OPTOELECTRONIC HIGH ORDER FEEDBACK NEURAL NETWORK (HOFNET)

Thesis by

Zhi Qiang MAO

Submitted to the University of London

for the Degree of Doctor of Philosophy

Department of Electronic and Electrical Engineering
University College London
Torrington Place, London WC1E 7JE
United Kingdom

January 1992

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisors Dr David Selviah and Professor John Midwinter for their kind supervision and good guidance. Without their stimulation, I would not have designed a novel neural network nor would have finished this thesis. I would also like to thank Madam Tao for help with making holograms, Ali Twaij for the helpful discussion about optical neural networks and English correction, Anthony Davis for English correction, Peter Grigson and Peter Leung for the computer help and discussion on neural networks, and Chris Carey for the laboratory setup. Of course, without the financial support from the British Council, the Chinese Government and the Optical Interdisciplinary Research Centre, I would not have come here to study and carry out research. Furthermore, I would like to give thanks to the STC (now called BNR) for providing the electrically addressed spatial light modulator, which is indispensable for our optical neural network system. At the end I should say 'thank you' to my wife for her constant encouragement.

To my Father

ABSTRACT

In this thesis, a High Order Feedback neural NET or HOFNET and its optoelectronic implementation are discussed in detail. The HOFNET implements 'high order' by using a feedback loop and normalisation in each iteration. The order of the HOFNET is not fixed but increases by one after each iteration. In the HOFNET system, the input pattern is inserted in the middle of the system and is correlated in every iteration, so the time varying noise in the input pattern will be averaged. This is confirmed by the mathematical analysis and computer simulations. An optoelectronic system of the HOFNET is set up by using a Fourier Transform hologram array for information storage and an electrically addressed spatial light modulator in the feedback loop. The gain for the compensation of the loss in the optical correlation system is obtained by using a computer and a spatial light modulator. We have stored initially 3 patterns and later 14 different patterns in the holograms for pattern recognition and with a partial pattern as an input, the system successfully recognised the input pattern and recovered the obscured parts after 2 (in 3 pattern case) and 3 or 4 (in 14 pattern case) iterations, respectively. In this primary system, the feedback is serial through a computer. We designed two optical parallel feedback systems for the HOFNET by using optoelectronic devices. One is based on an optical fibre amplifier. The other is more attractive and the design uses an optically addressed spatial light modulator in one feedback loop for providing gain and a multiple quantum well based SEED device in the other feedback loop for providing normalisation. Thorough analysis and computer simulation have been done for this net.

CONTENTS

TITLE 1

ACKNOWLEDGEMENTS 2

ABSTRACT 4

CONTENTS 5

DEFINITIONS & ABBREVIATIONS 8

Chapter One INTRODUCTION 9

Chapter Two NEURAL NETWORK MODELS 14
2.1 Hopfield Model 14
2.2 The Hamming Net 18
2.3 Matched Filter Model (Inner-Product Model) 20
2.3.1 Matched filer model 20
2.3.2 Feature extraction 22
2.3.3 Orthogonal coding 24
2.4 High-order Neural Networks 26
2.4.1 Outer-product model 26
2.4.2 Inner-product model 28
2.5 Multilayer Perceptron 29
2.6 Comments on Neural Networks 31

Chapter Three HOLOGRAPHY 33
3.1 Introduction 33
3.2 Holographic Principle 34
3.3 Fourier Transform Holograms (FTHs) 39

3.4 Van der Lugt Matched Filter (Spatial Correlation System) 42
3.5 Spectrum Correlation System 44
3.6 Holographic Memories 46

Chapter Four REVIEW OF OPTICAL NEURAL NETWORK SYSTEMS 52
4.1 Introduction 52
4.2 Optoelectronic Neural Network Using Vector-Matrix Multipliers 54
4.3 Optical Linear Neural Network Systems 56
4.3.1 Optical linear neural network based on outer product model 56
4.3.2 Optical linear neural network based on inner product model 59
4.4 Optical High Order Neural Network Systems 64
4.4.1 Optical neural network based on high order outer product model 65
4.4.2 Optical neural network based on high order inner product model 68
4.5 Conclusion 70

Chapter Five HIGH ORDER FEEDBACK NEURAL NETWORK (HOFNET) 71
5.1 Introduction 71
5.2 Principle of the HOFNET 73
5.3 Noise Analysis of the HOFNET 82
5.4 Computer Simulations 83
5.5 Conclusions 90

Chapter Six OPTICAL HOFNET SYSTEM 91
6.1 Introduction 91
6.2 Focused Optical HOFNET System 92
6.3 Lensless Optical HOFNET System 100
6.4 Feedback System 105
6.5 Experimental Results 106
6.6 Conclusions 109

Chapter Seven DESIGN OF HOFNET SYSTEM WITH OPTICAL PARALLEL FEEDBACK LOOP 114
7.1 Optoelectronic Devices 114
7.1.1 Optically addressed spatial light modulator (OASLM) 114
7.1.2 Multiple quantum well and D-SEED device 116
7.1.3 Optical fibre amplifier 118
7.2 HOFNET Design by Using OASLM and D-SEED Devices 120
7.3 HOFNET Design by Using Optical Fibre Amplifier 130
7.4 Summary 133

Chapter Eight CONCLUSIONS AND FUTURE WORK 134
8.1 Conclusions 134
8.2 Future Work 136

REFERENCES 138

APPENDIX 152
I. Fourier Transformation 152
II. Intersecting Points of Two Hyperbolic Curves 155

PUBLICATIONS 159

DEFINITIONS & ABBREVIATIONS

DEFINITIONS

N: number of neurons in one dimensions;
M: number of stored patterns;
θ_m: threshold in the Hamming net;
ξ: modulation or spatial carrier frequency of a hologram;
*: conjugate sign;
$*$: correlation;
$\otimes$: convolution;
x, y: coordinate pair in the spatial domain;
μ, ν: coordinate pair in the frequency domain;
f: focal length of a lens;
λ: wavelength;
F.T.: Fourier Transform symbol;
δ: Dirac delta function;
G: gain.

ABBREVIATIONS

EDFA: Erbium Doped Fibre Amplifier;
FTH: Fourier Transform Hologram;
HOFNET: High Order Feedback neural NET;
ITO: Indium Tin Oxide;
MQW: Multiple Quantum Well;
OASLM: Optically Addressed Spatial Light Modulator;
PCM: Phase Conjugate Mirror;
SEED: Self-Electrooptic Effect Device;
SLM: Spatial Light Modulator;
SNR: Signal to Noise Ratio.

Chapter One

INTRODUCTION

Artificial neural networks attempt to achieve good performance at analogous processing, such as pattern recognition, via dense interconnection of simple computational elements. An artificial neural network usually has three characteristics:

(a) It consists of a large number of simple processing units (neurons);

(b) each processing element is connected to many or possibly all of the others;

(c) the network is programmed to respond appropriately to inputs by adjusting the weights of the connections between neurons during the learning procedure.

The idea of artificial neural networks is from our present understanding of the biological nervous system -- brain. A number of mathematical models, which were called neural network models, were proposed to simulate the performance of the brain (much more simplified). They have great potential in areas such as speech and pattern recognition where many processings are pursued in parallel and high computational rates are required. Currently the best neural network systems are far from equalling human performance in such applications. Neural network models explore many hypotheses simultaneously using densely parallel nets composed of many computational elements connected by links with varying weights. Recently, there has been a great deal of theoretical progress to justify optimism about future applications, and this has focused attention on the hardware realization of neural network architecture.

Research into neural networks has a long history. Establishment of detailed mathematical models began nearly fifty years ago with the work of McCulloch and Pitts [McC43], Hebb [Heb49]. More recent work by

Hopfield [Hop82], Grossberg [Gro86], Rumelhard and McClellard [Rum86], Psaltis [Psa85, 86], Lee [Lee86] and Gile [Gil87] has led to a new resurgence of the field. This new interest is due to the development of new network topologies and algorithms, novel optical devices and systems.

Electronics and optics are the two approaches under consideration for the hardware realisation of neural network models. There are two basic components that need to be implemented: neurons and interconnections. The neurons are typically simple thresholding elements that can be implemented by a single switching device (i.e., transistor). The switching speed or the accuracy required for the neurons can be performed by current electronic technology. A practical neural computer may require millions of neurons operating in parallel. This requirement by itself is also achievable by electronics. However, each of the neurons must be connected to several thousand other neurons, and these interconnections must be modifiable so that learning can take place. While this dense connectivity is relatively difficult to achieve electronically, optics is particularly well suited for the realisation of interconnections. Optics can provide three dimensional interconnects [Goo89] and the connection weights can be modified easily by using photorefractive or liquid crystal materials [Psa88b; Sat91; Web91].

A variety of optical architectures for the realization of neural networks have been proposed [Psa85, 87, 88b; Jan88a; Whi88a, b; Sel89b; Yu90, 91] by using two basic optical devices: holograms (especially volume holograms) and spatial light modulators (SLMs) (electrically addressed or optically addressed). Spatial light modulators are used to simulate a two-dimensional array of neurons at the neural plane, whereas holograms and other passive optical elements comprise the optical interconnecting system. The basic optical neural computer architecture is shown in Fig 1.1. As shown in Fig 1.1, the optical interconnections in the neural network can be constructed in three dimensions, compared with two-dimensional implementation in electronics. A second advantageous feature is the relative ease with which learning can be accomplished in dynamic holograms in photorefractive crystals [Tan85; Psa88b; Her86]. There are other advantages of the optical implementation over the electronic one, such as parallel processing, however, learning and three-dimensional storage of information are the most important.

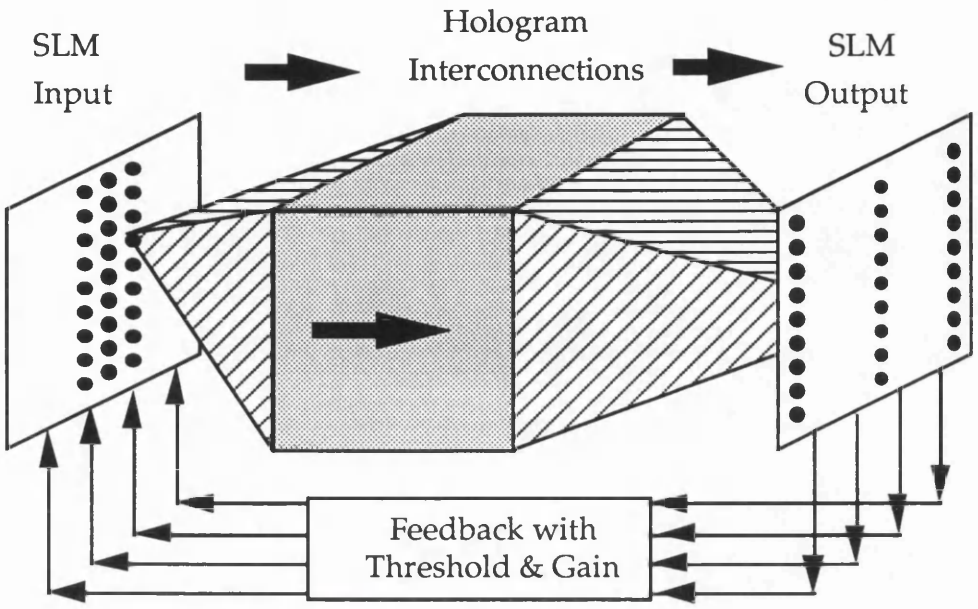


Fig. 1.1 Optical Neural Network Architecture

Hopfield proposed a new neural network model, based on the Hebbian learning rule in 1982 [Hop82]; it is called the "Hopfield model". Since then, many optical systems based on the Hopfield model (also called the outer product model) or its modified version: the inner product model [Pae87; Ath86] have been proposed and set up in laboratories. Such neural networks have a disadvantage that the memory capacity is very small ($M \approx 0.15\,N$, where M is the number of patterns stored in the system and N is the number of pixels (elements) in one pattern). If more patterns are stored in the network, the system performance will deteriorate sharply and will converge into spurious results or give incorrect answers. Based on the Hebbian learning rule and the Hopfield model, the high order neural networks were proposed [Lee86; Psa86, 88a; Gil87]. The motivation for investigating high order neural networks is the increase in storage capacity that results from the increase in the number of independent variables or degrees of freedom in the neural networks [Psa88a]. Several models have been proposed and optically implemented [Jan89; Lin89; Ath86]. Generally we can classify them into two kinds: outer product high order neural network and inner product high order neural network, the

details of which are covered in chapter two.

High order neural network models have a higher memory capacity, but their hardware, either optical or electronic, is more difficult to implement because the number of interconnections in the network increases exponentially [Gil87]. Optics can implement three dimensional interconnections, but the dynamic range of optical devices is quite small (only 27dB available recently in an Asymmetric FP modulator) [Whi89], which limits the high order implementation [Owe87]. In order to solve these problems, we propose a new High Order Feedback neural NET model, which is abbreviated to, HOFNET. Its optical or optoelectronic implementation requires neither a large number of interconnects nor optical devices with large dynamic range. The nonlinearity in this system is implemented by the feedback loop. The degree of nonlinearity is not fixed but equal to the number of iterations. In this system the input pattern is correlated with the stored patterns in every feedback loop, so if the input pattern or the system contains time-dependent noise, the noise will be averaged and can easily be suppressed in optics. The penalty for such advantages of the HOFNET is that the system is spatially variant. So the input pattern must be put exactly in the same position as reconstructed patterns from the hologram.

The layout of this thesis is as follows. In chapter two, we briefly describe several neural network models which are related to the HOFNET or are of help to understand the algorithm of the HOFNET. Then the important memory recording technique: holography, is introduced in chapter three. We emphasise the Fourier transform holograms, which are used in our optical HOFNET system. In this chapter, we also introduce the spectral correlation system and prove that the system can detect the similarity of two patterns. Following that, in chapter four, we review the optical neural network systems, set up by other institutions in recent years. From chapter five, we start to discuss the principle of the HOFNET, its optoelectronic system design and construction. Chapter five covers the details of the HOFNET: principle, mathematical analysis and computer simulations (noise and pattern recognition). Then, in chapter six, we describe an optical HOFNET system with serial electronic feedback (via electrically addressed spatial light modulator (E-SLM)), and some basic experimental results obtained from our experimental system,

demonstrating pattern recognition, are shown. Furthermore, the design of two more optical HOFNET implementations with optical parallel feedback loops is discussed in chapter seven. At the end, conclusions and future work proposals are given.

In this thesis, all the holograms are supposed to be thin unless they are specified. N is the number of neurons in one dimension, so if the patterns are two-dimensional, the number of neurons in one pattern is N^2. This thesis mainly concentrates on the system design. Therefore, although I have used and discussed briefly several optical or optoelectronic devices, such as the optically addressed spatial light modulator and multiple quantum well devices, I have not researched on them in depth.

Chapter Two

NEURAL NETWORK MODELS

2.1 Hopfield Model

The Hopfield neural network model was proposed by Hopfield in 1982 [Hop82] and is normally used with binary inputs. This net is based on the Hebbian learning rule [Heb49] and suitable for auto-associative memory and solving optimization problems. It consists of a large number of neurons and all the neurons are fully interconnected to each other although not to themselves. The strength of the interconnections between neurons are symmetric and updated according to the Hebbian learning rule, while the neurons are just the simple nonlinear thresholds, such as the sigmoid or clipped functional threshold [Lip87]. Three kinds of non-linear thresholds are commonly used in neural networks (Fig. 2.1):

(a) Sigmoid threshold

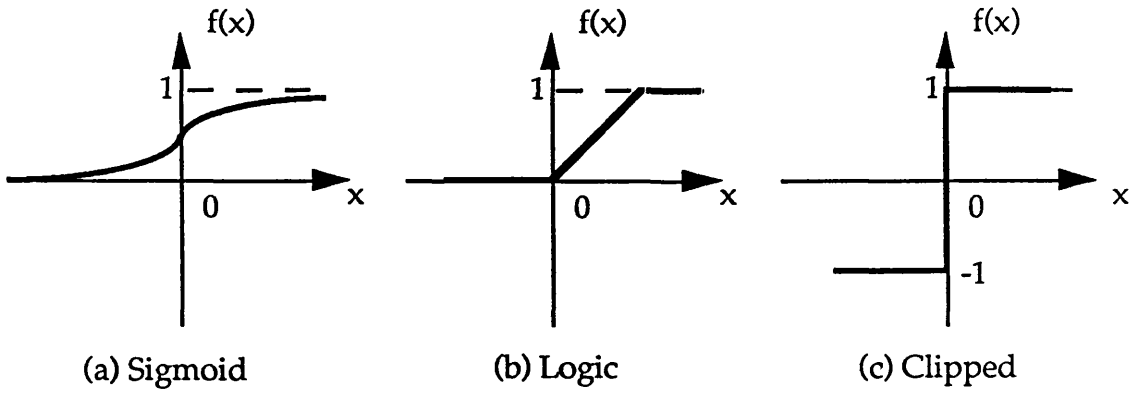
$$f(x)=\frac{1}{1+e^{-\alpha x}} \qquad \alpha \text{ is a constant} \tag{2.1}$$

(b) Logic threshold

$$f(x)=\begin{cases}1 & (x \geq 1)\\ x & (0 < x < 1)\\ 0 & (x \leq 0)\end{cases} \tag{2.2}$$

(c) Clipped threshold

$$f(x)=\begin{cases}1 & (x \geq 0)\\ -1 & (x < 0)\end{cases} \tag{2.3}$$



(a) Sigmoid (b) Logic (c) Clipped

Fig. 2.1 Nonlinear Thersholds

The outputs of the network are fed back to the neurons after a threshold to start a new iteration. So the Hopfield net can be considered to be composed of (1) Interconnections; (2) Threshold; (3) Feedback. The neuron-typed Hopfield net architecture is shown in Fig. 2.2. It may be regarded as an auto-associative system with N nodes containing binary inputs and outputs (bipolar or unipolar). Suppose a set of binary and bipolar vectors V(m) (m=1, 2, ..., M), each containing N elements $v_i(m)$ (i=1, 2, ... N), are stored in a matrix T_{ij}, according to the Hebbian learning rule, T_{ij} can be written as

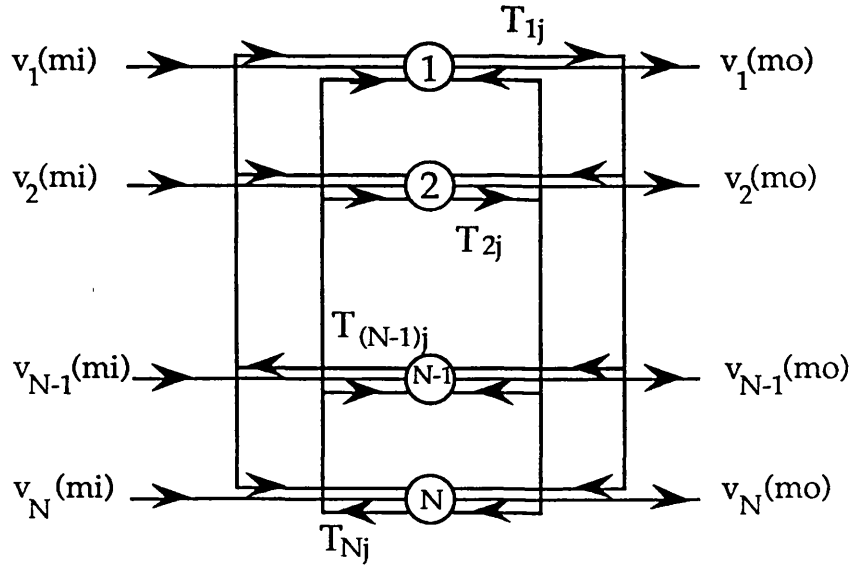


Fig. 2.2 Hopfield Neural Network Model

$$T_{ij} = \sum_{m=1}^{M} v_i(m)v_j(m) \quad (i \neq j)$$
$$T_{ii} = 0 \qquad (i = j) \tag{2.4}$$

where $T_{ii} = 0$ means that no neurons are connected to themselves. If the memory is addressed by one of the stored vectors, say v_i (mi), it yields the output as follows:

$$\begin{aligned} v_i(mo) &= \sum_{j=1}^{N} T_{ij}v_j(mi) \\ &= \sum_{m=1}^{M}\sum_{j=1}^{N} v_i(m)v_j(m)v_j(mi) \\ &= (N-1)v_i(mi) + \sum_{m\neq mi}^{M}\sum_{j\neq i}^{N} v_j(mi)v_j(m)v_i(m) \end{aligned} \tag{2.5}$$

$v_i(mo)$ is composed of the sum of two terms: signal and unwanted cross-talk. Suppose $v_i(m)$ has the same probabilities of +1 and -1, the sum of the unwanted cross-talk terms is $\sqrt{(M-1)(N-1)}$ [Far85]. So the signal to noise ratio is

$$\begin{aligned} SNR &= \frac{N-1}{\sqrt{(M-1)(N-1)}} \\ &\approx \sqrt{\frac{N}{M}} \qquad (N >> 1,\ M >> 1) \end{aligned} \tag{2.6}$$

If N is sufficiently larger than M, with high probability the elements of vector $v_i(mo)$ will take the sign of the corresponding elements $v_i(mi)$. Thresholding of $v_i(mo)$ will therefore give $v_i(mi)$

$$v_i(mi) = \text{sgn}\,[v_i(mo)] \tag{2.7}$$

With this model, the maximum number of stored patterns is limited to [Lee86]

$$M = \frac{N}{4 \log N} \tag{2.8}$$

or roughly 0.15N [Hop82], when N is small. If the stored patterns are orthogonal to each other, with careful selection, the memory capacity can be as high as half of the neurons of one pattern [Sel89a]. But to practical use, it is still a small number. Abu-Mostafa and St. Jacque [Abu85] have performed a theoretical analysis for storage capacity of object domain neural networks, they found the upper limit of the number of stored patterns is N.

When the memory is addressed by a binary vector that is not one of the stored vector, the net would give a binary valued vector output which is an approximation of the stored vector that is at the shortest Hamming distance from the input vector. The Hamming distance is defined as the number of pixels in one vector that are different from the corresponding ones in another vector. If this output vector is fed back and used as the input to the network, the new output is generally a more accurate version of the stored vector and continued iteration converges to the correct vector. If the stored vectors are unipolar binary (0, 1) $u_i(m)$, the above equations are then applicable with $[2u_i(m)-1]$ replacing $v_i(m)$ and $u_i(mi)$ replacing $v_i(mi)$. For such vectors, the SNR of the estimate $u_i(mi)$ can be shown to be lower by a factor of $\sqrt{2}$ [Far85; Hop82].

The Hopfield net can work synchronously and asynchronously. The energy of the net is defined as

$$E = -\frac{1}{2} \sum_{j \neq i}^{N} \sum_{i=1}^{N} T_{ij} v_i(mi) v_j(mi) \tag{2.9}$$

ΔE due to $\Delta v_i(mi)$ is given

$$\begin{aligned} \Delta E &= -\Delta v_i(mi) \sum_{j \neq i}^{N} T_{ij} v_j(mi) \\ &= -\Delta v_i(mi) v_i(mo) \\ &= -\Delta v_i(mi) \mathrm{sgn}(v_i(mi)) \end{aligned} \tag{2.10}$$

If $v_i(mi)$ changes from +1 to -1, $\Delta E=-(-1-1)x(-1)=-2<0$; if $v_i(mi)$ changes from -1 to +1, $\Delta E=-(+1+1)x(+1)=-2<0$. Thus, the algorithm for altering $v_i(mi)$ always causes E to be monotonically decreased. State changes will continue until a least (local or global) E is reached.

In the above discussion, $T_{ij}=T_{ji}$, that means the matrix is symmetric. An asymmetric T_{ij} can lead to the possibility that a minimum will be only metastable and will be replaced in time by another minimum. Additional asymmetric terms could be easily generated by a minor modification

$$\Delta T_{ij} = A \sum_{m=1}^{M-1} v_i(m+1)v_j(m) \qquad (2.11)$$

When they are added to T_{ij} and A is reasonably adjusted, the system would spend a while near V(m) and then leave and go to a point near V(m+1). In the Hopfield net, we set $T_{ii}=0$, which causes the net to evolve toward a local minimum of an energy function. Selviah [Sel87, 89b] and Gindi [Gin88] found that the net with $T_{ii}=M$ resulted in slightly improved performance compared with that of one with zeros in the diagonal. This is because the net evolves only when updates that lower the energy by a sufficient amount are accepted. If the diagonal elements are not zero and take, instead, a value of M, then the system is easier to implement optically, since all of the element's values are then completely determined by the Hebbian learning rule [Sel87, 90].

2.2 The Hamming Net

The Hamming net is also called the "Optimum minimum error classifier" [Lip87], which calculates the Hamming distance to the exemplar for each class and selects that class with the minimum Hamming distance. The diagram of the Hamming net is shown in Fig. 2.3. Weights and thresholds are first set in the left subnet such that the matching scores generated by the outputs of the middle nodes in Fig. 2.3 are equal to N minus the Hamming distances to the exemplar patterns. These matching scores will range from 0 to the number of elements in the input (N) and are highest for those nodes corresponding to classes with exemplars that best match the input. Usually a threshold of N/2 is chosen in these nodes. Thresholds and weights in the right MAXNET are fixed. All thresholds are set to zero

and weights from each node to itself are 1, while the weights between nodes are inhibitory with a value of -ε, where $\varepsilon < 1/M$ (M is the number of patterns stored in the net). After weights and thresholds have been set, a binary pattern with N elements is presented to the input of the net. It must be presented long enough to allow the matching score outputs of the left subnet to settle and initialize the output values of the MAXNET. The input is then removed and the MAXNET iterates until the output of only one node is positive. Classification is then complete and the selected class corresponds to that node with a positive output. The mathematical algorithms are represented as follows. In the left subnet (Fig. 2.3):

$$T_{im} = v_i(m), \qquad \theta_m = N/2 \tag{2.12}$$

where $1 \le i \le N, 1 \le m \le M$, $v_i(m)$ is one of the binary patterns stored in the interconnections. T_{im} is the connection weight from input i to node m and θ_m is the threshold in that node.

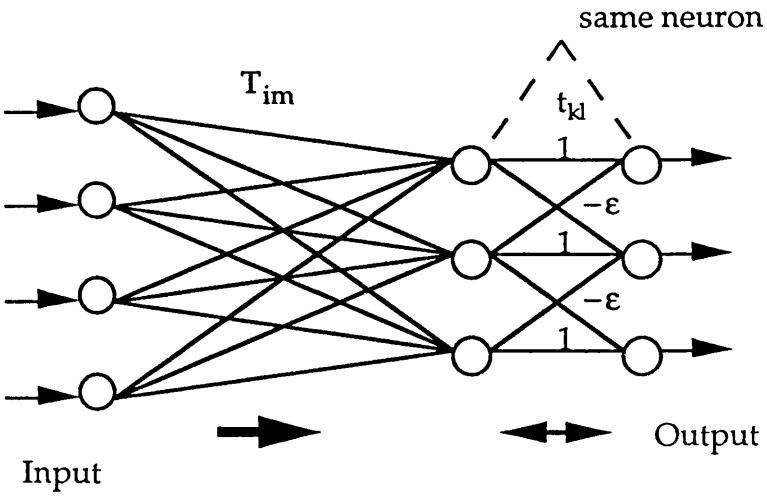


Fig. 2.3 Hamming Net Model

In the right subnet (MAXNET):

$$t_{kl} = 1 \quad (k=l)$$
$$t_{kl} = -\varepsilon \quad (k \neq l, \varepsilon < 1/M) \tag{2.13}$$

where t_{kl} is the connection weight from node k to node l and all the

thresholds in this subnet are zero. When an unknown input pattern $v_i(mi)$ is inserted in the left subnet, the output of node m in the MAXNET can be written as

$$o_m(0) = F_t(\sum_{i=1}^{N} T_{im} v_i(mi) - \theta_m) \tag{2.14}$$

where F_t is a logic threshold (Fig. 2.1(b)). Here and below it is assumed that the maximum input to this threshold never causes the output to saturate. After the nth iteration of the MAXNET, the output of the subnet can be written as

$$o_m(n) = F_t(o_m(n-1) - \varepsilon \sum_{k \neq m}^{M} o_k(n-1)) \tag{2.15}$$

This process is repeated until convergence at which point the output of only one node remains positive.

It can be proven that the MAXNET will always converge as long as the inhibitory interconnection weight ε is less than 1/M [Sel89b]. The Hamming net has a number of obvious advantages over the Hopfield net. It implements the optimal minimum error classifier when bit errors are random and independent, and thus performance of the Hopfield net must either be worse than or equivalent to that of the Hamming net in such situations. The Hamming net also requires fewer connections (M(M+N)) than the Hopfield net (N(N-1)). Finally, the Hamming net does not suffer from spurious output patterns which can produce a 'no-match' result [Lip87].

2.3 Matched Filter Model (Inner-Product Model)

2.3.1 Matched Filter Model

The matched filter network is mainly used for pattern recognition [Sel89b]. It is formally identical to the vector-matrix multiplication operation in the synchronous Hopfield neural network. It is well known that the best way to recognise the existence of a known signal in the presence of unknown,

additive noise is to perform the correlation between the input noise signal V(mi) and the known signal V(m) and to see if this is greater than some threshold:

$$V(mi)*V(m)=\sum_{i=1}^{N} v_i(mi)v_i(m) \qquad (2.16)$$

where * means correlation and the i subscript indicates the bit index. The matched filter model is shown in Fig. 2.4. Instead of representing the weighted interconnection net of a vector-matrix multiplier (Hopfield model), it is represented as a bank of correlating matched filters [Van64]. When an input pattern $v_i(mi)$ is passed through a matched filter with an impulse response of the pattern complex conjugate *V(m) (which, since we assume real signals throughout, is equal to V(m)), the output gives a measure of the closeness for the match of V(mi) and V(m). Hence, to distinguish which code vector V(m) is closest to the input vector V(mi), it is necessary to distribute the input code in equal measure to each of a bank of matched filters as shown in Fig. 2.4 and to search for the largest correlation output.

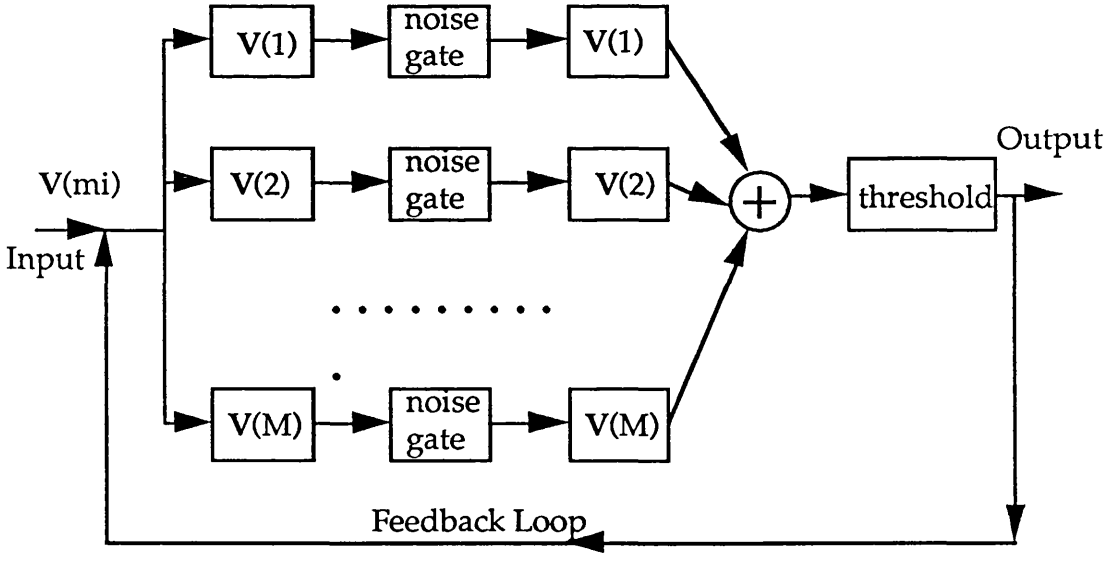


Fig. 2.4 Matched Filter Model

Only the magnitude of the central correlation spike is relevant for the degree of match between V(mi) and V(m) since the central spike corresponds to an exact overlap of the memorised codes and the input

code [Sel89b]. Thus, the correlation sidelobes and much of the associated noise can be gated out. In an optical implementation this function might be performed by one pinhole for each filter [Pae87]. Hence a set of Dirac delta functions can be formed which are scaled by magnitudes of the correlation peaks in each channel. These pass into a second bank of matched filters which have the impulse responses of the memorised codes V(m). Each channel has, therefore, two filters and the output from each of the second filters is the corresponding code scaled by the magnitude of the correlation peak. The scaled output codes from each channel are summed together before being supplied to the nonlinear threshold. The output after the threshold is generally closer to one of the stored patterns than the original input is. This is fed back to the input to start a new iteration to get an even more accurate output. After several iterations, a perfect output, which is one of the stored patterns, is given. Selviah et al [Sel89b] have proved the equivalence of the matched filter network and the vector matrix multiplication neural network (e. g. Hopfield model).

2.3.2 Feature extraction

Feature extraction means that one extracts useful information from a set of patterns, which is characteristic of that set of patterns. The main purpose of the feature extraction in pattern recognition is to encode or modify the stored patterns so that their similarities are suppressed and their differences are emphasised. Therefore, a greater separation in the correlation magnitudes is achieved. This corresponds to performing automatic feature extraction on the original stored patterns, V(m) ($v_i(m)$, i=1, 2,..., N; m=1, 2,..., M) to create a new set of feature patterns F(m) ($f_i(m)$, i=1, 2,...,N; m=1, 2,..., M) with the essential features of the original patterns which distinguish each from the rest of the set. A simplified feature extractor is shown in Fig. 2.5 [Sel89a]. A memory pattern V(k) is presented to the system and features of it are calculated and organized into the elements of a feature vector F(k). The feature pattern produced from a particular original pattern depends on the exact nature of the other original patterns in the set. One way in which this can be done is by using a modification of the Gram-Schmidt orthogonalisation procedure [Koh88]. This technique produces linear combinations of the original features that

are independent of each other. This is usually intended to improve discrimination, but its primary objective is to simplify the required post-processing by performing dimensionality reduction. One iterative feature extraction algorithm based on the modified Gram-Schmidt technique can be summarised as follows (Fig. 2.5):

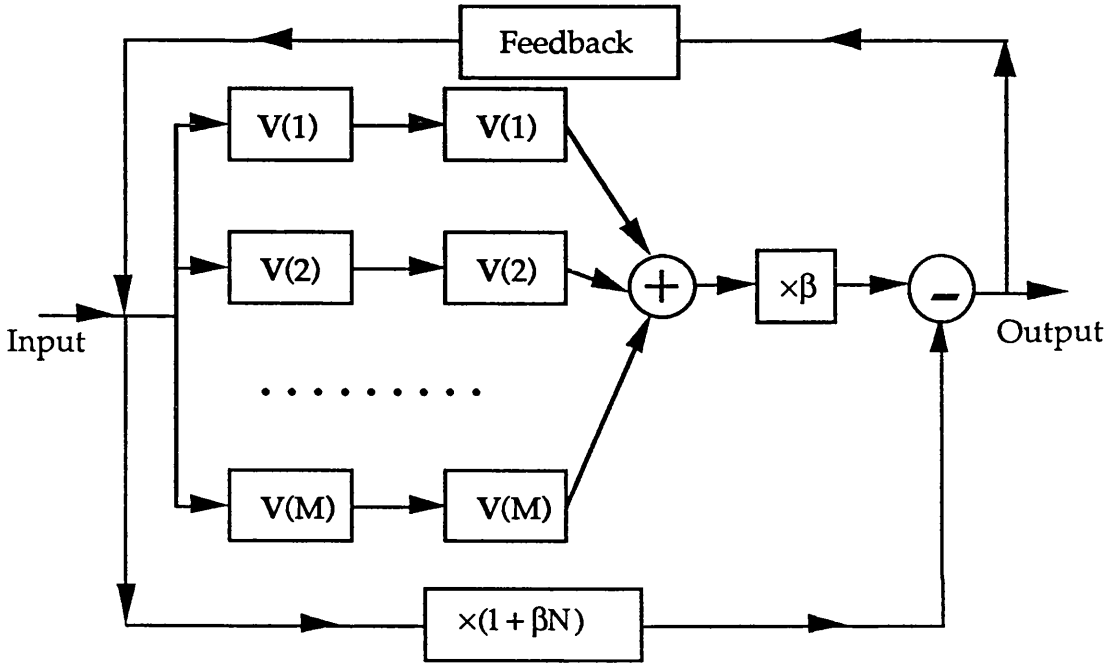


Fig. 2.5 Feature Extraction System

(1) Initialise the feature pattern, before any iterations, to be the original pattern

$$\mathbf{F}(m)^{(0)} = \mathbf{V}(m) \tag{2.17}$$

(2) At the nth iteration evaluate

$$\mathbf{G}(m)^{(n)} = \mathbf{F}(m)^{(n-1)}(1+\beta N) - \beta \sum_{m=1}^{M} (\mathbf{F}(m)^{(n-1)} \cdot \mathbf{V}(m))\mathbf{V}(m) \tag{2.18}$$

Repeat for all m where $\mathbf{G}(m)^{(1)}$, for example, is the unnormalised feature code after one iteration;

(3) Perform the nth normalisation

$$F(m)^{(n)} = \frac{G(m)^{(n)}N}{G(m)^{(n)} \cdot V(m)} \tag{2.19}$$

where $\beta<1/M$ determines the rate of convergence. Step (2) and (3) are repeated $1/\beta$ times to give final feature patterns. More accurate final feature patterns are obtained for smaller values of β.

2.3.3 Orthogonal coding

The orthogonal patterns are completely independent of each other. For a space spanned by N independent coordinates, the maximum number of orthogonal patterns is limited to N. The classical computational method for the evaluation of orthogonal projections is the Gram-Schmidt process [Koh88]. Let there be M distinct Euclidean vectors denoted by V(1), V(2), ..., V(M), a new orthogonal vector basis O(m) (m=1, ..., M) is defined by the recursion

$$O(1) = V(1)$$

.....

$$O(m) = V(m) - \sum_{i=1}^{m-1} \frac{(V(m), O(i))}{|O(i)|^2} O(i)$$

.....

$$O(M) = V(M) - \sum_{i=1}^{M-1} \frac{(V(M), O(i))}{|O(i)|^2} O(i) \tag{2.20}$$

where (V(m), O(i)) is the inner product of V(m) and O(i) and the sum must be taken over such terms for which $O(i) \neq 0$. According to the Gram-Schmidt process, there exists a set of orthogonal vectors, but the elements of the vectors can be any value. In the neural network, we are more interested in binary data. One way to produce N orthogonal binary and bipolar patterns is described as follows (let N be 16):

(1) Set

o1 = {1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1) (unit vector);

(2) Reverse the second half of the elements of the unit vector

o2 ={1, 1, 1, 1, 1, 1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1)

(3) Divide o1, o2 into two parts with equal length and reverse the second half of each part, we have

o3 = {1, 1, 1, 1, -1, -1, -1, -1, 1, 1, 1, 1, -1, -1, -1, -1)
o4 = {1, 1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1, 1, 1, 1, 1}

(4) Divide o1, o2, o3, o4 into four parts and reverse the second half of each part, we have

o5 = {1, 1, -1, -1, 1, 1, -1, -1, 1, 1, -1, -1, 1, 1, -1, -1}
o6 = {1, 1, -1, -1, 1, 1, -1, -1, -1, -1, 1, 1, -1, -1, 1, 1}
o7 = {1, 1, -1, -1, -1, -1, 1, 1, 1, 1, -1, -1, -1, -1, 1, 1}
o8 = {1, 1, -1, -1, -1, -1, 1, 1, -1, -1, 1, 1, 1, 1, -1, -1}

(5) Similarly repeat step (4), we can have the rest 8 vectors

o9 = {1, -1, 1,-1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1}
o10={1, -1, 1, -1, 1, -1, 1, -1, -1, 1, -1 1, -1, 1, -1, 1}
o11={1, -1, 1, -1, -1, 1, -1, 1, 1, -1, 1, -1, -1, 1, -1, 1}
o12={1, -1, 1, -1, -1, 1, -1, 1, -1, 1, -1, 1, 1, -1, 1, -1}
o13={1, -1, -1, 1, 1, -1, -1, 1, 1, -1, -1, 1, 1, -1, -1, 1}
o14={1, -1, -1, 1, 1, -1, -1, 1, -1, 1, 1, -1, -1, 1, 1, -1}
o15={1, -1, -1, 1, -1, 1, 1, -1, 1, -1, -1, 1, -1, 1, 1, -1}
o16={1, -1, -1, 1, -1, 1, 1, -1, -1, 1, 1, -1, 1, -1, -1, 1}

From the above algorithm, N must be an integer power of 2. If N does not satisfy this condition, N orthogonal patterns exist, but possibly not binary and bipolar. Fig. 2.6 is the 16 orthogonal patterns, each with 16 elements. Here white elements stand for +1, and black ones -1.

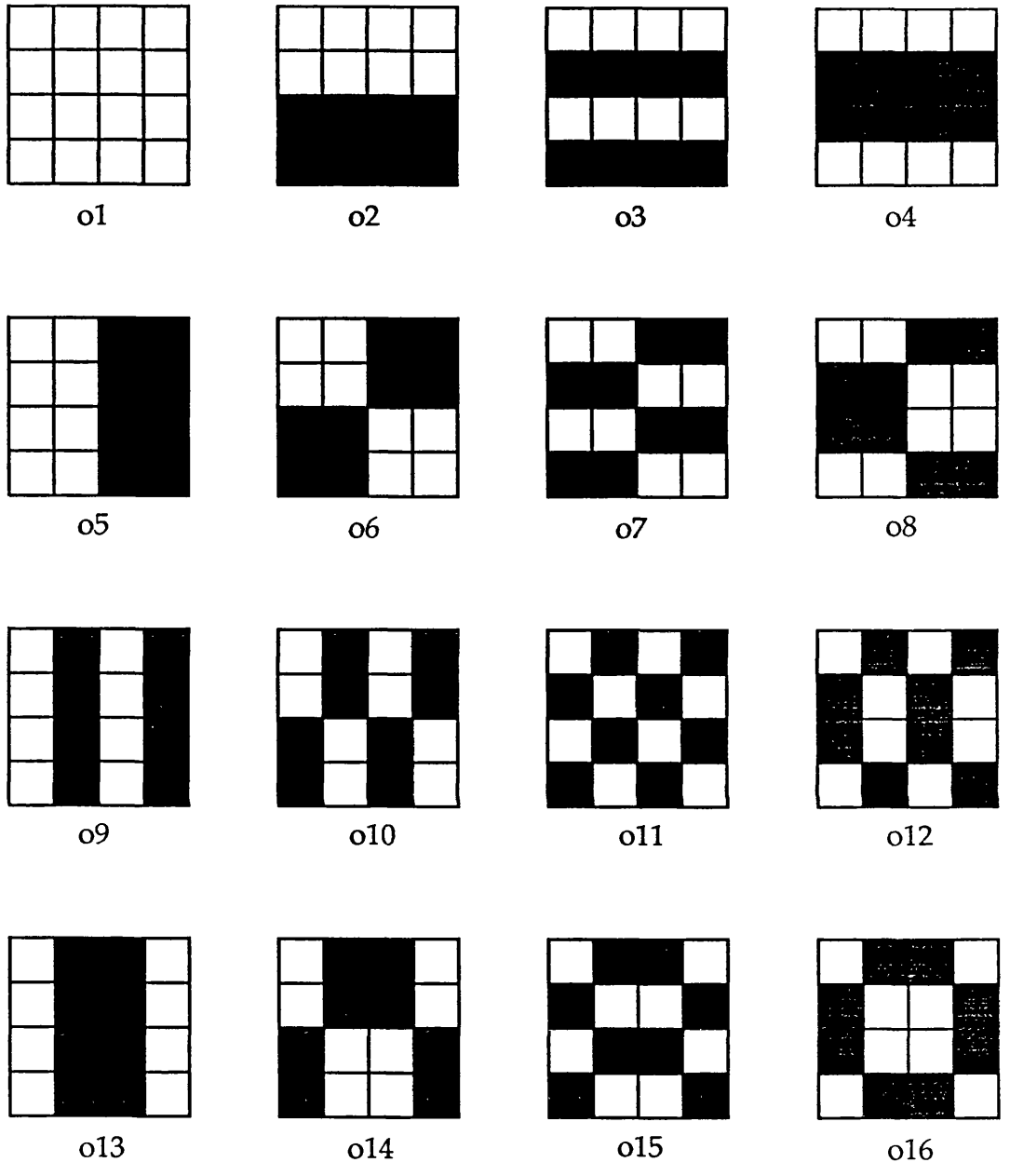


Fig. 2.6 Sixteen Orthogonal Binary and Bipolar Patterns
(each with 4x4 pixels)

2.4 High-order Neural Networks

2.4.1 Outer-product model

As mentioned above, the Hebbian-learning-rule based Hopfield model has

a limited memory capacity (N/4logN). When M exceeds the capacity, a lot of minimum energy states turn into spurious states and the network's ability to recognise the memorised patterns is drastically reduced. Furthermore, this net has a symmetric memory matrix, so it can only be used for an auto-associative memory system. A higher order neural network model has a much larger memory capacity by increasing the number of interconnections and can be used for hetero-associative memory [Lee86; Che86]. In such a model, the Hebbian matrix is replaced by a multiply associated high order correlation tensor

$$T_{i1i2..j1j2..k1k2..} = \sum_{m=1}^{M} u_{i1}(m)u_{i2}(m)...v_{j1}(m)v_{j2}(m)...w_{k1}(m)w_{k2}(m)... \tag{2.21}$$

where U(m), V(m), ..., W(m) belong to one of the stored pattern set. They are multiply associated in the sense that one could recall any pattern from the set by an input of partially correct remainder patterns in the set. Following the Hopfield net, we define an energy function

$$E = -\sum_{i1,i2..j1,j2..k1,k2..} (s_{i1}s_{i2}...t_{j1}t_{j2}...r_{k1}r_{k2}...T_{i1i2...j1j2...k1k2...}) \tag{2.22}$$

In this expression, we have set the threshold to zero for convenience. Note also that we have used a set of patterns (S, T, R...) with connections among neurons in different patterns, as well as among neurons within the same pattern. Different patterns may have different sizes.

The dynamic evolution of the neurons is governed by

$$s_{i1} = F_t\left(\sum_{i2..j1j2..k1k2..} T_{i1i2...j1j2...k1k2...}s_{i2}...t_{j1}t_{j2}...r_{k1}r_{k2}...\right) \tag{2.23}$$

The updating of the neurons can be done either asynchronously or synchronously. Such a net always converges to a point with minimum energy. For auto-associative recall, we have the tensor

$$T_{ijk\ldots} = \sum_{m=1}^{M} v_i(m)v_j(m)v_k(m)\ldots \tag{2.24}$$

Output of the neurons is

$$v_i(mo) = \sum_{j,k,\ldots} T_{ijk\ldots} v_j(mi)v_k(mi)\ldots \tag{2.25}$$

where $v_i(mi)$ is an input pattern and energy is

$$E = -\sum_{i,j,k\ldots} v_i(mi)v_j(mi)v_k(mi)\ldots T_{ijk\ldots} = -\sum_{m=1}^{M}\left(\sum_i v_i(mi)v_i(m)\right)^n \tag{2.26}$$

where n is the order of the correlations.

2.4.2 Inner-product model

In the outer-product model, the order of the tensor is very high according to the degree of the high order. It is difficult to implement it practically, either electronically or optically. In order to implement such a net, we re-write the evolution of the neurons for auto-associative memory and re-organise it as

$$\begin{aligned} v_i(mo) &= \sum_{(j,k,\ldots)}^{N}\sum_{m=1}^{M} v_i(m)v_j(m)v_k(m)\ldots v_j(mi)v_k(mi)\ldots \\ &= \sum_{m=1}^{M} v_i(m)\left(\sum_{j=1}^{N} v_j(m)v_j(mi)\sum_{k=1}^{N} v_k(m)v_k(mi)\ldots\right) \\ &= \sum_{m=1}^{M} v_i(m)\left(\sum_{j=1}^{N} v_j(m)v_j(mi)\right)^n \\ &= \sum_{m=1}^{M} C(m)^n v_i(m) \end{aligned} \tag{2.27}$$

where

$$C(m) = \sum_{j=1}^{N} v_j(m) v_j(mi) \tag{2.28}$$

is an inner-product between the stored vectors and the input vector and n is the order of correlation. If n = 1, it is equivalent to the matched filter model (linear inner-product model). Such high order nets can have a large memory capacity and can be implemented optically much more easily than outer-product high order nets [Pae87; Owe87].

From the above analysis, we can see an outer-product model can be converted into an inner-product model. They are equivalent in functioning characteristics. Generally speaking, inner-product models require a large number of interconnections,which can be implemented optically. The inner-product model, however, needs nonlinear neurons for its implementation, in which electronics has some advantage over optics.

2.5 Multilayer Perceptron

The idea of the multilayer perceptron was originally used to solve the problems of mapping from the similarity structure of the input to a different similarity structure of the output, such as the exclusive-or (XOR) problem. The algorithm of the multilayer perceptron is based on the so-called generalised delta rule (GDR) [Rum86]. This net first uses the input vector to produce its own output vector and then compares this with the desired output, or target vector. If there is no difference, no learning takes place. Otherwise the weights are changed to reduce the difference. If there are no hidden units, the net becomes the perceptron model [Koh88]. The rule for changing weights follows the presentation of input/output pair p

$$\Delta_p T_{ji} = \eta\left(t_{pj} - o_{pj}\right) i_{pi} = \eta \delta_{pj} i_{pi} \tag{2.29}$$

where t_{pj} is the target input for the *j*th component of the output vector for pattern p, o_{pj} is the *j*th element of the actual output pattern produced by presentation of input pattern p, i_{pi} is the value of the *i*th element of the input pattern, $\delta_{pj} = t_{pj} - o_{pj}$, and $\Delta_p T_{ji}$ is the change to be made to the weight from the *i*th to the *j*th unit following presentation of pattern p.

For the multilayer perceptron model (Fig. 2.7), there exists hidden units, but the learning algorithm can still be deduced from the above equation if we modify this equation a little. We rewrite the above equation as

$$\Delta_p T_{ji} = \eta \delta_{pj} o_{pi} \tag{2.30}$$

We use o_{pi} to replace i_{pi}. Here o_{pi} are the input elements to that layer. So for the first layer

$$o_{pi} = i_{pi} \tag{2.31}$$

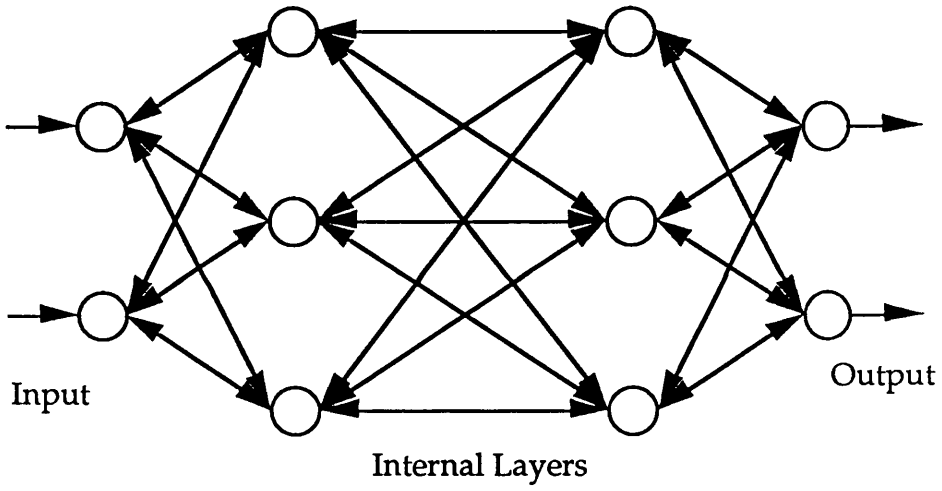


Fig. 2.7 Multiplayer Perceptron with Error Back Propagation

and for the internal layers, the o_{pi} is the output from the previous layer. The determination of the error signal is a recursive process which starts with the output units. If a unit is an output unit, its error signal is very similar to the standard delta rule. This is given by

$$\delta_{pj} = (t_{pj} - o_{pj}) F_t{}'_j(net_{pj}) \tag{2.32}$$

where $F_t{}'_j(net_{pj})$ is the derivative of the threshold function. A useful threshold function is the logistic activation function in which

$$o_{pj} = \frac{1}{1+e^{-\sum_i T_{ji} o_{pi} + \theta_j}} \tag{2.33}$$

where θ_j is a bias similar in function to a threshold and net_{pj} is

$$net_{pj} = \sum_i T_{ji} o_{pi} + \theta_j \tag{2.34}$$

It is interesting to note that the linear threshold function, on which the perceptron model is based [Rum86], is discontinuous and hence will not suffice for the GDR. The error signal for hidden units for which there is no specified target is determined recursively in terms of the error signals of the units to which it directly connects and the weights of those connections. That is

$$\delta_{pj} = F_t'_j(net_{pj}) \sum_k \delta_{pk} T_{kj} \tag{2.35}$$

whenever the unit is not an output unit.

So the application of the GDR involves two phases:

(1) The input is presented and propagated forward through the network to compute the output value opj for each unit. This output is then compared with the targets, resulting in an error signal for each output unit.

(2) A backward pass through the network is involved (analogous to the initial forward pass) during which the error signal is passed to each unit in the network and the appropriate weight changes are made.

In this net, not all weights need to be variable. Any number of weights in the network can be fixed.

2.6 Comments on Neural Networks

Neural networks try to simulate the functions of the brain. They are designed to be good at processing analog data, such as pattern recognition, speech processing etc. They do not do digital calculations as well as the present electronic computers. So they are complementary machines to electronic computers rather than better replacements. Electronics has been

shown not to be very successful for implementation of neural computers with a large number of neurons because of their planar connections due to the interference of two intersected electrical signals and it offers only a limited fanout and connectivity. So although it is useful for a small number of neurons, it does not scale as the number of neurons increases. There is no interference of two intersected optical beams in the free space, optics can provide a large number of very complex interconnections. Recently optical implementations of neural computers have been the hot subject, stimulated by the development of high quality spatial light modulators and some other high speed logic optoelectronic devices [Tan85; Whe88]. The difficulty in storing large information in a small area still exists because of lack of a suitable material like silicon in electronics. Photorefractive materials have been used as the information storage media and up to 500 patterns have been stored in a $LiNbO_3$ material recently [Mok91]. It is still too far from practical use. We look forward to the breakthrough in storage media. It might take a time for optical neural computers to emerge from the laboratory.

Chapter Three

HOLOGRAPHY

3.1 Introduction

The roots of holography can be traced back to work by Gabor [Gab48]. A hologram is an interference pattern which contains all the information about a surface or transparency in a coded form. The hologram is recorded on a photographic plate with high resolution. During exposure, this plate is located in the front of the object in such a way that it is illuminated not only by the light from the object (which is also called object beam), but also by another beam of light (reference beam) which is coherent to the object beam. So the interference fringes are formed in the plate. A special feature of the holographic method is the explicit or implicit interaction of two beams of mutually coherent light. A hologram does not contain any elements resembling the object. The recorded image of an object is made visible by illuminating its hologram with the reference beam. A hologram is characterised by a high degree of redundancy. It can be divided into several parts and reproduces the same image from any part. But the sharpness of the image and its three-dimensional nature become weaker and weaker as the size of the part is reduced [Sor71]. The term "holography" originates from the Greek work "holo" which means complete and "grapho" which means "writing", so the composite word means complete recording. Holograms were originally stimulated by trying to increase the resolution in electron microscopes [Gab48, 49]. It was found that if an object was recorded in a hologram by a beam with wavelength of λ_1 and reconstructed by another beam with wavelength of λ_2, the reconstructed image could be magnified by the order of λ_2/λ_1 [Yu73]. Later holographic technique was found to be useful for storing three-dimensional objects. Such a hologram can reveal a real three dimensional picture. When you change the viewing position, you can see the different

parts of the image stored in the hologram. Fourier transform holograms are especially useful to store information because of their information compression [Kni74]. A whole page of text can be stored in a very small hologram, say 200μm. As the hologram includes phase information as well as amplitude information, it is widely used for image processing systems, such as pattern recognition, image deblurring, and pseudo-colour encoding [Lee81]. There are two kinds of hologram, the thin or planar hologram and the thick or volume hologram. If the thickness of the recording gelatine, s, is much less than the fringe period of the hologram, d, say s<10d, it is called a planar hologram, otherwise it is called a volume hologram [Kog69; Sol81]. As an example, a holographic plate, Kodak 649F has the thickness of emulsion, about 15 μm. When it is used to record a transmission hologram, the fringe period, d, is several microns, such a hologram is regarded as a thin hologram. If, however, it is used to record a reflection hologram, in which d is less than one micron, such a hologram can be regarded as a thick hologram.

Recently, more interest has been directed towards holographic interconnections. It is stimulated by liquid crystal and photorefractive materials which can record real-time volume holograms. In the volume holograms, one-to-one interconnections are proposed by using special coding because of the Bragg angular selectivity [Psa88c; Mid89]. Recently optical interconnections have been implemented by the use of volume holograms in photorefractive materials [Yeh88; Bra89]. First, we discuss the basic principles of holograms. Then Fourier Transform holograms are introduced and based on this, the Van der Lugt matched filter correlation system [Van64; Col81] is described. After that, we discuss another correlation system, which we call the spectrum correlation system. It is used in our neural network (HOFNET) system. Finally, we cover the encoding methods for information storage and describe the memory capacity. In this chapter, all the holograms are assumed to be thin transmissive ones, unless they are specified.

3.2 Holographic Principle

A hologram records not only the amplitude distribution of an object as in normal photographs, but also the phase information. The first hologram was formed by Gabor [Gab48] using visible light. As highly coherent light --

laser was not available when Gabor made the first hologram, he made the hologram by using a highly transmissive object illuminated by a light source. The transmitted light consists of two components: (1) a strong, uniform plane wave directly passed; (2) a weak scattered wave generated by the image. So the object beam has the same angle as the reference beam. Such recording would relieve the requirement for light coherence, but the difficult is that the reconstructed image is superimposed on the background light [Goo69]. Because of this, such holograms are called on-line holograms. In order to overcome this difficulty, a hologram can be stored by using one beam illuminating the object and another beam as a reference beam, which has an angle to the object beam. Such recording requires a good coherent light source to make sure that at any point in the photographic plane, the light from the object is coherent with that from the reference. As in such a hologram, the object beam has a different direction from the reference, such a hologram is called an off-line hologram [Col71]. The first successful off-line hologram was made by Leith and Upatnieks [Lei62]. A general hologram recording system is shown in Fig. 3.1, a separate reference beam derived from the same coherent source is used to illuminate the photographic plate during the recording process, at an offset angle θ to the beam from the object. For simplicity, this reference beam is assumed to be a collimated beam with uniform intensity. In this system, the amplitude distribution in the recording plane is the Fresnel diffraction of the object, so such holograms are also called Fresnel holograms.

The complex amplitude from the object beam at any point (x, y) on the photographic plate can be written as

$$o(x,y) = |o(x,y)|\exp[-i\phi(x,y)] \tag{3.1}$$

and that of the reference beam is

$$r(x,y) = r\exp(i2\pi\xi x) \tag{3.2}$$

where o(x,y) is the object amplitude at point (x,y) in the photographic plane, $\phi(x,y)$ is the phase distribution, $\xi=\sin\theta / \lambda$ is the modulation or spatial carrier frequency, r is the constant amplitude of the reference beam and λ is the wavelength of the laser. Since the reference beam has uniform

intensity, only its phase varies across the photographic plate. Here we assume that the phase of the reference beam is zero at the origin of the coordinates. The resultant intensity at the photographic plate is

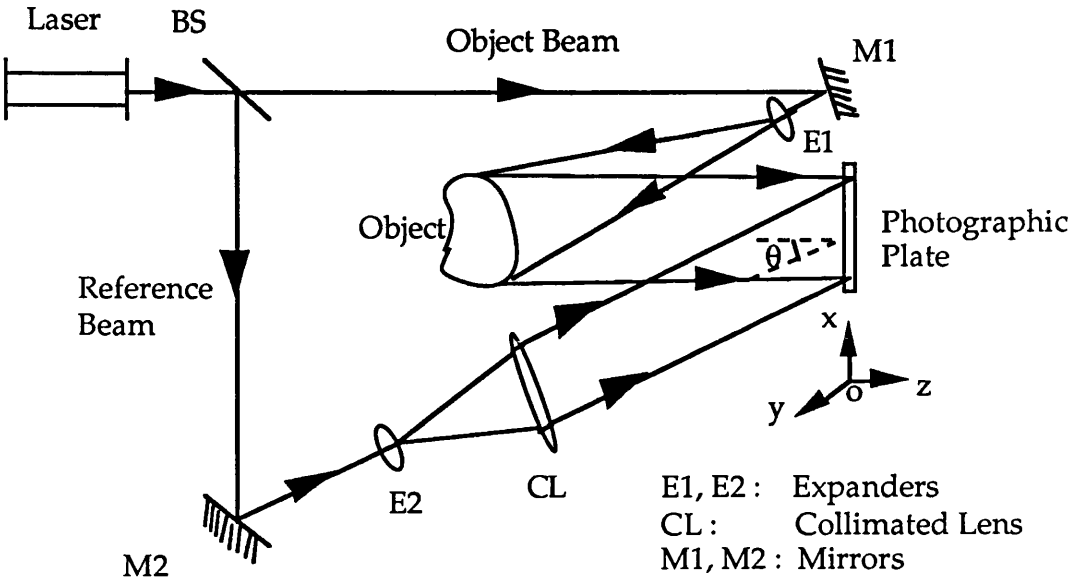


Fig. 3.1 Arrangement for Recording a Hologram

$$\begin{aligned} I(x,y) &= |r(x,y)+o(x,y)|^2 \\ &= |r(x,y)|^2 + |o(x,y)|^2 \\ &\quad + r|o(x,y)|\exp[-i\phi(x,y)]\exp(-i2\pi\xi x) \\ &\quad + r|o(x,y)|\exp[i\phi(x,y)]\exp(i2\pi\xi x) \end{aligned} \tag{3.3}$$

that is

$$I(x,y) = r^2 + |o(x,y)|^2 + 2r|o(x,y)|\cos[2\pi\xi x + \phi(x,y)] \tag{3.4}$$

The third term on the right hand side of the equation indicates that the amplitude and phase of the object are encoded, respectively. They are modulated by a set of interference fringes with a spatial carrier frequency equal to ξ.

It is assumed that this plate is processed in such a way that its amplitude transmittance t (the ratio of the transmitted amplitude to that incident on it) is a linear function of the intensity and can be written as

$$t = t_0 + \beta TI \tag{3.5}$$

where t_0 is a constant background transmittance, T is the exposure time and β is a parameter determined by the photographic material used and the processing conditions. Substitute Eq. (3.3) into Eq. (3.5), the amplitude transmittance of the hologram is, accordingly,

$$\begin{aligned} t(x,y) = t_1 + \beta T\{&|o(x,y)|^2 \\ &+ r|o(x,y)|\exp[-i\phi(x,y)]\exp(-i2\pi\xi x) \\ &+ r|o(x,y)|\exp[i\phi(x,y)]\exp(i2\pi\xi x)\} \end{aligned} \tag{3.6}$$

where $t_1=t_0+\beta Tr^2$.

To reconstruct the image, the hologram is illuminated once again, as shown in Fig. 3.2 with the same reference beam that was used to record it. The complex amplitude u(x,y) of the transmitted beam is, in this case also, the sum of four terms, each corresponding to one of the terms of Eq. (3.6), and can be written as

$$\begin{aligned} u(x,y) &= r(x,y)t(x,y) \\ &= u_1(x,y) + u_2(x,y) + u_3(x,y) + u_4(x,y) \end{aligned} \tag{3.7}$$

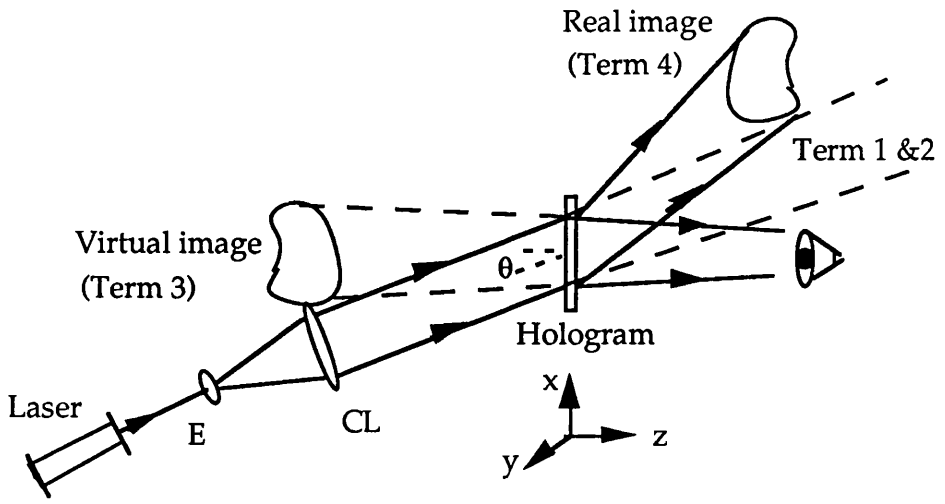


Fig. 3.2 Reconstruction of the Hologram

where

$$u_1 = t_1 r \exp((i2\pi\xi x)$$
$$u_2 = \beta T r |o(x,y)|^2 \exp(i2\pi\xi x)$$
$$u_3 = \beta T r^2 o(x,y)$$
$$u_4 = \beta T r^2 o^*(x,y)\exp(i4\pi\xi x) \qquad (3.8)$$

where * means complex conjugate.

The first term on the right hand side of Eq. (3.7), $u_1(x.y)$, is merely the attenuated reference beam, which is a plane wave directly transmitted through the hologram. This directly transmitted beam is surrounded by a halo due to the second term, $u_2(x,y)$, which is spatially varying. The angular spread of this halo is determined by the angular extent of the object. The third term, $u_3(x,y)$ is identical with the original object wave, except for a constant factor, and generates a virtual image of the object in its original position; this wave makes an angle θ with the directly transmitted light. Similarly, the fourth term, $u_4(x,y)$, gives rise to the conjugate image. However, in this case, the fourth term includes an exponential factor, $\exp(i4\pi\xi x)$, which indicates that the conjugate image is deflected off the axis at an angle approximately twice what the reference wave makes with it. Thus, even though two images - one conjugate (u_4) and one original (u_3) - are reconstructed in this setup, they are angularly separated from the directly transmitted beam and from each other, and if the offset angle θ of the reference beam is made large enough, it is possible to ensure that there is no overlap.

The minimum value of the θ is required to ensure that each of the images can be observed without any interference from the others, as well as from the directly transmitted beam and the halo of scattered light surrounding it. It is determined by the minimum spatial carrier frequency ξ for which there is no overlap between the angular spectra of the third and fourth terms and those of the first and second terms. Suppose the highest frequency in the spatial frequency spectrum of the object beam is ξ_m. According to the calculation [Har84], if the spatial carrier frequency ξ is chosen so that

$$\xi \geq 3\,\xi_m \qquad (3.9)$$

then u_3 and u_4 will not overlap u_1 and u_2. As $\xi = \sin\theta/\lambda$, so the minimum requirement for θ is

$$\theta \geq \sin^{-1}(3\lambda\,\xi_m) \qquad (3.10)$$

3.3 Fourier Transform Holograms (FTHs)

Another interesting and important category of hologram recording configurations is one in which the interference at the photographic plate is between the Fourier transform of the object and a parallel reference beam. Such holograms are called Fourier Transform Holograms (FTHs). A typical optical arrangement for recording a FTH [Col71] is shown in Fig. 3.3. The object, a transparency, located in the front focal plane of a Fourier transform lens FL, is illuminated by a parallel beam of coherent light. If the complex amplitude leaving the object plane is o(x,y), then the complex amplitude of the object beam at the photographic plate, located in the back focal plane of lens FL, is

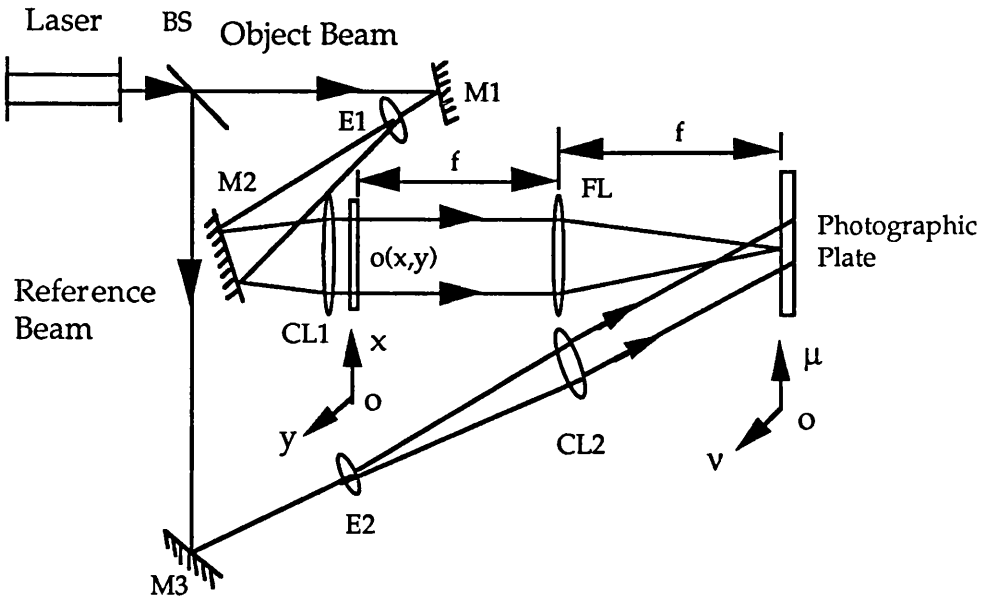


Fig. 3.3 Fourier Transform Hologram Recording Setup

$$O(\mu, \nu) = F.T.\{o(x,y)\} \tag{3.11}$$

where F.T. is the Fourier Transform function (see Appendix). The complex amplitude of the reference beam at the photographic plate is

$$R(\mu,\nu) = R\exp(i2\pi\xi\lambda f\mu) \tag{3.12}$$

where R is the constant amplitude, f is the focal length of the lens CL2, (x, y) is the coordinate pair in the object plane and (μ, ν) is the coordinate pair in the Fourier Transform plane or the photographic plane (see Fig. 3.3). The relation between (μ, ν) in the frequency domain and (x, y) in the spatial domain is

$$\mu = \frac{x}{\lambda f}\ ; \qquad \nu = \frac{y}{\lambda f} \tag{3.13}$$

For simplicity we suppose the reference amplitude R is 1. The intensity of the interference pattern formed by these two beams is, therefore,

$$\begin{aligned} I(\mu,\nu) &= |R(\mu,\nu) + O(\mu,\nu)|^2 \\ &= 1 + |O(\mu,\nu)|^2 + O(\mu,\nu)\exp(-i2\pi\xi\lambda f\mu) + O^*(\mu,\nu)\exp(i2\pi\xi\lambda f\mu) \end{aligned} \tag{3.14}$$

To reconstruct the image, the processed hologram is placed in the front focal plane of lens FL and illuminated with a parallel coherent beam with amplitude A as shown in Fig. 3.4. For simplicity, we suppose A = 1. It is assumed as before, that the amplitude transmittance of the processed hologram is a linear function of I(μ, ν), the complex amplitude of the light transmitted by the hologram is

$$\begin{aligned} U(\mu,\nu) &= t_0 + \beta T\, I(\mu,\nu) \\ &= (t_0 + \beta T) + \beta T\,|O(\mu,\nu)|^2 + \\ &\quad + \beta T\, O(\mu,\nu)\exp(-i2\pi\xi\lambda f\mu) + \beta T\, O^*(\mu,\nu)\exp(i2\pi\xi\lambda f\mu) \end{aligned} \tag{3.15}$$

The complex amplitude in the back focal plane of lens FL is the inverse

Fourier transform of U(μ, ν) (see Appendix)

$$\begin{aligned} u(x,y) &= \text{F.T. } \{U(\mu,\nu)\} \\ &= (t_0 + \beta T)\,\delta(x,y) + \beta T\, o(x,y) * o(x,y) \\ &\quad + \beta T\, o(x+\xi\lambda f, y) + \beta T\, o^*(-x+\xi\lambda f, -y) \end{aligned} \tag{3.16}$$

As shown in Fig. 3.4, the first term on the right-hand side of Eq. (3.16) comes to a focus on the axis, while the second term forms a halo around it. The third term corresponds to the original object shifted downwards by a distance $\xi\lambda f$ or $f\sin\theta$, while the fourth term is the conjugate of the original object inverted and shifted upwards by the same amount. Both the images are real and can be recorded on a photographic film placed in the back focal plane of lens FL.

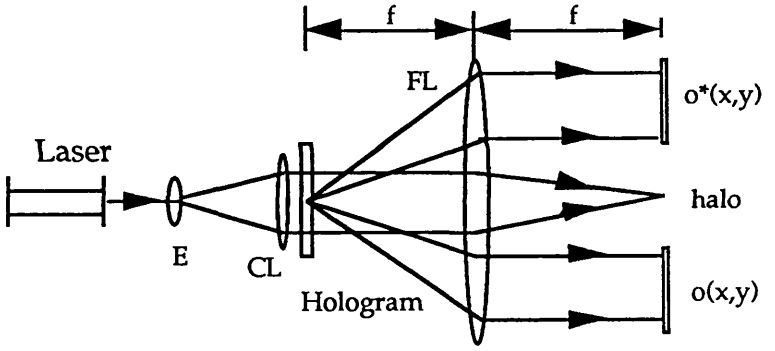


Fig. 3.4 Reconstruction of the Fourier Transform Hologram

Fourier transform holograms (FTHs) have a useful property called space invariance that the reconstructed image is stationary even when the hologram is translated in its own plane. This is because a shift of a function in the spatial domain only results in its Fourier transform being multiplied by a phase factor which is a linear function of the spatial frequency. This phase factor has no effect on the intensity distribution in the image. In the Fourier transform hologram, the object information which usually occupies a large area is compressed in the vicinity of the focal point and the area used to record the hologram is small compared to the object itself. The size of the FTH depends on the focal length f, the wavelength of the laser, the fineness structure of the object pattern and the

aberration of the FT system. So FTHs can be used for the purpose of high-density recording.

3.4 Van der Lugt Matched Filter (Spatial Correlation System)

The Van der Lugt matched filter [Van64; Col71] is used to recognise a pattern from several patterns by correlation. For example, if we want to recognise a pattern g(x,y) from a bank of patterns $f_i(x,y)$ (i=1, ..., M), we can correlate g(x,y) with all $f_i(x,y)$ and decide their similarities according to their correlation magnitudes. The larger the correlation peak is, the more similar will be the two patterns. One way to perform the correlation of two patterns f(x,y) and g(x,y) is by the multiplication of their Fourier Transform spectra. By the convolution theorem,

$$f(x,y)*g(x,y) = F.T.\{F(\mu,\nu)G^*(\mu,\nu)\} \tag{3.17}$$

where $F(\mu,\nu)$ and $G(\mu,\nu)$ are the Fourier transforms of f(x,y) and g(x,y) respectively. In optics, multiplication is much easier to implement and the Fourier transform can easily be performed by a lens. So instead of directly performing the correlation, we implement the multiplication and then Fourier Transform it. The problem, as you see from Eq. (3.17), is the recording of a conjugate Fourier transform spectrum of g(x,y). The system proposed by Van der Lugt for storing the $G^*(\mu,\nu)$ is the same as Fig. 3.3 with o(x,y) replaced by g(x,y) and the correlation implementation system is shown in Fig. 3.5. So the transmittance t of the processed hologram (see Eq. (3.5) becomes

$$\begin{aligned} t(\mu,\nu) = {} & (t_0 + \beta T) + \beta T|G(\mu,\nu)|^2 + \\ & + \beta T G(\mu,\nu)\exp(-i2\pi\xi\lambda f\mu) + \beta T G^*(\mu,\nu)\exp(i2\pi\xi\lambda f\mu) \end{aligned} \tag{3.18}$$

Here we still suppose the amplitude of the reference beam is 1. We insert an unknown pattern f(x,y) in the front focal plane of lens FL1 and at the back focal plane, where the FTH is placed, we get the Fourier transform of f(x,y), that is $F(\mu,\nu)$.

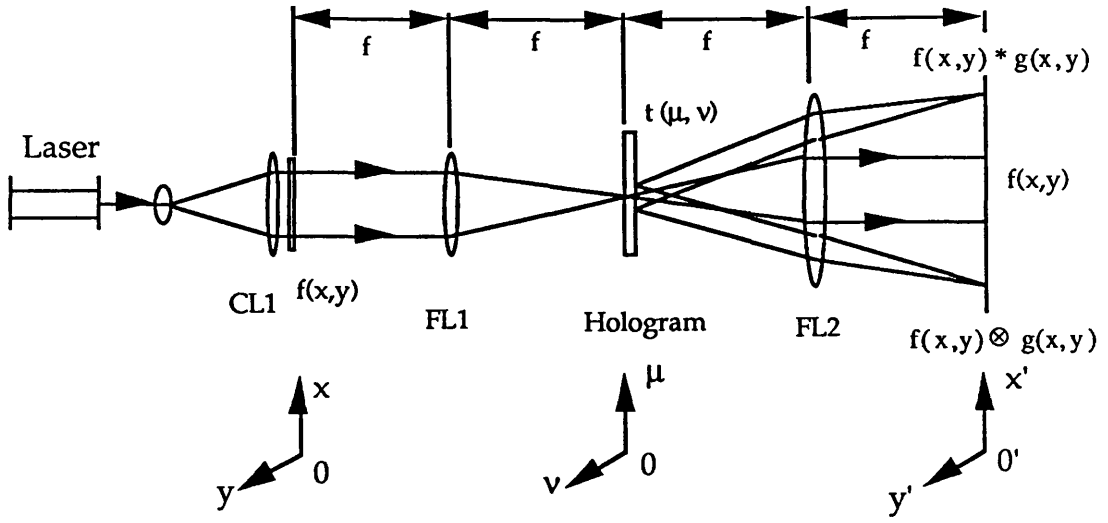


Fig. 3.5 Van der Lugt Matched Filter Correlation System

So after the hologram, the amplitude can be expressed as

$$\begin{aligned} U(\mu,\nu) &= F(\mu,\nu)\times t(\mu,\nu) \\ &= F(\mu,\nu)(t_0+\beta T)+\beta TF(\mu,\nu)|G(\mu,\nu)|^2 \\ &\quad +\beta TF(\mu,\nu)G(\mu,\nu)\exp(-2\pi\xi\lambda f\mu) \\ &\quad +\beta TF(\mu,\nu)G^*(\mu,\nu)\exp(2\pi\xi\lambda f\mu) \end{aligned} \tag{3.19}$$

At the back focal plane of lens FL2, the complex amplitude is

$$\begin{aligned} u(x',y') &= F.T.\{U(\mu,\nu)\} \\ &= (t_0+\beta T)f(x',y')+\beta Tf(x',y')\otimes g(x',y').g(x',y') \\ &\quad +\beta Tf(x',y')\otimes g(x',y')\otimes\delta(x'+\xi\lambda f,y') \\ &\quad +f(x',y').g(x',y')\otimes\delta(x'-\xi\lambda f,y') \end{aligned} \tag{3.20}$$

where (x′, y′) are coordinates in the back focal plane of lens FL2.

The first term is the attenuated image of the input pattern and the second term is the auto-correlation of g(x′,y′) (a little bit similar to δ−function) which is convoluted by the pattern f(x′,y′). It contributes to a halo superimposed with the first term. The third term is the convolution between the input pattern and the stored pattern, shifted downwards at $x'=-\xi\lambda f$ or $(-f\sin\theta)$. The fourth term is the correlation of f(x′,y′) and g(x′,y′) located upwards at $x'=\xi\lambda f$ or $(f\sin\theta)$. If f(x,y) is the same as g(x,y), then the

system performs the auto-correlation, which results in a very bright spot at the position $x'=f\sin\theta$. So we can use this method to detect the similarity of the input pattern f(x,y) and the stored pattern g(x,y) or for pattern recognition according to the brightness of the correlation spot.

As the Van der Lugt matched filter system uses correlation in the spatial domain, we call this system as "the spatial correlation system". By putting different filters in the filter plane (the back focal plane of lens FL1 or the front focal plane of lens FL2), the input pattern can be processed in the way you want, for example image addition or subtraction, deblurring etc., so such a system is also called a 4-f image processing system. Usually the Fourier transform hologram of an object occupies a very small area and the spectrum of the object consists of a very bright zero-order spot plus weak side-lobes. When the input pattern is rotated, its Fourier spectrum looks the same before, so it is difficult to align them. In the next section, we will describe "the spectrum correlation system", which occurs in the spectral domain.

3.5 Spectral Correlation System

The spectral correlation system is shown in Fig. 3.6. This is a spatial variant system. This time the hologram recorded in Fig. 3.3 is put in the front focal plane of the lens FL1 and f(x,y) is placed in the back focal plane of the lens FL1 or the front focal plane of the lens FL2. When a parallel beam reconstructs the hologram, suppose the hologram is processed as before, then the amplitude after the hologram can be written as

$$t(\mu,\nu) = t_0 + \beta T + \beta T\left\{|G(\mu,\nu)|^2 + G(\mu,\nu)\exp(-i2\pi\xi\lambda f\mu) + G^*(\mu,\nu)\exp(i2\pi\xi\lambda f\mu)\right\} \qquad (3.21)$$

The term which interests us is that containing the spectrum of the conjugate pattern, that is the last term on the right hand side of Eq. (3.21). We block all the other terms. At the back focal plane of lens FL1 where the pattern g*(x,y), which is the same as g(x,y) for a real pattern, is superimposed with pattern f(x,y). At the back focal plane of lens FL2, the correlation of the Fourier transforms of the two patterns is

$$F(\mu,\nu) * G(\mu,\nu) = F.T.\{f(x,y)g(x,y)\} \tag{3.22}$$

where * means correlation and f(x,y) and g(x,y) are let to be real patterns.

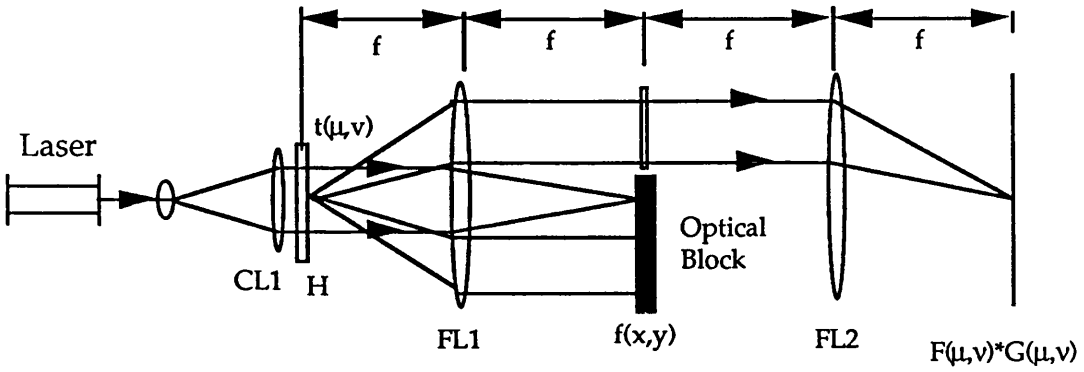


Fig. 3.6 Spectrum Correlation System

By the definition of the Fourier transform

$$F(\mu,\nu) * G(\mu,\nu) = \int\int f(x,y)g(x,y)\exp[-j2\pi(\mu x + \nu y)]dxdy \tag{3.23}$$

At the centre of the correlation plane, where $\mu=0$, $\nu=0$, we have

$$F(\mu,\nu) * G(\mu,\nu)\big|_{\mu=0,\nu=0} = \int\int f(x,y)g(x,y)dxdy \tag{3.24}$$

This equation is just Parseval's Theorem for real patterns [Sor71]. This is a correlation between the spectra of the two patterns, so we call it "the spectrum correlation". According to Eq.(3.3), the term in the right hand indicates the similarity of the two analogue patterns f(x,y) and g(x,y). So we can use the zero order of spectrum correlation to replace that of pattern correlation to indicate the similarity. The advantage of this replacement is that we can record the Fourier transform holograms with the same modulation frequency for all patterns to be stored during the learning procedure (see chapter 6). This eliminates the problem of diffraction efficiency variation due to the modulation frequency variation of the holograms. Another advantage is that we can implement the parallel correlation by using a spatially multiplexed hologram array, which

eliminates the cross-talk of the holograms if they are stored superimposedly.

3.6 Holographic Memories

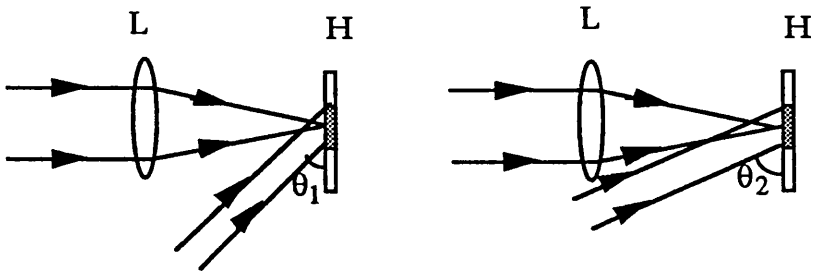
Holographic memories are an attractive candidate for large capacity neural-network-based computer storage systems, such as arithmetic databases. They offer the potentials of providing parallel associative search capability to the whole stored data base and perhaps can include data processing within the memory itself. A rapid retrieval of stored data is essential for neural network based systems. Generally the holographic memory capacity is limited by the holographic resolutions, while the resolutions are limited by the size of the hologram aperture, and the spatial frequency limit of the recording medium. Yu [Yu73] found that the minimum resolvable distance in holography is proportional to the wavelength of the coherent source used for the wavefront construction and to the distance between the object and the recording medium, and it is inversely proportional to the size of the hologram and the spatial frequency limit of the recording medium.

Usually, Fourier transform holograms are used to store information, since Fourier transform holograms have some advantages e.g., spatial invariance, spectrum processing, information compression. There are three kinds of holographic memories, specified respectively as:

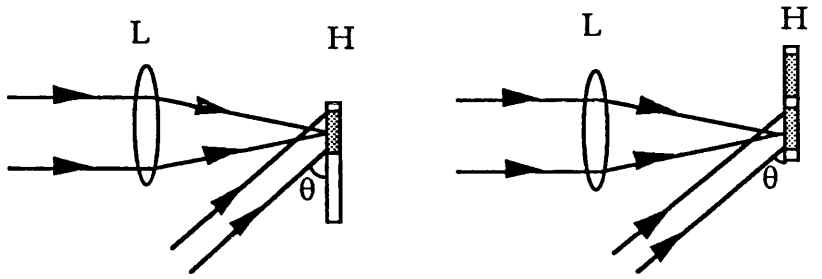
(1) Angularly multiplexed;
(2) Spatially multiplexed;
(3) Hybrid.

In the angularly multiplexed recording (Fig. 3.7(a)), the holograms are superimposed on the same area of the recording medium and to enable them to be distinguished, each hologram has a different modulation frequency so the angles between the reference beam and patterns to be stored must be set differently for each stored pattern. As all the holograms are superimposed together, there is cross-talk or interference between them. In the real time hologram case, such as in photorefractive materials, the new hologram storing possibly erases the previous ones, so all the patterns to be stored in the system have to be

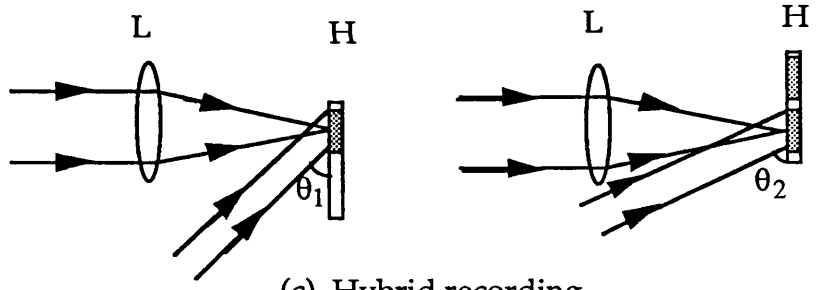
updated from time to time [Psa88b].



(a) Angularly multiplexed recording



(b) Spatially multiplexed recording



(c) Hybrid recording

Fig. 3.7 Holographic Memories

When such holograms are replayed by a parallel beam, there are differences between planar holograms and volume holograms (Fig. 3.8). In the planar hologram case, all the stored patterns will be reconstructed. If the modulation frequency and position and the wavelength of the reconstruction beam are the same as those of one of the stored holograms, then the pattern corresponding to that hologram will be replayed without distortion, but all the others will be distortedly reconstructed and shifted. For the volume hologram case (Fig.3.8b), however, the modulation frequency and the position of the reconstruction beams must be the same as or very close to those of one of the stored patterns. Otherwise, no

pattern will be replayed because of the selectivity of the Bragg diffraction angle [Col71], if the wavelength is kept unchanged for recording and reconstructing. If the wavelength is changed, then the modulation frequency and/or the position must be changed as well, in order to match the Bragg diffraction condition, but such changes usually result in aberrations in the reconstructed pattern [Col71].

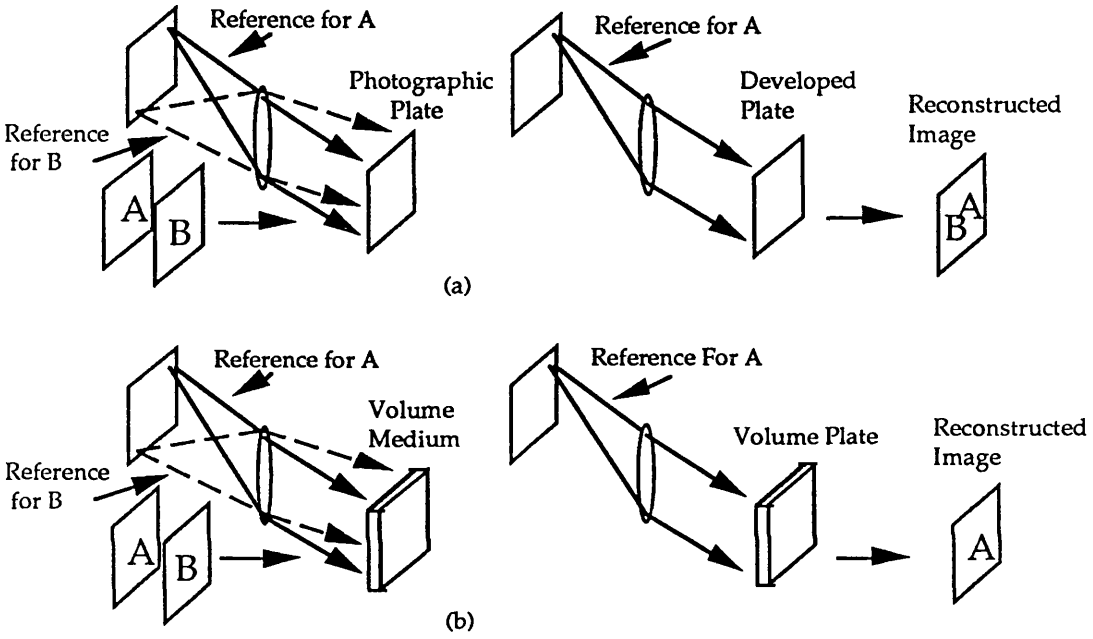


Fig. 3.8 Comparison of Planar and Volume Holograms
((a) Planar Hologram; (b) Volume Hologram)

In the spatially multiplexed recording system, the modulation frequencies for all the patterns are the same, while the positions on the recording plate are different. The advantage of such recording is that there is no cross-talk between the stored patterns and the number of stored patterns can be increased by using a larger recording material. Actually, the size of the recording plate is limited by the f-number of the lens, unless a lenslet array is used. In this recording system, there is no difference between the planar and volume holograms except for the diffraction efficiency and memory capacity. In the hybrid recording configuration, both frequency modulation and hologram position change are involved. Such a recording system is especially useful for volume holograms. In the volume medium, each small area can store a number of patterns by using angularly multiplexing and also we can store a series of such pattern

groups in other positions by using spatial multiplexing. Generally the larger the hologram, the better the resolution and the signal to noise ratio (SNR) [Sor71]. But actually the difference is very small unless the size of the hologram is comparable to the grating period. So we can divide the recording plate into a number of sub-sectors and in each sub-sector many holograms can be stored by using angular multiplexing, the number of stored patterns depending on the thickness of the recording medium and the wavelength.

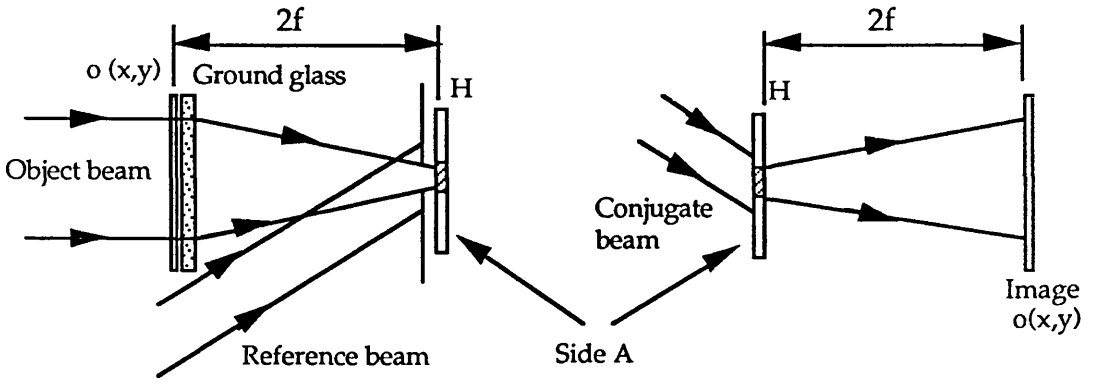


Fig. 3.9 Fresnel Holographic Memory

Although Fourier transform holograms are commonly used for pattern storage, Fresnel holograms can also be used for such purpose. A optical setup for pattern storage and its reconstruction system are shown in Fig. 3.9. The transparent pattern o(x,y) is illuminated by a parallel beam, a piece of ground glass is placed just after the transparency, so that the input pattern is encoded randomly. In the recording plate, which is put a distance 2f away behind the input pattern, the whole information is distributed over every physical point because of the ground glass scattering. So we use an aperture in front of the recording plate to confine the hologram in a small area. The smallest size is limited by the minimum resolvable structure of the input pattern, not by the f-number of the lens as in the FTHs' case. Another difference is that the system used here is an image system, rather than a Fourier transform system, so the amplitude distribution in the recording plane is the Fresnel of the input pattern. When we replay the hologram recorded in Fig. 3.9(a), we use the

conjugate beam of the reference beam used for hologram recording, so there will be a real pattern reconstructed in the plane 2f away from the hologram (Fig. 3.9(b). We have designed an optical HOFNET neural network system based on this hologram and the details will be covered in chapter 7.

Holograms store information associatively in the recording material, so they are very useful for the neural network memory. The memory capacity is an important factor for this kind of application. We will discuss this factor for the planar hologram and the volume hologram respectively. The memory capacity can be defined as the maximum bits that can be stored in a holographic material. If the hologram is thin, the theoretical memory capacity can be written as [Psa88c]

$$M_p \approx \frac{S}{\lambda^2} \tag{3.25}$$

where S is the area of the recording material and λ is the wavelength. If the hologram is thick, however, the memory capacity can be written as

$$M_v \approx \frac{V}{\lambda^3} \tag{3.26}$$

where V is the volume of the recording material. Equations (3.25) and (3.26) are used only for the comparison of thin holograms and thick holograms. For the real materials, the memory capacity will be reduced greatly by considering noise level, the dynamic range of the refractive index, cross-coupling etc. [Sol89]. Suppose V=1x1x1cm^3, λ is 0.5 μm, then M_v is on the order of 10^{12}. For the same memory capacity, if using a thin hologram, the area must be in the order of 10000 cm^2. If a pattern contains 100x100 bits, then 10^8 patterns can be stored. For information storage (including interconnections), volume holograms are more compact than planar holograms. Recently 500 holograms have been stored in a photorefractive material $LiNbO_3$ [Mok91]. Each hologram contains an image with 320 by 220 pixels, so the total memory capacity is

$$500 \times 320 \times 200 = 3.52 \times 10^7 \tag{3.27}$$

Considering some factors, which affect the memory capacity, such as noise,

refraction modulation depth, power [Sol90], this result is quite encouraging.

Chapter Four

REVIEW OF OPTICAL NEURAL NETWORK SYSTEMS

4.1 Introduction

Systems to simulate the operations of the human brain by emulating its anatomic structure are called artificial neural networks [See chapter two]. Like the brain, they are composed of a large number of simple processors that are extensively interconnected. Optical technology matches well to the architecture of a neural network system, because the technology's strengths lie in those areas that are indispensable to a neural network system for the interconnections of a large number of processing elements. The weaknesses of optics are in areas that are less critical for the functioning of a neural network system, such as the ability to perform intricate logic operations at the processor level. As optical processing elements communicate through light beams, they can be connected to one another without attaching a cumbersome wire between each pair of elements and they need not be confined to the restrictive planar configurations of silicon chips [Bel86]. Optical interconnections for implementation of neural networks are being considered as a means of relieving communication bottle necks encountered in VLSI chips [Goo84].

In neural network systems, information is associatively stored in interconnections between processors, rather than in the processors themselves [Lip87]. The large degree of parallelism and the three dimensional (3-D) interconnection capability, are required. Although electronics is very useful for a neural network system with a small number of neurons, it could not implement a large number of interconnections because of the interference between two intersecting connections. Optical systems are the best choice for the neural network implementations. Recently such optical implementations have been receiving considerable attention [Lin89; Hor90; Far89; Yu90; Jan88a;

Whi88a]. Many researchers have tried to modify the Hopfield model, so that it is suitable to be implemented practically (optically or opto-electronically) [Sha89; Gin88]. The optical requirements have been studied thoroughly [Yar86; Gho88; Oth89]. The most promising devices for establishing optical interconnections in optical neural systems are holograms [Psa88c]. Holograms are best known as a means of generating 3–D images, but more generally they represent an effective technique for recording and reconstructing the intensity of a light ray as well as the direction in which it was travelling. A planar hologram can direct any light beam on one side of it to any point on the other side, provided the total number of points and light beams do not exceed the number of resolvable spots on the film. A volume hologram made from a photorefractive crystal is even more prodigious in its capability to connect light emitters to light detectors, but the recording and encoding is also more complex. An optical neural network system is usually based on holographic techniques for interconnections and opto-electronic devices for logic processing, while their algorithm relies on the model [See chapter two]. They can have the ability of error correction and highly parallel information processing. The incorporation of non-linear feedback enhances dramatically error-correction capability and storage capacity of the holographic memory. In the neural networks, inhibitory interconnections (a bipolar interconnection matrix) are usually required. Although it is possible to implement negative interconnections optically by using phase encoding, it would make the system too complex to be practical. Several methods have been proposed and carried out. We describe them in the following sections.

As there are so many optical neural network systems that have been demonstrated in the world, it is difficult to cover all of them in one chapter. Here we mainly discuss those related to our neural network (HOFNET), that is those based on the Hebbian learning rule and roughly classify them into two kinds of nets: one is called the outer product model system (Hopfield model) and the other the inner product system (Matched filter model) [Sel89]. First, in section 4.2, we describe the first optical neural system based on the matrix-vector multiplication algorithm [Psa85]. In Sec. 4.3, linear optical neural network systems are presented [Jan88a; Pae87]. We will present them as the outer product model and the inner product model respectively. In Sec. 4.4 optical systems based on high order neural

networks which have a higher memory capacity are presented [Jan89; Owe86]. The layouts are similar to Sec. 4.3. A brief conclusion is given at the end of the chapter.

4.2 Opto-electronic Neural Network Using Vector-Matrix Multipliers

The first optical neural network system based on the Hebbian learning rule was implemented by using the vector-matrix multiplier [Goo78]. It is shown schematically in Fig. 4.1, in which the array of light-emitting diodes (LED's) represents N logic elements with binary states (LED on or off), which are to be interconnected in accordance with the Hopfield model (See section 2.2.1) [Psa85; Far85]. This is accomplished by the addition of nonlinear feedback (including feedback, thresholding, and gain) to the well-known optical vector-matrix multiplier [Psa87]. Gain is included to compensate for losses. The optical feedback has the thresholding (non-linearity) and the gain concentrated between the photo diode (PD) array and the LED array, which can be fabricated monolithically on GaAs or replaced by a optical transistor amplifier [Tan85].

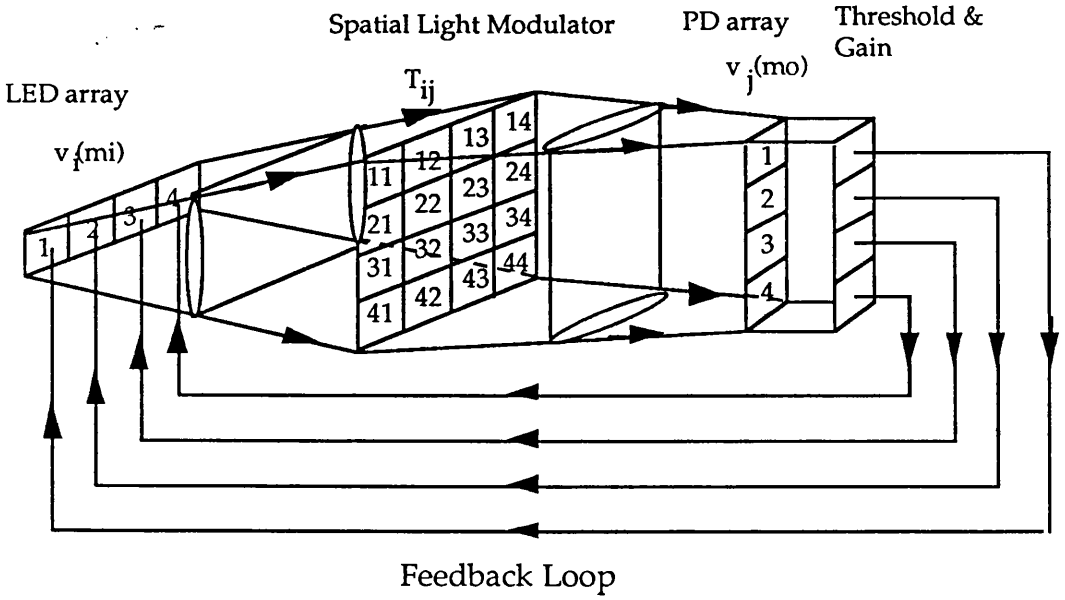


Fig. 4.1 Vector-Matrix Multiplier Based Optical Neural Network

Multiplication of the vector by the matrix in these schemes is

accomplished by horizontal and vertical cylindrical lenses. In opto-electronics, it is difficult to directly implement bipolar interconnections. There are two main methods to cope with this problem. One is to divide the interconnection matrix T_{ij} into two parts T_{ij}^+ and T_{ij}^-, so T_{ij} can be written as

$$T_{ij} = T_{ij}^+ - T_{ij}^- \tag{4.1}$$

where both T_{ij}^+ and T_{ij}^- are non-negative and $v_i(m)$ is one of the M stored patterns in the matrix. The other method is to use a DC bias and the details will be introduced in the next section. One way to implement the bipolar interconnection matrix is shown in Fig. 4.2. It is realized opto-electronically with incoherent light by assigning its negative and positive values to adjacent rows. Light passing through each row is focused onto adjacent pairs of photo diodes of the PD array that are electronically connected in opposition, as shown in Fig. 4.2 [Psa86c]. So the penalty is that the number of detectors is twice of the neurons (LEDs).

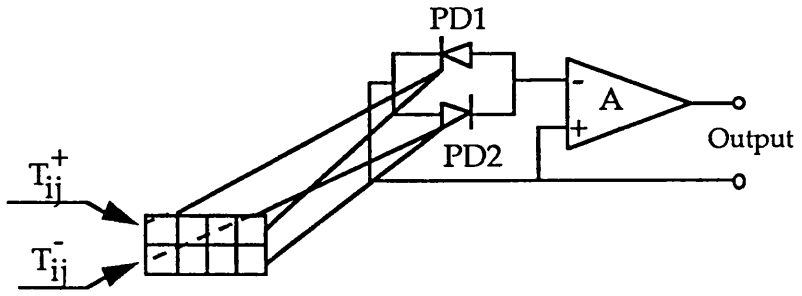


Fig. 4.2 Implementation of Bipolar Interconnections

Here the positive and negative elements of each row of the T_{ij} matrix are separated into two subrows, one for positive values and one for negative. The light transmitted through the two subrows is integrated horizontally with the aid of another set of anamorphic lenses and brought to focus on two adjacent photo diodes of the PD array connected in opposition. The output of the first diode pair circuit will be applied through an electronic thresholding circuit to the first element of the LED array, as shown in Fig. 4.1. Similar connections are made between other detector pairs of the photo detector array and corresponding elements in

the LED array. Thus if an input pattern $v_i(mi)$ is displayed in the LED array, the output can be written as

$$v_j(mo) = \sum_{i=1}^{N} T_{ij} v_i(mi)$$
$$= \sum_{i=1}^{N} (T_{ij}^{+} - T_{ij}^{-}) v_i(mi) \qquad (4.2)$$

where $v_i(mi)$ is the input pattern and the network fires according to whether $v_i(mo)$ exceeds the threshold or not.

Such a system is very useful for image processing and it can process information in parallel. In this system the input pattern is limited to one-dimension only, otherwise the interconnection matrix T_{ij} cannot be implemented in a plane. In the following sections, we will discuss some neural network systems, which can process two dimensional patterns.

4.3 Optical Linear Neural Network Systems

When we consider the possibility of optically storing 2-D patterns, each pattern containing NxN neurons, in a neural network system, we directly extend the Hopfield model to 2-Ds (See section 2.2.1), and the patterns are stored in a four-dimensional tensor. Since we have only two independent spatial variables to work with in an optical system, it is difficult to implement such multi-dimensional interconnections. One solution is direct: the N^4 interconnections are divided into N^2 2-D interconnections which can be processed by optics (outer-product model) [Yu90; Lu90; Yan90b; Lin89]. The other solution is to modify the Hopfield model so that it can be implemented by 2-D interconnections (inner product) [Psa85, 87; Pae87; Sof86; Owe86]. Here two kinds of implementation based on the outer product model and the inner product model are described respectively (see chapter two for the definition of outer product and inner product models).

4.3.1 Optical linear neural network based on the outer product model

Extending the 1-D N neuron Hopfield neural network into the 2-D NxN one, the algorithm equation becomes

$$v_{ij}(mo) = \sum_{k=1}^{N}\sum_{l=1}^{N} T_{lkij} v_{lk}(mi) \tag{4.3}$$

where T_{lkij} is

$$T_{lkij} = \sum_{m=1}^{M} v_{lk}(m) v_{ij}(m) \tag{4.4}$$

where v_{lk} (m) represents the state of the *lk*th neuron of the *m*th stored pattern, T_{lkij} is a 4-D interconnection weight matrix, $v_{lk}(mi)$ is an input pattern and lk and ij are the pixel positions in the patterns. In order to implement the 4-D interconnections, the matrix T_{lkij} is partitioned into an array of 2-D submatrices T_{11ij}, T_{12ij}, ..., T_{NNij}, in which each submatrix is an NxN size. Thus we can see that a 4-D associative memory matrix can be displayed as an N^2 2-D array representation.

To avoid using bipolar interconnections, one solution has already been mentioned in the above section. Another way is to add a constant dc component to the clipped binary and bipolar interconnection T_{lkij} to obtain a non-negative interconnection $\overline{T}_{lkij}$ and compensate it with an input-dependent thresholding operation as

$$v_{ij}(mo) = \sum_{l=1}^{N}\sum_{k=1}^{N} \overline{T}_{lkij} v_{lk}(mi) - \sum_{l=1}^{N}\sum_{k=1}^{N} v_{lk}(mi) \tag{4.5}$$

where

$$\overline{T}_{lkij} = T_{lkij} + 1 \tag{4.6}$$

Here for all l, k, i, j, T_{lkij} is assumed to be hard clipped, that means if it is more than zero, it takes +1; or else it takes -1. The other alternative, similar to this method, to avoid bipolar interconnections was described by White [Whi88a, b, c], in which the Hopfield model is rearranged to be in two parts to remove negative connections and does not require subtraction stages [Psa85] or a time varying threshold level [Jan88a, b], so it is ideally suitable for an optical implementation.

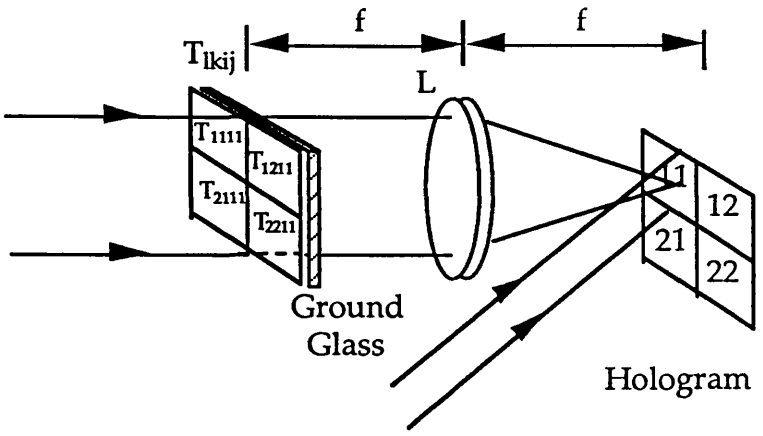


Fig. 4.3 One to N Holographic Interconnections

The architecture for implementing one to N interconnections holographically is shown in Fig. 4.3. For a fixed ij, say 11, the matrix T_{lk11} is arranged in a two dimensional array according to the subscript indices lk. A ground glass is put behind the input plate, which is inserted in the front focal plane of lens L, so that at the back focal plane in the position (11), all the information is collected. In Fig. 4.3 we use F.T. holograms, so we call it a focused system. Jang [Jan88a,b] implemented the same interconnection by using a unfocused system. . Fig. 4.4 is a focused system by using Fourier transform holograms. In each hologram element H_{ij}, all the interconnections T_{lkij} (l=1,...,N, k=1,...N) are recorded. So when it is reconstructed, 1 to N^2 interconnections are implemented.

In Fig. 4.4, each of the NxN multi-element hologram performs an interconnection of 1 to N^2 (in Fig. 4.4, N=2). Each pixel in an NxN input pattern can either illuminate (if the pixel is on) or not illuminate (if it is off) a corresponding hologram element located behind it. Each hologram element contains an NxN pattern; the full set of connection patterns is N^2xN^2. The light from each element of the NxN connection hologram is projected onto a single NxN detector array to carry out the summation. The thresholding operations enhance the power and the contrast of the output image. The zeroth-order diffracted beam is itself the binary pattern output and all the first-order diffracted beams represent the interconnections of the neuron elements. In the system both the zeroth-order diffracted beam and the first-order diffracted beam are utilized [Jan88a].

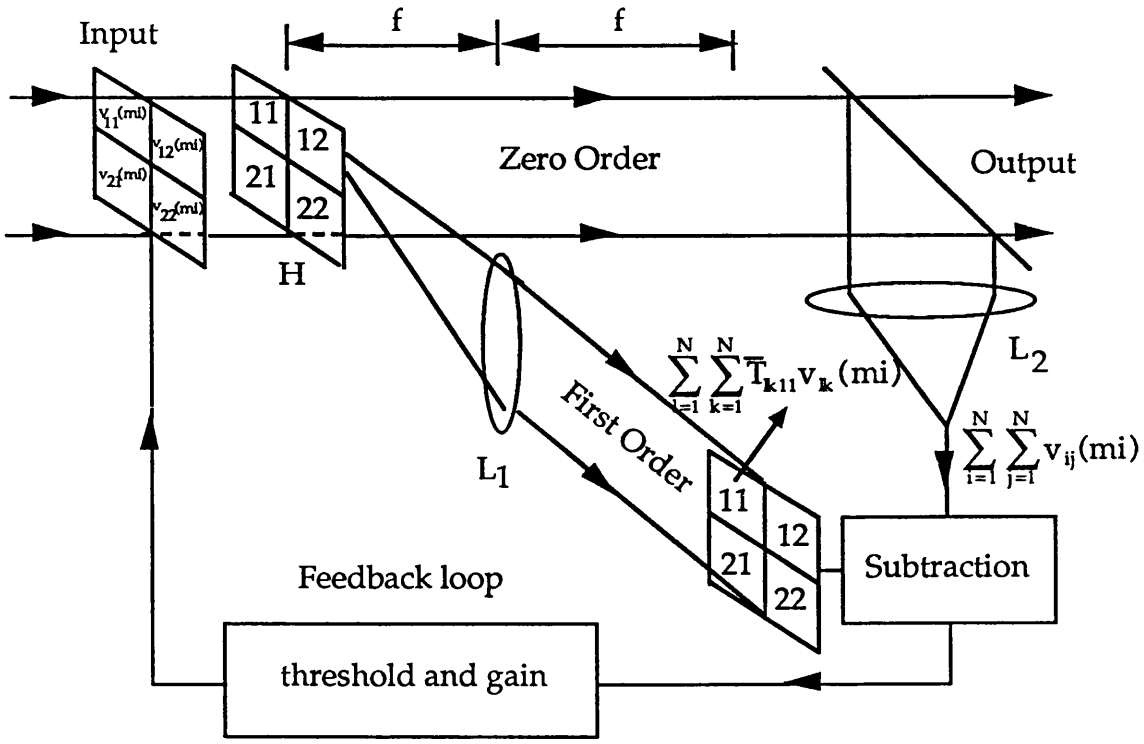


Fig. 4.4 2-D Optical Neural Network Based on Outer Product Model

Other examples of optical systems of the 2-D Hopfield model are constructed by using a high-resolution liquid-crystal television [Yan90; Lu90; Yu90]. In these systems a lenslet array is used to replace the hologram array to implement 1 to N^2 interconnections. Recently an associative memory system based on time encoding by using an echo correlator was proposed [Bel91].

4.3.2 Optical linear system based on the inner product model

The 2-D inner product model algorithm can be extended directly from the 1-D model described in chapter one. It is written as follows [Pae87]:

$$v_{ij}(mo) = \sum_{m=1}^{M}\left(\sum_{l=1}^{N}\sum_{k=1}^{N} v_{lk}(mi)v_{lk}(m)\right)v_{ij}(m) \qquad (4.7)$$

or its analog equivalent [Psa85]

$$f_{mo}(x,y)=\sum_{m=1}^{M}\left[\left(f_{mi}(x,y)*f_m(x,y)\right)\otimes f_m(x,y)\right] \tag{4.8}$$

where * means correlation and ⊗ means convolution; $f_{mi}(x,y)$ is an input pattern, $f_m(x,y)$ is one of the stored patterns (m=1,2,..., M) and $f_{mo}(x,y)$ is the output pattern. So the optical implementation of this system can be decomposed into three steps. Firstly, the inner product of the input pattern and each stored pattern is calculated. Actually it is the zeroth-order correlation which detects the similarity between the input pattern and the stored patterns [see chapter three]. This can be optically calculated by a 4-f image processing system by using matched filter technique. Secondly, each inner product or correlation should be multiplied or correlated by the stored patterns. Finally weighted summation is carried out over all the stored patterns with the weight equal to the zero-order correlation to give the final output.

The memory in this neural network can be formed by storing patterns in the form of a conventional Fourier transform hologram. All of the patterns to be stored are modulated by a different reference frequency (Fig. 3.6) which is equivalent to a different angle between the reference beam and each object beam. It is called angular multiplexed encoding. Such an arrangement gives the correlation outputs separately in the correlation plane so that individual processing for each channel can be made [Owe87]. For simplicity, we assume that the reference frequencies change by a constant increment between subsequent recordings. When we record the interference between all the patterns to be stored and the reference beam on a holographic plate at the Fourier transform plane, after processing, the amplitude transmittance of the hologram becomes

$$\begin{aligned} t_1(\mu,\nu) &= \left|\sum_{m=1}^{M} F_m(\mu,\nu)+\exp[-j2\pi\lambda f(\xi_m\mu+\zeta_m\nu)]\right|^2 \\ &= \sum_{m=1}^{M} F_m^*(\mu,\nu)\exp[-j2\pi\lambda f(\xi_m\mu+\zeta_m\nu)] + \text{other terms} \end{aligned} \tag{4.9}$$

where $F_m(\mu,\nu)$ is the Fourier transform of $f_m(x, y)$ and $\xi_m = \xi_0 + m\,\delta\xi$ and $\zeta_m = \zeta_0 + m\,\delta\zeta$ are the modulation frequencies which depend on the angle of incidence of the reference beam. $\delta\xi$ and $\delta\zeta$ are the frequency differences

between two adjacent recordings.

The whole optical neural network system is shown in Fig. 4.5. The input pattern is placed at plane P_1 and is Fourier transformed by lens L_1. Its Fourier transform illuminates the holographic memory placed at the Fourier plane P_2. The correlation of the input pattern and each stored pattern is formed at plane P_3 by lens L_2. The inner product values are sampled by an array of pinholes at plane P_3. Each pinhole is positioned exactly where each of the stored patterns was centred when the Fourier transform hologram was recorded. Therefore, if the input is one of the stored patterns centred on the optical axis, then a sharp auto-correlation peak will be formed at plane P_3 on one of the pinholes. Light emerging from each pinhole is retroflected by a phase conjugate mirror (PCM) [Ath86] or even a mirror placed immediately after the pinholes [Pae87], and the reflected light illuminates the hologram to form the reconstructed patterns of all of the memories at the output plane P_4. The reconstruction due to light from each pinhole is the entire composite memory shifted by the position of the pinhole. At the origin in the plane P_4 we obtain the superimposition of all of the memories. The strength with which each memory is represented in this superimposition is proportional to the value of the inner product between the input pattern and the corresponding stored pattern. A window is placed at P_4 to select only the desired central portion of the reconstructed patterns.

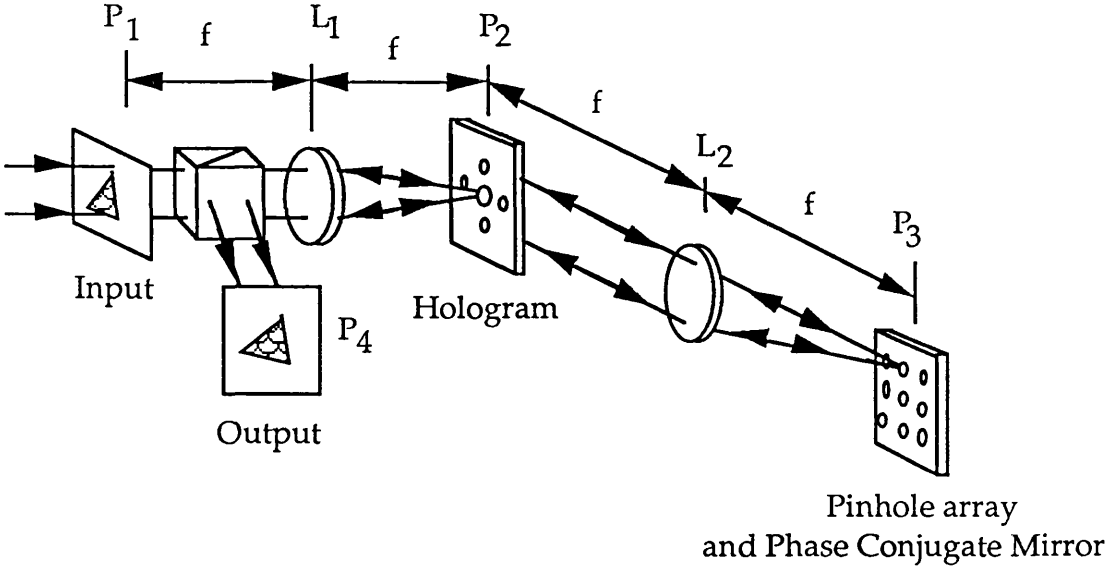


Fig. 4.5 Optical Linear Neural Network Based on Inner Product Model

In the output plane P_4, there are totally M^2 patterns reconstructed, if the patterns are stored in planar holograms. Each retroflected beam from each pinhole will reconstruct all the M patterns, but they are shifted from each other because of the thin hologram. If the modulation frequency difference between the adjacent holograms is big enough [see chapter three], then for each reconstruction beam, only one pattern is located in the centre and all the other patterns are located outside the centre pattern. All the other M(M-1) patterns are discarded and most of the light energy is wasted, so the reconstructed patterns are very weak. If, however, the patterns are stored in a volume hologram, then because of the Bragg diffraction angular selectivity, only M patterns are reconstructed and superimposed at the central area in the plane P_4. So in this system, volume holograms are preferred.

We now briefly describe the operation of this system analytically. Insert the input pattern $f_{mi}(x,y)$ into the plane P_1 in Fig. 4.5. Then, the term of interest in the amplitude of the light diffracted by the hologram is

$$t_2(\mu,\nu)=\sum_{m=1}^{M}F_{mi}(\mu,\nu)F_m^*(\mu,\nu)\exp[-2\pi\lambda f(\xi_m\mu+\zeta_m\nu)] \tag{4.10}$$

where $F_{mi}(\mu,\nu)$ is the Fourier transform of the input pattern $f_{mi}(x,y)$, t_2 is the amplitude distribution of one diffracted beam and (μ,ν) are the coordinates in the spectral plane. At the correlation output plane P_3, the light amplitude is the Fourier transform of Eq. (4.10)

$$\begin{aligned} t_3(x',y') &= \sum_{m=1}^{M}\left(f_{mi}(x',y')*f_m(x',y')\right)\otimes\delta(x'-\lambda f\xi_m, y'-\lambda f\zeta_m) \\ &= \sum_{m=1}^{M} g_m(x',y')\otimes\delta(x'-\lambda f\xi_m, y'-\lambda f\zeta_m) \end{aligned} \tag{4.11}$$

where $g_m(x', y')$ is the correlation of $f_m(x',y')$ with $f_{mi}(x',y')$ and (x', y') are the coordinates in plane P_3. The correlation output, which is sampled by the pinhole array and reflected by the PCM at plane P_3 will reconstruct the hologram again. The amplitude distribution $t_4(\mu, \nu)$ of the light travelling from right to left in Fig. 4.5, immediately to the left of P_2, is given by

$$t_4(\mu,\nu)=\sum_{m=1}^{M}\sum_{m'=1}^{M} g_m(0,0)F^*_{m'}(\mu,\nu)\exp\{-j2\pi\lambda f[\mu(\xi_m-\xi_{m'})+\nu(\zeta_m-\zeta_{m'})]\} \tag{4.12}$$

The light at the output plane P_4 is the Fourier transform of Eq. (4.12). The total light amplitude at P_4 can be written as

$$t_5(x,y)=\sum_{m=1}^{M}\sum_{m'=1}^{M} g_m(0,0)f_{m'}[x+\lambda f(\xi_m-\xi_{m'}),y+\lambda f(\zeta_m-\zeta_{m'})] \tag{4.13}$$

Note that in the above equation, unless m=m', the reconstructed patterns will emerge on off-axis positions. If the hologram is thick, then according to the Bragg diffraction angular selectivity, only other terms which satisfy m=m' come out. So if we put a window, whose size is equal to that of a stored pattern, in the centre of the plane P_4, then we have

$$t_6(x,y)=\sum_{m=1}^{M} g_m(0,0)f^*_m(x,y) \tag{4.14}$$

where t_6 is the amplitude distribution in the central window. If $f_m(x,y)$ is real,

$$f_m(x,y)=f^*_m(x,y) \tag{4.15}$$

So we have

$$t_6(x,y)=\sum_{m=1}^{M} g_m(0,0)f_m(x,y) \tag{4.16}$$

Compare the result in Eq.(4.16) with Eq. (4.7) or (4.8), we see that the optical system we described performs exactly the same as the inner product neural network system.

If the input pattern is more similar to a stored pattern than others, then that stored pattern will be displayed at the output plane, but at the same time all the other memories are also weakly read out. This crosstalk can be eliminated if the stored images are binary, in which case the output

can be thresholded and fed back to the input for multiple iterations.

An advantage of the holographic memory in neural networks is that a limited amount of shift invariance is possible if Fourier transform holograms are used [Owe86; Psa86; Ath86]. If thin holograms are used, the amount of permissible input pattern shift without ambiguity in pattern identification is limited by the angular separation between reference beams. As it is shown in Fig. 4.5, the output consists of all of the stored patterns. Assuming the input object is centred on-axis, the desired output will also be centred on-axis. The others, which are undesired, are positioned off-axis. The separation between these patterns is determined by the angular separation between the reference beams used to record the patterns. The correlation and convolution operations of the memory, as the input pattern is shifted the output pattern will also shift and remain coincident with the input pattern. Psaltis et al [Psa86] have analysed this problem and found that full shift invariance over the entire field of view without ambiguity is possible only if a single pattern is stored. As the number of stored patterns increases, the permissible shift decreases. If thick holograms are employed to record the information, an additional limit is involved in the shift invariance: the Bragg selectivity effect. Shifts in the input pattern result in shifts in the angular spectrum of plane waves incident on the hologram. If the shifts are too large, the Bragg condition will no longer be met and the pattern will not be reconstructed.

4.4 Optical High Order Neural Network Systems

Linear neural networks based on the Hebbian learning rule have low storage capacity [Psa88a; Lee86]. To overcome this limit, non-linear or high order neural network models have been introduced. One model called a nonlinear discriminant model or high order outer product model was proposed [Psa86; Lee86; Gil86] and some optical implementations of such systems are presented. Also based on the inner product model, Soffer et al [Sof86; Owe86] proposed a high order inner product model which has a nonlinearity in the correlation domain. Firstly, we will describe an optical quadratic neural network system based on a high order outer product model. This is the highest order to have been implemented before the work in this thesis. Then an optical system to implement high order neural networks based on the inner product model is presented.

4.4.1 Optical neural network based on high order outer product model

Neural network systems with order higher than two are very difficult to implement [Lam89; Liu86], unless the HOFNET model is used. We describe a quadratic neural network system based on holographic technique with 1-D patterns stored in the system as an example of high order neural network implementation. In the quadratic neural network with N neurons, N^3 interconnections are required. It can be realized with an optical vector-matrix multiplier and hologram array or microlens array for the interconnections [Jan89; Lin89, 90; Hor89].

The hologram array recording is shown in Fig. 4.6 and an optical high order neural network system is shown in Fig. 4.7. As in optics, only two independent spatial variables are available, it is difficult to implement N^3 interconnections between the neurons. The algorithm of the second order neural network is as follows

$$v_i(mo) = \sum_{j=1}^{N}\sum_{k=1}^{N} W_{ijk} v_j(mi) v_k(mi) \tag{4.18}$$

where W_{ijk} is the tensor where patterns are stored

$$W_{ijk} = \sum_{m=1}^{M} v_i(m) v_j(m) v_k(m) \tag{4.19}$$

Slight modifications of the above equation are made for optical implementation. First, clip the tensor as before, that means if the element of the tensor is more than zero, it takes 1 or else it takes -1. This modification allows the use of a binary spatial light modulator. Split the above equation into two equations

$$\begin{aligned} T_{ij} &= \sum_{k}^{N} \mathrm{sgn}(W_{ijk}) v_k(mi) \\ v_i(mo) &= \sum_{j}^{N} \mathrm{sgn}(T_{ij}) v_j(mi) \end{aligned} \tag{4.20}$$

where the sgn function is the same as the hard clipped function as shown in Fig 2.1. As it is done in the previous section, in order to avoid processing bipolar data, we add a constant dc to the $sgn(W_{ijk})$ and $sgn(T_{ij})$ respectively to get their corresponding interconnection tensor $\overline{W}_{ijk}$ and matrix $\overline{T}_{ij}$, where $\overline{W}_{ijk} = sgn(W_{ijk}) + 1$ and $\overline{T}_{ij} = sgn(T_{ij}) + 1$. The details can be seen in section 4.3.1. Now these non-negative interconnections can be easily implemented by using a 3-D interconnection method [Jan88a]. The matrix T_{ij} is the weighted summation and each term $\overline{W}_{ijk} v_k(mi)$ may be calculated by representing $v_k(mi)$ by an element of a binary unipolar 1–D SLM and performing interconnections with a subhologram recorded in Fig. 4.6.

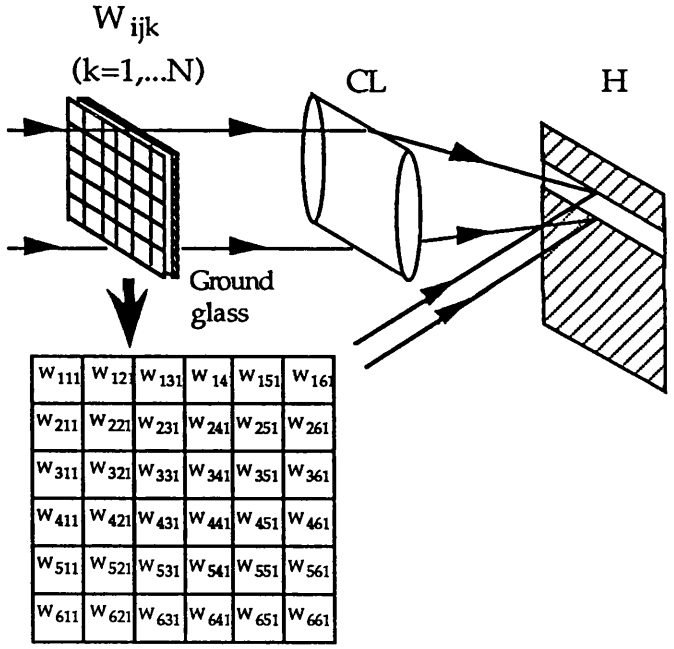


Fig. 4.6 Holographic Recording System for NxNxN interconnections

To obtain $\overline{T}_{ij}$, incoherent summation of such N projected patterns should be accomplished in the image plane of N subholograms as shown in Fig. 4.7. Each 2-D non-negative interconnections $\overline{W}_{ijk}$ that can be implemented by a 2-D SLM (the ON state means 2 and the OFF state means 0) is recorded in the corresponding hologram element. The ground glass in Fig. 4.6 is used not only to overcome the dynamic range problem of hologram plates but to randomise the phases so that incoherent addition can be performed in the image plane. The thresholding level is

the total number of 1's in an input pattern (see section 4.3.1). A 2-D detector array with at least N^2+1 elements detects the image of $\overline{T}_{ij}$ and the threshold level; the thresholding operation is executed electronically to obtain $\overline{T}_{ij}$. The non-negative matrix is displayed in a programmable 2-D SLM and used in the next step.

Step 2 is simply a vector-matrix multiplication. The total system is shown in Fig. 4.7, where the vector-matrix multiplication is accomplished. A parallel beam, which is modulated by an input pattern displayed on a 1-D spatial light modulator is used to reconstruct the hologram recorded in Fig. 4.6. The first order diffraction implements the summation of $\Sigma\overline{W}_{ijk}v_k(mi)$ which, after thresholding, is displayed on a 2-D SLM to represent $\overline{T}_{ij}$. The undiffracted light (zero-order) passing through the hologram is used to illuminate the 2-D SLM to implement the summation of $\Sigma\overline{T}_{ij}v_j(mi)$. This light beam passes a cylinder lens and is collected by a 1-D detector array. This final detected pattern is thresholded and fed back to the input 1-D SLM to start a new iteration.

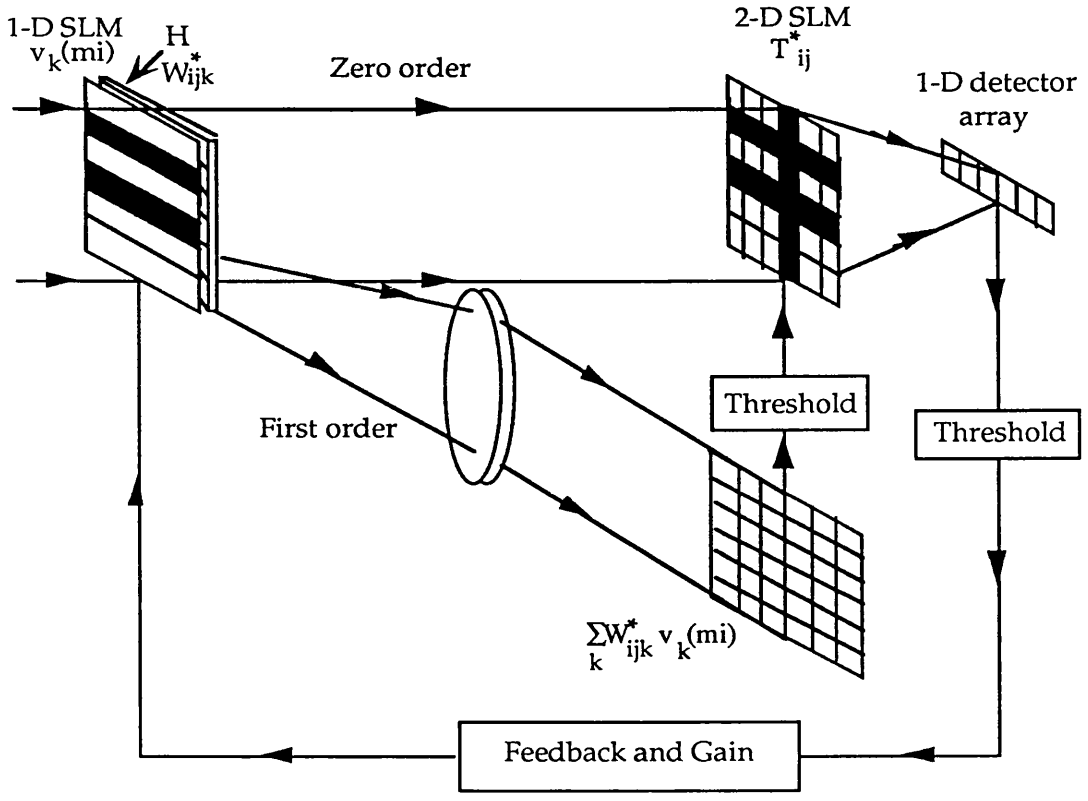


Fig.4.7 Optical neural network based on high order inner product model

4.4.2 Optical neural network based on high order inner product model

The configuration of an optical associative neural system based on the high order inner product model is shown in Fig. 4.8. It is implemented by putting non-linear devices, say phase conjugate mirrors (PCMs) in the correlation domain of a conventional neural network system [Owe87; Ath86], e.g., the one described in Sec. 4.3.2. The memory is recorded utilising angular multiplexing of training patterns, as shown in Fig. 2.1. The optical feedback and nonlinearity in the correlation domain are used to improve system performance.

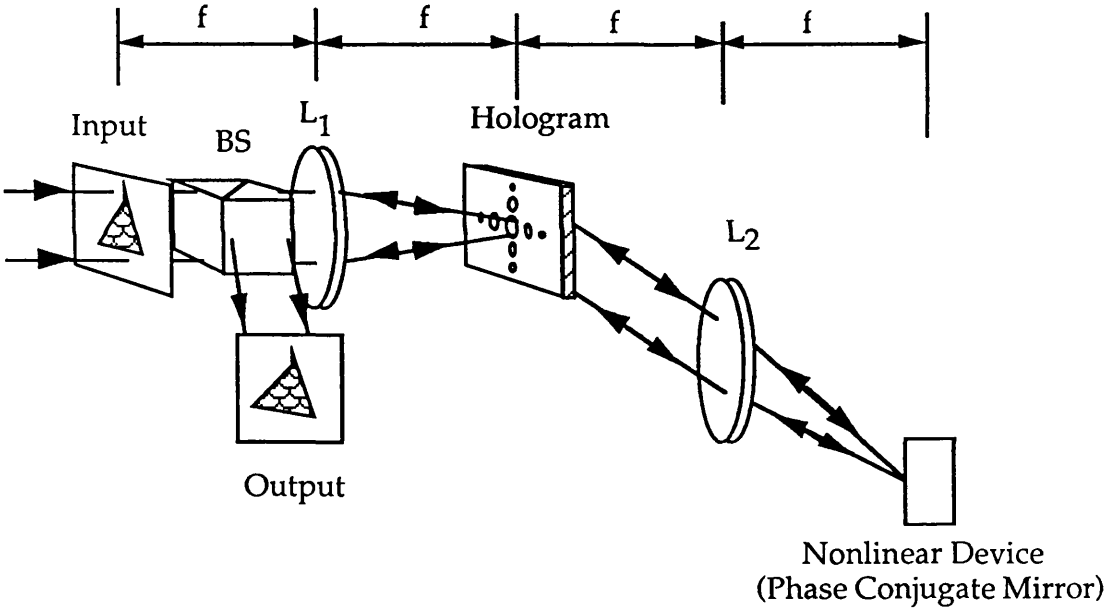


Fig. 4.8 Optical Neural Network Based on High Order Inner Product Model

The non-linear device, e.g., PCM is positioned in the optical loop of the system (Fig. 4.8). When a partial or distorted object addresses the hologram, a set of partially reconstructed reference beams is produced. Each reconstructed beam is weighted by the correlation of the input object with the stored object associated with that particular reference beam. The reference beams are detected by the nonlinear devices and the outputs are retroflected back toward the hologram, which reconstructs the complete stored objects. The replayed objects are phase conjugated by the PCM and

the process is iterated until a self-consistent solution of the system is found. These self-consistent solutions or eigenfuntions of the system are simply the stored objects in the hologram. Non-linearities in the back focal plane of lens L_2 tend to form regions of attraction in state space around the stored objects and increase the memory capacity greatly, as discussed in section 2.3.1.

Generally speaking, a high order inner product neural network model is more easily implemented optically or opto-electronically than an outer product one with the same order. In the inner product model system, the large number of interconnections in the outer product model system corresponds to the requirements for large dynamic ranges of devices because of the use of non-linear devices. Otherwise, the correlation peaks will be saturated and selection of the maximum correlation peaks becomes impossible. Therefore, although the non-linearity in the back focal plane of lens L_2 could enlarge the difference of the correlation peaks, it also makes the second largest one saturate to the same value as the largest one, both equal to the possible output of the devices, that is the saturated output (Fig. 4.9). In order to solve this problem, we designed a novel inner product system, in which the non-linearity is increased by one after every iteration, and during each iteration, the correlation peaks are normalised. So the requirement for the large dynamic ranges for devices is alleviated. The details are presented in chapters six and seven.

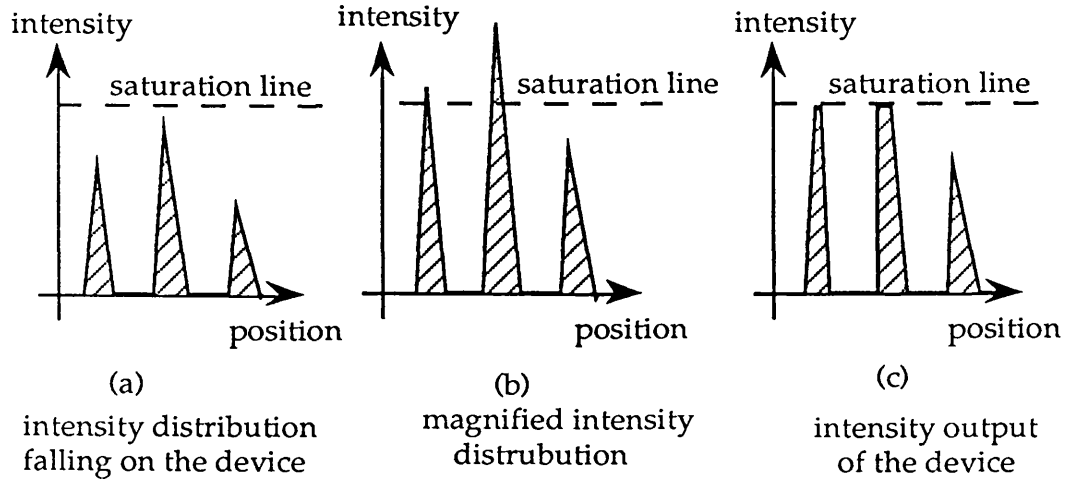


Fig. 4.9 Saturation Problem

4.5 Conclusion

We have briefly discussed the optical systems of neural networks based on the Hebbian learning rule. Because of the limit of the thesis length, it is very difficult for us to cover all of the optical neural network systems here. Another class of neural networks, which have a learning capability, have been causing more attention recently [Yos89a, b; Psa90a; Lu90;]. Applications of neural networks include pattern recognition [Son88; Psa90a], word break [Pae89], image processing [Yeh89] etc.

Chapter Five

HIGH ORDER FEEDBACK NEURAL NETWORK

(HOFNET)

5.1 Introduction

We showed earlier (see chapter two), that high order neural networks have a larger memory capacity than linear neural networks having the same number of neurons [Lee86; Psa88a]. The increase in storage capacity of high order neural nets is a result of the increased number of independent parameters or degrees of freedom needed to describe a higher order associative mapping [Psa88a]. Although a two-dimensional array of neurons can interconnect with another two-dimensional array of neurons by using volume holograms (three-dimensional interconnects), the independent interconnections are limited to N^3, where N is the number of pixels that are available in one dimension, rather than N^4 [Mid88; Psa88c]. That means we could not connect directly N^4 independent interconnects; only N to N^2 or $N^{3/2}$ to $N^{3/2}$ connections can be performed. Special selection of N^3 independent interconnects from the whole N^4 has been discussed [Mid88; Psa88c]. Therefore, for a Hebbian-learning-rule based neural network, the maximum order of the nets with one-dimensional N neurons is two, which requires N^3 interconnections, if we use one hologram to perform the interconnection. For a neural network with the order higher than two, say, three, with one-dimensional neurons (N), the full interconnections are N^4. It is very difficult to implement either optically or electronically such complicated interconnects, when the number of neurons N is very large. One solution to this problem is to divide N^4 interconnections into N times of N^3 or N^2 times of N^2 interconnections, which results in a complex system [Yu90; Lu90; Jan89]. The other solution is to modify the high order neural network so that the

high order could be implemented by using nonlinear devices [Owe87; Ath86]. Such nets are called inner-product high order neural nets (see section 2.4.2). In the inner product high order neural nets, the high order is performed by a nonlinear response of a device in the correlation domain, so the dynamic range of the device used in the system should be large enough to accommodate the nonlinear output. Because of the limited dynamic range of optical devices, typically less than 30dB [Sel90], it is difficult to implement a net optically with an order higher than two. So up to now, only second order neural network systems have been demonstrated optically [Liu89; Jan89; Hor90].

Another problem about the high order neural net is noise tolerance. In the inner product model, the system indicates the similarity according to the size of the correlation peaks of the input pattern and stored patterns. As the input pattern is presented only once in such a system, the noise in the input pattern will be frozen. When the input pattern is similar to two stored patterns, the difference between the correlations is very small. Quite often, the correlation between the noised input pattern and the most similar stored pattern is less than that between the noised input pattern and the second most similar stored pattern, which will result in a wrong answer or converging into a local minimum. Based on the Hopfield model and using the advantage of the Hamming net that it always converges to the correct answer (no local minimum) [Lip87], we propose a novel High Order Feedback neural NET, which we call abbreviatedly HOFNET. The HOFNET is very similar to the matched filter net [Sel89 or see section 2.4] and consists of a bank of correlators. The nonlinearity is used to enlarge differences of the correlation outputs by using normalisation in each iteration. The nonlinearity is implemented by a feedback loop, instead of nonlinear devices, while the feedback loop is necessary for a Hebbian-learning-rule based net to have an error correction capability. So we do not require nonlinear devices to implement the high order HOFNET. In section 5.2, we discuss the algorithm of the HOFNET and show how the nonlinearity is implemented by a feedback loop. Noise analysis of the HOFNET is discussed in section 5.3. Computer simulations for noise and pattern recognition are described in section 5.4. Conclusions are given at the end.

5.2 Principle of the HOFNET

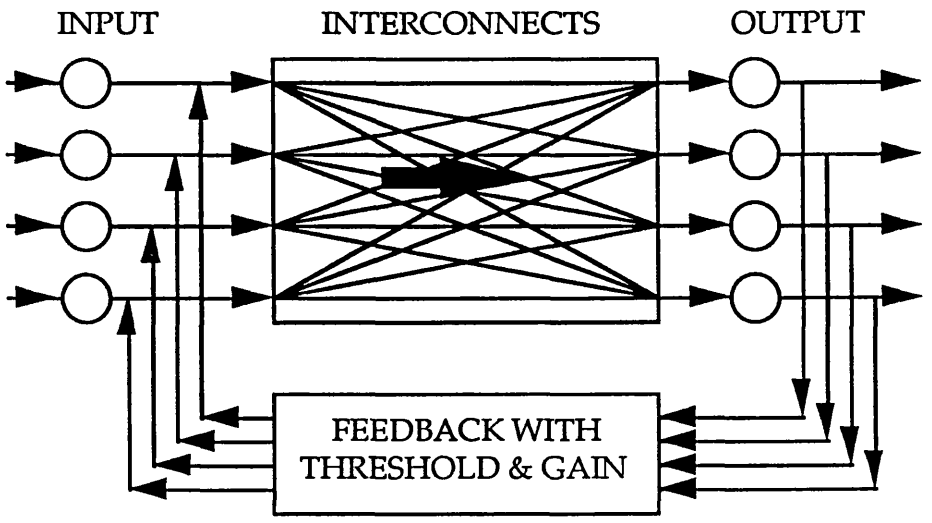


Fig. 5.1 Conventional Neural Network

In the conventional neural network (Fig. 5.1), when an input pattern is inserted into the system, it is interconnected to all the stored patterns. After processing (summation, threshold and gain), the network gives an output, which is fed back as a new input pattern to start a new iteration. So the input pattern is used only at the beginning of the iterations. If the input pattern contains time-varying noise, the noise will be "frozen", which might result in a wrong answer or spurious convergence. In our HOFNET system, however, the input pattern is correlated in every iteration, so the time-varying noise in the input pattern will be averaged. Before we discuss the HOFNET in detail, we give an example to see how the noise affects the result. Suppose we store two patterns A and B in a neural network system, an input pattern C is a little more similar to the pattern A than to the pattern B and their normalized noise-free correlations are 0.8 (A & C) and 0.75 (B & C) respectively (1.0 means completely correlated). If the noise level is 5% of the correlation outputs, then the correlation between the patterns A and C is ($0.80 \pm 5\% \times 0.80$), that is from 0.84 to 0.76 and the correlation between patterns B and C is ($0.75 \pm 5\% \times 0.75$), that is from 0.7875 to 0.7125. There is an overlap between these two correlation variations, so possibly at one time the correlation

between patterns B and C is larger than that between patterns A and C (Fig.5.2). If the input pattern is sampled at this time, a wrong answer will be given.

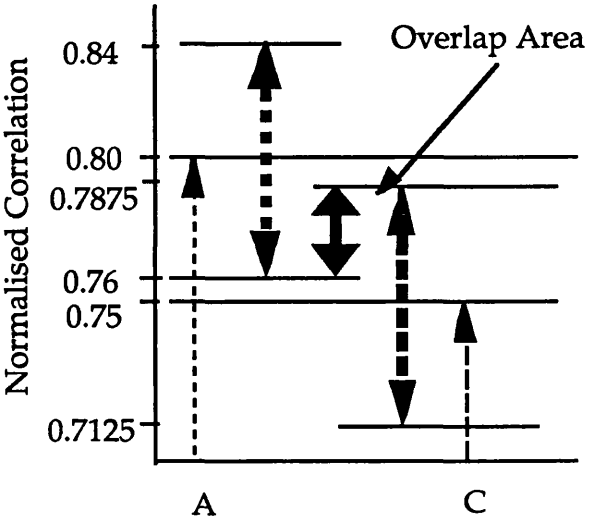


Fig. 5.2 The Effect on the Correlation Amplitudes of Noise

The design of HOFNET is to try to average the noise and enlarge the differences of the correlations between the input pattern and stored patterns by giving the power to the correlation outputs and normalising them. It has some similarity with the matched filter neural network model [Sel89]. The diagram of the matched filter net is shown in Fig. 5.3. It consists of M parallel correlator channels, where M is the number of known exemplar patterns stored in the learning procedure. Each channel acts as an individual correlator which correlates the input pattern, $v_i(mi)$, with one of the stored patterns, **V**(m) (m=1, ..., M) (which contains N elements $v_i(m)$, i=1, 2, ... , N). The correlation output at each channel reconstructs the corresponding stored pattern in a weighted way. All the reconstructed patterns are summed and thresholded to give an output, which, generally speaking, is more similar to a stored pattern than the input pattern. The output pattern is fed back to start a new iteration.

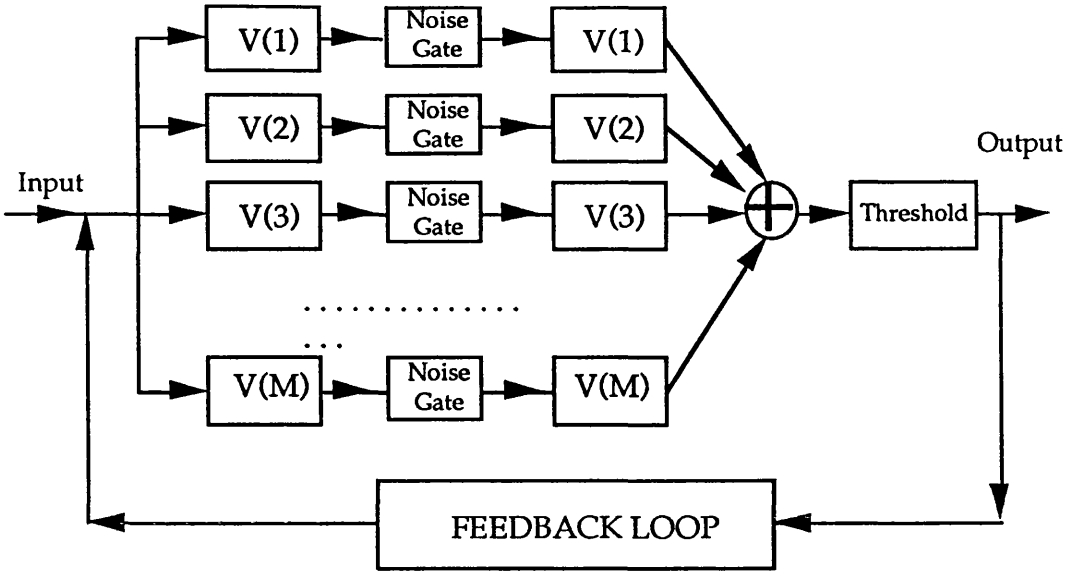


Fig. 5.3 Matched Filter Model of Hopfield Net

The High Order Feedback neural NETwork, or HOFNET is very similar to the matched filter model except for the use of nonlinearity to the correlation outputs. The diagram of the HOFNET in parallel to Fig. 5.3 is shown in Fig. 5.4. It also consists of M parallel channels. Each channel acts as an individual correlator which correlates the input pattern, $v_i(mi)$, with one of the stored patterns, V(m) (m=1, ..., M). Here in Fig. 5.4, the column (a) contains all the stored patterns V(m), C(m) is the correlation output between the input pattern and stored patterns. The column (b) is a nonlinear function to give the *n*th power to the correlation C(m). The column (c) has two functions: (1) normalisation; (2) threshold. It normalises the correlation output $(C(m))^n$ from the column (b) and sets the threshold as: if the normalised correlation K(m) is larger than N/2, keep its original value, or else it is set to be zero. The normalized outputs reconstruct the stored patterns in a weighted way to give an output. The weight K(m), apparently, depends on the correlation value C(m) and the nonlinear degree n. In the HOFNET system, the nonlinear function is implemented by the feedback loop and the n is not fixed but increases by one after every iteration. Take the previous example,

$$C(1) = 0.80\,; \quad C(2) = 0.75 \tag{5.1}$$

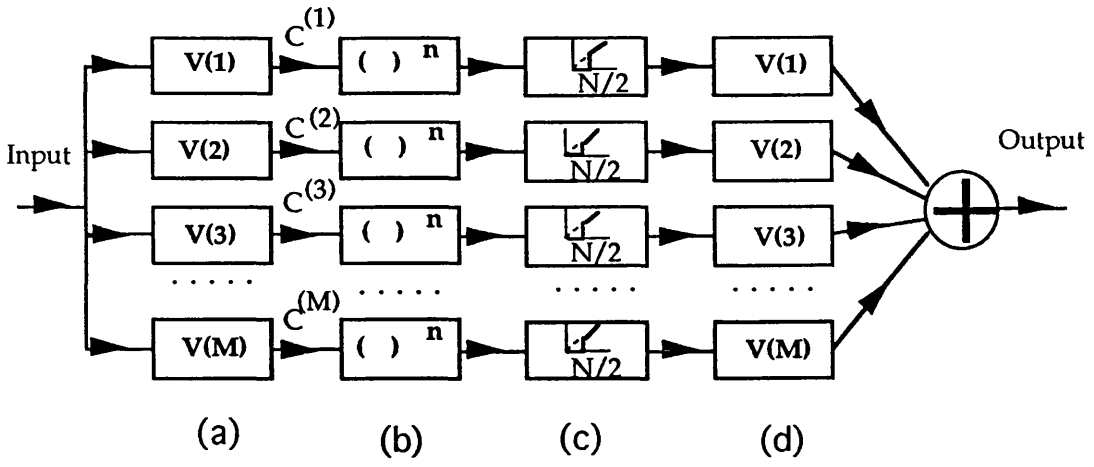


Fig. 5.4 Matched Filter Model of HOFNET

The normalised correlation from the column (b) is

$$K(i) = N \times (C(i))^n / (C_{max})^n \quad (i = 1, 2) \qquad (5.2)$$

where C_{max} is 0.8 and N is 2. Table 1 lists the correlation from the column (b) $(C(i))^n$ and K(i) for different n.

Table 1 Correlation Change with the Nonlinearity n

n	1	2	3	4	5	6	8	10	11
$(C(1))^n$	0.8	0.84	0.51	0.41	0.33	0.26	0.17	0.11	0.086
$(C(2))^n$	0.75	0.56	0.42	0.32	0.24	0.18	0.10	0.056	0.042
K(1)	2	2	2	2	2	2	2	2	2
K(2)	1.98	1.75	1.65	1.54	1.45	1.36	1.19	1.05	0

From the above table, we see that when the nonlinearity gets higher (n is larger), the normalised correlation difference becomes larger as well. When n is 11, the smaller one is less then half of N, so is thresholded. The minimum nonlinearity n required to select the largest correlation and threshold all the others depends on the closeness of the correlation

outputs and number of patterns stored in the system. The closer the correlation outputs are, or the greater the number of the stored patterns is, the larger n is required. At the same time the dynamic range of the nonlinear device must be very large. The HOFNET, however, implements the nonlinear function by the feedback loop, so this requirement does not exist in the HOFNET system.

One channel of the HOFNET is shown in Fig. 5.5. Each channel performs the same processing functions starting with the multiplication of an input constant signal level, A, by the Fourier transform of one of the known stored patterns, $FT(v_i(m)) = V(m)$. The output $AV(m)$ is inverse Fourier transformed to give $Av_i(m)$, which is then multiplied by the complex conjugate of the input pattern, $v_i(mi)$, which for real input patterns is just $v_i(mi)$ itself. This is followed by a further Fourier transform which, by the convolution theorem, yields the required correlation (denoted by *). Finally we filter out the side-lobe of the correlation and let the zero-order pass through an amplifier of gain g to give $Ag(V(m)*V(mi))$, which is fed back to the start of the loop to continue similar iterations. The output, taken before the multiplication by the input pattern, is given on repeated iterations by

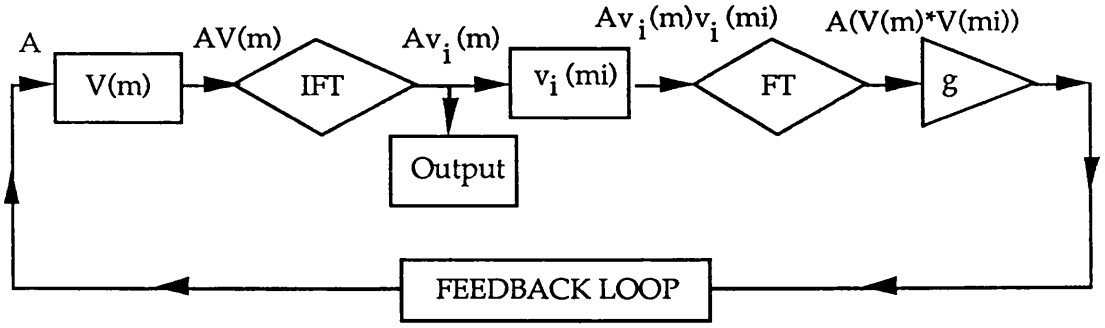


Fig. 5.5 Single Channel of HOFNET

Before iteration: $Av_i(m)$,

After 1st iteration: $A[g(V(m)*V(mi))]v_i(m)$,

After 2nd iteration: $A[g(V(m)*V(mi))]^2v_i(m)$,

... ...,

After nth iteration: $A[g(V(m)*V(mi))]^nv_i(m)$.

So at the output plane after the nth iteration, we have the summation of the outputs from all channels

$$\begin{aligned} v_i(mo) &= \sum_m A\left(gV(m) * V(mi)\right)^n v_i(m) \\ &= A\sum_m \left(C^{(m)}\right)^n v_i(m) \end{aligned} \tag{5.3}$$

where $C^{(m)}$ is

$$C^{(m)} = gV(m) * V(mi) \tag{5.4}$$

It is the inner-product or the zero-order correlation of the Fourier spectra of the input pattern and the stored patterns. We have proved in chapter three that it also indicates the similarity of the input pattern and the stored patterns. Actually it is equal to

$$C^{(m)} = g\sum_{i=1}^{N} v_i(m)v_i(mi) \tag{5.5}$$

The difference of the HOFNET with the high order matched filter is that the order of the nonlinearity of the network is not fixed but increases by unity on each iteration until the strongest signal is selected and all the others are suppressed. Each iteration performs an independent calculation of the correlation which is multiplied by the product of all of the earlier similar calculations circulating in the loop. So time varying noise within the net or on the input pattern tends to be averaged out. The loop gain is given by

$$G = g(V(m)*V(mi)) \tag{5.6}$$

If G is unity in one channel, then the signal in that channel circulates at constant amplitude. If G is greater than unity the signal magnitude increases as the loop iterates, increasing at higher rates the closer the input pattern is to the known stored pattern in that channel [Sel90]. If G is less than unity the signal in that channel decreases, ultimately vanishing or

iterating at noise level. The value of G (and hence g) determines the radius of attraction of the net around each of the stored patterns. By setting G correctly we can arrange that only input patterns sufficiently similar to the known stored pattern in the channel persist, circulating within the loop.

To see this clearly consider the case when the known stored patterns (having N pixels or bits) are all N of the orthogonal members of a binary (+1, -1) set [Sel90], so that they are all equal Hamming distance apart (the Hamming distance is N/2. If the input pattern, also assumed to be binary (+1, -1), is less than N/4 bits different from a given pattern, s_1, say, then the correct answer will be obtained if it converges to s_1. If it is more than N/4 bits different, it will be closer to another of the orthogonal stored patterns and so it should not converge to s_1. Therefore, we need to set the radius of attraction area with the pattern s_1, to N/2. We want patterns with correlations greater than N/2 to persist and those less to vanish. This can be achieved by setting (V(m)*V(mi)) = N/2 for G = 1 giving a fixed gain of g = 2/N [Sel90]. If any binary input pattern is put into the net, each channel will iterate but only one pattern will persist. Because only the channel with the pattern stored which is most similar to the input pattern will have a gain greater than loss and all the other channels have the gain less than loss. Therefore, all others will decay to zero on sufficient repeated iterations so giving the correct answer.

In practical systems, it is not convenient to implement inhibitory interconnections, especially in optics. We can use phase patterns to perform negative interconnects, but the phase patterns are invisible and it is not easy to keep the phase at exactly 180^o shift for minus one interconnects. In our system, however, there are no direct inhibitory interconnections. We selects the maximum correlation by feedback loop, normalisation and threshold.

Compare the output v_i(mo) after the nth iteration with an inner product high order neural network system which has a nonlinear response in the correlation plane [Owe87] (See section 3. 1. 2). The nonlinear function can be written as

$$f(x) = x^k \qquad (5.7)$$

where x is the correlation output or input to the nonlinear device, and k is the nonlinearity of the device response used in the system. We conclude that our system is equivalent to the inner product high order net and the order of the HOFNET is not fixed but increases by one after every iteration. After the nth iteration, the HOFNET has the same order as the inner product net with the nonlinearity k=n.

Rewrite equation (5.3) as

$$v_i(mo) = \underbrace{A(C(mi))^n v_i(mi)}_{\text{(Signal)}} + \underbrace{A \sum_{m \neq mi} (C(m))^n v_i(m)}_{\text{(Noise)}} \tag{5.8}$$

The signal to noise ratio is

$$SNR = \frac{|C(mi)|^n}{\sqrt{\sum_{m \neq mi} \left[(C(m))^n\right]^2}} \tag{5.9}$$

Assume that the binary and bipolar $v_i(m)$ have the values of +1 or -1 with equal probability. Expressions for C(m) according to the phase diffused model [Kra82] are given as

$$C(m) = \sqrt{\frac{2N}{3}} + N\delta(m - mi) \tag{5.10}$$

where $\delta(m - mi)$ is Dirac-Delta function. Therefore the SNR can be written as

$$SNR \approx \sqrt{\frac{N^n}{M}} \tag{5.11}$$

Here the definition of noise has some difference from the usual one. The noise mentioned here is actually the cross-correlation between the input pattern and the stored patterns except the one which is most similar to the input pattern. If n=1, then the net becomes a linear neural network, such as the Hopfield net. Suppose when SNR = X, the network can converge

into a correct answer. For the linear neural network (n=1),

$$SNR \approx \sqrt{\frac{N}{M}} \tag{5.12}$$

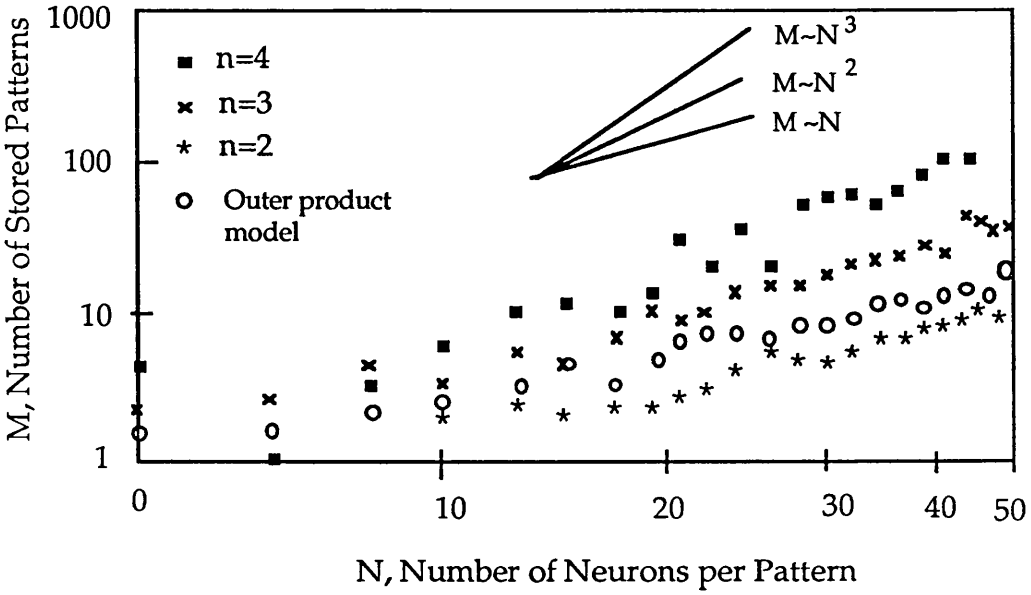


Fig. 5.6 Comparison of Storage Capacity for Various Nonlinearities

So we can have the $M=N/X^2$. When N is not very large, we can have approximately $M \approx 0.15N$. From the same level of noise, for a nonlinear network with order n, the memory capacity can be written as

$$M \approx 0.15N^n \tag{5.13}$$

It is clear M is proportional to N^n. Thus without considering limits set by the available dynamic range of devices, the storage capacity can be increased by increasing the number of iterations. The improvement in storage capacity of the non-linear network is analogous to the improvement in storage capacity found by others for high order discriminant models based on the outer product neural models [Lee86; Psa86]. Owechko did some calculations for nonlinear holographic associative memory systems based on the inner product net and confirmed the above mathematical analysis (Fig. 5.6) [Owe87].

5.3 Noise Analysis of the HOFNET

In section 5.2, we discussed the principle of the HOFNET with no noise involved. In practice, noise exists everywhere, so now we consider the case which involves both system noise $n_S(\tau, i)$ and input noise $n_I(\tau, m)$, where τ is the time and the system noise is also a function of the channel m (m=1, ..., M) and input noise a function of the pixel i (i=1,..., N). The system can be represented in Fig. 5.7. The input pattern $v_i(mi)+n_I(\tau_0, i)$ correlates with all the stored patterns **V**(m) (m=1, ..., M) to give correlation outputs:

$$C^{(1)}(m)=[v_i(mi)+n_I(\tau_0 ,i)]*v_i(m)+n_S(\tau_0, m) \qquad (5.14)$$

where τ_0 is the time when the system starts to operate and $C^{(1)}(m)$ is the first correlation from channel m.

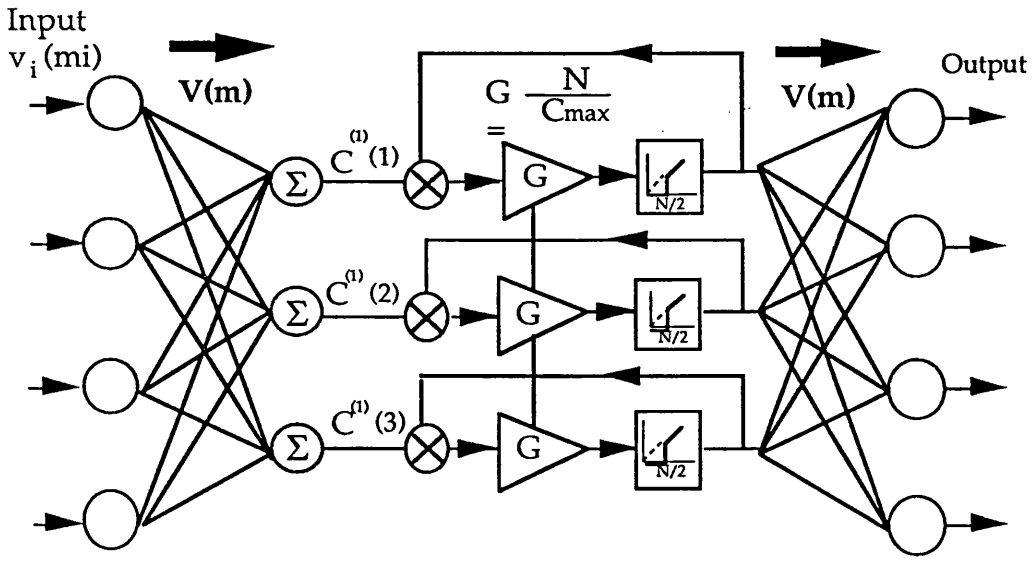


Fig. 5.7 HOFNET with Time Dependent Noise

$C^{(1)}(m)$ goes to an amplifier and is thresholded and fed back to multiply the new correlation outputs at time τ_1, so

$$C^{(2)}(m)=C^{(1)}(m)\,[v_i(mi)+n_I(\tau_1, i)]*v_i(m)+n_S(\tau_1, m) \tag{5.15}$$

Consequently, after the *n*th iteration, we have

$$C^{(n)}(m)=C^{(n-1)}(m)[v_i(mi)+n_I(\tau_{(n-1)},i)]*v_i(m)+n_S(\tau_{(n-1)},m) \tag{5.16}$$

The output $v_i(mo)$ at the output plane is

$$v_i(mo)=\sum_{m=1}^{M} C^{(n)}(m)v_i(m) \tag{5.17}$$

Insert $C^{(1)}(m)$, ..., $C^{(n)}(m)$ into Eq. (5.17), $v_i(mo)$ can be written as

$$\begin{aligned} v_i(mo) = &\sum_{m=1}^{M}\left(\sum_{i=1}^{N}\left(v_i(m)v_i(mi)\right)^n\right)v_i(m) \\ &+\sum_{m=1}^{M}\left\{\sum_{i=1}^{N}\left(v_i(m)*v_i(mi)\right)^{n-1}\left[\sum_{j=0}^{n-2}\left(\sum_i v_i(m)n_I(t_j,i)+n_S(t_j,m)\right)\right]\right\}v_i(m) \\ &+\sum_{m=1}^{M} n_S(t_{n-1},m)v_i(m) \end{aligned} \tag{5.18}$$

From the above equation, if we suppose that $n_S(\tau, m)$ and $n_I(\tau, i)$ are random and time-varying noises and the stored patterns and the input pattern are binary (+1, -1) with equal probability of +1 and -1, then all the terms in the right hand side of Eq. (5.13) are zero except for the first term. Therefore, if the noises (system noise and input noise) are time varying, then at the output plane, the noises will be averaged out. The result will not be affected, as the same noise level will be added to all the channels.

5.4 Computer Simulations

The purpose of this design is to find a system which is noise tolerant and overcomes the limited dynamic range of the devices, as optical systems produce noise and optical devices have a limited dynamic range (now 27dB is obtainable by an asymmetric FP MQW modulator [Whi89]). Here

we did some computer simulation for noise analysis. The architecture was simulated to assess the tolerance to both time independent noise and time varying noise. The inputs consisted of patterns formed on a grid of 8x8=64 pixels. We defined the characters 0 to 9 and the letters A to Z on this grid (Fig. 5.8). For differing numbers of patterns stored, M = 5, 10, 20, 30, we input noisy patterns and found the probability of convergence. For example, we entered the pattern for a letter A but having 3 pixels reversed (i.e. +1 replaced by -1 or vice versa). The positions of these pixels were chosen randomly numerous times and the probability of correct recall was calculated. This was repeated for all the memorised characters. These results describe the behaviour for time independent noise. In addition we input patterns in which, using our earlier example, the 3 pixels chosen change in time so that a different random choice of pixels was made at the next time step. This was used to assess the averaging ability of the net design and to check that time dependent noise was not a problem.

The results are shown in Fig. 5.9. These show clearly that as the input noise was increased, in terms of the number of reversed pixels, that the probability of correct recall dropped quite suddenly after a certain noise threshold. The net could tolerate more time dependent noise than time independent noise as a result of the modifications made. If no special modifications have been made we might expect identical curves for both types of noise of equal magnitude. Fig. 5.10 plots the amount of input noise against the number of stored patterns assuming an 70% probability of correct recall. This shows that the tolerance to input noise drops with the number of stored patterns but that more time dependent noise can be tolerated particularly for memory capacities less than 0.5 N. The improved performance with respect to time dependent noise is particularly important for optical systems with high levels of internal system noise.

One application of neural networks is pattern recognition. The HOFNET can be used to recognise a pattern from a highly distorted pattern, especially when the noise is time-dependent. First, as in the previous section, we trained the HOFNET with digits 0 to 9 and capital letters A to Z and then gave a distorted input to see the procedure of convergence. Fig. 5.11 is an example. A pattern with 8 pixels reversed (1 changes to 0 vice versa) is inserted into the system, the network will start to iterate. After the first iteration, the output is similar to B or 8, rather than S. This is because the B is very similar to 8, 6, 9 3 etc., so it is a local

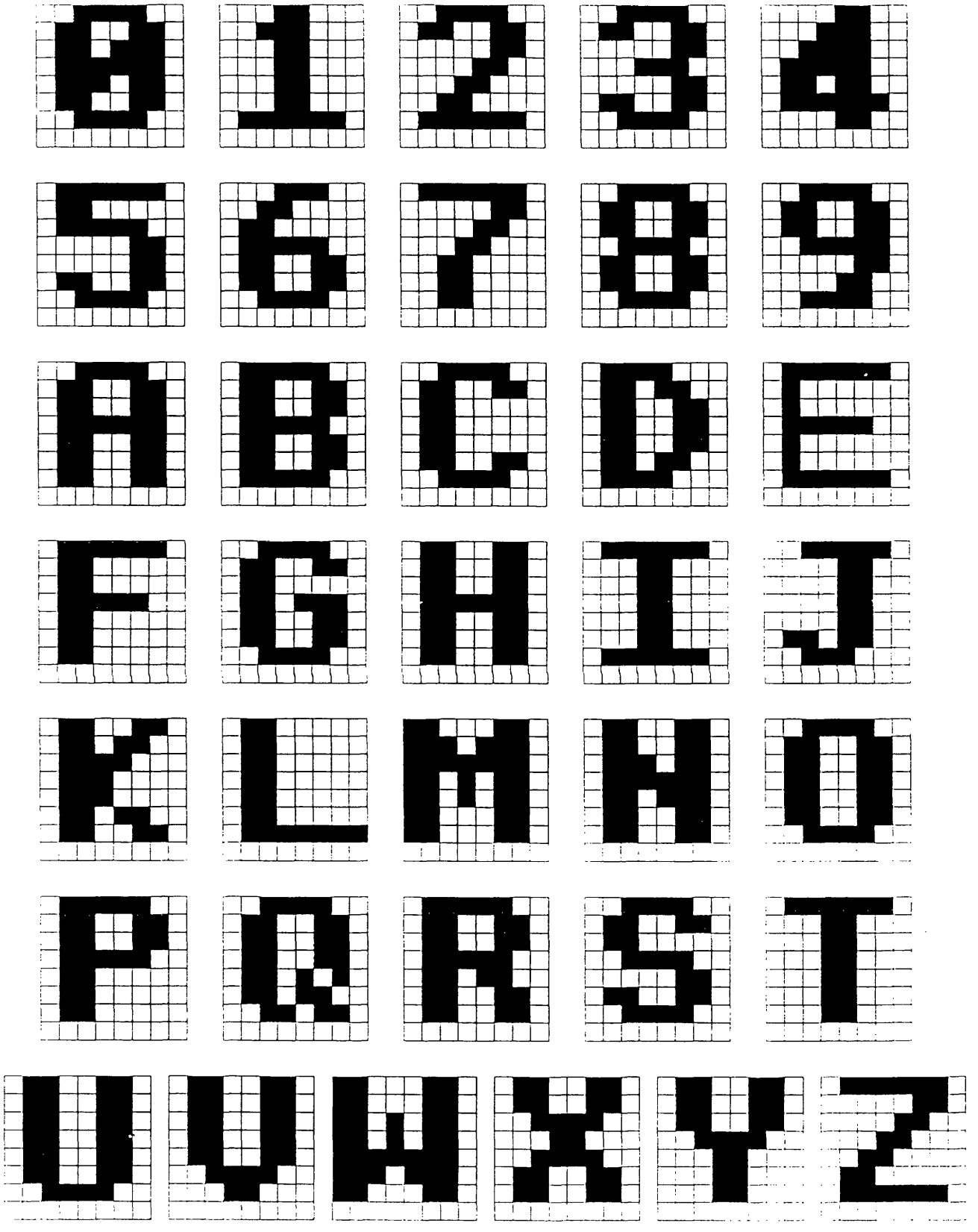
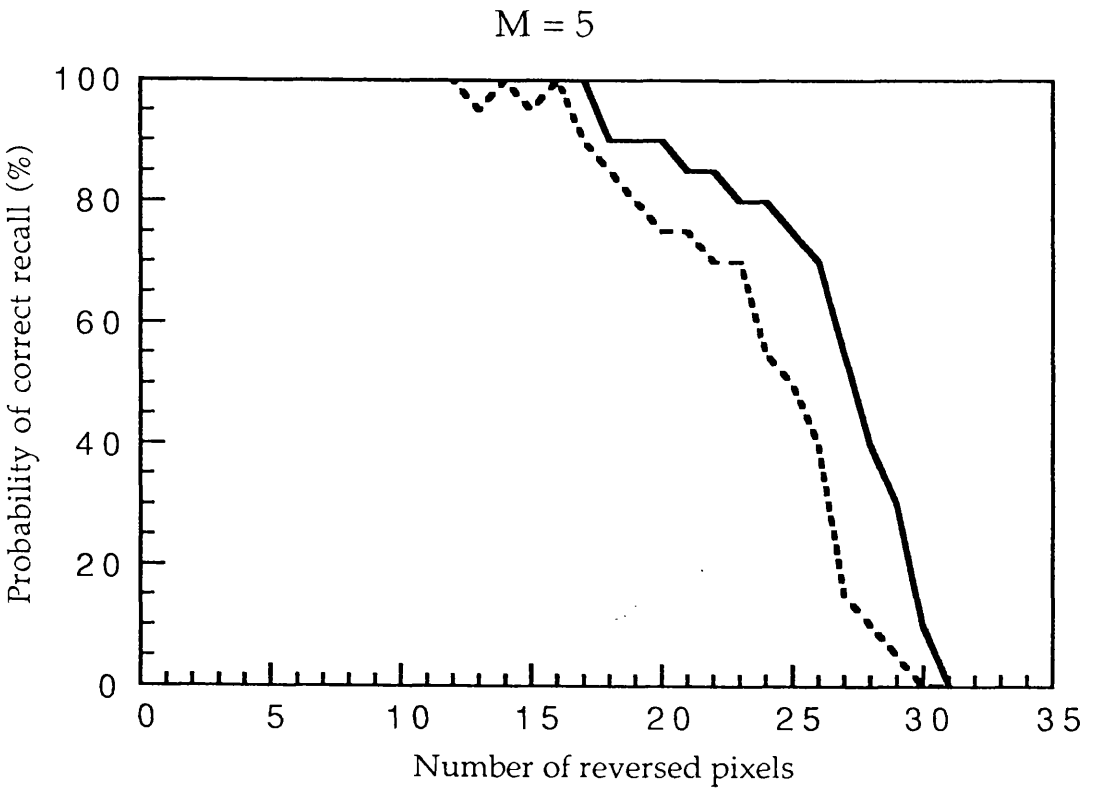
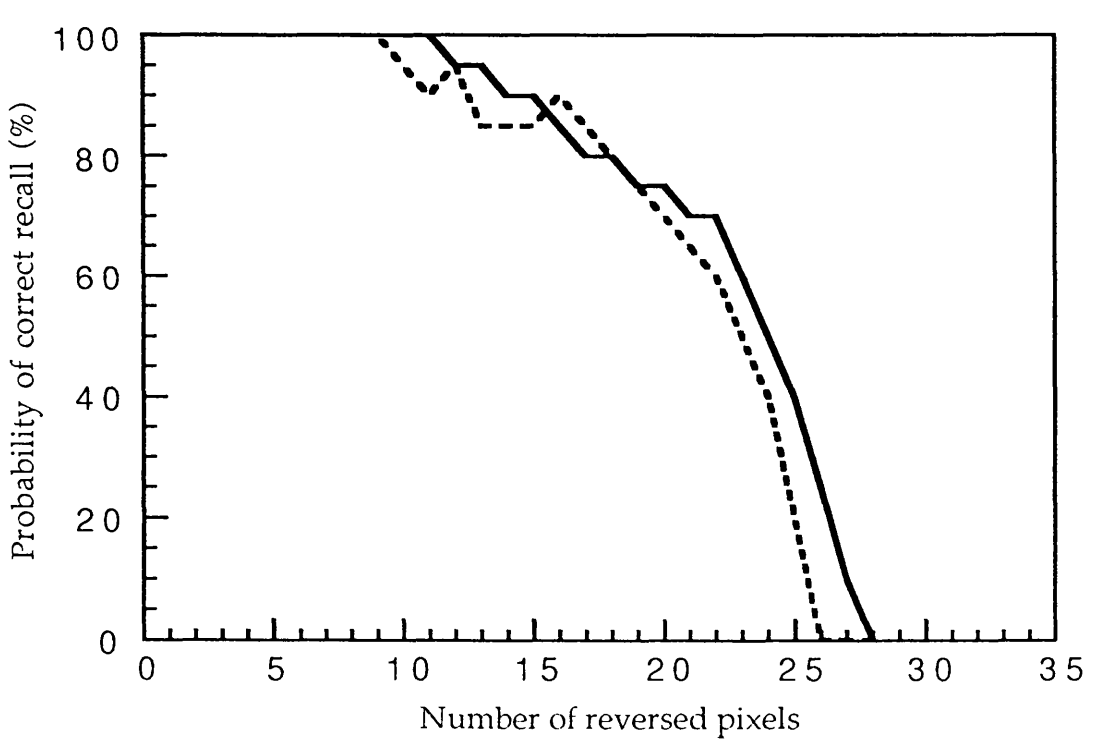


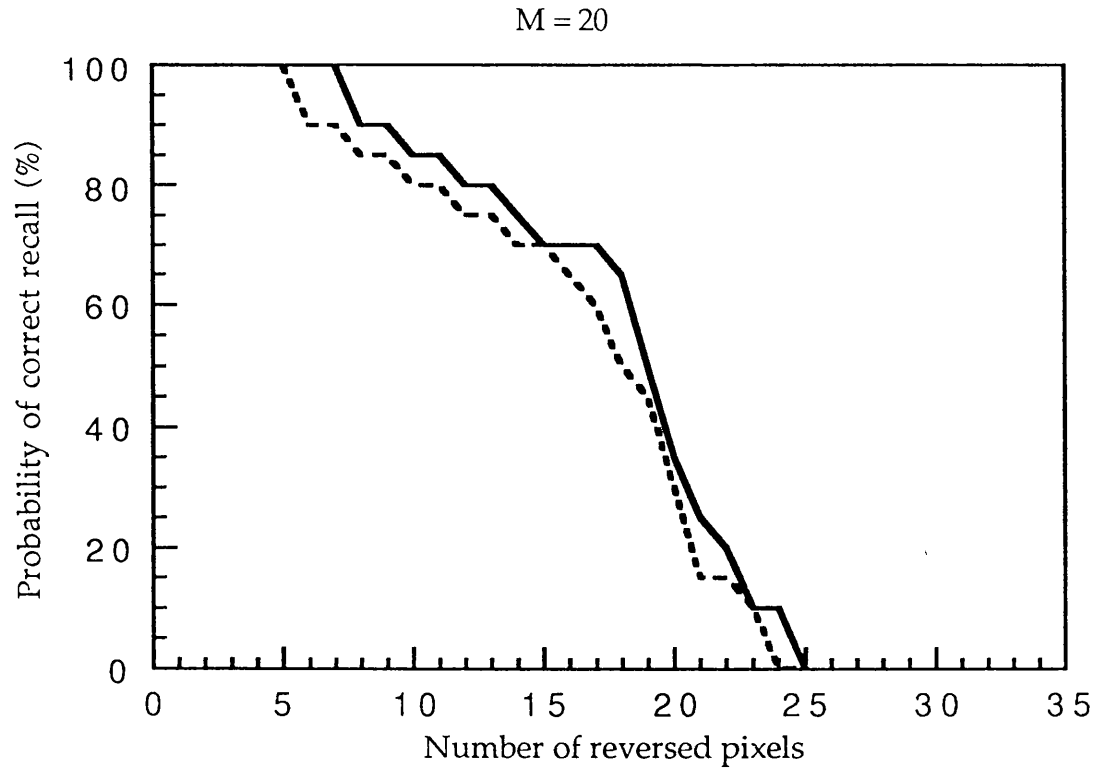
Fig. 5.8 36 Patterns Stored in the HOFNET (N=64, M=36)



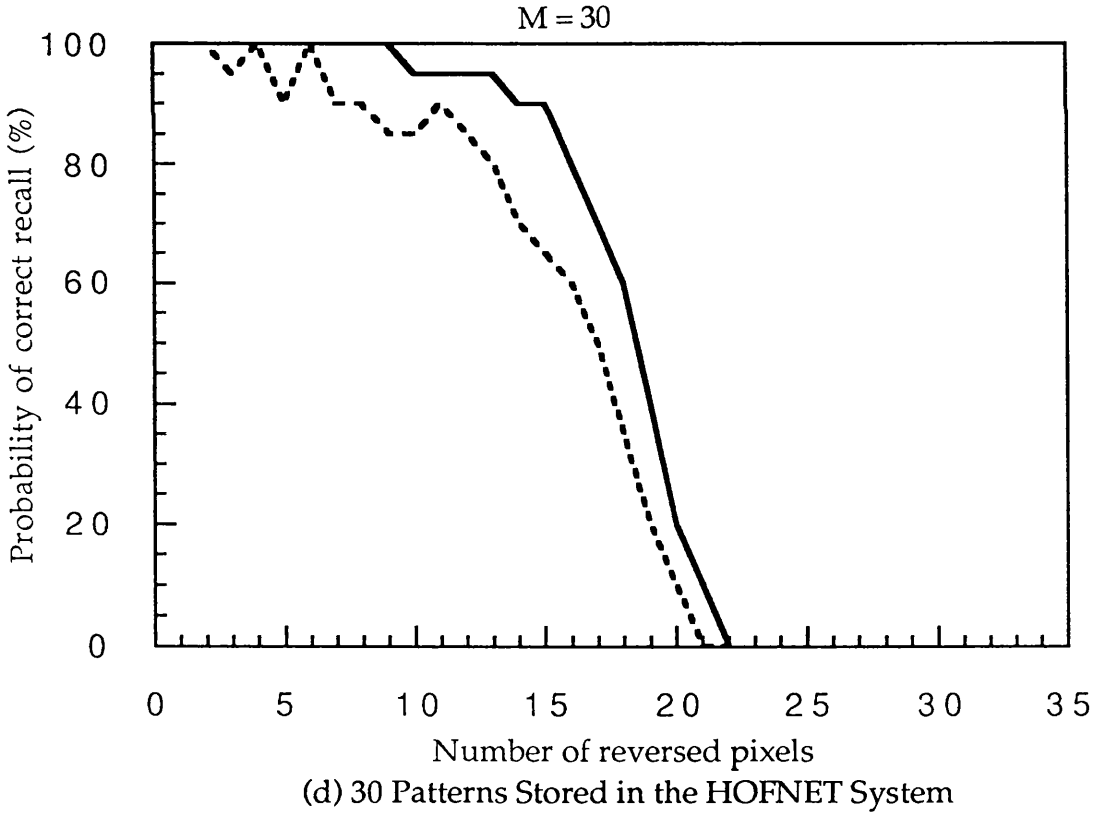
(a) 5 Patterns Stored in the HOFNET System



(b) 10 Patterns Stored in the HOFNET System



(c) 20 Patterns Stored in the HOFNET System



(d) 30 Patterns Stored in the HOFNET System

Fig. 5.9 Probability of Correct Recall as a Function of Input Noise (N=64)
____ Time dependent noise - - - - - Time independent noise

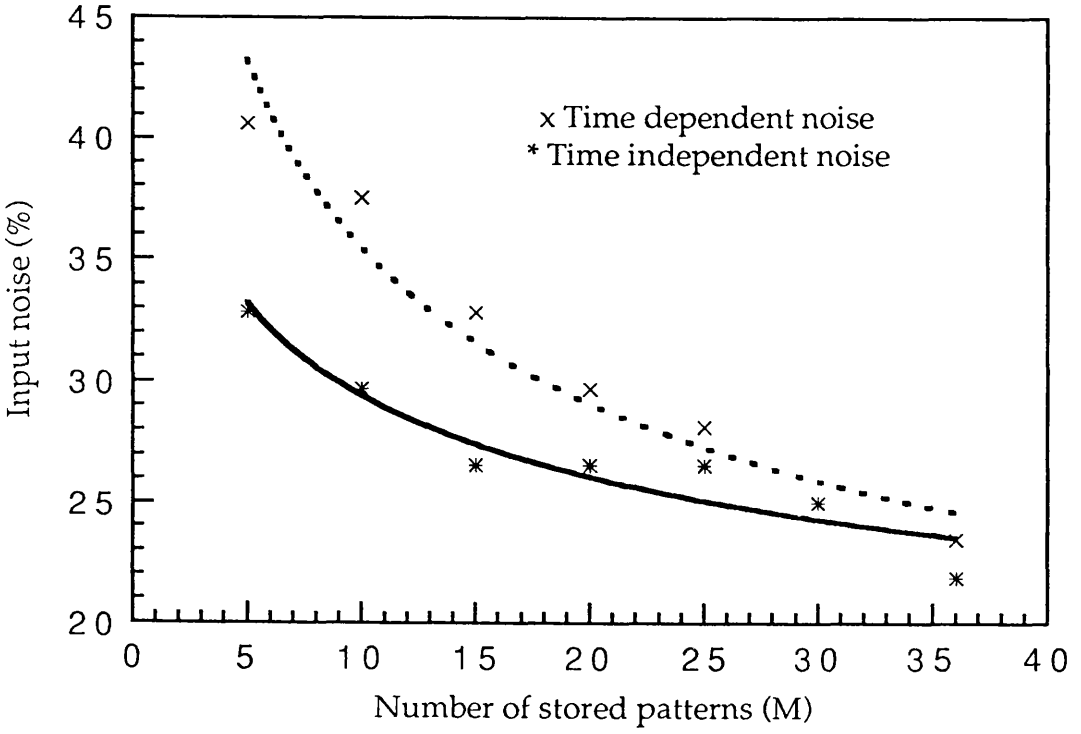


Fig. 5.10 Maximum Input Noise for Correct Recall as a Function of Memory Capacity (N=64)

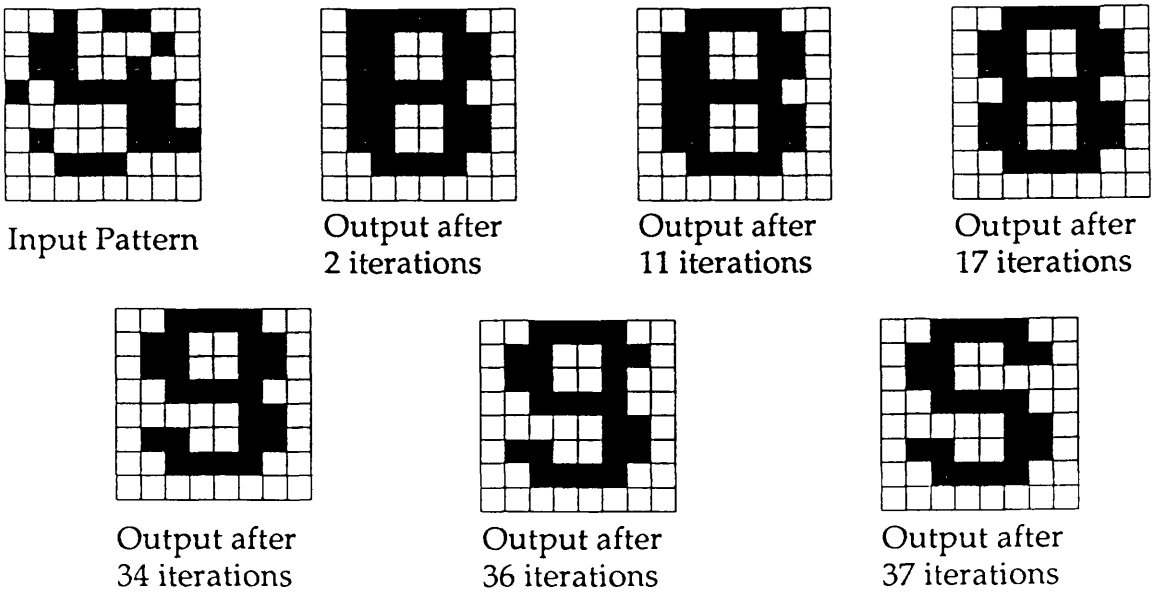


Fig. 5.11 Pattern Recognition Procedure

minimum.

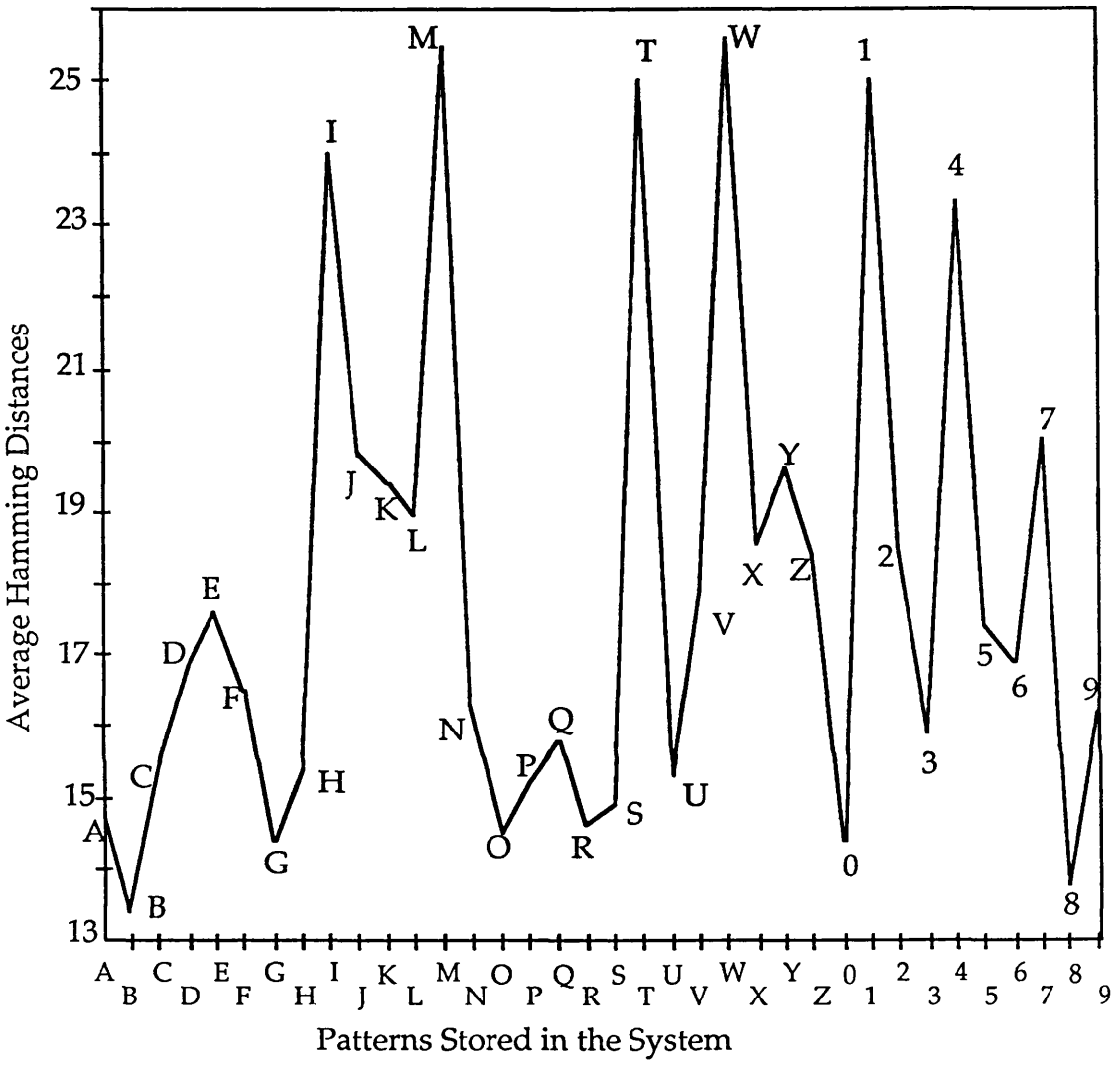


Fig. 5.12 Average Hamming Distances of Stored Patterns

Fig. 5.12 is the average Hamming distances for all the stored patterns. As defined before (see chapter two), the Hamming distance is the number of corresponding different pixels of two patterns. We calculate the Hamming distances between one pattern and all the other patterns and average it over the number of patterns. The Hamming distance of the pattern B or 8, as shown in Fig. 5.12, is very small, which means that they are very similar to other patterns, e.g. 6, 9, 3. Therefore, when a pattern is inserted into the system, the system tends to recognise it as a more stable pattern B or something similar to it. To avoid this problem, Twaij et al

[Twa91] proposed a feature extraction learning algorithm which tries to enlarge the difference of two patterns and suppress their similarity. So the patterns after feature extraction, have almost the same Hamming distance. The local minimum problem is removed. As the iteration proceeds, the output starts to change, getting closer to S. After 37 iterations, the correction is given.

5.5 Conclusions

The HOFNET system has some advantages over other neural network systems. Firstly, in the HOFNET system, the input is correlated in each iteration, so the system is more noise-tolerant. Also in the HOFNET system, the nonlinearity is implemented by a feedback loop, rather than nonlinear devices, so the high order is not limited by the dynamic range of nonlinear devices. Although the HOFNET system is designed specifically for optical implementation, it can also be used for electronic construction. In our system, no direct inhibitory interconnects make it easier to implement, both optically and electronically.

Chapter Six

OPTICAL HOFNET SYSTEM

6.1 Introduction

In this chapter, we discuss the optical or optoelectronic implementation of the HOFNET system. The optical HOFNET system has some differences from conventional ones. First we modify the usual formulation of the holographic memory in neural networks. We calculate the similarity of the input pattern and the stored patterns according to the zero-order spectrum correlation or the multiplication of the input pattern and the stored patterns (see section 3.4). The amplitude or intensity distribution of this correlation is usually different from that of pattern correlation (spatial correlation) (see section 3.3), but if we consider only the zero-order, as we discussed previously, it indicates the similarity of the input pattern with each of the stored patterns. We have designed two optical HOFNET systems, which we call a focused system and a lensless system, respectively and we have optically implemented the focused system. In the focused system, the patterns are stored in the form of Fourier transform spectra and a lens is used to make the Fourier transform, while in the lensless system, the patterns are stored directly in the hologram (Fresnel hologram), no lenses are used in the recording system. Actually, the lensless system is only a pseudo-spectrum correlation, as in this system Fresnel diffraction theory is used. The details are discussed in section 6.3. The focused system requires only spatially multiplexed holograms in order to cope with correlation outputs separately, that means that all the holograms are recorded in the same way (spatially multiplexed), so it is easy to eliminate the problem of the diffraction efficiency varying, as the diffraction efficiency is usually the function of the modulation frequencies. The lensless system, however, involves both spatial and angular changes. When we record a pattern in the form of holograms, a

diaphragm is used in front of the recording plate to limit the size of the hologram. As soon as the pattern is stored, another pattern is put in the exact position as the previous one and the diaphragm is moved to a different position while the reference beam is unchanged. But moving the diaphragm causes the incident angle of the object beam with the holographic plate to change as well, so the holograms are recorded in the hybrid way (see section 3.5). The details are discussed in section 6.3. More importantly, we increase memory capacity by using non-linearity in the correlation plane, which is implemented by a feedback loop, instead of nonlinear devices in that plane [Owe87]. The feedback loop is necessary for a Hebbian-learning-rule based neural net in order to make the system have error correction capability. In section 6.2 and 6.3, we discuss the implementation and design of the HOFNET. In both systems, an electrically addressed ferroelectric liquid crystal spatial light modulator is used in the feedback loop, which is discussed in section 6.4. Some experimental results are given in section 6.5. Comments on these systems are given at the end.

6.2 Focused Optical HOFNET System

The optical HOFNET system contains two main parts: correlation or inner product sub-system and feedback sub-system. The inner products of the input pattern and the stored patterns indicate the similarities of them. We have proved that the zero order of the spatial or spectrum correlation of two patterns is equivalent to the inner product of the two patterns. Here the spectrum correlation system (Fig. 3.6) is selected for such a task and patterns are stored in holograms in a spatially multiplexed way (Fig. 6.1 or see section 3.5). The patterns are stored in the form of their Fourier transform spectra, which are implemented by a lens, so we call this recording the "focused" recording, and consequently call the system based on such a hologram the "focused" optical system. Here we let all the patterns be amplitude modulated binary unipolar patterns (black and white). As soon as one Fourier transform hologram is recorded, the recording plate is moved laterally so that another hologram is recorded in a different position in the holographic plate, while the angles between the reference beam and the object beams keep unchanged. The whole focused optical HOFNET system is shown in Fig. 6.2 and the photograph of the real

system is shown in Fig. 6.3.

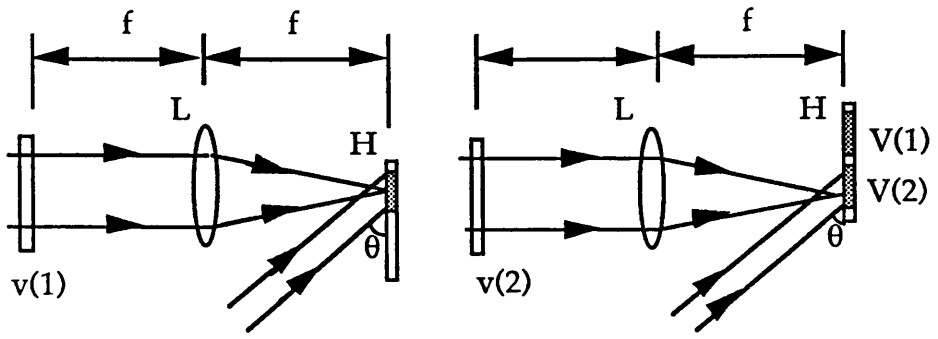


Fig. 6.1 Spatially Multiplexed Hologram Recording

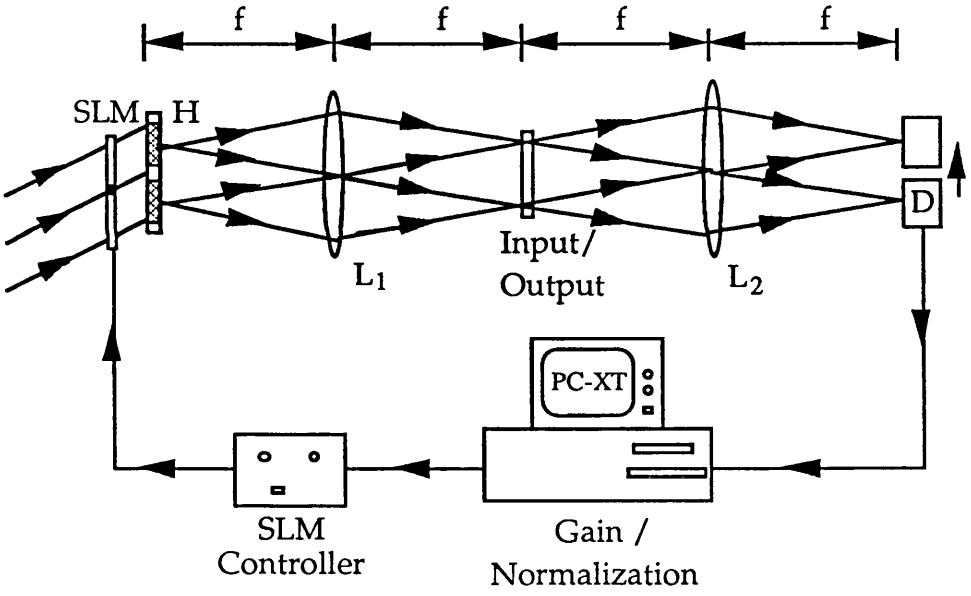


Fig. 6.2 Focused Optical HOFNET System

The spatially multiplexed hologram array H recorded in Fig. 6.1 is replayed by a parallel beam which is spatially modulated by a spatial light modulator (SLM). The SLM is an electrically addressed ferroelectric chiral smectic C liquid crystal one produced by STC Ltd, now called BNR. The thickness of the liquid crystal layer is 1.7 μm. The SLM is divided into M "subsectors", each corresponds to a hologram element recorded on the plate. According to the translation invariance property of the Fourier

transform hologram and the recording condition during the learning procedure, all of the reconstructed patterns will be superimposed at the back focal plane of lens L_1. If an amplitude modulated binary input pattern is inserted in this plane and superimposed on the reconstructed patterns, at the back focal plane of lens L_2, the correlation peaks which indicate the similarities between the input pattern and the stored patterns are obtained. These correlation peaks are fed back to control the transmittance of their corresponding subsectors in the SLM.

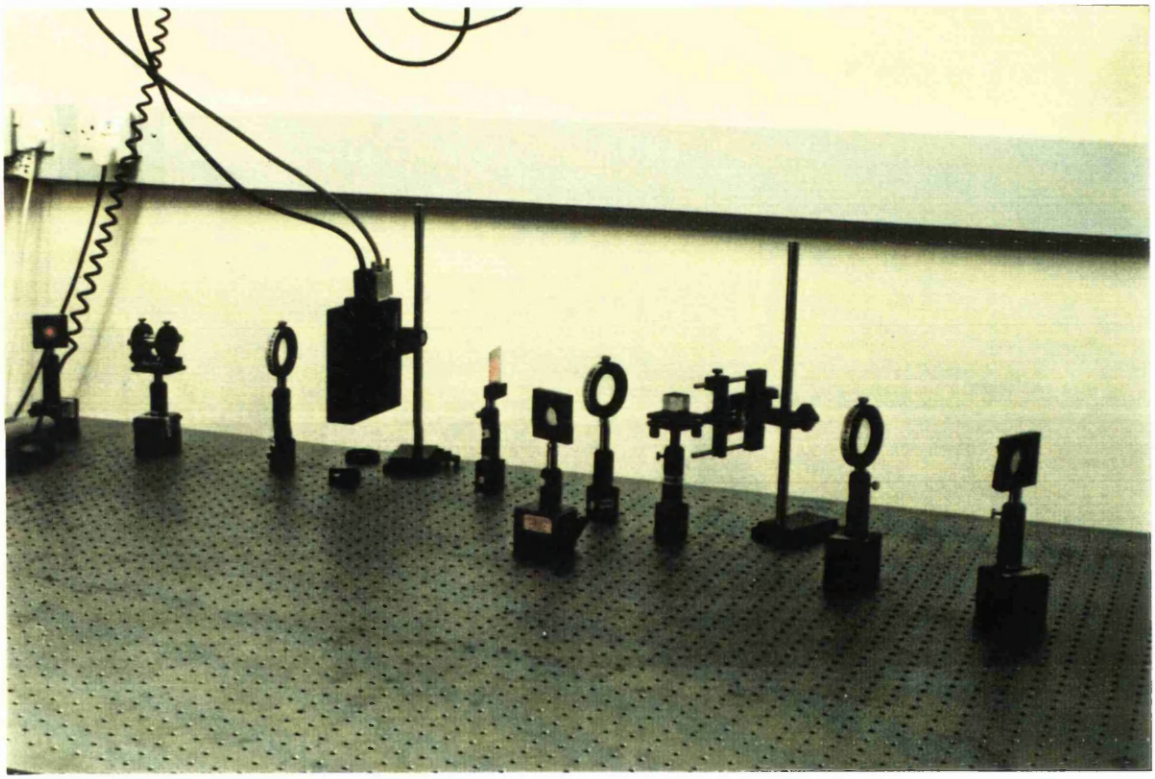


Fig. 6.3 Photograph of the Optical HOFNET System

The SLM consists of binary, unipolar 128x128 pixels in an area of about 2cm x 2cm, each pixel having a size of 165 μm x 165μm. The average contrast ratio is about 13dB, including the gaps between adjacent pixels. For a single pixel, however, the on-off contrast can be as high as 25dB. In our system, 4x4 or 16 patterns are stored in the hologram array and the size of each hologram element is 3mm in diameter and the spacing between two adjacent holograms is 5mm (Fig. 6.4). So the SLM is divided into 4x4 subsectors each with the size of 5mm x 5mm, so that each subsector matches the size of one hologram. Each subsector, therefore, contains

32x32 binary pixels (black and white) (Fig. 6.5). The SLM is used to control the transmitted light passing through it by setting its transmittance of each subsector to be proportional to the correlation output of the corresponding channel. As the correlation outputs are changed continuously, we, ideally, require a grey level SLM. However, since this was not available at our present laboratory, we simulated grey levels on the binary SLM by further subdividing the subsector into 8x8 units, each containing 4x4 pixels. This resulted in 16 simulated grey level steps. This scheme gives a more uniform transmittance than if the subsector were not subdivided.

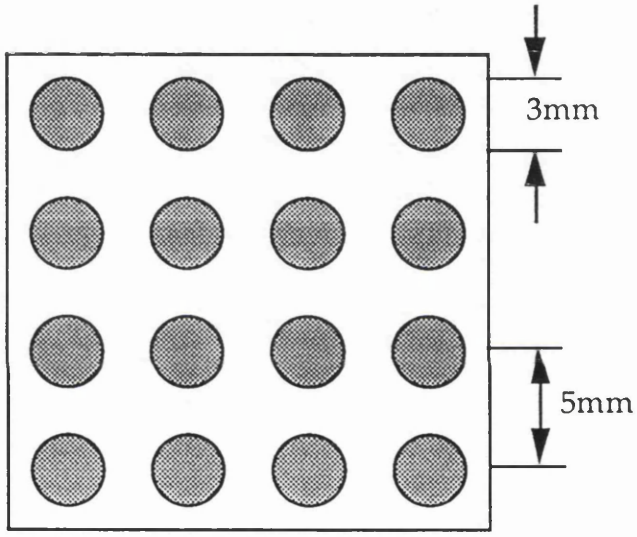


Fig. 6.4 Spatially Multiplexed Hologram Array

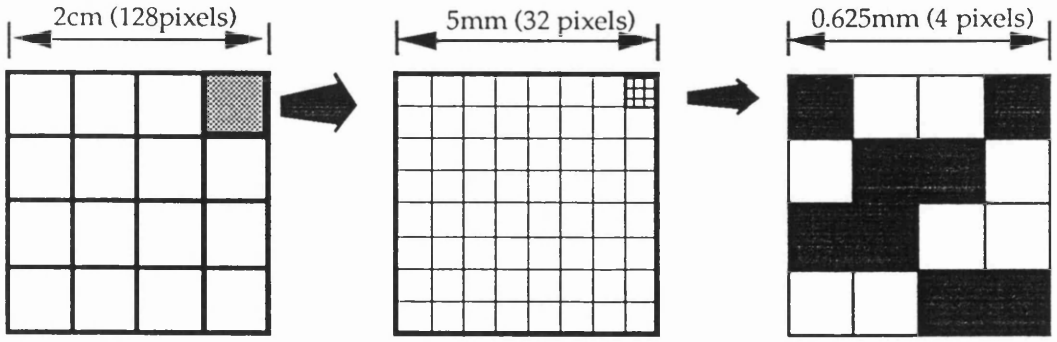


Fig. 6.5 Grey Level Simulation by the Binary SLM

Before the system starts to iterate, the transmittance of all of the subsectors of the SLM is set to be the highest (completely transparent). So the light intensity falling on the hologram elements is uniform. The correlation outputs, however, depend on the similarity of the input pattern and the stored patterns in corresponding channels. These outputs are fed back to control the transmittance of subsectors of the SLM via a computer. The transmittance of a subsector is set to be proportional to the correlation output in the corresponding channel. On the second iteration, the intensity of the reconstruction laser beam falling on the hologram elements is no longer uniform, but is modulated by the SLM. The pattern which is most similar to the input pattern will be illuminated by the same light as during the first iteration (the corresponding subsector of the SLM is normalised to have the highest transmittance), thus the correlation peak in that channel will keep unchanged. The other hologram elements are illuminated by weakened light, so the correlation outputs in corresponding channels decrease compared to those during the first iteration. After several iterations, the weaker correlations will fall below a threshold assuming the maximum correlation is normalised to keep constant in each iteration. Fig. 6.6 shows an example of the procedure of the maximum correlation output selection.

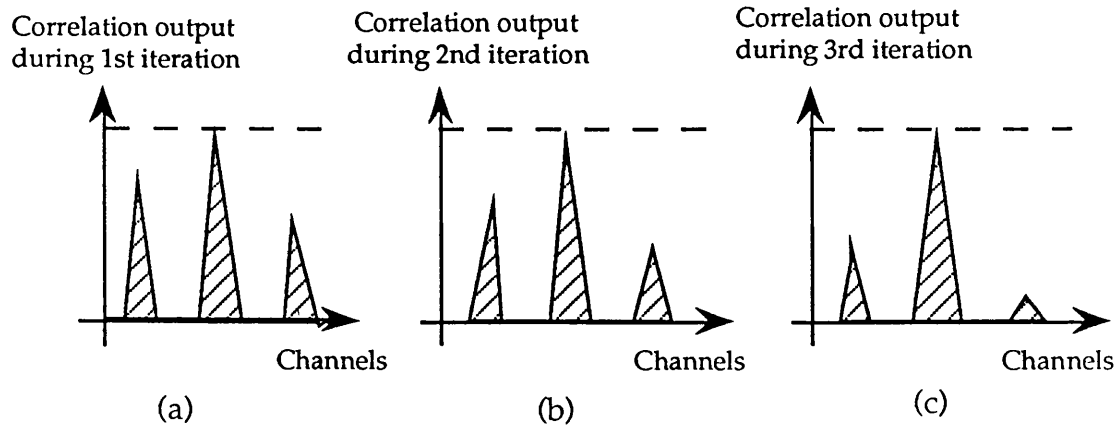


Fig. 6.6 Maximum Correlation Output Selection Procedure

Suppose the training patterns we discuss here are binary (the patterns can be analogue, the supposition of binary patterns is only for easy comparison with neural networks), so the set of patterns stored in the

hologram can be written as

$$v(m) = \{v_1(m),\ldots,v_i(m),\ldots,v_N(m)\} \quad (m=1,\ldots,M) \tag{6.1}$$

Suppose the Fourier transform of v(m) is V(m'), then

$$V(m') = \{V_1(m'),\ldots,V_k(m'),\ldots,V_N(m')\} \qquad (m'=1,\ldots,M) \tag{6.2}$$

For simplicity, we let the interval distance between two adjacent sub-holograms be 1 and only consider one dimension. So the hologram array H can be expressed as

$$g(x) = V(m') \otimes \delta(x-m') \tag{6.3}$$

where delta δ is the Kronecker delta function and x is spatial position in the hologram plane. In the back focal plane of lens L_1, we obtain the Fourier transform of g (x), where an input pattern v(mi) (v_1(mi), ..., v_i(mi), ..., v_N(mi)) is injected, so at the back focal plane of lens L_2, we have the Fourier transform of the multiplication of the input pattern v(mi) and stored patterns v(m) (m=1, 2, ..., M). If we put a pinhole array at this plane to threshold all the high orders of the correlation, after the pinhole array, we have the amplitude distribution on this plane

$$C(x)^{(1)} = \sum_{i=1}^{N} v_i(m)v_i(mi) \otimes \delta(x-m) \tag{6.4}$$

where $C(x)^{(1)}$ ($C(x)^{(1)} = 0$, unless x=m) is the correlation output or inner product during the first iteration. $C(x)^{(1)}$ is fed back to adjust the SLM transmittance so that the transmittance of each subsector is proportional to its corresponding zero order correlation value. Now the light intensity falling on the holograms is modulated. At the end of this iteration and after the pinhole, we will get

$$C(x)^{(2)} = \left(\sum_{i}^{N} v_i(m)v_i(mi)\right)^2 \otimes \delta(x-m) \tag{6.5}$$

where $C(x)^{(2)}$ is the correlation output during the second iteration. Similarly, after the *n*th iteration, we have an output at the back focal plane of lens L_2

$$C(x)^{(n)} = \left(\sum_{i=1}^{N} v_i(m) v_i(mi) \right)^n \otimes \delta(x-m) \tag{6.6}$$

At the input/output plane we have an output

$$\begin{aligned} v_i(mo) &= \sum_{m=1}^{M} C(x)^{(n)} v_i(m) \\ &= \sum_{m=1}^{M} \left(\sum_{i=1}^{N} v_i(m) v_i(mi) \right)^n v_i(m) \end{aligned} \tag{6.7}$$

By comparing this equation with the high order inner product model of Owechko et al [Owe87], where the output is

$$v_i(mo) = \sum_{m=1}^{M} f(C(m)) v_i(m) \tag{6.8}$$

The nonlinearity function f(x) is

$$f(x) = x^n$$

Comparing equations (6.7) and (6.8), we know our system is equivalent to the inner product model system with a nonlinearity in the correlation plane. The signal to noise ratio of such a system can be written as [Owe87]

$$SNR = \frac{|C(i_0)^{(n)}|}{\sqrt{\sum_{i \neq i_o} |C(i)^{(n)}|^2}} \tag{6.9}$$

where i_0 corresponds to the channel with maximum correlation. Here the SNR means the ratio of the inner product of the input pattern with the most similar stored pattern, to the square root of the sum of inner products of the input pattern with the other stored patterns.

As it is discussed in chapter five, suppose individual pixels of the patterns have the values +1 and -1 with equal probability, we can calculate $C(i)^{(n)}$ [Kra82]

$$C(i)^{(n)} = \left(N\delta(i-i_0)+\sqrt{2N/3}\right)^n \tag{6.10}$$

where δ is the Kronecker delta function. Substituting equation (6.10) into equation (6.9) and approximating, we have

$$SNR \approx \left(\sqrt{3N/2}\right)^n \Big/ \sqrt{M} \qquad (N, M >> 1) \tag{6.11}$$

From equation (6.11), we can see that the memory capacity is roughly proportional to N^n [Owe86] (or see section 2.4). So the theoretical storage capacity of high order neural networks depends on the order of the nonlinearity (the value of n in equation (6.7)). In our system, n is equal to the number of iterations. So the nonlinearity is not fixed but increases by one on each repeated iteration. It does not require many iterations to ensure that the memory capacity limited by the net architecture is greater than the memory capacity limited by optical components. Another interpretation is that for the same memory capacity repeated iterations reduce the error rate since this results in a greater differential gain and more averaging to reduce noise.

In the above discussion, we did not consider the noise problem. Now we start to discuss this problem. If the system and the input pattern contain time-varying noise, then the equation (6.4) becomes, for the first correlation calculation,

$$\begin{aligned} C(x)^{(1)} &= \sum_{i=1}^{N} v_i(m)\left(v_i(mi)+n_I(\tau_0)\right) \otimes \delta(x-m)+n_S(\tau_0) \\ &= C(\tau_0) \end{aligned} \tag{6.12}$$

where $C(\tau_0)$ is the correlation between the input pattern and the stored patterns with noise, $n_I(\tau_0)$ is the time–varying input noise at time instant τ_0 and $n_S(\tau_0)$ is time–varying system noise power at time instant τ_0,

which may be caused by scattering or the detectors. So $C(x)^{(1)}$ is the function of time. When $C(x)^{(1)}$ is fed back to control the SLM transmittance of the subsector, we will get after the second iteration

$$C(x)^{(2)} = C(x)^{(1)}\left(\sum_{i=1}^{N} v_i(m)\left(v_i(mi) + n_I(\tau_1)\right) \otimes \delta(x-m) + n_S(\tau_1)\right)$$
$$= C(\tau_0)C(\tau_1) \qquad (6.13)$$

where $C(\tau_1)$ is the correlation sampled at time τ_1. Similarly, after the nth iteration, we have at the back focal plane of lens L_2

$$C(x)^{(n)} = C(x)^{(n-1)}\left(\sum_{i=1}^{N} v_i(m)\left(v_i(mi) + n_I(\tau_{n-1})\right) \otimes \delta(x-m) + n_S(\tau_{n-1})\right)$$
$$= C(\tau_0)C(\tau_1)\ldots C(\tau_{n-1}) \qquad (6.14)$$

where $C(\tau_{n-1})$ is the correlation sampled at time τ_{n-1}. Now $C(x)^{(n)}$ is the multiplication of correlations sampled at different times $\tau_0, \tau_1, \ldots, \tau_{n-1}$. So if the noise is time dependent, then the noise will be averaged. Compared to those systems where the input pattern is used only at the beginning [Psa85; Owe87], this system is more tolerant to the time–varying noise.

6.3 Lensless Optical HOFNET System

Usually in optical neural networks, the patterns are stored in the form of Fourier transform holograms [see section 3.1 or Pae87]. Actually, Fresnel holograms can also be used to store information or implement interconnections [Jan88a, b; Jan89]. Here we introduce a novel system, where the Fresnel holograms are used for pattern storage. The holograms are recorded in the hybrid way, that means the modulation frequencies and positions for the stored patterns in the hologram array are different from each other [see chapter three]. The recording system is shown in Fig. 6.7. The ground glass is used to diffuse the phase of the pattern, so that any point in the recording area contains all the information from the pattern. As soon as one pattern is stored, the aperture in front of the hologram to limit the size of the hologram element is moved laterally and another

pattern is stored in a different position. The difference, compared to the Fourier transform holograms, is the aperture moving, rather than the holographic plate moving. It results in a modulation frequency change of the recorded pattern as well as a position change. The distance between the input pattern $o_1(x,y)$ and the holographic plate is set to be d_1, and there is no lens used in the system. So the recording pattern is the Fresnel diffraction of the input pattern. Because of the ground glass, at any point in the recording plate, the amplitude can be written as

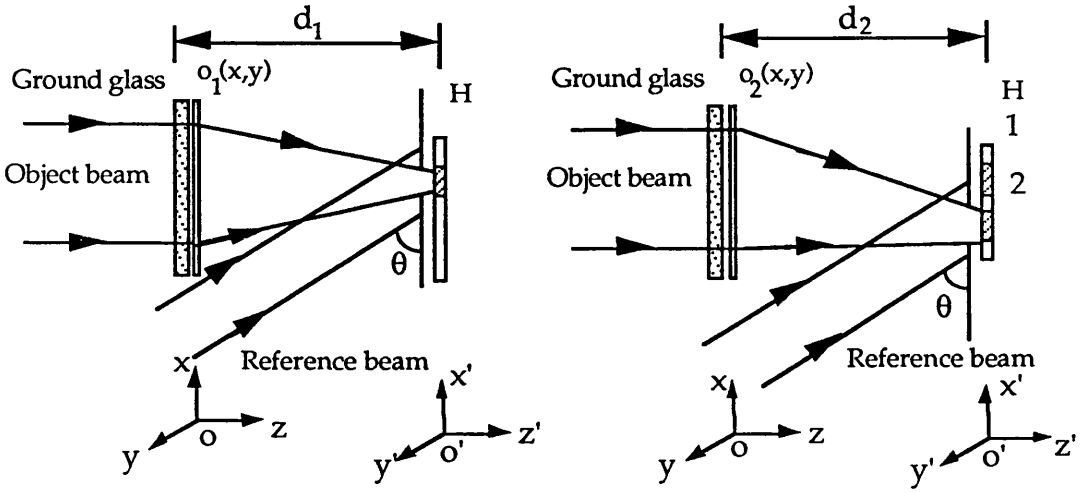


Fig. 6.7 Lensless Holographic Memory Recording

$$O(x',y') = o_1(x,y)\frac{\exp(-j2\pi \mathbf{k}\cdot\mathbf{r})}{r} \tag{6.15}$$

where k is the propagation vector and k is equal to $1/\lambda$; r is the vector pointing to (x′, y′) from (x, y) and r is

$$r = \sqrt{(x'-x)^2 + (y'-y)^2 + (d_1)^2} \tag{6.16}$$

where (x, y) and (x′, y′) are the coordinate pairs in the object plane and the recording plane, respectively. Suppose, for simplicity, the amplitude of the parallel reference beam is 1 and the reference beam lies in parallel to the x′o′z′ plane, then the reference is

$$R = \exp(-j2\pi\xi x') \tag{6.17}$$

where $\xi = \frac{\sin\theta}{\lambda}$ is the carrier frequency, so at the recording plate, the total intensity distribution is

$$\begin{aligned} I(x',y') &= \left|O(x',y') + R\right|^2 \\ &= \left|o_1(x',y')\right|^2 + 1 + o_1(x,y)\frac{\exp(-j2\pi\mathbf{k}\cdot\mathbf{r})}{r}\exp(j2\pi\xi x') \\ &\quad + o_1^*(x,y)\frac{\exp(j2\pi\mathbf{k}\cdot\mathbf{r})}{r}\exp(-j2\pi\xi x') \end{aligned} \tag{6.18}$$

Suppose the holographic plate is linearly processed (see chapter two), when the processed hologram is read by the conjugate reference beam, then the fourth term in Eq. (6.18) which we are interested in, becomes

$$o'_1(x,y) = o_1^*(x,y)\frac{\exp(j2\pi\mathbf{k}\cdot\mathbf{r})}{r} \tag{6.19}$$

If M patterns $o_i(x,y)$ (i=1, ..., M) are stored in the holographic plate at different positions (x_i', y_i'), then the total amplitude is

$$A(x,y) = \sum_{i=1}^{M} o_i^*(x,y)\frac{\exp(j2\pi\mathbf{k}\cdot\mathbf{r}_i)}{r_i} \tag{6.20}$$

where r_i is

$$r_i = \sqrt{(x_i'-x)^2 + (y_i'-y)^2 + d_1^2} \tag{6.21}$$

All these patterns are superimposed together. The whole lensless optical HOFNET system is shown in Fig. 6.8. A lens with focal length f is put behind the hologram, d_1 away. The hologram is reconstructed by a broad parallel beam, so that all the stored patterns are replayed. An input pattern g(x,y) is put just in front of the lens, which is multiplied by the replayed patterns. At the image plane of lens L (d_2 away from the lens L), a series of correlations between the input pattern and the stored patterns are performed by the integral lens L. Now we prove this by mathematical

analysis. For simplicity suppose $d_1 = d_2 = 2f$. The reconstructed patterns are superimposed at the plane a distance 2f away from the hologram. If an input pattern g(x,y) is inserted in this plane, then the amplitude distribution just after lens L is

$$A_1(x,y) = \sum_{i=1}^{M} o_i^*(x,y) \frac{\exp(j2\pi \mathbf{k}\cdot\mathbf{r}_i)}{r_i} g(x,y) \exp\left[\frac{-j\pi k(x^2+y^2)}{f}\right] \qquad (6.22)$$

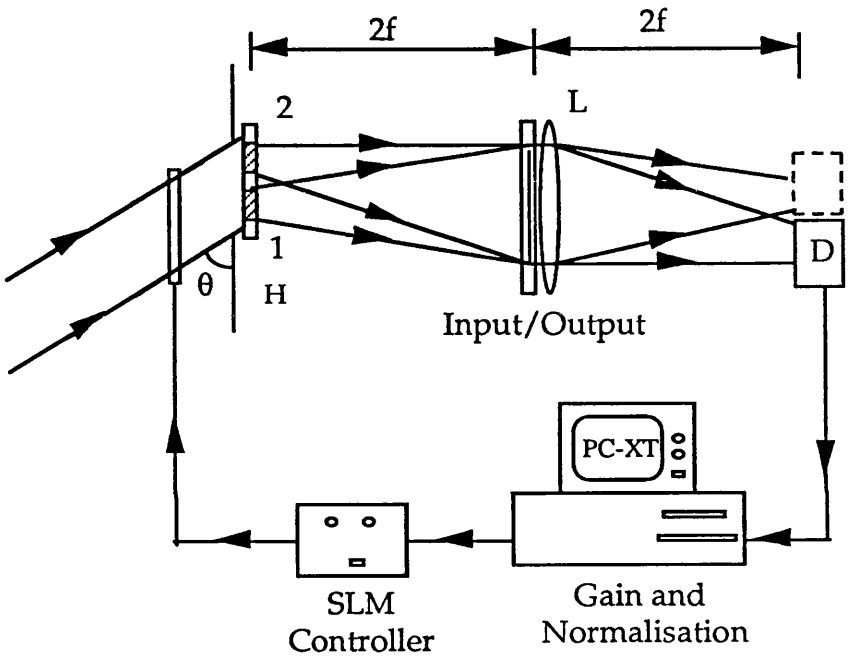


Fig. 6.8 Lensless Optical HOFNET System

Suppose $2f >> x, y, x', y'$, expand r_i as

$$r_i = 2f + \frac{(x_i'-x)^2 + (y_i'-y)^2}{4f} \qquad (6.23)$$

Then $A_1(x,y)$ can be written as

$$A_1(x,y) = c\sum_{i=1}^{M} o_i^*(x,y) \exp\left\{\frac{j\pi k}{2f}\left[(x_i'-x)^2 + (y_i'-y)^2\right]\right\} \times \times \exp\left[\frac{-j\pi k}{f}(x^2+y^2)\right] \qquad (6.24)$$

where $c = \frac{\exp(j4\pi kf)}{2f}$ is a constant.

At the image plane, we have the Fresnel diffraction of the amplitude $A_1(x,y)$

$$A_2(x_o,y_o) = c\frac{\exp(j4\pi kf)}{j2\lambda f}\sum_{i=1}^{M}\iint o_i^*(x,y)g(x,y)\exp\left\{\frac{j\pi k}{2f}\left[(x_i'-x)^2+(y_i'-y)^2\right]\right\}\times$$
$$\times\exp\left[\frac{-j\pi k}{f}(x^2+y^2)\right]\exp\left\{\frac{j\pi k}{2f}\left[(x_o-x)^2+(y_o-y)^2\right]\right\}dxdy$$

$$= c'\sum_{i=1}^{M}\iint o_i^*(x,y)g(x,y)\exp\left\{\frac{j\pi k}{f}\left[(x_i'+x_o)x+(y_i'+y_o)y\right]\right\}dxdy \tag{6.25}$$

where

$$c' = c\frac{\exp(j4\pi kf)}{j2\lambda f}\exp\left[\frac{j\pi k}{2f}(x_o^2+y_o^2)\right]\exp\left[\frac{j\pi k}{2f}(x'^2+y'^2)\right] \tag{6.26}$$

So at the position $x_o = -x_i'$, $y_o = -y_i'$, we have

$$A_2(x_o,y_o) = c'O_i\left(\frac{k}{2f}x_o,\frac{k}{2f}y_o\right)*G\left(\frac{k}{2f}x_o,\frac{k}{2f}y_o\right)\otimes\delta\left(\frac{k}{2f}x_o+\frac{k}{2f}x_i',\frac{k}{2f}y_o+\frac{k}{2f}y_i'\right) \tag{6.27}$$

where O_i and G are the Fourier transforms of o_i and g respectively. So the intensity at position $x_o = -x_i'$; $y_o = -y_i'$ indicates the similarity between the input pattern and the stored pattern located at position (x_i, y_i) in the recording plate (see section 3.4). If the patterns stored in the hologram are real patterns, then equation (6.27) becomes

$$A_2(x_o,y_o)\Big|_{\substack{x_o=-x_i'\\ y_o=-y_i'}} = \iint o_i(x,y)g(x,y)dxdy \tag{6.28}$$

So this system can be used in the HOFNET implementation to detect the similarity of the input pattern and the stored pattern. In condition to the focused system, we may also use this system for the implementation of the HOFNET. The feedback loop is the same as the one described in the previous section.

6.4 Feedback System

In our system, the non-linearity increases as the number of iterations increases. This can alleviate the requirements for large dynamic ranges of devices used in the system [Owe87]. For a system with non-linearity of order 'n', suppose the maximum correlation and minimum correlation are specified respectively as C_{max} and C_{min}, then the dynamic range of the device used in the system, must be at least

$$10n\log_{10}\left(\frac{C_{max}}{C_{min}}\right)(dB) \qquad (6.29)$$

Otherwise, the maximum correlation will be clipped or the minimum correlation will be undetectable. In our system, however, n is not a constant but increases gradually and we normalise the correlation peaks in each iteration, so the requirements of the dynamic ranges of devices will be softened. Suppose after the *k*-th iteration, the correlation peaks are $C(i)^{(k)}$, we normalise them according to the following formulae

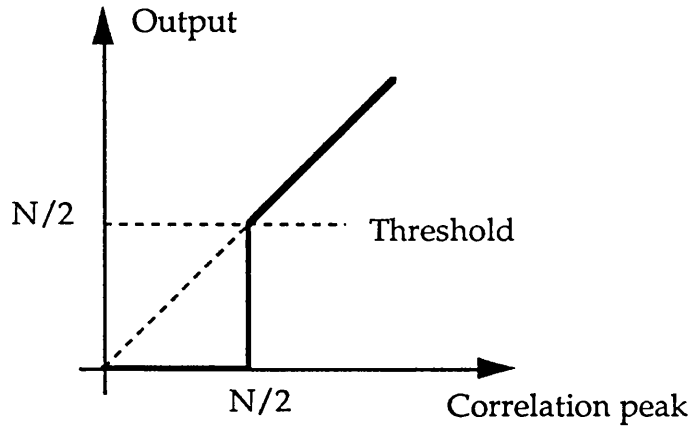


Fig. 6.9 Threshold Function for Correlation Peaks

$$C'(i)^{(k)} = \frac{C(i)^{(k)}}{\sqrt{\sum_{i=1}^{M} C^2(i)^{(k)}}} \times N \qquad (6.30)$$

where N is the total number of pixels of a pattern, and setting the threshold as (Fig. 6.9)

$$C'(i)^{(k)} = \begin{cases} C'(i)^{(k)} & (C'(i)^{(k)} \geq T) \\ 0 & (C'(i)^{(k)} < T) \end{cases} \qquad (6.31)$$

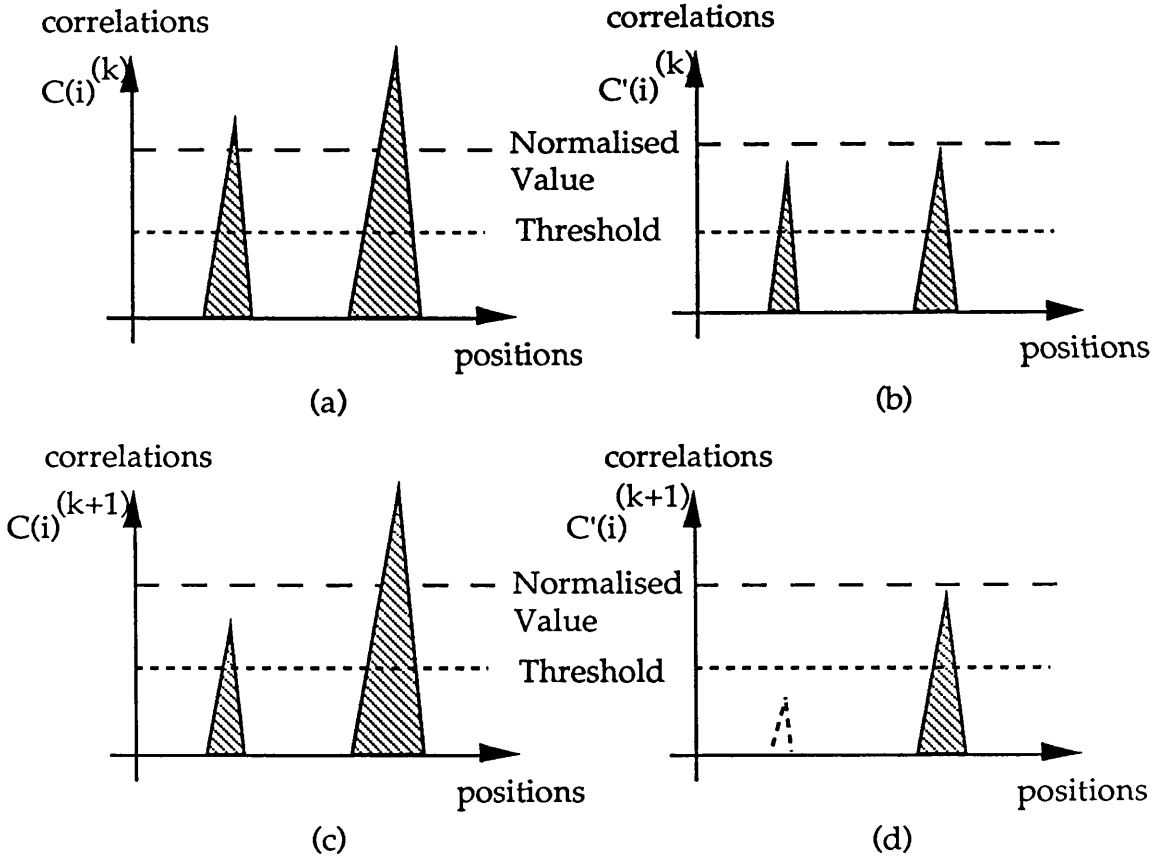


Fig. 6.10 Normalisation of the Correlation Peaks

The transmittance of each subsector of the SLM is adjusted according to Eq. (6.31). The whole normalisation procedure is shown in Fig. 6.10.

6.5 Experimental Results

Practically, it is difficult to realize uniform diffraction efficiency. Originally, we used ordinary slide films with plastic substrate as input transparencies to record holograms. The diffraction efficiencies were found to vary over a large range, from less than 4% to more than 14%. This was much larger than expected and far beyond the fluctuation range of errors allowed in the experiment. This phenomenon was found to be caused by a polarisation rotation effect caused by the birefringence of the plastic transparency substrate. When light passed through the slide films, the polarisation rotation angle varied across one pattern and from one pattern to another. When such films were replaced by glass holographic plates, much better results were obtained as the birefringence of the glass is negligible. As the efficiency also depends on exposure time, emulsion evenness, the form of the pattern itself, processing procedures as well as the modulation frequency, it is very difficult to keep them all exactly the same. In our preliminary experiment, 14 different patterns were stored with a variation in diffraction efficiency from 13% to 15%. The holograms were arranged in a 4x4 array with a space between two adjacent holograms of about 5mm, which matched the subsector spacing of the spatial light modulator (SLM). The diameter of each hologram circular spot was chosen randomly as 3mm. The variation of diffraction efficiencies is compensated by changing the transmittance of the SLM subsectors. The multiplication of the efficiency of a hologram element by the transmittance of the corresponding subsector is the same for all stored patterns before the system starts to iterate. Therefore if no input pattern is inserted in the correlation system, the outputs in the output plane are the same for all channels.

The optical HOFNET system is designed so that the same input pattern appears in each of the channels. This is achieved by allowing the channels to intersect at the input pattern while maintaining the individual identity of each channel by angular multiplexing. Such an arrangement has the advantage that noise in the input pattern or the system noise is exactly transferred into each channel so no discrepancies between channels due to non-uniform or noisy input devices such as SLM, can occur. After several iterations, the time-dependent noise will be averaged out. In our initial experiment, a single detector preceded by a

pinhole was moved to detect the zero order correlations. The electronic detector output was fed back via an IBM/XT computer to control the intensity of the light incident on each hologram independently by means of the liquid crystal SLM. For initial system assessment three orthogonal training patterns (Fig. 6.11(a)) were stored. Optically it is difficult to deal with negative amplitudes so all of the data was made non–negative and the negative value -1 is replaced by zero. In the orthogonal pattern set, if the corresponding elements of two patterns are the same, their product is +1 (fully correlated), otherwise it is -1 (uncorrelated). The algorithm is

$$1x1=1,\ (-1)\times(-1)=1;\ 1\times(-1)=-1,\ (-1)\times 1=-1$$

If, however, -1 is replaced by 0, then the above algorithm becomes

$$1\times 1=1,\ 0\times 0=0;\ 1\times 0=0,\ 0\times 1=0$$

Now if product 1 still means fully correlated, and product 0 means uncorrelated, the correlation of two zeros (fully correlated) cannot be detected. In order to solve this problem, each 16 pixel (4x4) binary pattern was reproduced twice, side by side, one copy being made the negative of the other, so the number of pixels is doubled. So in Fig. 6.11(a), the lower part of each pattern is the negative of the upper part. The algorithm for the multiplication of two pixels of separate patterns is

$$1x1+0x0=1,\ 0x0+1x1=1,\ 1x0+0x1=0,\ 0x1+1x0=0$$

where 1 and 0 represent on and off pixels respectively. Now the products indicate similarity (if the two pixels are the same (black or white), the product is 1, otherwise it is 0). So the actual patterns consisted of 8x4 pixels. When one quarter of an input pattern was obscured, the net converged to the complete correct pattern in 2 iterations (Fig. 6.11(b)).

Later we increased the number of stored patterns to 14 different patterns, each with 32 pixels (so the actual patterns had 64 pixels). Fig. 12(a) is the patterns to be stored in the hologram and Fig. 12(b) is the patterns reconstructed from the hologram. Because the hologram is recorded in the silver halide plate and bleached to increase the diffraction efficiency, the

noise is very high. Actually Fig. 12(b) is the patterns stored in the system. When an input is inserted into the net, the net could recognise the pattern in 2 or 3 iterations (Fig. 6.13). This input pattern is from Fig. 12(a). Although it is a 'perfect' pattern, compared to Fig. 12(b), which contains actual stored patterns, it is a distorted pattern. If we insert an input with about 1/8 obscured into the net, the net could recognise the pattern and restore the hidden part in 3 or 4 iterations, depending on the injected pattern and the system alignment (Fig. 6.14).

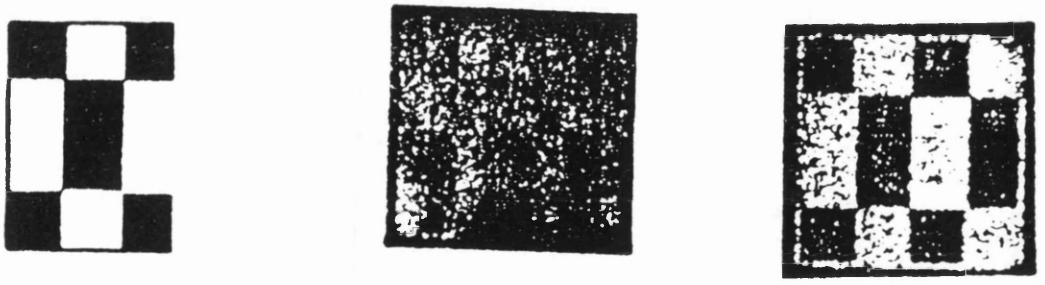
As shown in Fig. 12(b), the reconstructed patterns from the hologram are seriously noise distorted. These original patterns in the high contrast slide films are copied to a holographic plate in order to avoid the birefringence. The contrast of the holographic plate, however, is not very high (for the Agfa plate, γ is about 4) and also the imaging system used for copying is not in very high quality. These make the quality of the patterns in the holographic plate decrease: the contrast is low and the edges of the patterns are not sharp. When we store these patterns in a hologram, we use a silver halide plate to record the hologram and in order to increase the diffraction efficiency, the plate is bleached to become a phase hologram, which increases the noise greatly by random scattering. Therefore, if we copy the original patterns in the slide films to a very high contrast photographic plate and use phase recording materials, such as dichromated gelatin, we believe the quality of the patterns stored in the hologram can be increased greatly.

6.5 Conclusions

We have demonstrated the use of two dimensional arrays of Fourier Transform holograms in an optoelectronic implementation of a high order feedback net (HOFNET). The nonlinearity in this net is produced by means of feedback and not by optical material non-linearities and so the net is not limited by the switching speed, non-uniformity or availability of large arrays of optoelectronic nonlinear neurons. The patterns used in the system are amplitude modulated only, so the loss is quite high. If phase patterns or polarised encoded patterns are used, the loss will be reduced. Also the patterns are recorded in silver halide holographic plate and bleached to increase the diffraction efficiency, so the noise level is high. If we use photorefractive material or liquid crystal to record the hologram,

the diffractive efficiency can be increased and the loss level will be reduced.

(a) Patterns Stored in the System



Input Output before iteration Output after 2 iterations

(b) Pattern Recognition Procedure

Fig. 6.11 Experimental Result of Pattern Recognition (3 patterns stored)

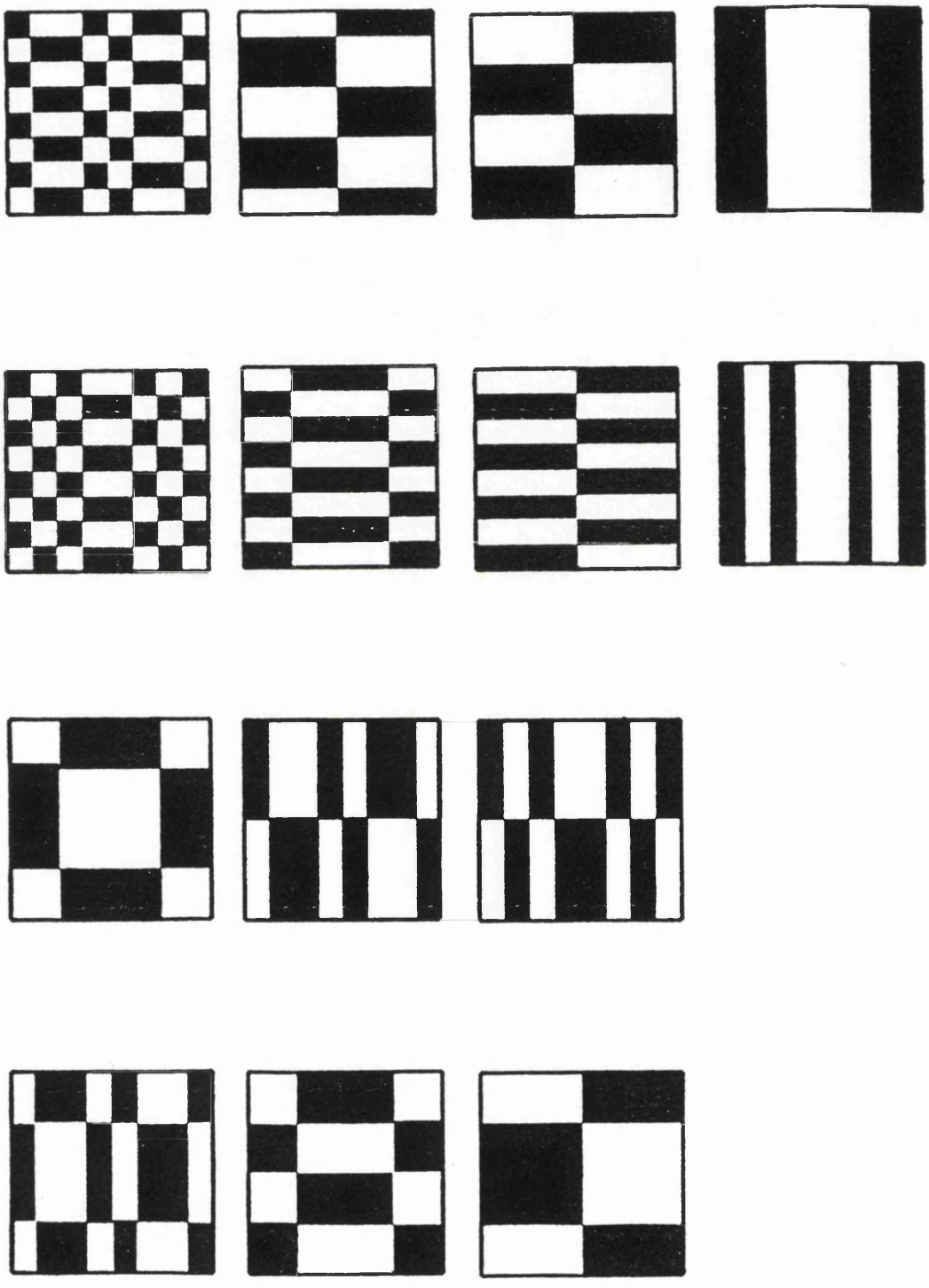


Fig. 6.12(a) Patterns to Be Stored in the System

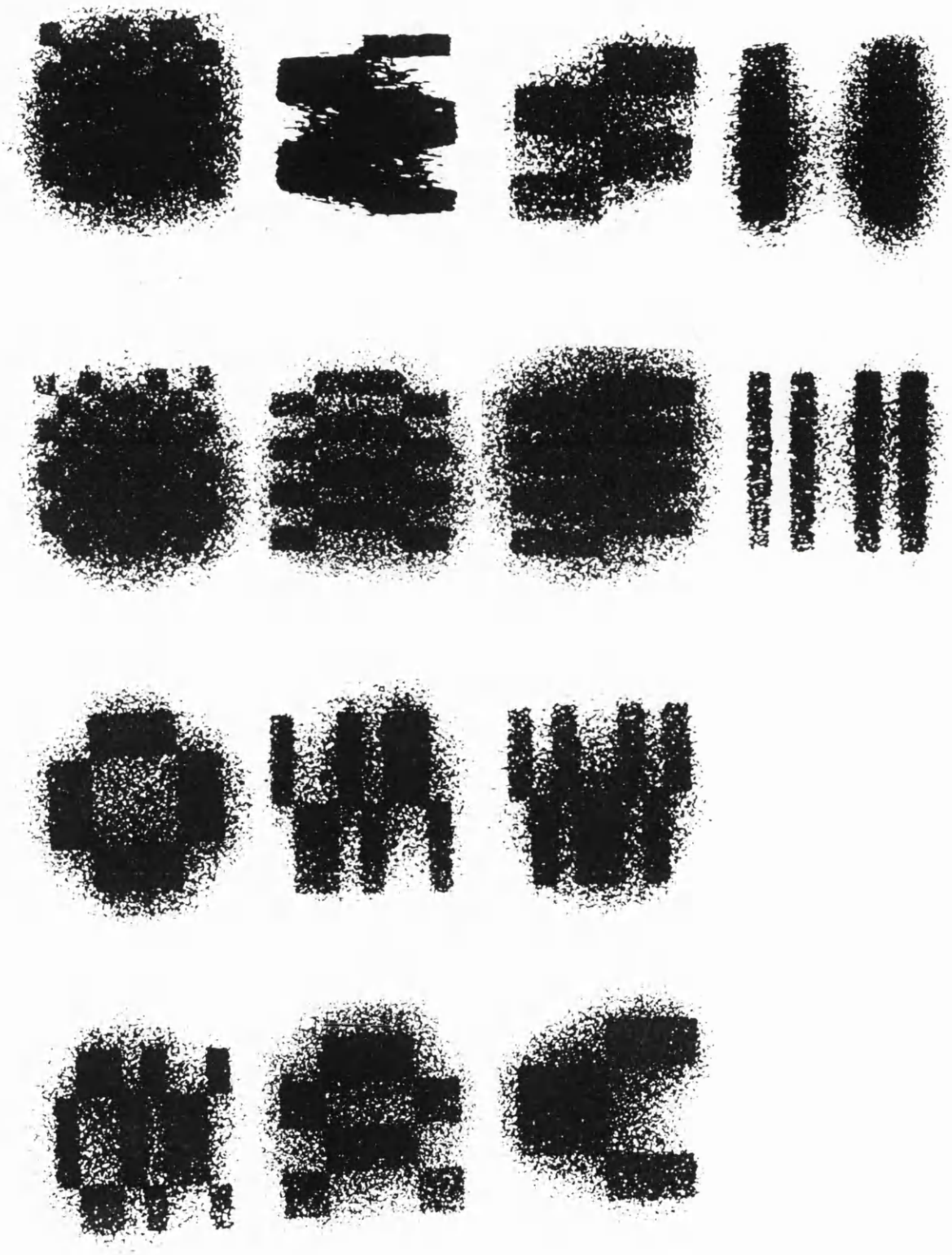
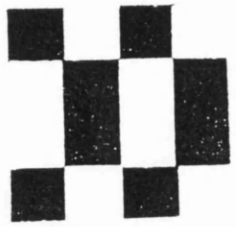
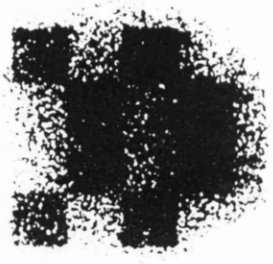


Fig. 6.12(b) Patterns Actually Stored in the System



Input Pattern

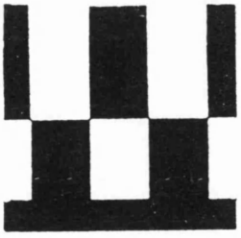


Actually Stored Pattern

Output Before Iteration

Output After 3 Iterations

Fig. 6. 13 Pattern Recognition with Full Input Pattern



Input Pattern

Actually Stored Pattern

Output Before Iteration

Output After 1st Iteration

Output After 2nd Iteration

Output After 3rd Iteration

Fig. 6.14 Pattern Recognition with Partial Input Pattern

Chapter Seven

DESIGN OF HOFNET SYSTEM WITH OPTICAL PARALLEL FEEDBACK LOOP

In the above chapter, we described a HOFNET system using an electrically addressed spatial light modulator in the feedback loop. In that system, the normalisation of the correlation peaks is performed by a computer. Therefore, it is not a completely optical system. Here we introduce two completely optical systems by using opto-electronic devices. In section 7.2, we first describe the devices that are used in the design of HOFNET systems: optically addressed spatial light modulators (OASLM), self-electrooptic effect devices (SEEDs), especially D-SEED devices and optical fibre amplifiers (OFAs). Then an optical HOFNET system, in which gain is supplied by an OASLM and normalisation is performed by a SEED device is discussed in details in section 7.3. Section 7.4 describes another HOFNET system using an optical fibre amplifier in the feedback loop, which performs both gain and normalisation. Summary is given at the end.

7.1 Optoelectronic Devices

7.1.1 Optically addressed spatial light modulator (OASLM)

An OASLM uses either photoconductors or photocathodes as the charge-generation element and liquid crystals, electrooptic crystals or magneto-optic materials as its light-modulating elements [Chi90; Sel89]. The fundamental structure of an OASLM is shown in Fig. 7.1. It is composed of the photoconductor (recently amorphous silicon or hydrogenated amorphous silicon are common), the modulation medium (for example, nematic or ferroelectric smectic liquid crystal) and the light blocking layer/mirror, which are sandwiched between two transparent conducting

electrodes (indium tin oxide or ITO). The light blocking layer acts to separate the read and write beams, reducing the effect of the read beam on the photoconductor. Information is transferred from one beam of light called the write light to another light beam called the read light. The read light may have different coherence, wavelength or divergence. As it can be seen in Fig. 7.1, the photoconductor, light blocking layer and liquid crystal are in series between the two electrodes. The resistance of the light blocking layer and liquid crystal are roughly constant. When a write light falls on the photoconductor, the resistance of the photoconductor changes, so the voltage falling on the liquid crystal will change inversely if the a.c. voltage across the two electrodes is kept constant. The response of the liquid crystal to an applied voltage is for the reorientation of the molecules to occur, consequently, changes the polarisation of the read light. So the polarisation of the read light also indicates the voltage change on the liquid crystal or on the photoconductor, therefore indicates the intensity change of the write light. If we put a polariser in front of the read light, the change of polarisation of the read light becomes the intensity modulation. So the read light is modulated by the write light.

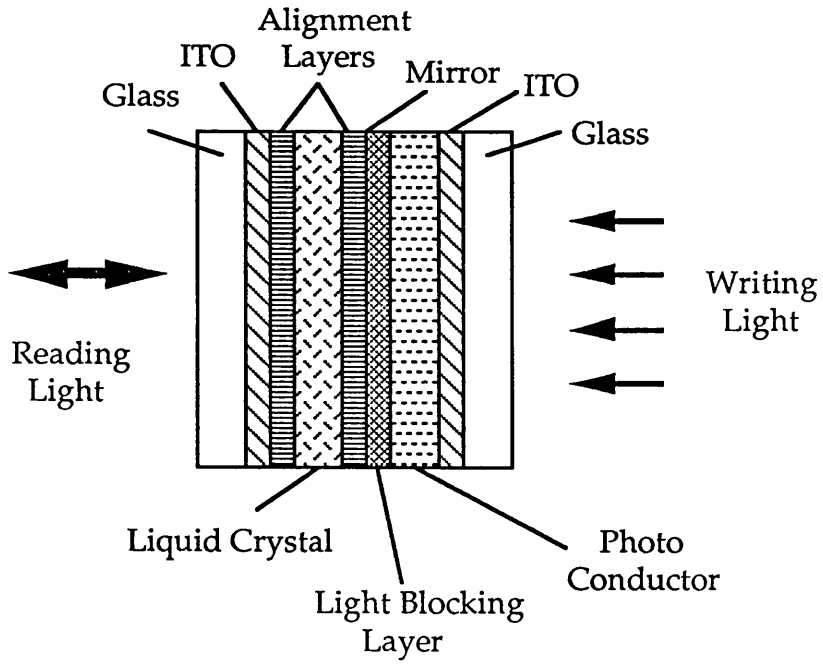


Fig. 7.1 Fundmental Structure of an OASLM

Most devices of this type reported to date use a rubbed polymer alignment for the liquid crystal which has the disadvantages of high levels of defects (leading to a low contrast ratio) and/or poor bistability. Crossland et al [Cro90] introduced a proprietary silicon oxide evaporated alignment to induce a quasi-bookstack geometry which gave a device with good bistability. Most OASLMs used in reflection employ light blocking layers to prevent the read light from washing out the image by reducing the conductivity of the photosensor. This, however, is unnecessary in a bistable device since it ceases to be photosensitive when the drive voltage is removed, but retains the stored information for interrogation. The read beam is modulated so that no read light falls on the device while the image is being written. Once an image has been written, a read beam of any reasonable intensity can be used without degrading the stored image.

7.1.2 Multiple quantum well and D-SEED device

Multiple quantum well (MQW) structures consist of alternate thin layers of GaAs and AlGaAs or some other materials with different band gaps. One property of the MQW structure is the existence of clearly-resolved exciton absorption peaks near the optical absorption edge at room temperature. An MQW p-i-n diode layer structure is shown in Fig. 7.2 [Mil84, 85]. When an reverse electric field is applied perpendicular to the quantum well layers, the whole absorption edge moves to lower photo energies and the excitons remain resolved to high fields. This effect in the MQW structures can be explained through a mechanism called the Quantum-Confined Stark Effect (QCSE) [Mil85]. Because the QCSE gives a shift of a large absorption, modulators can be made with only microns of materials. Such structures can also function as optical detectors at the same time as modulators. This simultaneous operation is the basis of the self-electrooptic device (SEED). The behaviour of the SEED depends greatly on the nature of the electric circuit and on the sign of the feedback. Here we discuss a D-SEED system which we will use in our HOFNET system design.

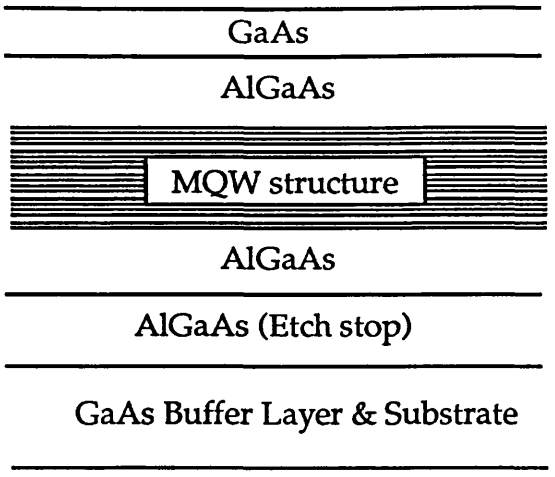


Fig. 7.2 Multiple Quantum Well Structure

The D-SEED system is shown in Fig. 7.3, where a photodiode detector is included in the circuit in series with the MQW p-i-n diode. When a light with intensity I_1 falls on the detector, it produces a photocurrent i_D, which is almost independent of reverse bias supply voltage and is linearly proportional to the amount of light shining on it I_1. If another laser beam I_2 shines on the MQW p-i-n structure, the absorption in the MQW generates count-flowing current i_{MQW}. When i_{MQW} is less than i_D, the action of the total current is to charge the capacitance of the MQW structure, the voltage across the MQW starts to rise, which in turn causes i_{MQW} to rise. Conversely, if i_{MQW} exceeds i_D, the voltage starts to fall, which in turn makes the i_{MQW} decrease. So in the end an equibrium will be reached when $i_D = i_{MQW}$. Because the absorbed power by the MQW is proportional to i_{MQW}, there is a linear relation between the light I_1 and the light I_2. The output of the MQW p-i-n diode I_3 can be written as

$$I_3 = I_2 - kI_1 \qquad (7.1)$$

where constant k depends on the electric circuit and the quantum efficiency of the photodiode detector and MQW structure. This idea was first proposed by Professor John Midwinter for image subtraction and later was found by our discussion to be useful for implementing normalization in the optical HOFNET system.

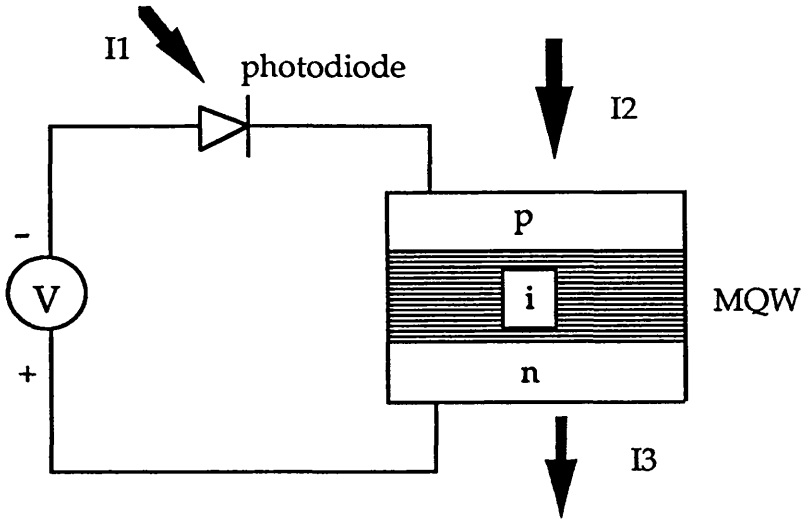


Fig. 7.3 Schematic Diagram for the D-SEED Configuration

7.1.3 Optical fibre amplifier

Recently there is considerable interest in using rare earth doped fibres to amplify weak signals. In an erbium doped fibre amplifier (EDFA), the gain more than 35 dB at various wavelengths has been obtained [Lam89; Urq88]. Erbium doped silica fibres can be made to lase at an important telecommunications window, and so the greatest interest has been at about 1500 to 1600 nm. Pumping is possible for the pump wavelengths: 532, 800, 980, 1500 nm etc. It might be thought of as an advantage to use semiconductor lasers at about 800 nm to pump the fibre amplifier. But for Er^{+} doped fibre amplifiers, the excited-state absorption (ESA) of the pump light impedes severely the use of high power semiconductor laser [Lam89]. The ESA transitions have the effect of depleting the pump light resulting in a reduction of pumping efficiency and hence gain. The other pump bands at 670 nm and 532 nm, although exhibiting reduced ESA, are also impeded by the large Stokes shift between pump and signal wavelengths, which results in reduced efficiency. So the better pump wavelengths are 980 nm and 1490 nm. Gain efficiencies of 3.9 dB/mW and 2.1 dB/mW have been demonstrated for 978 nm diode laser and 1493 nm F-centre laser pumped EDFA. As an example, an experimental set up for a 978 nm diode

laser pumped amplifier is shown in Fig. 7.4 [Lam89].

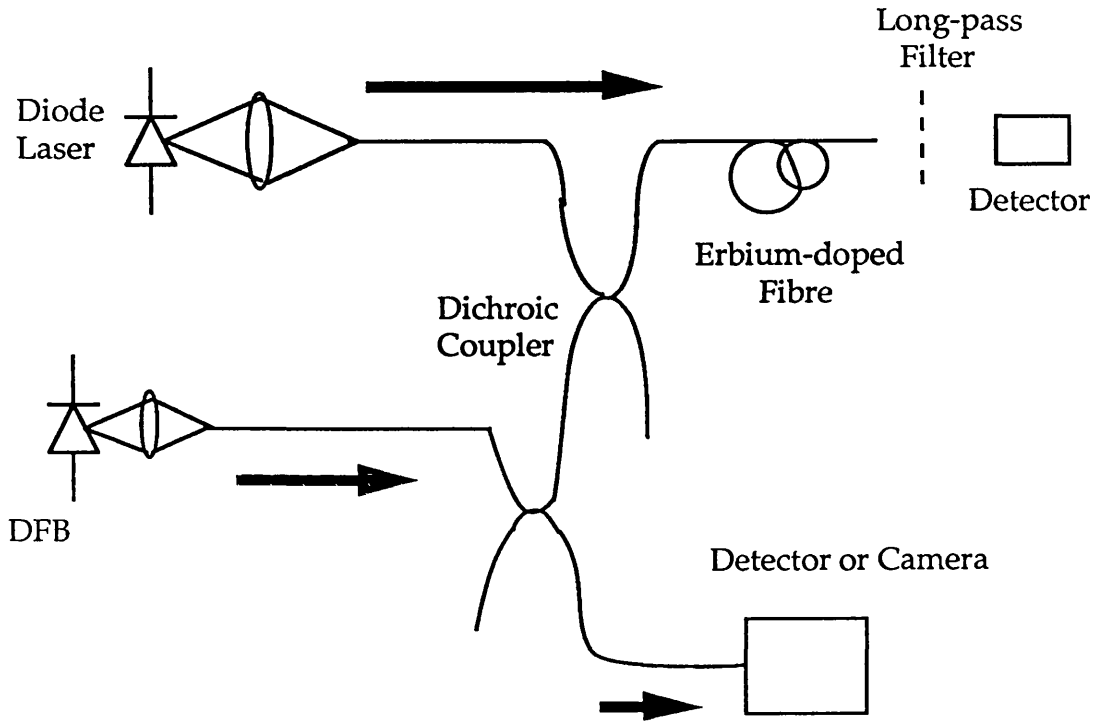


Fig. 7.4 Experimental Setup of Optical Fibre Amplifier

The output from the strained quantum well diode pump laser was launched into the wedge tipped fibre. The pump light was multiplexed with the signal light from a DFB laser operating at 1535 nm via a dichroic coupler and launched into the amplifier fibre. Usually for a fixed pump power, the gain drops as the input signal increases. An example is shown in Fig. 7.5. Here a 20 mW pump laser with wavelength 978 nm is used to pump the signal with wavelength 1535 nm from the DFB laser. A gain as high as 40 dB can be obtained. The gain is usually reduced for longer pump wavelengths and increasing the pump power reduces the gain sensitivity to pump wavelength and increases the gain [Lam89].

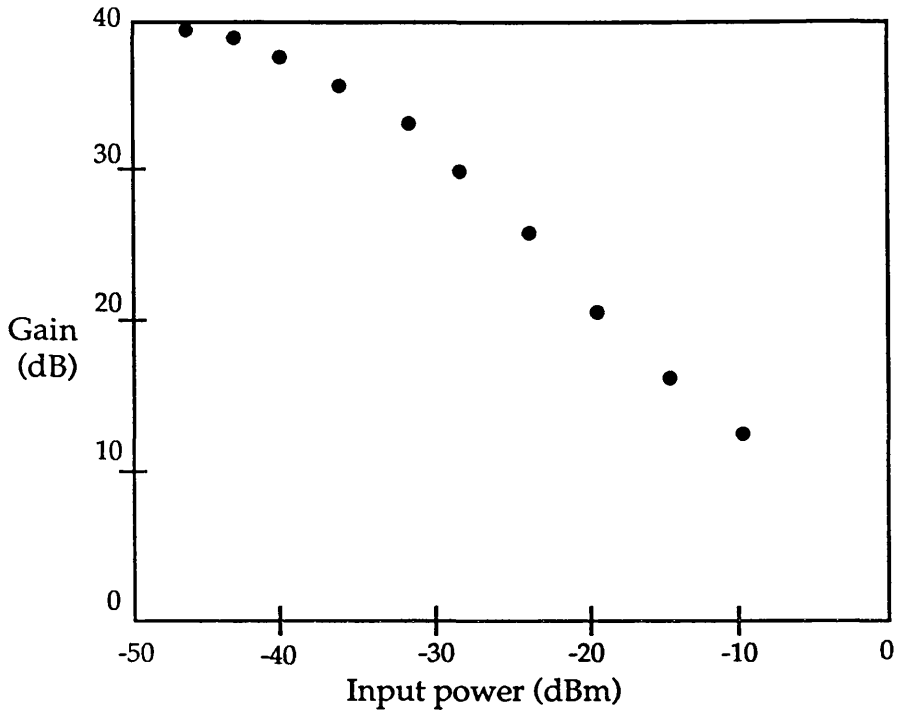


Fig. 7.5 Gain vs Input Signal for Optical Fibre Amplifier

7.2 HOFNET Design by Using OASLM and D-SEED Devices

A HOFNET system consists of two main subsystems: one is called correlation system and the other feedback system. Correlation system is composed of passive optical elements, such as lens, mirrors, which inevitably result in energy loss. Also the HOFNET is a non-linear system which requires normalisation to the correlation peaks at each iteration in order to avoid saturation. So the feedback system must perform two tasks: supplying gain and doing normalisation. Fig. 7.6 is a real optical HOFNET system, where the gain is provided by the optically addressed spatial light modulator (OASLM) and the normalisation is performed by the D-SEED device (system).

A parallel beam from a laser is modulated by the multiple quantum well (MQW) modulator and then falls on the OASLM via a polarised beam splitter. The reflected beam, which is modulated by the feedback

from the correlation output, reconstructs the hologram H. At the back focal plane of lens L_1, the reconstructed patterns stored in the hologram are multiplied by the input pattern, so at the back focal plane of lens L_2, we get the correlation output which is divided into two parts by another beam splitter: one is fed back by an imaging system to illuminate the OASLM as the write beam, the other is collected by a large photodiode detector to control the MQW modulator.

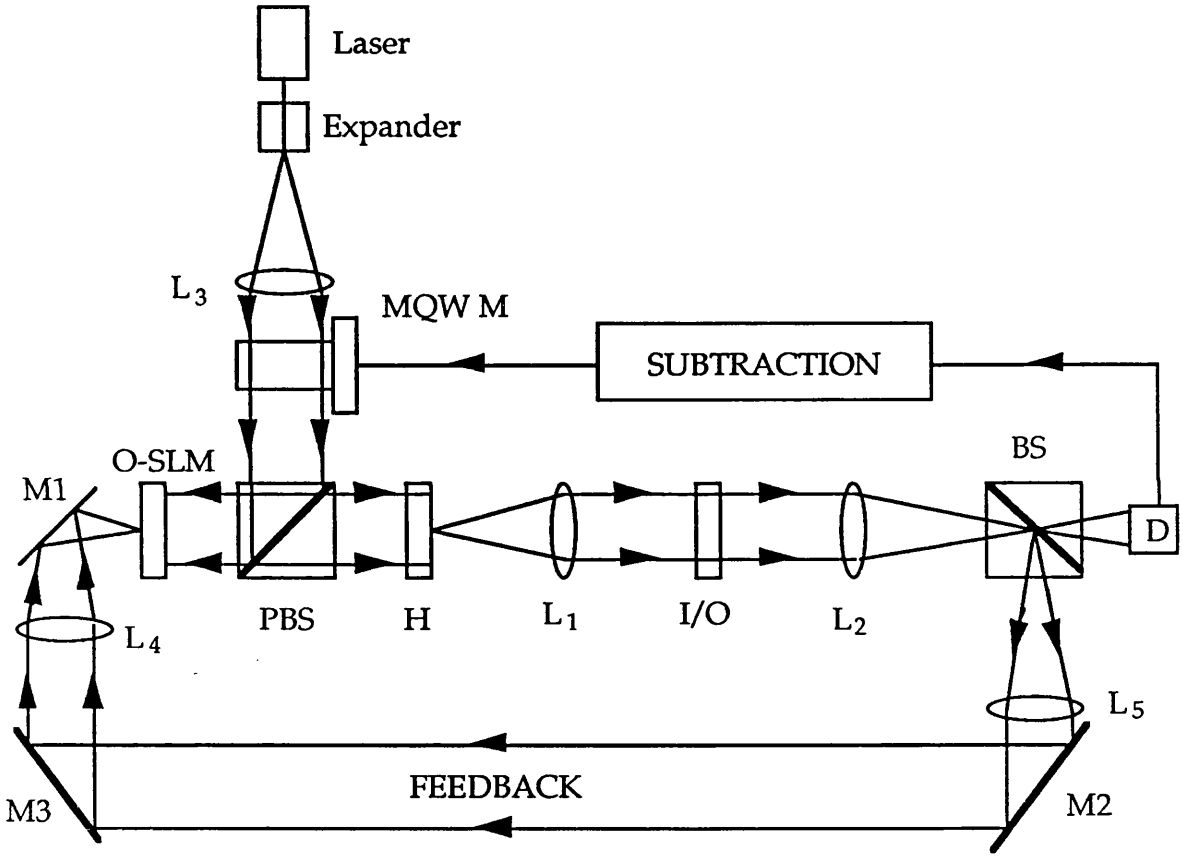


Fig. 7.6 Real Optical HOFNET System

Now we describe the operation of this system analytically to prove that the gain in the channel with the largest correlation is higher than that in the other channels. Suppose the collimated light after lens L_3 has the intensity I_o, at the beginning no feedback signal is from the photodiode detector, so the transmitted intensity is still I_0 (for simplicity the absorption and scattering are not considered). Before the iteration, we use I_0 to replay

the hologram H. At the correlation plane, we have the correlation outputs I_i (i = 1, 2,..., M), which indicate the similarity between the input pattern and the corresponding stored patterns. As passive elements are used in the system, we have

$$I_o > I_i \qquad (i = 1, 2, \ldots, M) \tag{7.2}$$

Suppose the beam splitter is half divided, then 0.5 I_i is fed back to control the OASLM, which we suppose has a gain of G. The other half of I_i are collected by the photodiode detector to control the MQW modulator. Let the efficiency of the photodiode be η, then after the first iteration, the light after the modulator becomes $\left(I_o - \frac{1}{2}\eta\sum_{i=1}^{M} I_i\right)$, which is modulated by the signals from the feedback via the OASLM. Now the light to replay the hologram becomes

$$I_i' = \left(I_o - \frac{1}{2}\eta\sum_{i=1}^{M} I_i\right) G\frac{1}{2} I_i \tag{7.3}$$

At the back focal plane of lens L_2, we have

$$I_i^{(1)} = \left(I_o - \frac{1}{2}\eta\sum_{i=1}^{M} I_i\right) G\frac{1}{2} I_i I_i \tag{7.4}$$

Similarly, after the second iteration, at the back focal plane of lens L_2, we get

$$I_i^{(2)} = \left(I_o - \frac{1}{2}\eta\sum_{i=1}^{M} I_i^{(1)}\right) G\frac{1}{2} I_i^{(1)} I_i \tag{7.5}$$

. . . . . .

After n-th iteration, we have

$$I_i^{(n)} = \left(I_o - \frac{1}{2}\eta\sum_{i=1}^{M} I_i^{(n-1)}\right) G\frac{1}{2} I_i^{(n-1)} I_i$$

$$= \left(I_o - \frac{1}{2}\eta \sum_{i=1}^{M} I_i^{(n-1)} \right) G \frac{1}{2} \left(I_o - \frac{1}{2}\eta \sum_{i=1}^{M} I_i^{(n-2)} \right) G \frac{1}{2} I_i^{(n-2)} I_i^2$$

$$= \cdots\cdots\cdots$$

$$= \left(G\frac{1}{2} \right)^n I_i^{n+1} \left(I_o - \frac{1}{2}\eta \sum_{i=1}^{M} I_i^{(n-1)} \right) \left(I_o - \frac{1}{2}\eta \sum_{i=1}^{M} I_i^{(n-2)} \right) \cdots$$

$$\left(I_o - \frac{1}{2}\eta \sum_{i=1}^{M} I_i^{(1)} \right) \left(I_o - \frac{1}{2}\eta \sum_{i=1}^{M} I_i \right) \tag{7.6}$$

Suppose $I_\alpha > I_i$ $(i \neq \alpha)$

$$\frac{I_\alpha^{(n)} / I_i^{(n)}}{I_\alpha^{(n-1)} / I_i^{(n-1)}} = \frac{I_\alpha^{n+1} / I_i^{n+1}}{I_\alpha^{n} / I_i^{n}} = \frac{I_\alpha}{I_i} > 1 \tag{7.7}$$

That is

$$\frac{I_\alpha^{(n)}}{I_i^{(n)}} > \frac{I_\alpha^{(n-1)}}{I_i^{(n-1)}} \tag{7.8}$$

or

$$\frac{I_\alpha^{(n)}}{I_\alpha^{(n-1)}} > \frac{I_i^{(n)}}{I_i^{(n-1)}} \tag{7.9}$$

Equation (7.8) means the ratio of the intensities between the maximum and the others at a certain time is always larger than that at the previous time. So the difference between the largest correlation and the others will become larger and larger. Another expression of Equation (7.9) says that the gain in the maximum correlation channel is always larger than the others.

The above discussion only proves that the system will always

converge into the maximum correlation, while all the other correlation outputs will eventually drop below the threshold. But we have not discussed how the system converge into the correlation output. Does the convergence occurs smoothly or with oscillation? As an example, we consider a HOFNET system with four channels (M = 4). Let the quantum efficiency of the SEED device be 50% ($\eta = 50\%$) and at the beginning $I_o = 1$, the correlation outputs before iteration, as an example are 0.80, 0.85, 0.90, 0.95 respectively (because passive elements are used in the correlation system, I_i must be less than I_o). So output intensity of a channel after n–th iteration, $I_i^{(n)}$ can be written as according to Eq. (7.6)

$$I_i^{(n)} = \left(1 - \frac{1}{4}\sum_{j=1}^{4} I_j^{(n-1)}\right)\frac{G}{2} I_i^{(n-1)} I_i \qquad (7.10)$$

We draw two sets of curves with different gain G as 4 and 7 respectively (Fig. 7.7 and Fig. 7.8). From the curves we can see when G = 4, the correlation outputs change smoothly. When G = 7, however, the correlation outputs change with oscillations, although the maximum can be selected eventually. Now we discuss the condition that the system can converge smoothly.

If the gain supplied by the OASLM is large enough to compensate for the loss in the correlation system, the system can always select the maximum output and suppress the others. We suppose after $n = N_o$, only the maximum remains to iterate, so equation (7.10) becomes

$$I_i^{(n)} = \left(1 - \frac{1}{4} I_j^{(n-1)}\right)\frac{G}{2} I_i^{(n-1)} I_i \quad (n>N0) \qquad (7.11)$$

or

$$I_i^{(n)} = \frac{G}{2} I_i - \left(2 - I_i^{(n-1)}\right)^2 \frac{G}{8} I_i \qquad (7.12)$$

As intensity must be more than zero. From Eq. (7.11) we have $I_j^{(n)} < 4$, while from Eq. (7.12) we know the maximum of $I_j^{(n)}$ is $0.5GI_j$, therefore we have

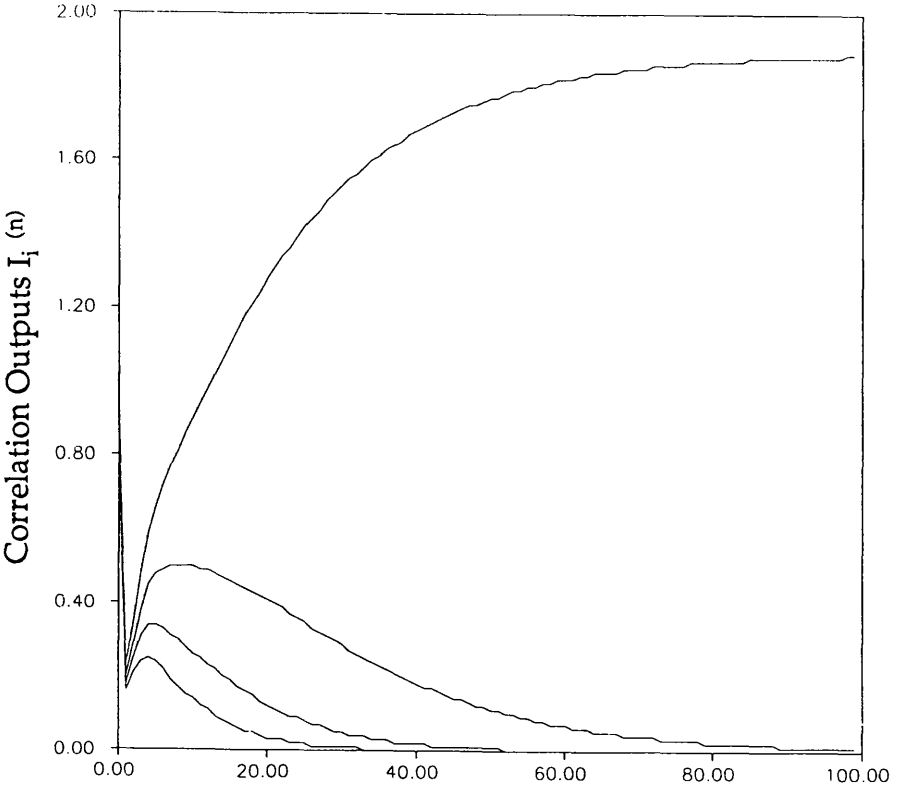


Fig. 7.7 Convergence Procedure with Gain G=4

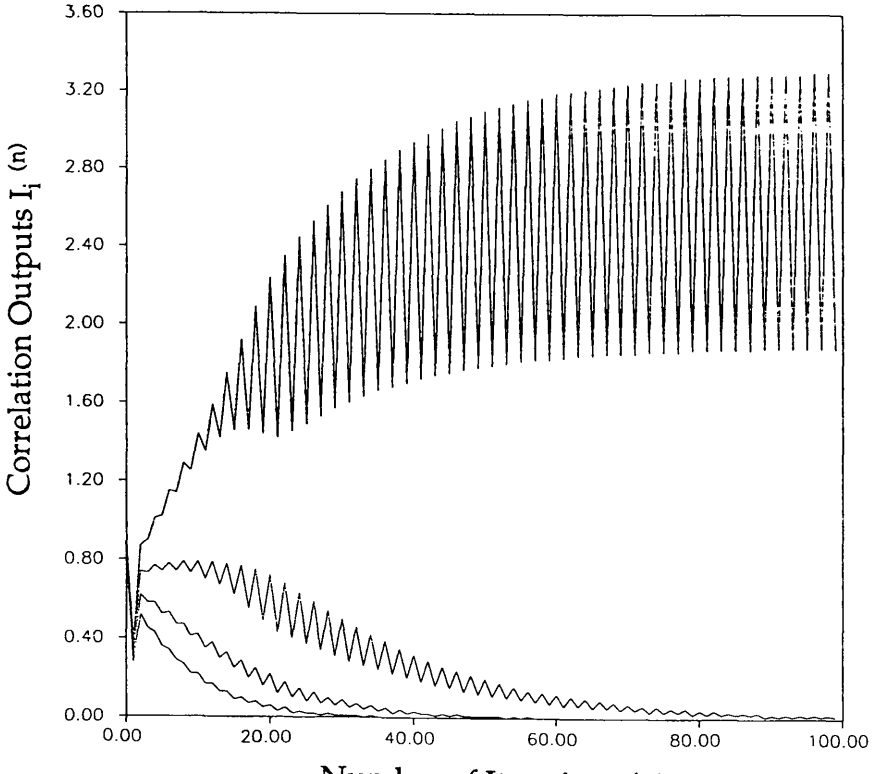


Fig. 7.8 Convergence Curve with Gain G=7

$$G < 8/I_j \tag{7.13}$$

Fig. 7.9 is the curve of equation (7.12) with $I_i^{(n-1)}$ as the x-axis, and the maximum value of $I_i^{(n)}$ is $GI_i/2$. Similarly after one more iteration, we have

$$I_i^{(n+1)} = \frac{G}{2} I_i - \left(2 - I_i^{(n)}\right)^2 \frac{G}{8} I_i \tag{7.14}$$

Draw equation (7.14) in Fig. 7.10, with $I_i^{(n+1)}$ as the x-axis. Apparently, if the system is stable, there must be

$$I_i^{(n+1)} = I_i^{(n)} = I_i^{(n-1)} \tag{7.15}$$

and superimpose them together, we will see:

(1) If $GI_i/2$ is very large, then the two curves have three intersection points, A, B, C, but only point B satisfies the equation (7.15) (Fig. 7.11);

(2) If $GI_i/2$ decreases gradually until only one intersection point B exists with equation (7.14) satisfied (Fig. 7.12);

(3) If $GI_i/2$ is very small, however, the intersection point B corresponds to minus intensity, which has no physical meaning (Fig.7.13).

Therefore, from the above discussion, there must exist a range of $GI_i/2$, within which, the two curves have only one point in the first area. From the mathematical analysis (see Appendix), we could find that if

$$1 < GI_i/2 < 3 \tag{7.16}$$

Only one positive point exists, that means the system converges gradually (stably). So the gain range is

$$\frac{2}{I_i} < G < \frac{6}{I_i} \tag{7.17}$$

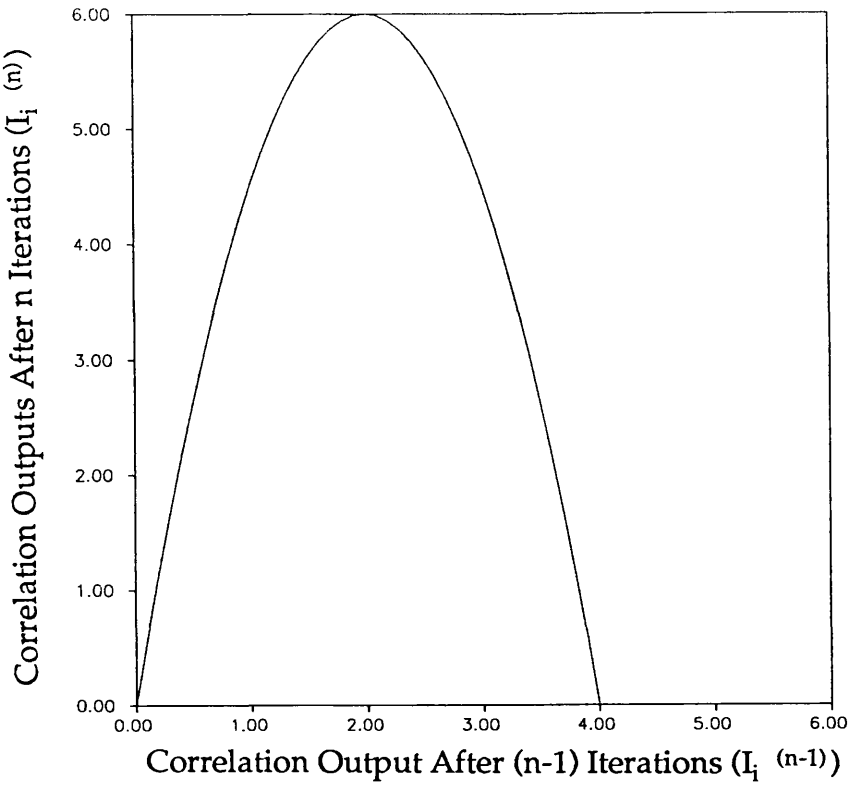


Fig. 7.9 Relation of Two Consequent Correlation Outputs

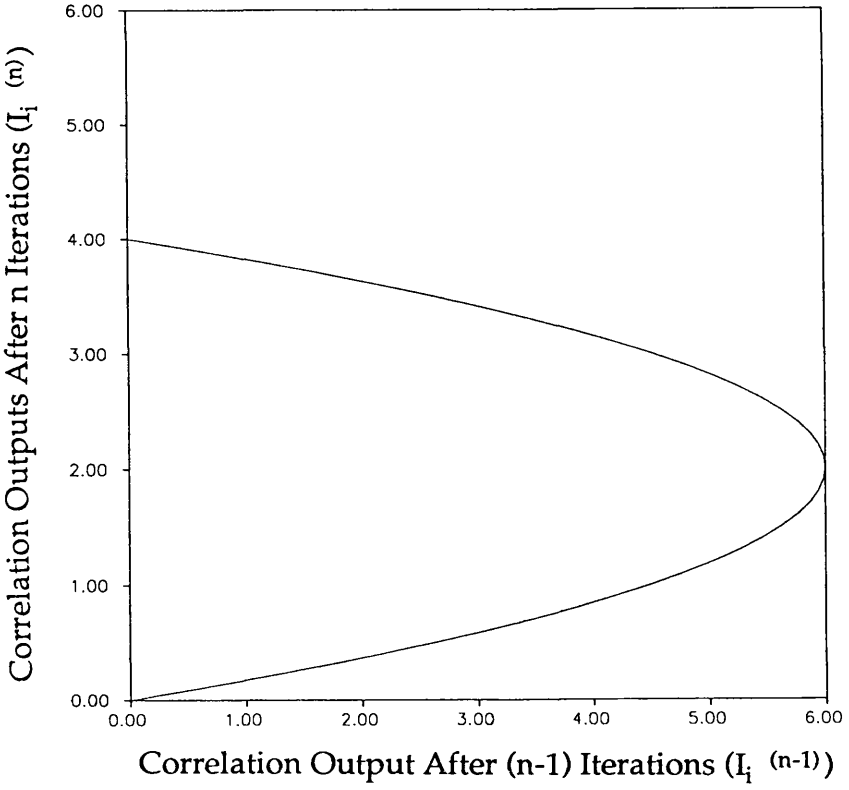


Fig. 7.10 Relation of Two Consequent Correlation Outputs

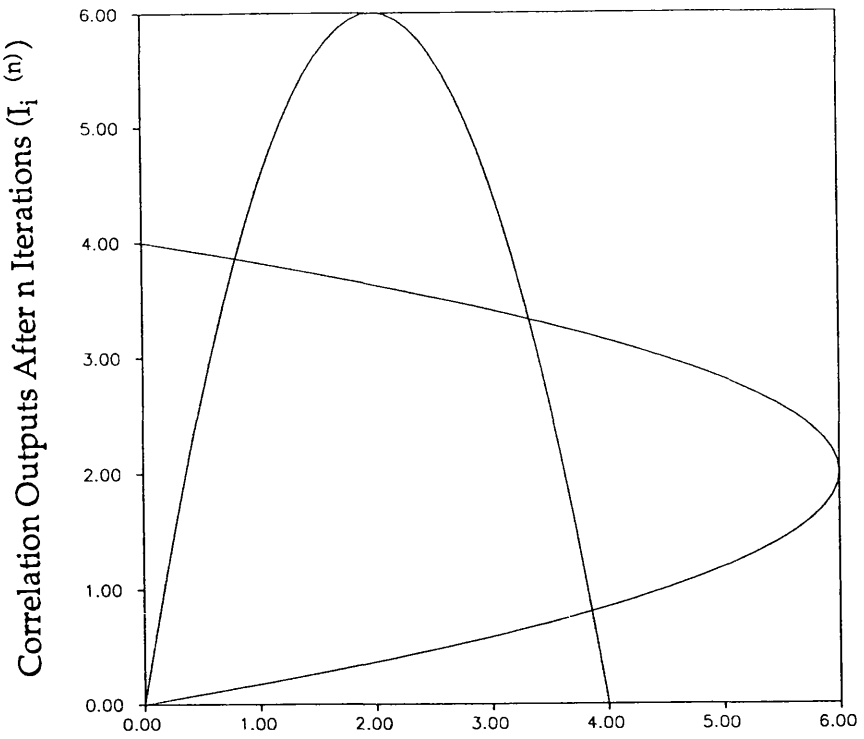


Fig. 7.11 Relation of Three Consequent Correlation Outputs with Large Gain

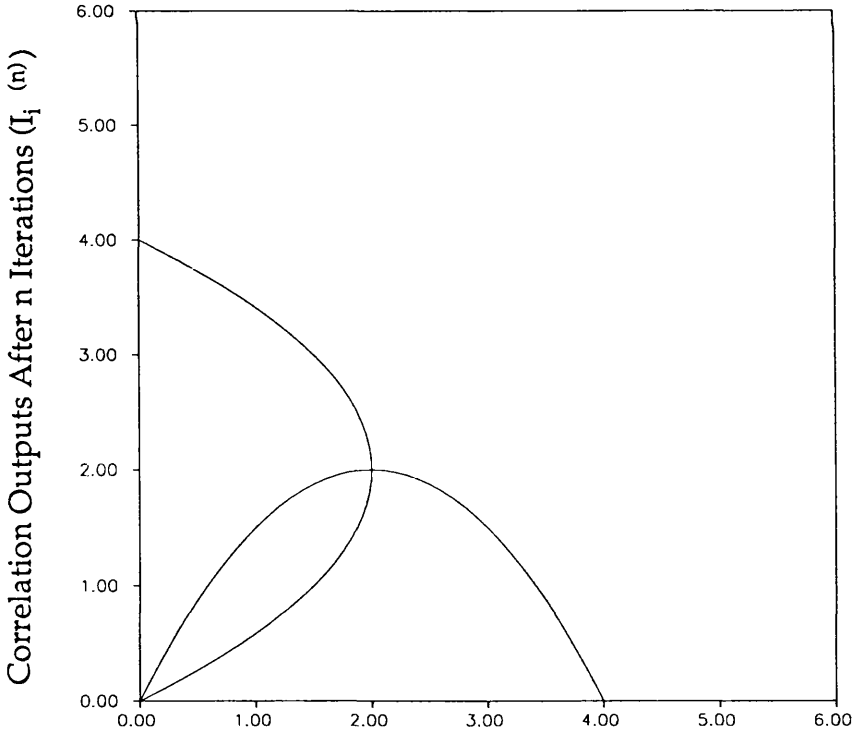


Fig. 12 Relation of Three Consequent Correlation Outputs with Medium Gain

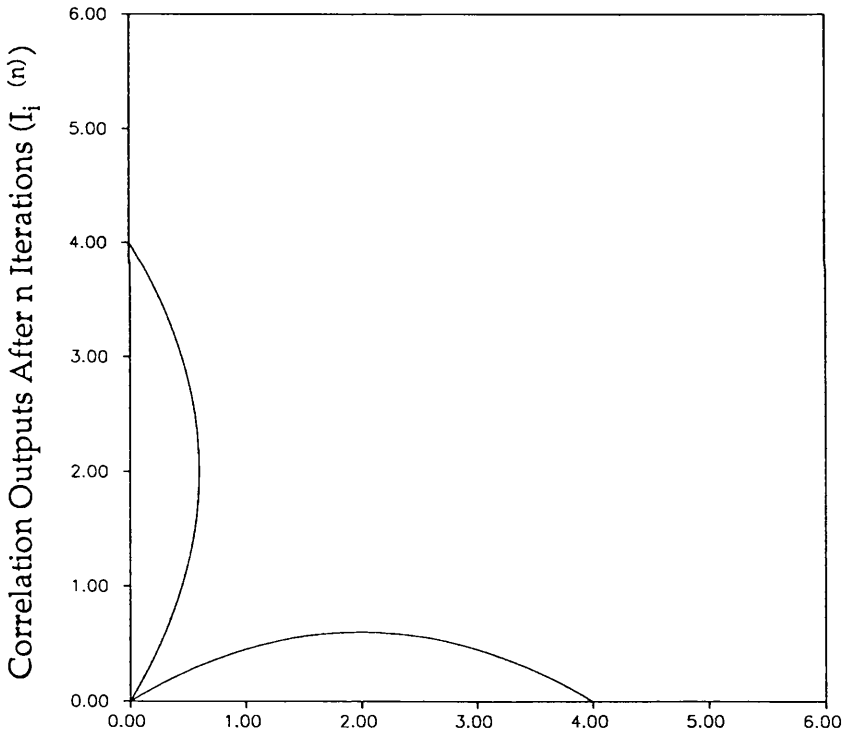


Fig. 7.13 Relation of Three Consequent Correlation Outputs with Small Gain

If, for example, the I_i are between 0.6 and 1.0, then if the gain G is between 3.2 and 6, the system will converge stably. That is why when G = 4, the system converges smoothly, while when G = 7 (less than 8, see Eq. (7.13)), the system converges oscillatedly (Fig. 7.8).

7.3 HOFNET Design by Using Optical Fibre Amplifier

Research on rare earth doped fibre amplifiers has recently received growing attention in the technical literature. The whole HOFNET system with rare earth doped fibre amplifier in the feedback loop is shown in Fig. 7.14. The fibre is transversely coupled to the pump so that the output signal from the correlation system experiences an amplification. In order to compensate for the loss in the correlation, the fibre amplifier must have at least a gain which is large enough to compensate for loss. But if the gain is constant, as the loss in the correlation system varies at different times, all the correlation peaks will be amplified to be saturated or diminished if the gain is not enough. Thus the maximum correlation cannot be selected. But the optical fibre amplifier has a gain which depends on the input signal. If the signal is weak, the gain is usually large, as the input signal increases, the gain will drop gradually as shown in Fig. 7.5. This feature can be used in our HOFNET system to supply both gain and normalisation. Now we discuss how this feature works in our system.

Now we prove that after a certain number of iterations, the signal in one channel which corresponds to the largest correlation will increase, while the signals in all the other channels will decay. Suppose we couple all the correlation peaks into an optical fibre amplifier in some encoding method, such as wavelength coding or mode coding. In order to explain more clearly, we suppose only two correlation peaks to be coupled into the amplifier and we specify them respectively as A_1 and A_2 and $A_1 > A_2$. At the beginning the total intensity is

$$I_1 = A_1 + A_2 \tag{7.18}$$

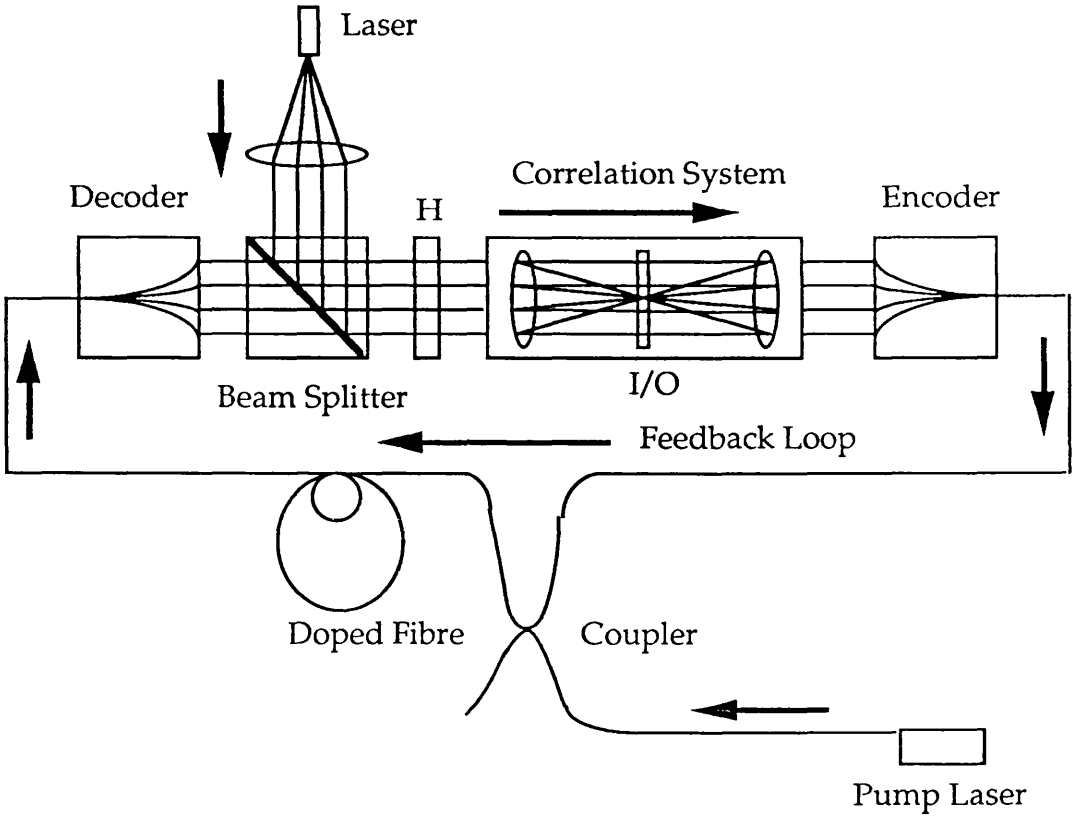


Fig.7.14 HOFNET System with Optical Fibre Amplifier

Now I_1 is coupled into the amplifier and amplified to be $g(I_1)(A_1+A_2)$, where $g(I_1)$ is the gain at the input intensity of I_1. Then the amplified signal is decoupled and fed back to start the second iteration. At the correlation plane, we have

$$I_2 = g(I_1)(A_1^2+A_2^2) \tag{7.19}$$

The output after the fibre amplifier becomes

$$g(I_1)\,g(I_2)(A_1^2+A_2^2) \tag{7.20}$$

Similarly, after the n-th iteration, the output becomes

$$g(I_1)\,g(I_2)\ldots g(I_n)(A_1^n+A_2^n) \tag{7.21}$$

Suppose after n-th iteration, the system will be stable (such an n always exists), then the gain obtained from the fibre amplifier would compensate for the loss in the correlation system and the output becomes constant. Therefore, we have

$$g(I_1)...g(I_n)\left(A_1^n + A_2^n\right) = g(I_1)...g(I_n)g(I_{n+1})\left(A_1^{n+1} + A_2^{n+1}\right) \tag{7.22}$$

In other words,

$$A_1^n + A_2^n = g(I_{n+1})\left(A_1^{n+1} + A_2^{n+1}\right) \tag{7.23}$$

Now if we can prove that $A_1 g(I_{n+1}) > 1$ and $A_2 g(I_{n+1}) < 1$, that means the signal in the channel corresponding to the larger correlation will get amplified while the others will decay, then we can say this system works. We rearrange equation (7.23) as

$$\frac{A_1^n}{A_2^n} = \frac{g(I_{n+1})A_2 - 1}{1 - g(I_{n+1})A_1} \tag{7.24}$$

or

$$\frac{A_1^n}{A_2^n} + 1 = \frac{g(I_{n+1})(A_2 - A_1)}{1 - g(I_{n+1})A_1} \tag{7.25}$$

In the left hand of equation (7.25) and according to our supposition $A_1 > A_2$, the numerator in the right hand of this equation is negative, so the denominator must be negative, that is

$$g(I_{n+1})A_1 > 1 \tag{7.26}$$

Similarly, it is not difficult to have

$$g(I_{n+1})A_2 < 1 \tag{7.27}$$

The above two equations mean that the signal in the larger correlation channel will be magnified (Eq. (7.26)) and the signal in the smaller

correlation channel will decay (Eq. (7.27)). So the fibre amplifier can select the maximum correlation.

7.4 Summary

From the above discussion, we know that the implementation of HOFNET requires two subsystems: correlation sub-system and feedback sub-system. The correlation subsystem is usually performed by the well-know image processing system -- 4-f system by using holographic technique [Col71]. The feedback system, however, should provide both gain to compensate for the loss in the correlation system and normalisation to select the maximum correlation output (Fig. 7.15). In section 7.2, we use OASLM to provide gain and D–SEED system to provide normalisation. In section 7.4, however, the optical fibre amplifier with nonlinear response provides both gain and normalisation, making the system more compact. Actually, all the systems with such characteristics can be used for such a task, e.g., the phase-conjugate mirror.

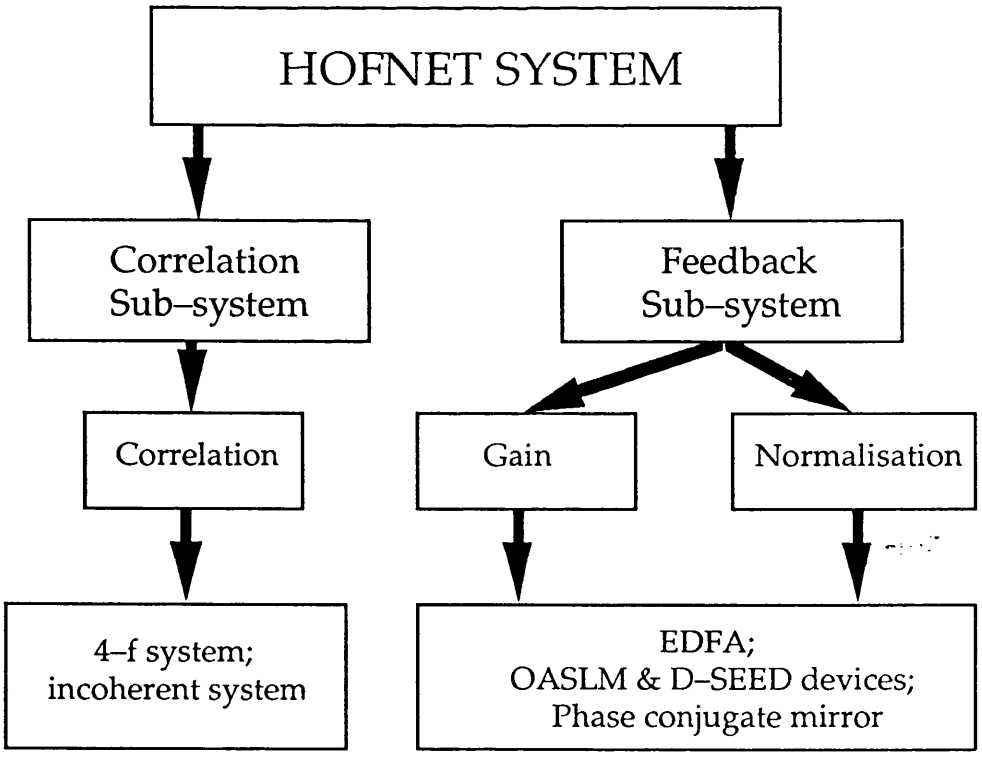


Fig. 7.15 HOFNET Neural Network System

Chapter Eight

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

In this thesis, we have discussed in detail the algorithm of the HOFNET, and optical implementation of the HOFNET system with serial electronic feedback via a computer. In chapter seven, we also designed two novel optical HOFNET systems with an optical parallel feedback loop. In this thesis, I mainly discovered the following ideas:

(a) We place the input pattern in the middle of the system and it is correlated in each iteration, making the system more tolerant of input and system time-dependent noises. This is confirmed by the computer simulations (chapter five);

(b) The high order of the neural network (HOFNET) is implemented by a feedback loop. Therefore, neither a very large number of interconnections nor nonlinear optoelectronic devices with large dynamic ranges are required in the system;

(c) The patterns are stored in the hologram array in the spatial multiplexing way, so the variation of diffraction efficiency for all holograms is greatly reduced (theoretically no variation at all).

(d) A binary spatial light modulator is used to simulate grey levels in the implementation of the HOFNET system. It is used to control the light falling on the hologram array.

From the above discussions, we summarise the advantages of the optical HOFNET system over conventional neural network systems as

follows:

(i) During the pattern learning or recording procedure, all the patterns are recorded in the same way, that is the angles between the object beams and the reference beam remain unchanged. This results in uniformity of diffraction efficiency of the holograms, as the efficiency is the function of the angle between the object and the reference.

(ii) The input pattern is in the middle of the correlation system and is multiplied in each iteration. So this is a true feedback system. If the input pattern or the system contains time-dependent noise, the noise will be averaged out.

(iii) In conventional correlation systems, it is necessary to align the spectrum of the input pattern with the spectra recorded in the hologram, but the spectrum of a pattern usually occupies a very small area and the intensity distribution of spectra of different patterns looks similar, consisting of a large zero-order and weak side-lobes. Moreover, the correlation intensity is very sensitive to the alignment, so it is very difficult to obtain an accurate correlation peak. In our system however, only pattern alignment is required, which is easier than spectrum alignment in most cases.

(iv) There is no cross-coupling between holograms, as the holograms are separated from each other in the recording plate. Therefore the memory capacity or the number of holograms can be increased, at least in theory, by increasing the size of the recording plate, although the size is limited by optical components, i.e. the f-number of lenses.

The disadvantage of the HOFNET is that the correlation system is space variant. The position of the input pattern is fixed, unlike the conventional correlation system in which the input pattern can be translated in the input plane. This problem can be partly solved by using the Circular Harmonic Component (CHC) filters and Mellin transform [Ars89].

The maximum number of patterns that can be stored in the system depends on the recording materials, the relative aperture of the Fourier

transform lens, the aberration of the lens, and the type of light source (Gaussian beam in most cases). Furthermore, the number is also limited by the type of patterns stored in the system (orthogonal or real patterns), the minimum detectable signal, which a detector can pick up from noise. Therefore, although, theoretically, we can store 2^N patterns, the actual number is much below this limit. According to a simple calculation, storing a few thousand patterns seems not difficult in our system.

8.2 Future Work

Although we have done some research on the HOFNET system, it is just the beginning of deep research into the system and the application, and the experimental results are very preliminary. As far as we see, we can do at least the following work or improvement on our present HOFNET system.

(a) Increase of diffraction efficiency and reduction of noise: As you might notice, our experimental results are very noisy. This is because the recording material is silver halide and holograms are bleached to increase the diffraction efficiency, which is still low compared to other phase recording materials, such as dichromated gelatin. The bleaching of holograms increases the noise level caused by grain scattering and possibly the grating distortion. So a better recording material should be used. Dichromated gelatin or photo-polymer is a good candidate for such use, because holograms recorded in these materials are phase-typed, that means the holograms are in the form of refractive index modulation and/or thickness modulation, changing the phase according to intensity. In these materials, the diffraction efficiency can be high (theoretically 100%) and the noise level is very low [Smo90]. Further research on recording material might go to real-time recording -- photorefractive material or liquid crystal [Web91, Tao91], where real-time updating is possible. As in our system, the holograms are separated from each other, updating one does not affect the others. So it is very useful for adaptive neural systems. Another possible solution to the reduction of noise is to use specially polarised light to make and reconstruct holograms.

(b) Implementation of the HOFNET systems designed in chapter seven: Chapter seven described only the theoretical design and we simplify performances of optoelectronic devices. Some details which might affect implementation of the system have not been considered fully, such as gain fluctuation, amplification saturation etc.

(c) Flexibility of the system: The present system is size, rotation and position dependent, which greatly limits its practical use. Such problems exist in most of the signal processing systems, especially in the application for pattern recognition. Although there are some methods to cope with these problems, they are still in a very preliminary stage and there is a lot to do [Ars89, Mic89]. Selviah also proposed a method that all the possible sizes, rotations and positions of a pattern are stored in a volume hologram.

(d) Increase of the number of stored patterns: In our HOFNET system, we only stored 16 patterns for demonstration. Actually much more patterns can be stored in the system. Recently 500 holograms have been stored in a 1cm x 1cm x 1cm photorefractive material. After simple calculation, it seems not difficult to store 10000 patterns, each with 10000 pixels in volume holograms [Sol90].

At the moment the optical HOFNET system is at the demonstration level. In the future, we hope a compact practical HOFNET system could be manufactured. When an input pattern with noise is inserted into the device, a perfect pattern appears at the output plane in a very short time (less than a millisecond).

REFERENCES

[Abu85] Abu-Mostafa, Y. S. and St. Jacques, J. N. (1985), "Information capacity of the Hopfield model", IEEE Trans. Inf. Theory **IT-31**, pp 461-464;

[Abu87] Abu-Mostafa, Y. S. and Psaltis, D. (1987), "Optical neural computer", Scientific American, March, pp 66-73;

[Agr88] Agranat, A., Neugebauer, C. F. and Yariv, A. (1988), "Parallel optoelectronic realization of neural networks models using CID technology, Appl. Opt., **27**, pp 4354-4355;

[Als91] Alspector, J. (1991), "Parallel implementation of neural networks: electronics, optics, biology", Optical Computing (IOCC'91), Salt Lake City, USA, pp 286-290;

[Ami85] Amit, D. J., Gutfreund, H. and Sompolinsky, H. (1985), "Spin-glass models of neural networks", Phys. Rev., **32**, pp 1007-1011;

[Ars89] Arsenault, H. H., (1989), "Distortion invariant pattern recognition using circular harmonic matched filter", Chaper 10, Optical Processing and Computing (ed. Arsenault, H. H., Szoplik, T. and Markow, B.) Academic Press, Inc;

[Ath86] Athale, R. A., Szu, H. H. and Friedlander, C. B. (1986), "Optical implementation of associative memory with controlled nonlinearity in the correlation domain", Opt. Lett., **11**, pp 482-484;

[Bar89] Barnard, E. and Casasent, D. (1989), "Optical neural net for matrix inversion", Appl. Opt., **28**, pp 2499-2504;

[Bar90] Barnes, N. M., Healey, P., McKee, P., O'Neill, A. W., Rejman-Greene, M. A. Z., Scott, E. G., Webb, R. P. and Wood, D. (1990), "High speed opto-electronic neural network", Electronics Letters, **26**, pp 1110-1112;

[Bay87] Bayley. J. S. and Fiddy, M. A. (1987), "On the use of Hopfield model for optical pattern recognition", Opt. Commu., **64**, pp 105-110;

[Bel86] Bell, T. E (1986), "Optical computing: A field in flux", IEEE Spectrum, **23**, pp 34-57;

[Bel91] Belov M. N. and Manykin, E. A. (1991), "Optical associative memory based on echo correlator", Opt. Lett., **16**, pp 327-329;

[Bia88] Bian, S., Xu, K. and Hong, J. (1988), "Optical associative memory model with threshold modification using complementary vector", SPIE Optical Computing, **963**, pp 560-563;

[Bra91] Brady, D. J. & Psaltis, D. (1991), "Holographic interconnections in photorefractive waveguides", Appl. Opt., **30**, pp 2324-2333;

[Bra42] Bragg, W. L. (1942), "The X-ray microscope", Nature, **149**, pp 470-471;

[Cau87] Caulfield, H. J. (1987), "Parallel N^4 weighted optical interconnections", Appl. Opt., **26**, pp 4039-4045;

[Che86] Chen, H. H., Lee, Y. C., Sun, G. Z., Lee, H. Y., Maxwell, T. and Giles, C. L. (1986), "High order correlation model for associative memory", AIP Conference Proceedings, New York, American Institute of Physics, pp 86 - 99;

[Che88] Chevallier, R. C. (1988), "Frequency multiplexed raster scheme of an optical neural network: shift invariant recognition", SPIE Optical Computing, **963**, pp 522-526;

[Chi90] Chittick, R. C., Crossland, W. A. and Brocklehurst, J. R. (1990), "Development and applications of a truly bistable optically addressed spatial light modulator", private communication;

[Col71] Collier, R. J., Burckhardt, C. B. & Lin, L. H. (1971), "Optical holography", Chapters 4, 8, 12, New York: Academic Press;

[Col88] Collings, N., Crossland, W. A., Chittick, R. C. and Bone, M. F. (1988), "The novel application of the electroclinic electro-optic effect to light valve technology", SPIE Optical Computing, **963**, pp 46-50;

[Cro90] Crossland, W. A., Vass D. G. (1990), "Liquid crystal spatial light modulators for optical interconnects of space switching", Optical Connecting and Switching (IEE Digest), May 1990;

[Cro91] Crossland, W. A., Birch, M. J., Davey, A. B. and Vass, D. G. (1991), "Ferroelectric liquid crystal/silicon VLSI backplane technology for smart spatial light modulators", IEE Colloquim, Digest No: 1991/158, pp 7/1-7/4;

[Dic91] Dickey, F. M. and Romero, L. A. (1991), "Normalized correlation for pattern recognition", Opt. Lett., **16**, pp 1186-1188;

[Dow90] Dowie, J. D., Hine, B. P., Reid, M. B. (1990), "An adaptive holographic implementation of a neural network", Proc. SPIE, **1247**, pp 274-283;

[Far85] Farhat, N. H., Psaltis, D., Prata, D. and Paek, E. (1985), "Optical implementation of the Hopfield model", Appl. Opt., **24**, pp 1469-1475;

[Far87] Farhat, N. H. and Psaltis, D. (1987), "Optical implementation of associative memory based on models of neural networks", Optical Signal Processing (ed. Horner, J. L.), Academic Press, pp 129-162;

[Far89] Farhat, N. H. (1989), "Optoelectronic neural networks and learning machines", IEEE Circuits and Devices Magazine, Sept., pp 32-41;

[Fis87] Fisher, A. D, Loppincott, W. L. and Lee, J. N. (1987), "Optical implementations of associative networks with versatile adaptive learning capabilities", Appl. Opt., **26**, pp 5039-5054;

[Gab48] Gabor, D. (1948), "A new microscopic principle", Nature, **191**, pp 777-778;

[Gab49] Gabor, D. (1949), "Microscopy by reconstructed wavefronts", Proceedings of the Royal Society A, **197**, pp 454-487;

[Gho88] Ghosh, A. and Chiun, T. (1988), "Almost necessary conditions for optical neural networks", Appl. Opt., **27**, pp 5002-5003;

[Gil87] Giles, C. L. and Maxwell, T. (1987), "Learning, invariance, and generalization in high-order neural networks", Appl. Opt., **26**, pp 4972-4978;

[Gin88] Gindi, G. R., Gmitro, A. F. and Parathasarathy, K. (1988), "Hopfield model associative memory with nonzero-diagonal terms in memory matrix", Appl. Opt., **27**, pp 129-135;

[Goo68] Goodman, J. W. (1968), "Introduction to Fourier Optics", McGraw-Hill Book Company;

[Goo78] Goodman, J. W., Dias, A. R. and Woody, L. M. (1978), "Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms", Opt. Lett., **2**, pp 1-3;

[Goo84] Goodman, J. W., Levnberger, F. I., Kung, S. Y. and Athale, R. A. (1984), "Optical interconnections for VLSI systems", Proc. IEEE, **72**, pp 850-866;

[Goo89] Goodman, J. W. (1989), "Optics as in interconnect technology", Chapter 1, Optical Processing and Computing (ed. Arsenault, H. H. et al), Academic Press, Inc;

[Gro86] Grossberg, S. (1986), " The adaptive brain I: Cognition, learning, reinforcement and rhythm; the adaptive brain II: vision, speech, language and motor control", Elsevier/North-Holland, Amsterdam;

[Gro89] Groot, P. J and Noll, R. J. (1989), "Adaptive neural network in a hybrid optical/electronic architecture using lateral inhibition", Appl. Opt., **28**, pp 3852-3859;

[Har84] Hariharan, P. (1984), "Optical holography", Chap. 2 & 4, Cambridge: Cambridge University Press;

[Heb49] Hebb, D. O. (1949), "The organization of behaviour", John Wiley & Sons, New York;

[Hen89] Henderson, G. N., Walkup, J. F. and Bochove, E. J. (1989), "Optical quadratic processor using four-wave mixing in $BaTiO_3$", Opt. Lett., **14**, pp 770-772;

[Her86] Herriau, J., Delboulbe, A., Juignard, J., Roosen, G. and Pauliat, G. (1986), "Optical-beam steering for biver array using dynamic holography", J. Lightwave Tech., **LT-4**, pp 905-907;

[Hom90] Homma, J., Kosugi, Y. and Sato T. (1990), "Evaluation and improvement of associative memory for pattern recognition", Appl. Opt., **29**, pp 1675-1681;

[Hop82] Hopfield, J. J. (1982), "Neural networks and physical system with emergent collective computational abilities", Proc. Natl. Acad.

Sci. USA, **79**, pp 2554-2558;

[Hop86] Hopfield, J. J. and Tank, D. W. (1986), "Computing with neural circuits: a model", Science, **223**, pp 625-633;

[Hop89] Hoptroff, R. G. and Hall, T. J. (1989), "Learning by diffusion for multilayer perceptron", Electronics Letters, **25**, pp 531-533;

[Hor90] Horan, P., Uecker, U. and Arimoto, A. (1990), "Optical implementation of a second-order neural network discriminator model", Japanese Journal of Applied Physics, **29**, pp 361-365;

[Hou91] Houselander, P. (1991), "The theory, design, unification and implementation of a class of artificial neural network", Ph. D. thesis;

[Ito90] Itoh, F., Kitayama, K. and Tamura, Y. (1990), "Optical outer-product learning in a neural network using optically stimulable phosphor", Opt. Lett., **15**, pp 860-862;

[Jan88a] Jang, J., Jung, S., Lee, S. and Shin, S. (1988), "Optical implementation of the Hopfield model for two-dimensional associative memory", Opt. Lett., **13**, pp 248-250;

[Jan88b] Jang, J., Shin, S. and Lee, S. (1988), "Optical implementation of quadratic associative memory with outer-product storage", Opt. Lett., **13**, pp 693-695;

[Jan89] Jang. J., Shin S. and Lee, S. (1989), "Programmable quadratic associative memory using holographic lenslet arrays", Opt. Lett., **14**, pp 838-840;

[Joh89] Johnson, R. V. and Tanguay, Jr., A. R. (1989), "Fundamental physical limitations of the photorefractive grating recording sensitivity", Chapter 3, Optical Processing and Computing (ed. Arsenault, H. H. et al), Academic Press Inc;

[Kam89] Kamemeru, S., Kakuta, M. and Shimizu, I. (1989), "Optical pattern recognition with object-multiplexed reflection-type matched spatial filter", Opt. Commu., **69**, pp 211-215;

[Kel91] Keller, P. E. and Gmitro, A. F. (1991), "Design and demonstration of an opto-electronic neural network using fixed planar holographic interconnects", Optical Computing (IOOC'91), Salt Lake City, USA, pp 80-83;

[Kis90] Kiselev, B. S., Kulakov, N. Yu, Mikaelyan, A. L. and Shkitin, V. A. (1990), "Optical realization of a high-order associative memory", Sov. J. Quantum Electron., **20**, pp 736-737;

[Kni74] Knight, G. R. (1974), "Page-oriented associative holographic memory", Appl. Opt., **13**, pp 904-912;

[Kog69] Kogelnik, H. (1969), "Coupled wave theory for thick hologram gratings", Bell System Technical Journal, **48**, pp 2909-2947;

[Koh88] Kohonen, T. (1988), "Self-organization and associative memory', 2nd Edition, Springer-Verlag, Berlin;

[Kra82] Kral, E. L., Walkup, J. F. and Hagler, M. O. (1982), "Optical frequency shifter for heterodyne interferometry using counter rotating wave plates", Appl. Opt., **21**, pp 1281-1288;

[Kyu88] Kyuma, K., Ohta, J., Kojima, K. and Nakayama, T. (1988), "Optical neural networks: system and device technologies", SPIE Optical Computing, **963**, pp 475-484;

[Lal89] Lalanne, P., Chavel, P. and Taboury, T. (1989), "Optical inner-product implementation of neural networks models", Appl. Opt., **28**, pp 377-385;

[Lam89] Lam, T. Y. D. and Carroll, J. E. (1989), "Double layer associative memories", Opt. Commu., **70**, pp 293-298;

[Lam89] Laming, R.I., et al (1989), "Efficient pump wave length of erbium-doped fibre optical amplifier", Electronics Letters, **25**, pp 12-14;

[Lee81] Lee, S. H. (1981), "Optical information processing: fundamentals", Berlin: Springer-Verlag;

[Lee86] Lee, Y. C., et al (1986), "Machine learning using a higher order correlation network", Physica, **22D**, North-Holland, pp 276-306;

[Lei62] Leith, E. N. & Upatnieks, J. (1962), "Reconstructed wavefronts and communication theory", J. O. S. A., **52**, pp 1123-1130;

[Lin89] Lin, S. and Liu, L (1989), "Opto-electronic implementation of a neural network with a third-order interconnection for quadratic associative memory", Opt. Commu., **73**, pp 268-272;

[Lin90] Lin, S. and Liu, L. (1990), "Shift, rotation and scale invariant optical associative memory", Optical Computing (IOCC'90), Japan, pp 169-172;

[Liu86] Liu, H. K., Kung S. Y. and Davis, J (1986), "A real-time optical associative retrieval technique", Opt. Eng., **25**, pp 118-122;

[Lip87] Lipman, R. P., (1987), "An introduction to computing with neural nets", IEEE ASSP Magazine, April, pp 5-22;

[Loh90] Lohmen, A. (1990), "Opto-electronic chip implementation of a quadratic associative memory, Opt. Lett., **15**, pp 279-281;

[Lu90] Lu, T., Yu, F. T. S. and Gregory, D. A. (1990), "Self-organizing optical neural network for unsupervised learning", Opt. Eng., **29**, pp 1107-1113;

[Mad85] Mada, H. (1985), "Architecture for optical computing using holographic associative memories", Appl. Opt., **24**, pp 2063-2066;

[Man91] Mankin, E. A. and Belov, M. N. (1991), "High-order neural networks and photon-echo effect", Neural Networks, **4**, pp 417-420;

[McC43] McCulloch, W. S. and Pitts, W. (1943), "A logical calculus of the ideas imminent in nervous activity", Bulletin of Mathematical Biophysics, **5**, pp115-133;

[Mid88] Midwinter, J. E. and Selviah, D. R. (1988), "Digital neural networks, matched filters and optical implementation" in Chapter 13, Neural Computing, Kagan Page Ltd Publisher;

[Mil84] Miller, D. A. B., Chemla, D. S. and Damen, T. C. (1984), "Optical-level shifter and self-linearized optical modulator using a quantum-well self-electro-optic effect device", Opt. Lett., **9**, pp 567-569;

[Mil85] Miller, D. A. B., Chemla, D. S., Damen, T. C., Wood, T. H., Burrus, C. A., Gossard, A. C. and Wiegmann, W. (1985), "The quantum well self-electrooptic effect device: opto-electronic bistability and oscillation, and self-linearized modulation", IEEE J. Quantum Electron, **QE-21**, pp 1462-1475;

[Mod89] Moddel, G., Johnson, K. M., Li, W. and Rice, R. A. (1989), "High-

speed binary optically addressed spatial light modulator", Appl. Phys. Lett., **55**, pp 537-539;

[Mok91] Mok, F. H., Tackitt, M. C. & Stoll, H. M. (1991), "Storage of 500 high resolution holograms in a $LiNbO_3$ crystal", Opt. Lett., **16**, pp 602-604;

[Nei91] Neifeld, M. A. and Psaltis, D. (1991), "Closed loop optical disk based associative memory", Optical Computing (IOCC'91), Salt Lake City, USA, pp 276-279;

[Nij89] Nijhuis, J. A. G. and Spanmenburg, L. (1989), "Fault tolerance of neural associative memories", IEE Proc. pt. E, **136**, pp 389-394;

[Oh88] Oh, S., Yoon T. and Kim, J. (1988), "Associative memory model based on neural networks: modification of Hopfield model", Opt. Lett., **13**, pp 74-76;

[Oht89] Ohta, J, Tai, S., Oita, M., Kuroda, K., Kyuma, K. and Hamanaka, K. (1989), "Optical implementation of an associative neural network model with a stochastic process", Appl. Opt., **28**, pp 2426-2428;

[Owe87] Owechko, Y., Dunning, G. J., Marom, E. and Soffer, B. H. (1987), "Holographic associative memory with nonlinearities in the correlation domain", Appl. Opt., **26**, pp 1900-1910;

[Pae87] Paek, E. G. and Psaltis, D. (1987), "Optical associative memory using Fourier transform holograms", Opt. Eng., **26**, pp 428-433;

[Pae89] Paek, E. G. and von Lehmen, A. (1989), "Holographic associative memory for word-break recognition", Opt. Lett., **14**, pp 205-207;

[Pae91] Paek, E. G. and Jung, E. C. (1991), "simplified holographic associative memory using enhanced non-linear processing with a thermoplastic plate", Opt. Lett., **16**, pp 1034-1036;

[Pet90] Peterson, C., Redfield, S., Keeler, James and Hartman, E. (1990), "Optoelectronic implementation of multilayer neural networks in a single photorefractive crystal", Opt. Eng., **29**, pp 359-368;

[Psa85] Psaltis, D. and Farhat, N. (1985), "Optical information processing based on associative-memory model of neural nets with thresholding and feedback", Opt. Lett., **10**, pp 98-100;

[Psa86] Psaltis, D. and Park, C. H. (1986), "Nonlinear discriminant functions and associative memories", Neural Networks for Computing Conference, Snowbird UT, pp 370-375;

[Psa87] Psaltis, D. and Hong, J. (1987), "Shift-invariant optical associative memories", Opt. Eng., **26**, pp 10-15

[Psa88a] Psaltis, D., Park, D. C., and Hong, J. (1988), "High order associative memories and their optical implementations", Neural Networks, **1**, pp 149-163;

[Psa88b] Psaltis, D., Brady, D. and Wagner, K. (1988), "Adaptive optical networks using photorefractive crystals", Appl. Opt., **27**, pp 1752-1759;

[Psa88c] Psaltis, D., Gu, X. G. and Brady, D. (1988), "Fractal sampling grids for holographic interconnections", SPIE Optical Computing, **963** Set. D., pp 468-474;

[Psa89] Psaltis, D; Brady, D., Gu, X. G. and Hsu, K. (1989), "Optical implementation of neural computers", Chapter 8, Optical Processing and Computing (ed. Arsenault, H. H. et al), Academic Press Inc;

[Psa90a] Psaltis, D. and Quio, Y. (1990), "Optical neural networks", Optics & Photonics News, Dec., pp 17-20;

[Psa90b] Psaltis, D., Brady, D., Gu, X. and Lin, S. (1990), "Holography in artificial neural networks", Nature, **342**, pp 325-329;

[Psa91] Psaltis, D. (1991), "Learning in optical neural networks", Optical Computing (IOOC'91), Salt Lake City, USA, pp 266-267;

[Rec88] Recce, M. and Treleavan, P. (1988), "Computing from the brain", New Scientist, May, pp 61-64;

[Rob91] Robinson, M. G. and Johnson, K. M., Jared, D., Doroski, D., Wichart, S. and Moddel, G. (1991), "Custom designed electro-optic components for optically implemented, multi-layer neural networks", Optical Computing (IOOC'91), Salt Lake City, USA, pp 84-87;

[Ros61] Rosenblatt, F. (1961), "Principles of neurodynamics: perceptrons and the theory of brain mechanisms", Spartan Books, Washington DC;

[Rum86] Rumelhart, D. E. and McClelland, J. L. (1986), "parallel distributed processing", Vol. 1, Chaps 5-8, The MIT Press;

[Sato91] Sato, K., Higuchi, K., and Katsumal, H. (1991), "Holographic television by liquid crystal device", IEE Proc. **342**, pp 20-23;

[Sel87] Selviah, D. R. and Midwinter, J. E. (1987), "Pattern recognition using opto-electronic neural networks", IEE Colloquium, **105**, pp 6.1-6.4;

[Sel88] Selviah, D. R. and Midwinter, J. E. (1988), "Spatial light modulator", Patent No: 8813483.8 (Aug. 1988);

[Sel89a] Selviah, D. R. and Midwinter, J. E. (1989), "Memory capacity of a novel optical neural net architecture", Optical Computing Conference, Foulouse, France, pp195-201;

[Sel89b] Selviah, D. R. and Midwinter, J. E. (1989), "Correlating matched-filter model for analysis and optimisation of neural networks", IEE Proc. Pt F, **136**, pp 143-148;

[Sel90a] Selviah, D. R. and Midwinter, J. E. (1990), "Extension of the Hamming neural network to a multilayer architecture for optical implementation", The First IEE Int. Conf. on Art. Neural Networs, Proc. **313**, pp 280-283;

[Sel90b] Selviah, D. R. and Midwinter, J. E. (1990), "Memory capacity of a novel optical neural net architecture", Optical Computing Int. Symp., Foulouse, France, pp 195-201;

[Sel90c] Selviah, D. R., Mao, Z. Q., Tao, S. and Midwinter, J. E. (1990), "Optoelectronic high order neural network", Electronics Letters, **26**, pp 1954-1955;

[Sha89] Shariv, I. and Firesem, A. A. (1989), "All-optical neural network with inhibitory neurons", Opt. Lett., **14**, pp 485-487;

[Sha91] Shariv, I., Gila, O. and Friesem, A. A. (1991), "All-optical bipolar neural network with polarization-modulating neurons", Opt. Lett., **16**, pp 1692-1694;

[Sin90] Singh, J. and Hong, S. (1990), "Optoelectronic architecture for associative memory applicable to 2-D pattern comparison", Appl. Opt., **29**, pp 1682-1688;

[Smo90] Smothers, W. K., Monroe, B. M. Weber, A. M. and Keys, D. E., (1990), "Photopolymers for holography", SPIE OE/Laser Conference Proc., Jan. 1990, pp315-328;

[Sof86] Soffer, B. H., Dunning, G. J., Owechko, Y. and Marom, E. (1986), "Associative holographic memory with feedback using phase-conjugate mirrors", Opt. Lett., **11**, pp 118-120;

[Sol81] Solymar, L. & Cooke, D. J. (1981), "Volume holography & Volume gratings", New York: Academic Press;

[Sol89] Solymar, L. (1989), "Reconfigurable optical interconnections in photorefractives", private communication;

[Son88] Song, S. H. and Lee, S. S. (1988), "Experimental and numerical analyses of diffused light holographic associative memory", Appl. Opt., **27**, pp 3590-3591;

[Son90] Song, S. H., Park, S. and Lee, S. S. (1990), "Optical implementation of a quadratic associative memory by using the polarization-encoding process", Opt. Lett., **15**, pp 1389-1391;

[Sor71] Soroko, L. M. (1971), "Holography and coherent optics", translated from Ruaaian by Albin Tybulewicz, Moscw: Nauka Press;

[Ste90] Stepien, P., Smolinska, B. and Gniadek, K. (1990), "All-optical associative memory in the coherent and in the intensity implementation. A Comparison", Optik, **84**, pp 39-42;

[Sun90] Sun, Y. and Zhang, Y. (1990), "An optical neural network used for content addressable two-dimensional associative memory", Optik, **85**, pp 43-46;

[Tak90] Takino, M., Ohtsu, N. and Yatagai, T. (1990), "Shift-invariant associative memory based on anamorphic incoherent correlator", Japanese J. of Appl. Phy., **29**, pp L1317-L1320;

[Tan85] Tanguay Jr., A. R. (1985), "Materials requirements for optical processing and computing devices", Opt. Eng., **24**, pp 2-18;

[Urq88] Urquhart, P. (1988), "Review of rare earth doped fibre lasers and amplifiers", IEE Proceedings, Pt. J., **135**, pp 385-407;

[Val83] Valley, G. C. and Klein, M. B. (1983), "Optical properties of photorefractive materials for optical data processing", Opt. Eng., **22**, pp 704-711;

[Van64] Vander Lugt, A. (1964), "Signal detection by complex spatial filtering", IEEE Transactions on Information Theory, **IT-10**, pp 139-145;

[Wag87] Wagner, K. and Psaltis, D. (1987), "Multilayer optical learning networks", Appl. Opt., **26**, pp 5061-5076;

[Wag91] Wagner, K. and Slagle, T. (1991), "Competitive optical learning with winner-take-all modulators", Optical Computing (IOCC'91), Salt Lake City, USA, pp 280-283;

[Wan91] Wang, X. M., Mu, G. G. and Zhang, Y. X. (1991), "Optical associative memory using an orthogonalized hologram", Opt. Lett., **16**, pp 100-102;

[War87] Warde, C. and Fisher, A. (1987), "Spatial light modulator: Applications and functional capabilities", Optical Signal Processing (ed. Horner J. L.), Academic Press, pp 478-523;

[Web91] D. J. Webb and Solymar L. (1990), "Comparison of three wave mixing in photorefractive and nematic liquid crystals", IEE Proc. **342**, pp108-112;

[Wei90] Weible, K. J., Pedrini, G., Xue, W. and Thalmann, R. (1990), "Optical implementation of a neural network associative memory using diffraction gratings", Japanese J. of Appl. Phy., **29**, pp L1301-L1303;

[Wei91] Weible, K. J., Collings, N., Xue, W., Pedrini, G. and Dandliker, R. (1991), "Experimental comparison of different associative memory techniques implemented optically by the same system architecture", Optical Computing (IOCC'91), Salt Lake City, USA, pp 92-95;

[Whe88] Wherrett, B. S. and Tooley, F. A. P. (ed.) (1988), "Optical computing", Proc. of the Thirty-fourth SUSS in Physics, Edinburgh;

[Whi88a] White, H. J., Aldridge, N. B. and Lindsay, I. (1988), "Digital and analogue holographic associative memories", Opt. Eng., **27**, pp 30-37;

[Whi88b] White, H. J. and Wright, W. A. (1988), "Holographic implementation of a Hopfield model with discrete weightings", Appl. Opt., **27**, pp 331-338;

[Whi88c] White, H. J. (1988), "Experimental results from an optical implementation of a single neural network", SPIE Optical Computing, **963**, pp 570-573;

[Whi89] Whitehead, M., Rivers, A., Parry, G., Roberts, J. S. and Button, C. (1989), "Low-voltage multiple quantum well reflection modulator with on:off ratio>100:1", Electronics Letters, **25**, pp 884-885;

[Wil89] Wilde, J. and Hesselink, L (1989), "Implementation of dynamic Hopfield-like networks using photorefractive crystals", Optical Computing (IOCC'89), Salt Lake City, UT, pp 10-13;

[Yan90a] Yang, G. G. (1990), "Optical neural network", a review, private communication;

[Yan90b] Yang, X., Lu, T. and Yu, F. T. S. (1990), "Compact optical neural network using cascaded liquid crystal television", Opt. Letter., **29** , pp 5223-5225;

[Yan91] Yang, G. G. (1991), "Performance analysis of optical associative memory with translation invariance", Optik, **87**, pp 63-67;

[Yar86] Yariv, A., Kwong, S. and Kyuma, K. (1986), "Demonstration of an all-optical associative holographic memory", Appl. Phys. Lett., **48**, pp 1114-1116;

[Yeh88] Yeh, P., Chiou, A. E. T. & Hong, J. (1988), "Optical interconnection using photorefractive dynamic holograms", Appl. Opt., **27**, pp 2093-2096;

[Yeh89] Yeh, P., Chiou, A. E., Hong, J., Beckwith, P., Chang, T. and Khoshnovisa, M (1989), "Photorefractive nonlinear optics and optical computing", Opt. Eng., **28**, pp 328-343;

[Yos89a] Yoshinaga, H., Kitayama, K. and Hara, T. (1989), "All-optical error-signal generation for backpropagation learning in optical multilayer neural networks", Opt. Lett., **14**, pp 202-204;

[Yos89b] Yoshinaga, H. and Kitayama, K. (1989), "Experimental learning in an optical perceptronlike neural network", Opt. Lett., **14**, pp 716-718;

[Yu73] Yu, F. T. S. (1973), "Introduction to Diffraction, Information Processing, and Holography", The MIT Press, Part III: Holography.

[Yu90] Yu, F. T. S., Lu, T. and Yang, X. (1990), "Optical neural network with pocket-sized liquid-crystal televisions", Opt. Lett., **15** , pp 863-865;

[Yu91] Yu, F. T. S, Yang, X. and Lu, T. (1991), "Space-time-sharing optical neural network", Opt. Lett., **16**, pp 247-249;

[Zha91a] Zhang, L., Robinson, M. G. and Johnson, K. M. (1991), "Optical implementation of a second-order neural network", Opt. Lett., **16**, pp 45-47;

[Zha91b] Zhang, W., Itoh, K., Tanida, J. and Ichioka, Y. (1991), "Hopfield model with multistate neurons and its optoelectronic implementation", Appl. Opt., **30**, pp 195-200;

APPENDIX

I. FOURIER TRANSFORMATION

1. Definition

By definition, the Fourier transform of a function f(x,y) is an integral operation:

$$F(\mu,\nu)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\exp[-i2\pi(\mu x+\nu y)]dxdy \qquad (I1)$$

The inverse Fourier transform is defined as

$$f(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(\mu,\nu)\exp[i2\pi(\mu x+\nu y)]d\mu d\nu \qquad (I2)$$

These relationships can be expressed more briefly with the aid of the Fourier transform operators F.T., $F.T.^{-1}$

$$F(\mu,\nu) = F.T.\ \{f(x,y)\}$$
$$f(x,y) = F.T.^{-1}\ \{F(\mu,\nu)\} \qquad (I3)$$

The function $F(\mu,\nu)$ is known as the Fourier transform of the function f(x,y), while the function f(x,y) is called the inverse Fourier transform of the function $F(\mu,\nu)$. The difference between the direct and inverse Fourier transforms lies in the different signs of the exponential functions in Eqs (I1) and (I2). The difference between F.T. and $F.T.^{-1}$ is very often ignored in the optical Fourier transform by a lens.

2. Properties of the Fourier Transform

(a) Linearity

If $F_1(\mu,\nu)$ and $F_2(\mu,\nu)$ are the Fourier transforms of the functions $f_1(x,y)$ and $f_2(x,y)$, respectively and a_1 and a_2 are arbitrary complex constants, the linear relation can be written as

$$f(x,y) = a_1 f_1(x,y) + a_2 f_2(x,y)$$
$$F(\mu,\nu) = a_1 F_1(\mu,\nu) + a_2 F_2(\mu,\nu)$$
$$F(\mu,\nu) = \text{F.T.}\{f(x,y)\} \qquad \text{(I4)}$$

This property plays a very important role in coherent optics and holography.

(b) Symmetry

Suppose the Fourier transform of the function $f^*(x,y)$ is $G(\mu,\nu)$, then

$$G(\mu,\nu) = F^*(-\mu,-\nu) \qquad \text{(I5)}$$

If $f(x,y)$ is a real function, i.e. $f(x,y) = f^*(x,y)$, we have

$$F^*(-\mu,-\nu) = F(\mu,\nu) \qquad \text{(I6)}$$

(c) Shift Theorem

If $f(x,y)$ is shifted by (x_o, y_o), then

$$\text{F.T.}\{ f(x-x_o, y-y_o) \} = \exp[-i2\pi(\mu x_o+\nu y_o)]\ \text{F.T.}\{ f(x,y) \} \qquad \text{(I7)}$$

That means a shift of the function $f(x,y)$ results in only the phase change of the Fourier transform and does not alter its modulus.

(d) Frequency Shift

If $F(\mu,\nu)$ is shifted by (μ_o,ν_o), then

$$F(\mu-\mu_o,\nu-\nu_o) = F.T.\{ f(x,y) \exp[2\pi(\mu_o x+\nu_o y)] \} \quad (I8)$$

3 Convolution and Correlation

The convolution of the two functions $f_1(x,y)$ and $f_2(x,y)$ is defined as

$$f(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_1(\alpha,\beta) f_2(x-\alpha, y-\beta) d\alpha d\beta \quad (I9)$$

It can be denoted by the operational symbol $\otimes$

$$f(x,y) = f_1(x,y) \otimes f_2(x,y) \quad (I10)$$

The correlation of the two functions $f_1(x,y)$ and $f_2(x,y)$ is defined as

$$f_{12}(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_1(\alpha,\beta) f_2^*(x+\alpha, y+\beta) d\alpha d\beta \quad (I11)$$

Similarly, it can be denoted by the operational symbol $*$

$$f_{12}(x,y) = f_1(x,y) * f_2(x,y) \quad (I12)$$

Here it should be noted that whereas the convolution operation is commutative relative to the interchange of the function $f_1(x, y)$ and $f_2(x, y)$, the correlation does not have the same property. Actually

$$f_{12}(x,y) = f_{21}^*(-x,-y) \quad (I13)$$

and the relation between correlation and convolution is

$$f_1(x,y) * f_2(x,y) = f_1(x,y) \otimes f_2^*(-x,-y) \quad (I14)$$

So for real patterns, it is easy to commute between correlation and convolution.

4 Relation between Correlation/Convolution and Fourier Transform

Suppose $F_i(\mu,\nu) = F.T.\{f_i(x,y)\}$ (i=1, 2), then we have

$$F_1(\mu,\nu)F_2(\mu,\nu) = F.T.\{f_1(x,y) \otimes f_2(x,y)\} \tag{I15}$$

$$F_1(\mu,\nu)F_2^*(\mu,\nu) = F.T.\{f_1(x,y) * f_2(x,y)\} \tag{I16}$$

From Eq. (I16), we can deduce the famous Parseval Equation

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_1(x,y)f_2^*(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F_1(\mu,\nu)F_2^*(\mu,\nu)d\mu d\nu \tag{I17}$$

If $f_1(x,y)=f_2(x,y)$, then we have

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |f_1(x,y)|^2 dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |F_1(\mu,\nu)|^2 d\mu d\nu \tag{I18}$$

This means energy conservation.

II INTERSECTING POINTS OF TWO HYPERBOLIC CURVES

The two hyperbolic curves are drawn in Fig. A1 and their mathematical forms are:

$$y = K(-x^2 + 4x) \tag{II1}$$
$$x = K(-y^2 + 4y) \tag{II2}$$

where K is a parameter which decides how many intersecting points there are of the two curves, except for the origin of coordinates, as shown in Fig. A1.

Insert equation (II1) into equation (II2), we have

$$x = K\left[-K(-x^2 + 4x) + 4\right]K(-x^2 + 4x) \tag{II3}$$

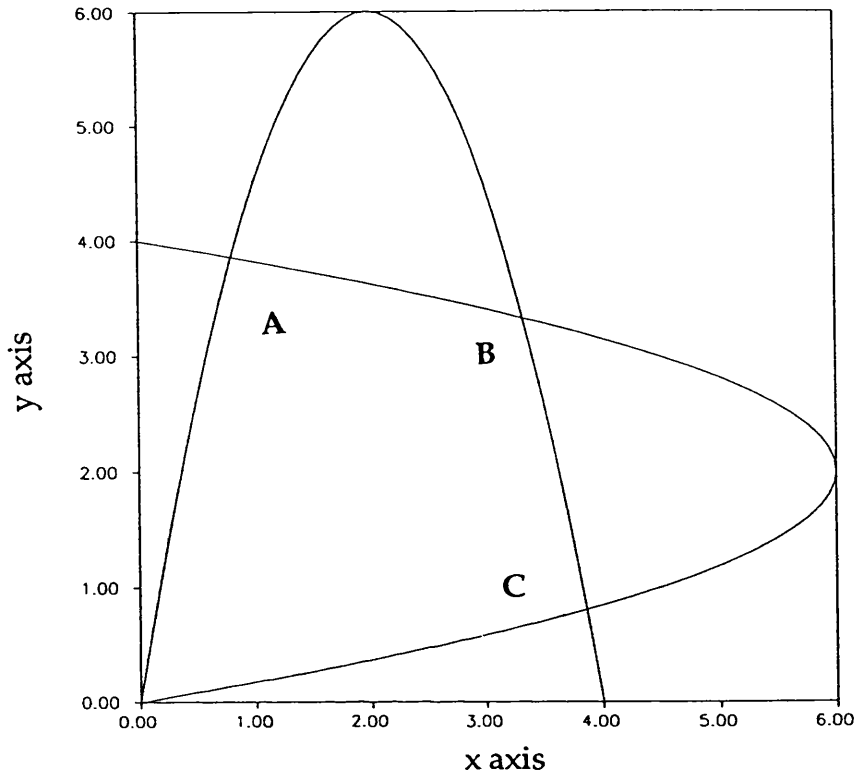


Fig. A1 Two Intersecting Hyperbolic Curves

Rearrange it as

$$x^3 - 8x^2 + \left(16 + \frac{4}{K}\right)x - \frac{16}{K} + \frac{1}{K^3} = 0 \qquad \text{(II4)}$$

As can be seen from Fig. (A1), there always exists a solution (an intersecting point) where $x = y$, excluding the origin where x=y=0 (point B in Fig. (A1)). We can get this solution very easily by setting

$$y_B = x_B = K\left(-x_B^2 + 4x_B\right)$$
$$x_B = \frac{4K-1}{K} \qquad \text{(II5)}$$

We rewrite equation (II4) as

$$\left(x - \frac{4K-1}{K}\right)\left(x^2 + ax + b\right) = 0 \qquad \text{(II6)}$$

where a, b are parameters, which are decided by comparing it with equation (II4). Expanding equation (II6) and write it in the same format as equation (II4), we have

$$x^3+\left(a-4+\frac{1}{K}\right)x^2+\left(b-4a+\frac{a}{K}\right)x+b\left(\frac{1}{K}-4\right)=0 \qquad \text{(II7)}$$

Compare equations (II7) and (II4), we have

$$\begin{aligned} &a-4+\frac{1}{K}=-8 \\ &b\left(\frac{1}{K}-4\right)=\frac{1}{K}\left(\frac{1}{K^2}-16\right) \end{aligned} \qquad \text{(II8)}$$

So we have

$$\begin{aligned} &a=-\left(4+\frac{1}{K}\right) \\ &b=\frac{1}{K}\left(4+\frac{1}{K}\right) \end{aligned} \qquad \text{(II9)}$$

If we want only one intersection point except the origin (others are not stable), then equation

$$x^2+ax+b=0 \qquad \text{(II10)}$$

should have no real solutions, which results in

$$a^2-4b<0 \qquad \text{(II11)}$$

that is

$$\left(4+\frac{1}{K}\right)^2-\frac{4}{K}\left(4+\frac{1}{K}\right)<0$$

Therefore, K must satisfy

$$K < \frac{3}{4} \qquad \text{(II12)}$$

Furthermore, we know the solution (coordinates of the point B) corresponds to the intensity (see chapter 7), so the solution x_B must be more than zero, that is

$$\frac{4K-1}{K} > 0$$

$$K > \frac{1}{4} \qquad \text{(II13)}$$

Combine (II12) and (II13) we have

$$\frac{1}{4} < K < \frac{3}{4} \qquad \text{(II14)}$$

Compare equation (II1) with equation (7.20) in the real system in chapter 7, we have

$$\frac{1}{4} < \frac{GI_i}{8} < \frac{3}{4}$$
$$\frac{2}{I_i} < G < \frac{6}{I_i} \qquad \text{(II15)}$$

This is the requirement that gain must satisfy. If the gain is too small ($G<2/I_i$), then the loss is more than gain, so the energy will decay and vanish eventually. If the gain is too large ($G>6/I_i$), however, there will be three intersecting points excluding the origin, the system will not stabilise but vibrate.

PUBLICATIONS

1. H. Facanha, D. R. Selviah, K. Steptoe and Z. Q. Mao, "Design of Fresnel holograms for optical interconnection of VLSI", 2nd International Conference on Holographic Systems, Components and Applications, Bath, U.K., Sept. 1989, IEE Proc. **311**, pp213-217;

2. D. R. Selviah, Z. Q. Mao and J. E. Midwinter, "Optoelectronic high order feedback neural net", Electronics Letters, **26**, 11, 1990, pp1954-1955;

3. Z. Q. Mao, D. R. Selviah, S. Q. Tao and J. E. Midwinter, "Optical implementation of high order neural network (HOFNET)", presented at the FOCUS-ESPRIT Conference, London, U.K., January 1991;

4. Z. Q. Mao, D. R. Selviah, S. Q. Tao and J. E. Midwinter, "Holographic high order associative memory system", 3rd International Conference on Holographic Systems, Components and Applications, Edinburgh, U.K., Sept. 1991, IEE Proc. **342**, pp132-136;

5. D. R. Selviah, Z. Q. Mao and J. E. Midwinter, "An Optoelectronic High Order Feedback NET (HOFNET) with variable non-linearity", 2nd International Conference on Artificial Neural Networks, Bournemouth, U.K., Nov. 1991, IEE Proc. **349**, pp59-63;

6. D. R. Selviah, Z. Q. Mao and J. E. Midwinter, "Optical neural network with optoelectronic feedback", presented at the IOP Conference on Optoelectronic Devices and Systems, 3rd December 1991;

7. Z. Q. Mao, D. R. Selviah and J. E. Midwinter, "Optical high order neural network (HOFNET) based on parallel correlators", to be submitted to Applied Optics;

8. Z. Q. Mao, D. R. Selviah and J. E. Midwinter, "Design of optical HOFNET by using image subtraction system for normalisation", the IOP Conference on Optoelectronic Neural Networks, Oxford, U.K., June 1992;

9. Z. Q. Mao, D. R. Selviah and J. E. Midwinter, "Optical high order neural network (HOFNET) and its application to pattern recognition", submitted to the International Artificial Neural Networks, Brighton, U.K., Sept. 1992.