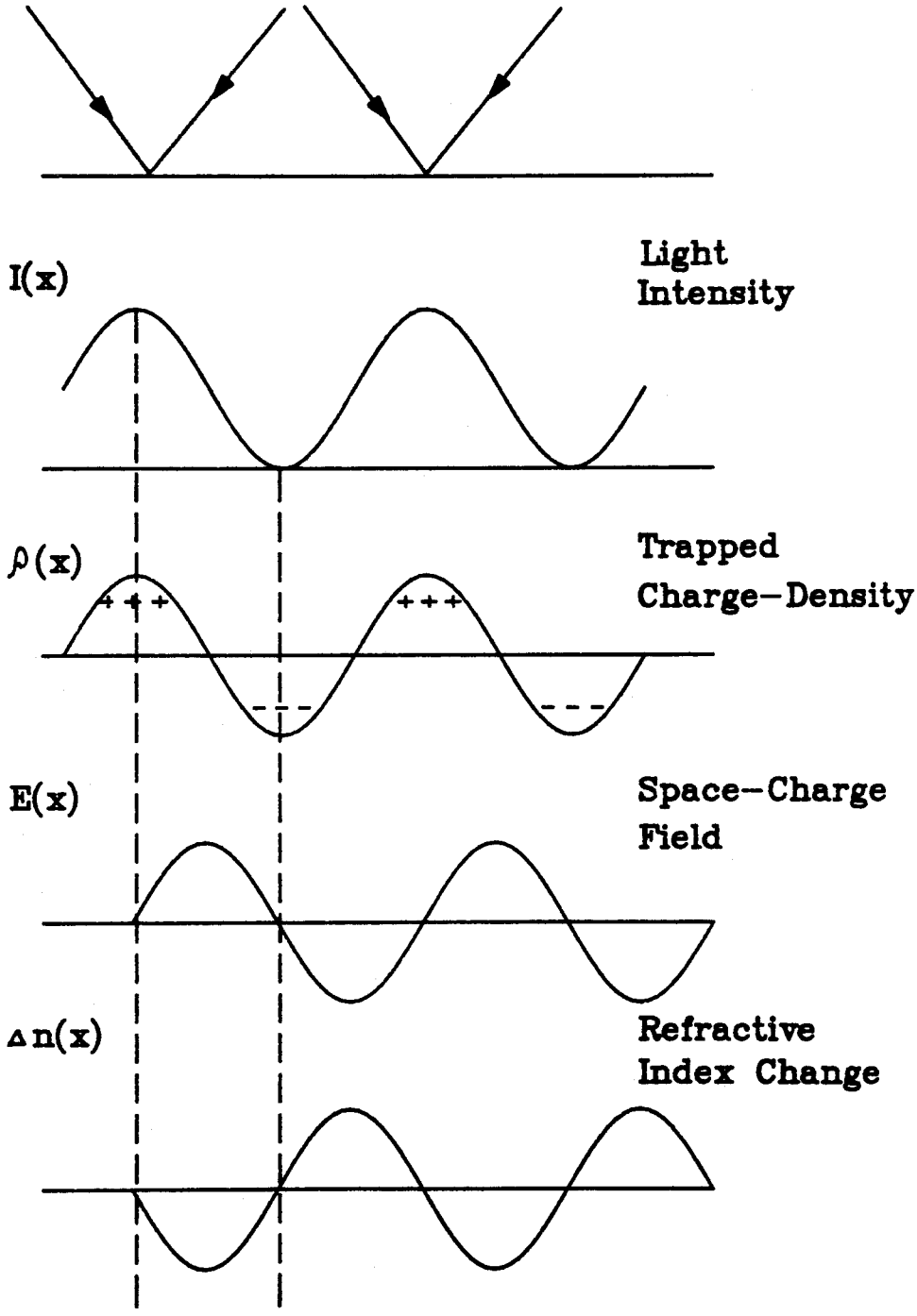


Figure 5.1: The Photorefractive Effect.

As shown in the figure, under illumination of an interference pattern, the trap sites in the crystal are ionized and electrons are excited to the conduction band (or holes into the valence band, depending on the specific material), leaving positively ionized donors. These electrons move to the regions where the light intensity is low through either diffusion or drift process. In the dark region, these electrons are re-trapped to produce a negative stored charge locally. Thus, a space-charge field corresponding to the optical pattern is built up in the crystal. The space charge field then modulates the refractive index of the crystal through the linear electro-optic effect. As a result, a phase hologram is formed.

We now consider using photorefractive gratings for interconnections in optical neural networks. The basic architecture of the optical system is shown in Fig. 5.2. As shown in the figure, the difference of this architecture from that of the planar holographic interconnections is the use of volume holograms. The crystal that we use in our experiment has a thickness of 3 mm, while the fringe spacing of the interference patterns is of the order of μm. Thus, the hologram is a three dimensional system of layers corresponding to a periodic variation of refractive index. Information is distributed and stored in the entire volume of the crystal. In principle, the number of gratings that we can store in a crystal is proportional to the volume divided by the wavelength cubed. We use the 514 nm line of an argon laser in our experiments. This implies that with an appropriate arrangement of the interconnected neurons, as many as 10^{10} independent interconnections may be stored in a single crystal. This is three orders of magnitude higher than that of planar holograms.

In what follows we consider the problem of how to design sampling grids for volume interconnections. We consider the case shown in Fig. 5.2. Suppose that two neurons a and b are associated with each other and that the holographic interconnection is recorded in the photorefractive crystal. The training process is similar to that of planar holographic interconnections (Fig. 2.13.), except that now the grating

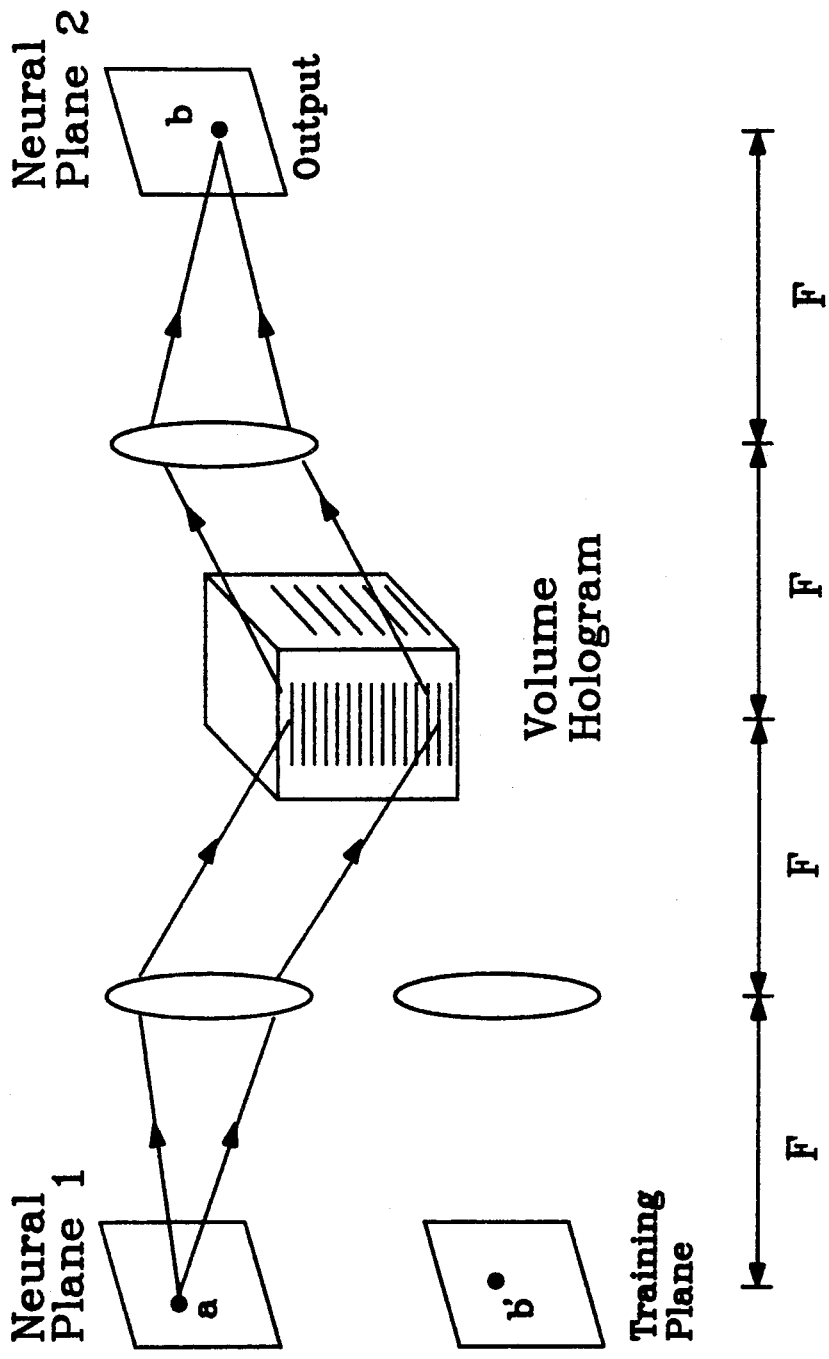


Figure 5.2: The Volume Holographic Interconnection between **a** and **b**.

is recorded in a volume. Let wave vector $\mathbf{k}_a$ represent the plane wave originating from the neuron a and $\mathbf{k}_b$ represent the plane wave originating from the neuron b; then, the grating formed by these two waves has a direction given by the vector

$$\mathbf{K} = \mathbf{k}_b - \mathbf{k}_a, \tag{5.1}$$

where $k_a = k_b = \frac{2\pi}{\lambda}$ and $K = \frac{2\pi}{\Lambda}$. λ is the light wavelength and Λ is the grating spacing. The relationship between Λ and λ is

$$2\Lambda sin\theta = \frac{\lambda}{n}, \tag{5.2}$$

where 2θ is the angle between the writing beams and n is the refractive index of the crystal. Eq. 5.2 is called the Bragg condition since it is similar to the Bragg condition of X-ray diffraction in solids. This is the greatest distinction between a volume and a planar grating. In reconstructing a wave, the diffracted light amplitude is a maximum only when the reading beam satisfies the Bragg condition. If the reading beam deviates from the Bragg angle, the diffraction efficiency falls off inversely proportional to the angle deviation squared. For the unslanted gratings, the angle resolution for the half-power diffraction is found to be

$$\Delta\theta_{1/2} \approx \frac{1}{2}\frac{\Lambda}{d}, \tag{5.3}$$

where d is the crystal thickness [8] . In our system $\Lambda \approx \mu m$ and $d = 3\ mm$; thus $\Delta\theta_{1/2} \approx 0.17\ mrad$. In our experiments where we use a Fourier-transform lens with a focal length of 25 cm, the above angular deviation corresponds to a displacement of 42.5 μm in the input neural plane. This means that as long as other neurons are separated from neuron a by more than 42.5 μm, there is no read-out from the grating K. Thus, this grating specifies a unique interconnection between a and b. This implies that volume interconnections are not shift-invariant. However, the Bragg condition is valid only in a direction parallel to the grating direction

K. It does not hold in the direction perpendicular to K. In other words, volume interconnections are shift-invariant in the perpendicular direction [9].

The above characteristics of volume interconnections can be shown using the **k**-space representation [4], as shown in Fig. 5.3. In the figure, the vector $\mathbf{k}_a$ is drawn with its origin at the center of the sphere, with magnitude equal to $\frac{2\pi}{\lambda}$ and in the direction of that of the plane wave from the neuron a. Similarly, the vector $\mathbf{k}_b$ represents the plane wave originating from neuron b. The grating vector K is drawn as a vector with its origin at the tip of $\mathbf{k}_a$ and its end at the tip of $\mathbf{k}_b$, and its magnitude equal to $\frac{2\pi}{\Lambda}$. Since any $\mathbf{k}_a$ and $\mathbf{k}_b$ that satisfy this condition (viz., the Bragg condition) give a reconstructable wave, the tips of these vectors form two circles on the **k**-sphere, as shown in Fig. 5.3(b). Now suppose the reading beam comes from a neuron that is shifted away from neuron a in a direction perpendicular to that of the grating vector **K**. In **k**-space this corresponds to a wave vector $\mathbf{k}_1$ with its tip lying on the bottom circle. As shown in Fig. 5.3(c), the vectorial sum of $\mathbf{k}_1$ and **K** still lies on the upper grating plane. Thus, the diffracted beam goes to neuron c which is in a position shifted away from neuron b in a direction perpendicular to **K**. Therefore, volume interconnections have the same degeneracy as that of planar interconnections. However, if the reading beam comes from a neuron which is shifted in the direction of **K**, then the vectorial sum of $\mathbf{k}_2$ and **K** is not on the upper circle. This is shown in Fig. 5.3(d). The Bragg condition is no longer satisfied and there is no diffracted wave. Thus, we do not need sampling grids in this direction as long as the separation between neurons is sufficiently large, such that the angle between the directions of the waves from these neurons is larger than the angular resolution given by Equ.(5.3). In summary, a sampling grid for photorefractive interconnection can be obtained from that for the planar interconnection, simply by adding all the shifted versions along the grating vectors with spacings larger than the angular resolution given by Equ.(5.3). Fig. 5.4 shows a sampling. In the figure (a) is the sampling grids

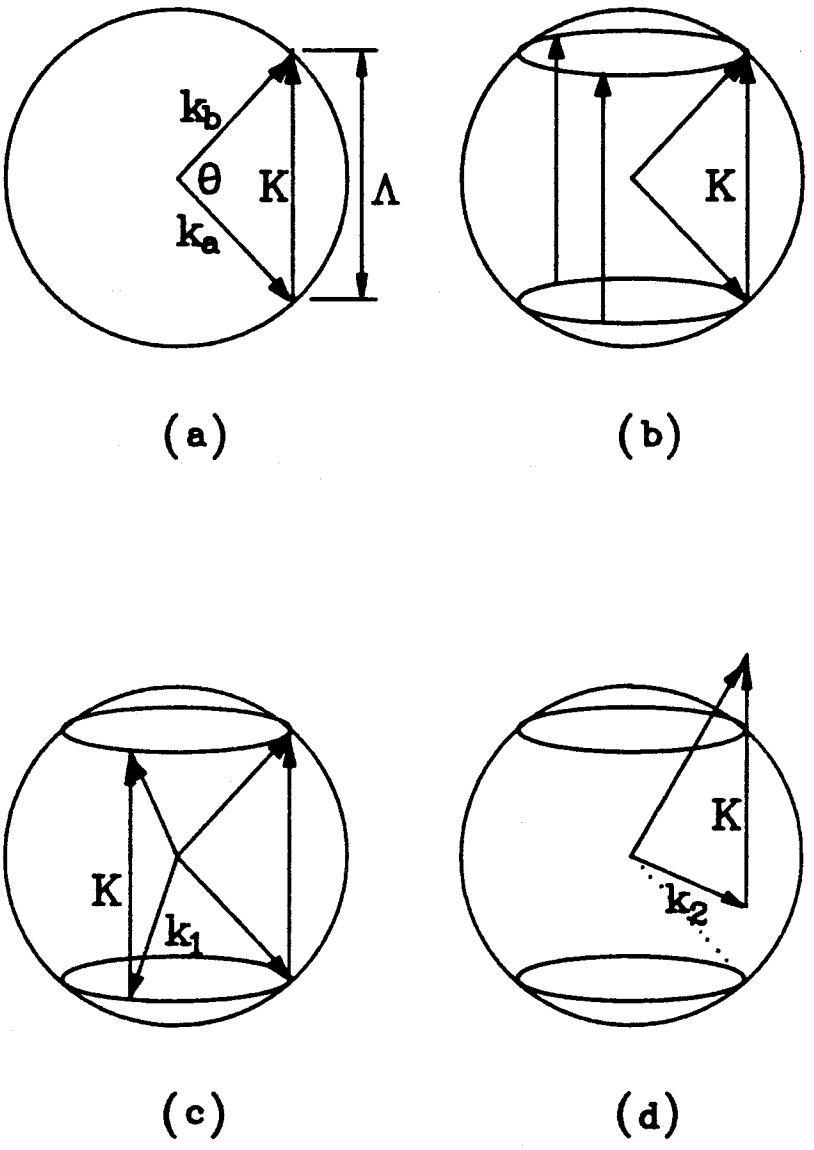


Figure 5.3: The volume holographic grating.

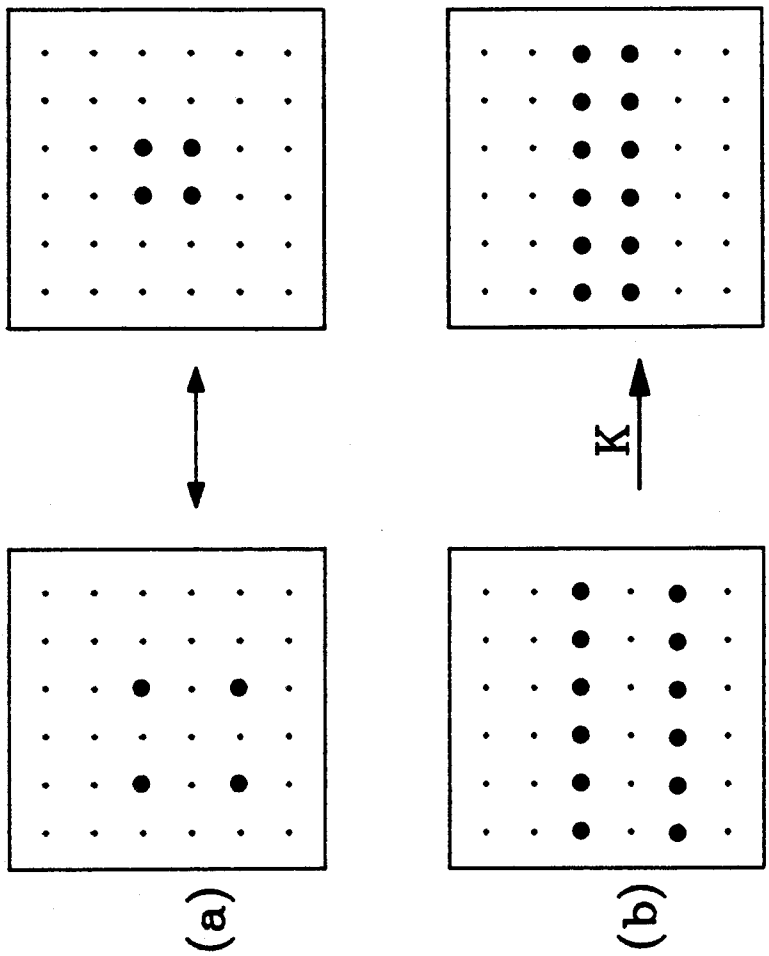


Figure 5.4: The Sampling Grids for Volume Interconnections. (a) The sampling grids for planar interconnections. (b) The sampling grids derived from (a) for volume interconnections.

for planar holographic interconnections (Fig. 4.5) and (b) is for volume holograms. As shown in the figure, (b) is obtained by adding all the shifted versions of (a) along the **K** direction.

5.3 Photorefractive Associative Memories with Maximum Storage and Maximum Efficiency

As was described in Section 5.1, the multiple exposure technique provides a possibility for obtaining maximum storage capacity in an optical associative memory. The problem is that writing a new hologram in a photorefractive crystal causes the charge patterns of previous holograms to decay exponentially in time. The grating is erased and the diffraction efficiency is decreased as the number of exposure increases. Finally, the diffracted light will become too weak to detect. Thus, the decay of the gratings limits the number of exposures that can be recorded in a photorefractive crystal, which in turn limits the storage capacity of the associative memory. We present in this section a copying method that allows us to recover the diffraction efficiency of each hologram in an SBN crystal to that obtainable from a single-exposure.

The temporal behavior of photorefractive holograms may be described in the simplest case by a growth in the amplitude of the space charge density proportional to $(1-e^{-\alpha It})$ during recording, and a decay proportional to $e^{-\alpha It}$ during the recording of successive holograms. I is the recording intensity and α depends on crystal parameters (e.g., doping concentration, orientation, etc.). The amplitude of the space charge corresponding to the m^{th} hologram when M holograms are recorded is

$$A_m = A_0(1 - e^{-\alpha It_m})exp(-\sum_{m'=m+1}^{M} \alpha It_{m'}), \tag{5.4}$$

where A_0 is the saturation amplitude and t_m is the recording time of the m^{th} hologram. If we record M holograms and require $A_m = A_{m+1}$ for all m, then it can be

shown [7] that the exposure schedule for each hologram is given by

$$t_m = \frac{1}{\alpha I} ln\left(\frac{m}{(m-1)}\right) \qquad m > 1. \tag{5.5}$$

We have been able to record as many as 111 holograms in SBN using this scheme. Fig. 5.5 shows reconstructed images from the hologram. In this experiment, the holograms are recorded between a reference plane wave and a series of signal plane waves incident at different angles. Thus, these holograms can be read out in one time by the reference wave. It is seen that the uniformity of the holograms is good. However, the diffraction efficiency of the holograms decreases very much. It can be shown that the amplitude of the space charge is

$$A_m = \frac{A_0}{M} \qquad m = 1, \cdots, M, \tag{5.6}$$

where A_0 is the saturation amplitude for the single grating. The experimental result of a multiple exposure hologram is shown in Fig. 5.6. We used the exposure schedule given by Eq. 5.5. The diffracted light read out from the hologram is detected by a CCD camera and is shown on an oscilloscope. It can be seen that the diffraction efficiency of the double-exposure hologram is $\frac{1}{4}$, and the triple-exposure hologram is $\frac{1}{9}$ that of the saturation diffraction efficiency of a single hologram. As discussed in Section 5.1,in order to achieve maximal storage capacity of the optical associative memory, we may need to make hundreds of exposures in a recording medium. The diffraction efficiency will then drop by a factor of more than 10^4. The signal might be too weak to be detected by optical neurons. In the following we present a method for increasing the diffraction efficiency of multiply exposed photorefractive holograms by periodic copying [10, 11].

The architecture of the system is shown in Fig. 5.7 and the photograph of the optical setup is shown in Fig. 5.8. We first describe the operation of the system. A series of holograms between a reference plane wave and a set of signal beams is

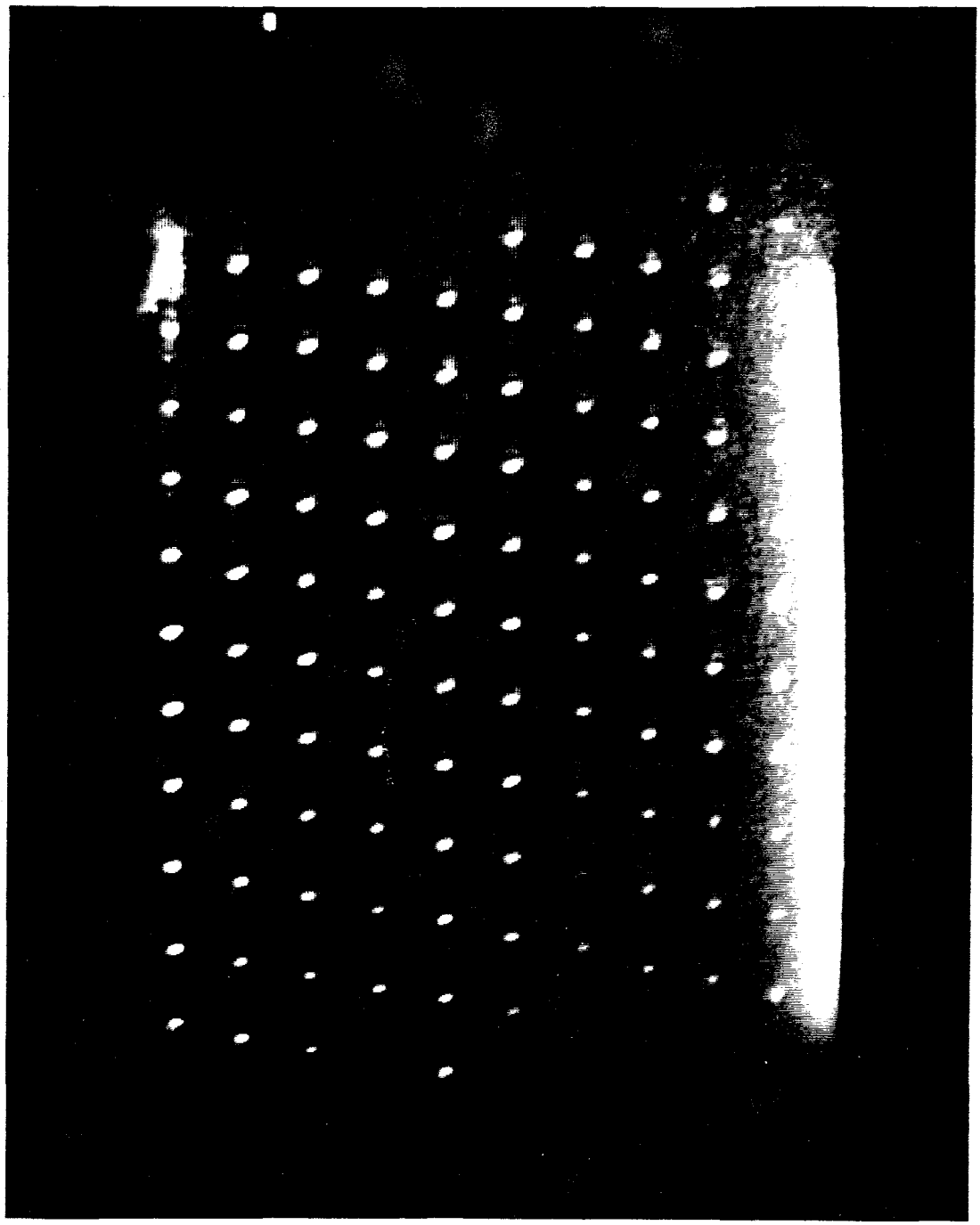
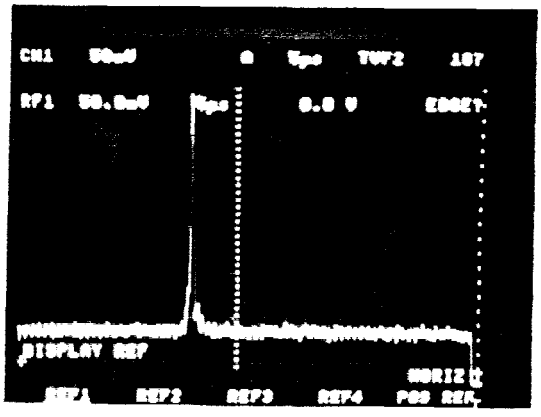
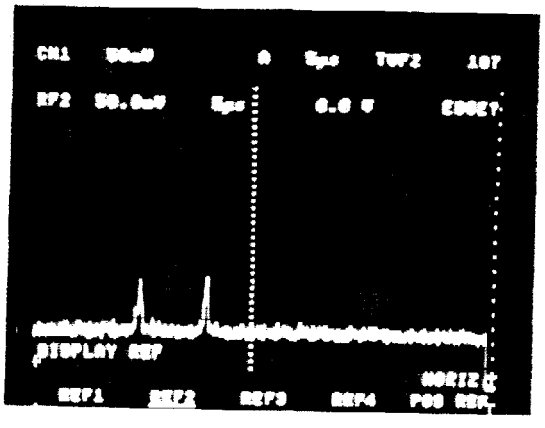


Figure 5.5: The Reconstructed Image of 111 Volume Gratings.

(a)



(b)



(c)

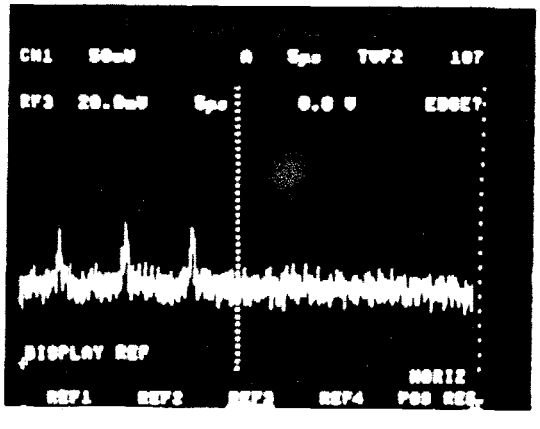


Figure 5.6: The Interconnection Strength of a Multiply Exposed Hologram in an SBN Crystal. (a) Single exposure. (b) Double exposure. (c) Triple exposure.

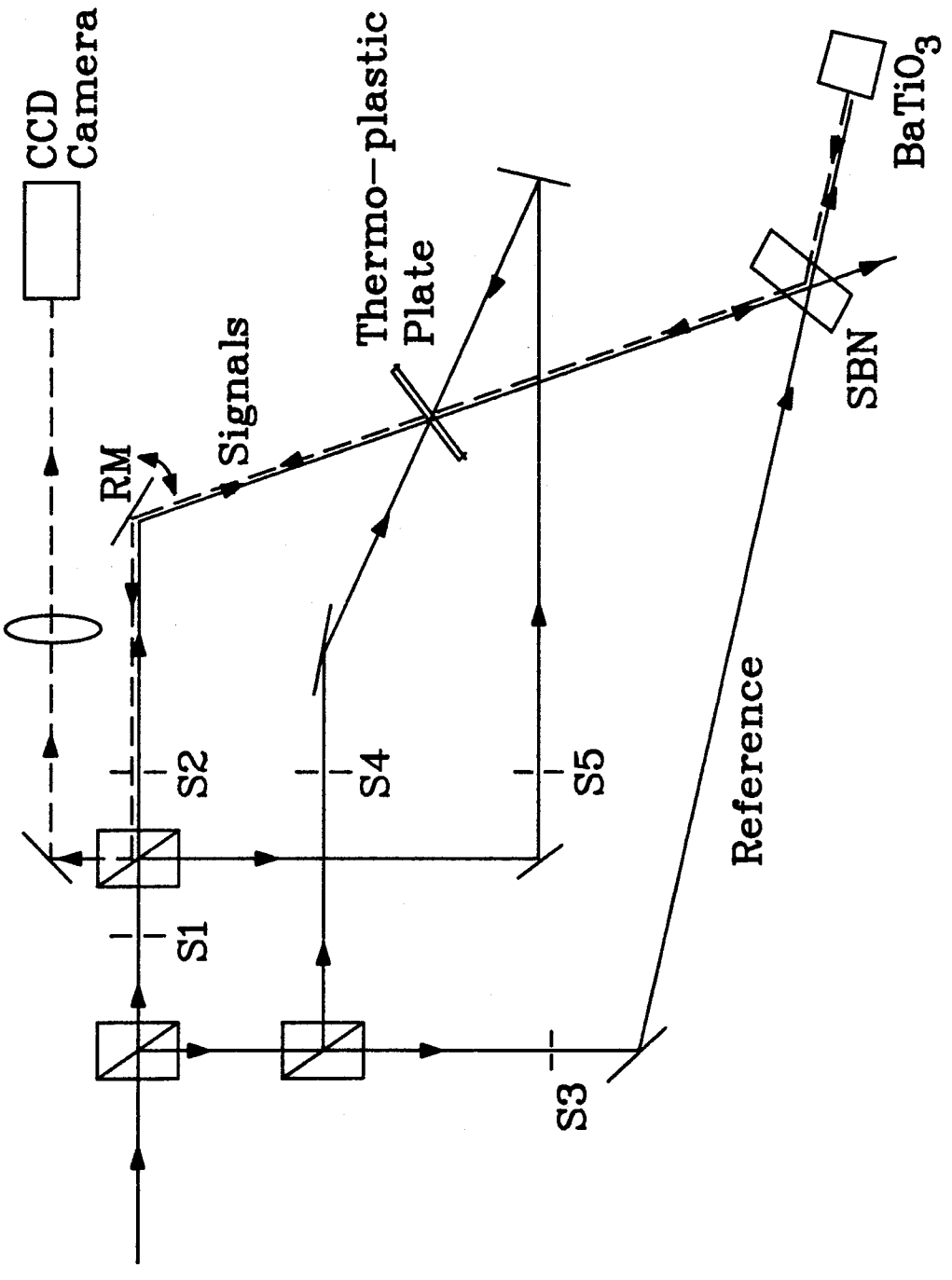


Figure 5.7: Schematic Diagram for the Periodic Refreshing Multiple Exposure Experiment.

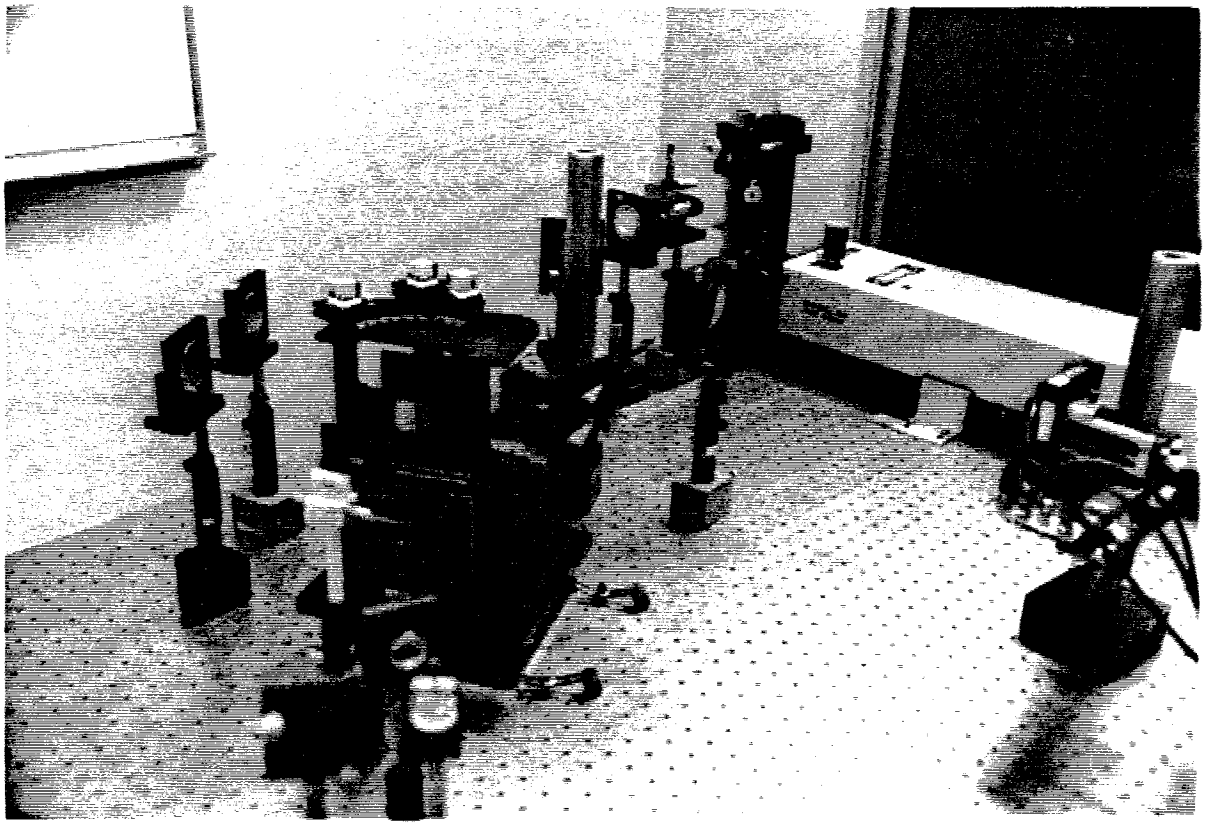


Figure 5.8: Photograph of the Setup for the Periodic Refreshing Multiple-Exposure Experiment.

recorded in an $SBN:Ce$ crystal. Shutters S_4 and S_5 are closed during this operation. In the experiment the signal beams are plane waves incident at different angles by rotating the mirror RM. While the use of plane waves as signal beams allows us to monitor the diffraction efficiency of the recorded holograms easily, the analysis and results apply equally well to arbitrary signal beams. The diffraction efficiency of the recorded holograms is monitored continuously, using the phase conjugate of the reference beam. The path of the diffracted conjugate wave to an 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. When the diffraction efficiency of the photorefractive holograms begins to be unacceptably low, the recorded holograms are copied to a second holographic medium, which in our experiment is a thermoplastic plate. The thermoplastic plate hologram is formed using the diffracted phase-conjugate reference wave and a back-traveling reference wave. Shutters S_2 and S_4 are closed. The hologram written on the plate is copied back to the SBN with the original total intensities in the signal and reference beams. The original reference beam and the wave diffracted from the thermoplastic by the counter-propagating wave from $S4$ are used to create this hologram. Shutters S_1 and S_5 are closed during this step. The result is a rejuvenated hologram of each of the signal beams in the SBN. The intensity diffraction efficiency of each hologram is now proportional to $\frac{1}{M}$, which is M times improved as compared to the previous $\frac{1}{M^2}$, but is M times less than that of a single hologram. The reduction factor of $\frac{1}{M}$ from that of a single hologram is inherent since in copying back from the thermoplastic plate, the total intensity in the signal beam is shared by M signals. Thus, the modulation depth of each signal is reduced by a factor of $\frac{1}{\sqrt{M}}$ and the intensity diffraction efficiency of each hologram is $\frac{1}{M}$. However, the total diffraction efficiency, summed over M holograms, is restored to its saturation value. Therefore, this copying scheme allows us to record information in a multiple exposure hologram with the maximum

diffraction efficiency that can be obtained in a single exposure.

We now describe how to increase the storage capacity in a photorefractive holographic associative memory using the system of Fig. 5.7. We begin by recording a series of M_1 holograms on the SBN following the schedule of Eq. 5.5. The amplitude of space charge of each hologram is $\frac{1}{M_1}$ of the saturation amplitude and the diffraction efficiency in intensity is $\frac{\eta_0}{M_1^2}$ at this point, where η_0 is the saturation diffraction efficiency for a single hologram. We copy the summed holograms in the SBN to the thermoplastic plate and back to the SBN using the procedure described in the previous paragraph. The amplitude of each hologram is then $\frac{A_0}{\sqrt{M_1}}$. We begin recording another series of new holograms on the SBN. Since this time the starting amplitude of the M_1 holograms is not at the saturation level A_0, the exposure schedule is different from that of the first M_1 exposures. It can be shown [11] that

$$t_m = \frac{1}{\alpha I} ln\left(\frac{1+(m-1)\chi}{1+(m-2)\chi}\right) \qquad m > 1, \tag{5.7}$$

where $\chi = \frac{A(0)}{A_0}$ represents the amplitude level of the starting hologram. In this case $\chi = \frac{1}{\sqrt{M_1}}$. We make M_2 exposures in the second cycle. At the end of the second cycle the amplitude of the holograms recorded in the SBN is

$$A_m = A_0 \frac{\chi}{1+M_2\chi} \qquad m = 1, \cdots, M_1 + M_2, \tag{5.8}$$

where A_0 is the saturation amplitude of the single grating. We select M_2 such that the total signal diffracted from the $M_1 + M_2$ holograms in the SBN to the thermoplastic plate equals that of the M_1 holograms in the first cycle, i.e., $\frac{\eta_0}{M_1}$. Physically, this means that we set a minimum signal level corresponding to the sensitivity of optical neurons. Once the total signal decreases to this level, we make a refreshening copy before recording more memories. The condition for selecting M_2 is

$$\frac{(M_1+M_2)}{(\sqrt{M_1}+M_2)^2} = \frac{1}{M_1}. \tag{5.9}$$

After M_2 exposures we copy to the thermoplastic, then back to the SBN and again make new holograms until the total diffraction efficiency falls to $\frac{\eta_0}{M_1}$ once more. From here on the process may proceed indefinitely. Each time $M = \sum M_i$ holograms are copied back and forth, the diffraction efficiency for each hologram is restored to $\frac{\eta}{M}$.

In fact, there is one constraint on the allowable number of refreshing cycles. In order to copy from one hologram to another, it must be possible to diffract enough energy from the source hologram to record the destination hologram. As the number of recording cycle increases, the diffraction efficiency of each of the photorefractive holograms decays. Furthermore, the total energy which may be diffracted from a photorefractive hologram is limited by the decay of the hologram on readout. This sets an upper limit for the number of refreshening cycles unless we add another beam amplification mechanism [12] between the two holographic media. In our experiment, the energy diffracted from the SBN is ample for recording the thermoplastic hologram. The energy that may be diffracted back from the thermoplastic is also relatively high, so no trouble is encountered in the copying processes.

Fig. 5.9 shows experimental results for recording 25 holograms in a SBN crystal. In the figure, the solid line represents the theoretical $\frac{1}{M^2}$ decay in the diffraction efficiency per hologram when the exposure schedule Eq. 5.5 is followed. The *'s are experimental data points for the mean diffracted power of the stored holograms. The dashed line shows the theoretical path followed by the diffraction efficiency per hologram with periodic copying when $M_1 = 5$. Thermoplastic holograms were made after 5, 10, and 15 exposures. The #'s represent the experimental data points. The dotted line corresponds to the theoretical decrease in diffraction efficiency proportional to $\frac{1}{M}$. Both experimental results agree well with theoretical predictions. Fig. 5.10 shows the image of 15 holograms reconstructed from the SBN before and after the refreshing cycle. The 15 images are monitored by a CCD camera. The

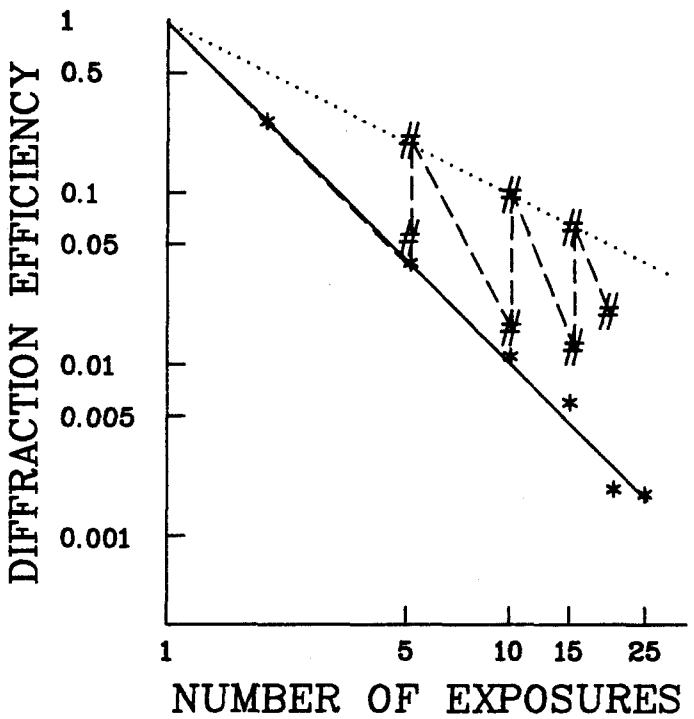
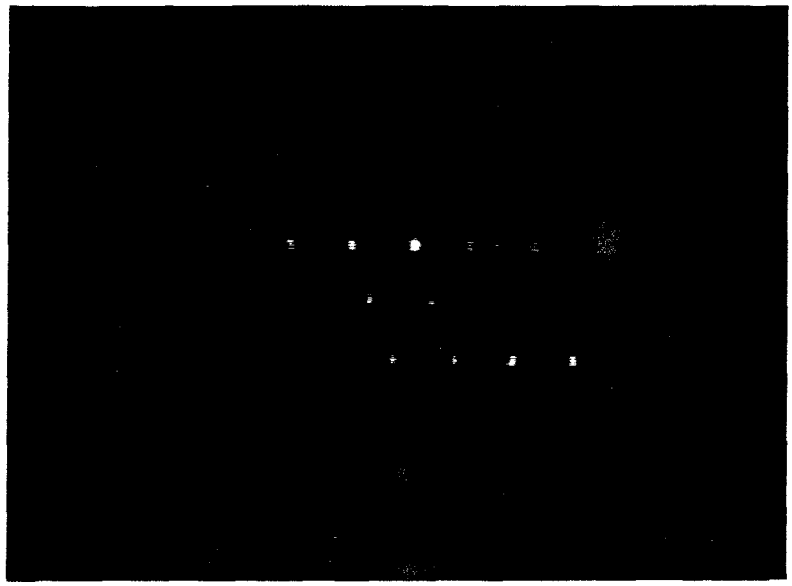


Figure 5.9: Experimental Results of the Multiply Exposed hologram.

(a)



(b)

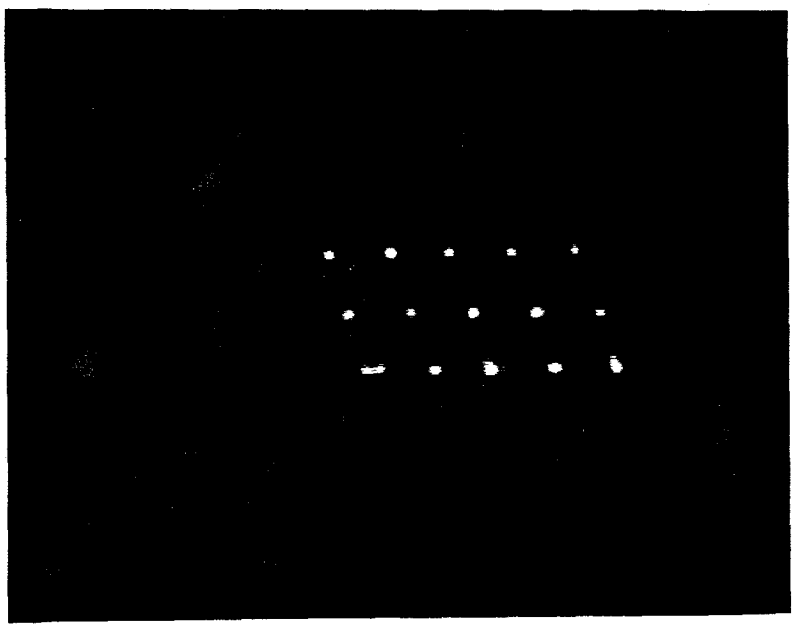


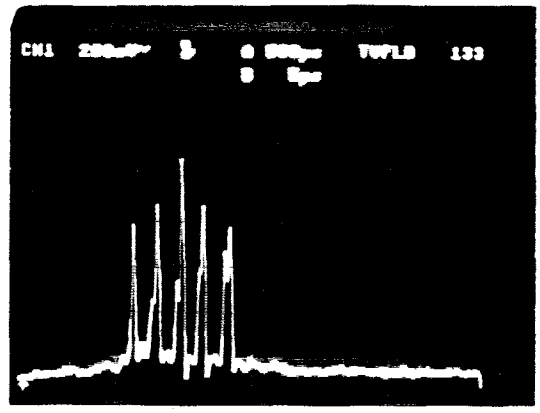
Figure 5.10: Experimental Results of the Multiply Exposed hologram. (a) Before refreshing. (b) After refreshing.

CCD signals are arranged on three video lines on a digital oscilloscope. Fig. 5.11 shows the amplitude of these holograms. It is seen that the holograms are not very uniform. The nonuniformity is partly due to accumulation of scattered noise in the SBN through repeated copying. The loss of image fidelity sets another limit on the total number of exposures that can be made.

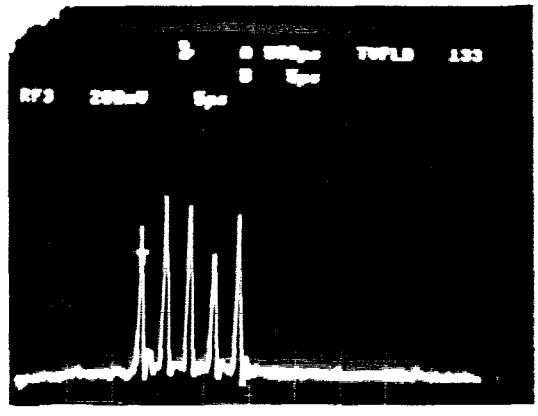
5.4 Conclusion

In this chapter we have presented the use of photorefractive crystals in optical associative memories. The designing rules for sampling grids for volume holographic interconnections are presented. While volume holographic interconnections are not shift invariant along the grating direction because of the Bragg condition, it is shift-invariant along the perpendicular direction. The sampling grids for this case can be obtained by adding all the shifted versions of the planar holographic interconnections along the grating direction. An exposure scheme for a multiple-exposed photorefractive hologram is presented. This allows the hologram to have the same diffraction efficiency as that using a single hologram for storing the same information. The basic idea of the scheme is to use two holographic media to construct separate memories which periodically refreshes each other. This scheme allows the optical associative memory to increase storage capacity and diffraction efficiency.

(a)



(b)



(c)

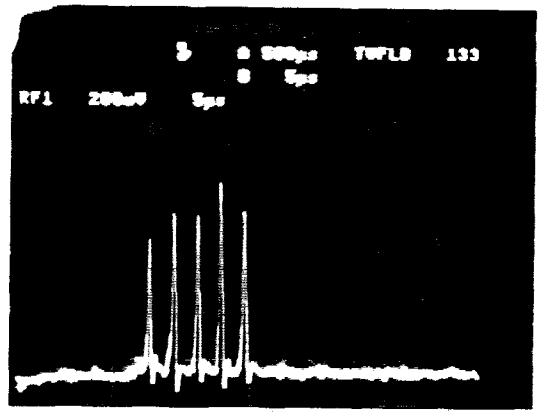


Figure 5.11: Amplitudes of the Reconstructed Images from the Refreshed Hologram of Fig. 5.10. Each line corresponds to one line of the image.

5.5 References for Chapter 5

1. S. Venkatesh and D. Psaltis, "Linear and Logarithmic Capacities in Associative Neural Networks," *IEEE Trans. Inform. Theory.* Vol. **IT-35**, No. 3, June (1989).

2. R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. Venkatesh, "The Capacity of the Hopfield Associative Memory," *IEEE Trans. Inform. Theory*, Vol. **IT-33**, No. 4, 461-482, July (1987).

3. D. Psaltis, X. G. Gu, and D. Brady, "Fractal Sampling Grids for Holographic Interconnections," SPIE Vol. 963-70, 1988 ICO Topical Meeting on Optical Computing, Toulon, France.

4. D. Psaltis, J. Yu, X. G. Gu, and H. Lee, "Optical Neural Nets Implemented with Volume Holograms," in *Technical Digest of Topical Meeting on Optical Computing*, Optical Society of America, Washington, DC (1987).

5. D. Psaltis, X. G. Gu, H. Lee, and J. Yu, "Optical Interconnections Implemented with Volume Holograms," to be published.

6. D. Psaltis, D. Brady, and K. Wagner, "Adaptive Optical Networks using Photorefractive Crystals," *Appl. Optics*, Vol. **27**, No. 9, 1752, May (1988).

7. K. Hsu, D. 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).

8. H. Kogelnik, "Coupled Wave Theory for Thick Hologram Gratings," *Bell Sys. Tech. Jour.*, **48**, 2909 (1969).

9. J. W. Yu, " Optical Processing using Photorefractive Crystals, " *Ph.D Dissertation*, California Institute of Technology, March (1988).

10. D. Brady, K. Hsu, and D. Psaltis, "Multiply Exposed Photorefractive Holograms with Maximal Diffraction Efficiency," Paper PDP2, Topical Meeting on Optical Computing, Optical Society of America, Salt Lake City, Feb. 27-March 1, 1989.

11. D. Brady, K. Hsu, and D. Psaltis, "Periodically Refreshed Multiply Exposed Photorefractive Holograms," submitted to *Opt. Lett.*, 1989.

12. J. P. Huignard and A. Marrakchi, "Coherent Signal Beam Amplification in Two-Wave Mixing Experiments with Photorefractive $Bi_{12}SiO_{20}$ Crystals," *Opt. Comm.*, **38** 249 (1981).