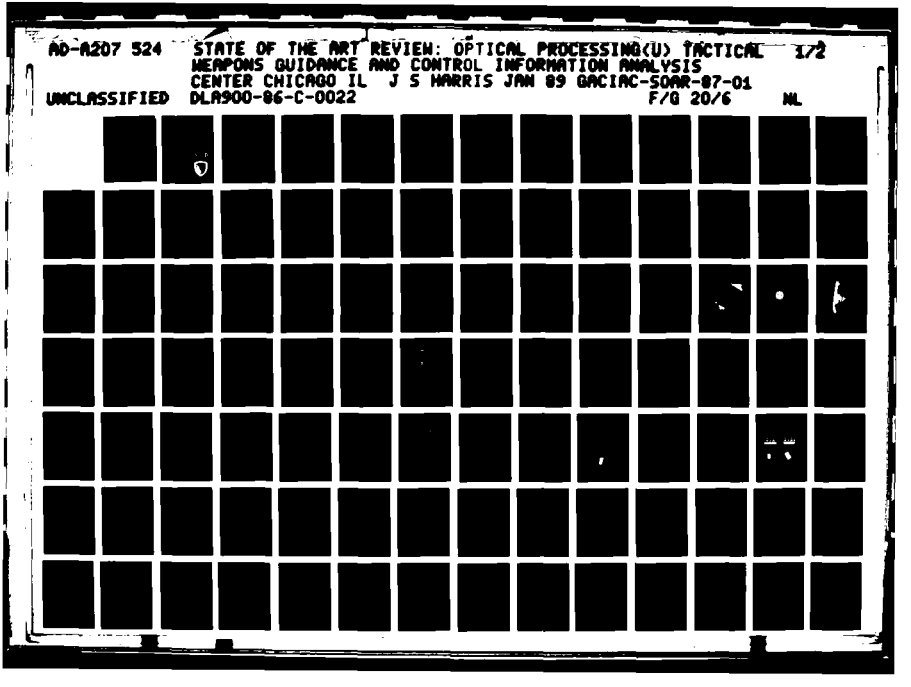
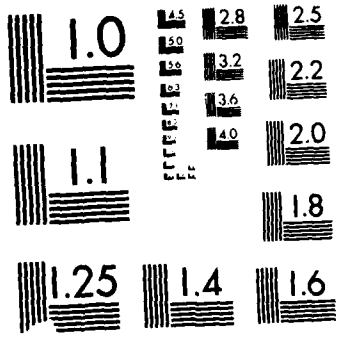

AD-A207 524 STATE OF THE ART REVIEW: OPTICAL PROCESSING(U) TACTICAL 1/2
WEAPONS GUIDANCE AND CONTROL INFORMATION ANALYSIS
CENTER CHICAGO IL J S HARRIS JAN 89 GACIAC-SOAR-87-01
UNCLASSIFIED DLA900-86-C-0022 F/G 20/6 NL



TACTICAL WEAPON
GACIAC
GUIDANCE & CONTROL
INFORMATION ANALYSIS CENTER

GACIAC SOAR 87-01

State of the Art Review
OPTICAL PROCESSING

J. S. Harris

January 1989

Approved for public release:
Distribution unlimited

Published by GACIAC
IIT Research Institute
10 West 35th Street
Chicago, Illinois 60616-3799

NOTICES

State-of-the-Art Review. This proceedings has been published by the Tactical Weapon Guidance and Control Information Analysis Center (GACIAC) as a service to the defense community. GACIAC is a DoD Information Analysis Center, administered by the Defense Technical Information Center, operated by IIT Research Institute under Contract No. DLA900-86-C-0022. GACIAC is funded by DTIC, DARPA, and U.S. Army, U.S. Navy, and U.S. Air Force Laboratories/ Controlling Activities having an interest in tactical weapon guidance and control. The Director of GACIAC is Dr. Robert J. Heaston. The Contracting Officer is Mrs. S. Williams, DESC, Dayton, Ohio. The Contracting Officer's Technical Representative is Mr. H. C. Race, U.S. Army Missile Command, ATTN: AMSMI-RD-SM, Redstone Arsenal, Alabama 35898-5246.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS Approved for public release Distribution Unlimited
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution Unlimited. Copies only from GACIAC.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) GACIAC SOAR-87-01	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION GACIAC IIT Research Institute 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U.S. Army Missile Command ATTN: AMSMI-RD-SM
6c. ADDRESS (City, State, and ZIP Code) 10 West 35th Street Chicago, IL 60616-3799	7b. ADDRESS (City, State, and ZIP Code) Redstone Arsenal, AL 35898-5246
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Missile Command 8b. OFFICE SYMBOL (If applicable) AMSMI-RD-SM	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DLA900-86-C-0022
8c. ADDRESS (City, State, and ZIP Code) Redstone Arsenal, AL 35898-5246	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 65802S; PROJECT NO 1.0; TASK NO; WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Optical Processing

12. PERSONAL AUTHOR(S)
J.S. Harris

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
State-of-the-Art	FROM 2-87 TO 1-89	January 1989	125

16. SUPPLEMENTARY NOTATION This report is only available from GACIAC. Reproduction is not authorized except by specific permission. 410948. $50.00

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
12	07	
09	05	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Optical Processors, Optical Correlation, Optical Data Processing, Optical Imaging, State-of-the-Art, Signal Processing, Radar, Synthetic Aperture Radar (cont)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Optical processing is a rapidly developing area of technology that is based on the ability of an optical system to perform, in a straightforward manner, mathematically complex operations on light passing through it. This technology uses light waves to do the work of electrons. The great strength of optical processing is performing complex mathematical transformations on large masses of data in parallel. This State-of-the-Art Review describes, in detail, optical processing using, in addition to text, figures and mathematical equations. Applications of optical processing are discussed in individual chapters. The applications include signal processing as applied to radar and one-dimensional signal analysis; image processing as applied to real-time processing, two-dimensional processing, correlators, and pattern recognition; and algebraic processing as applied to algebraic signal processing operations, Stanford optical matrix-vector multiplier, systolic array processing, and matrix-matrix multiplication systems. The final chapter discusses bistable optical devices for optical signal processing applications.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Howard C. Race	22b. TELEPHONE (Include Area Code) (205) 876-3171 22c. OFFICE SYMBOL AMSMI-RD-SM

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

18. SUBJECT TERMS (cont)

Acoustooptical Processors, Image Processing, Pattern Recognition, Algebra, Matrix Methods, Optical Modulators, Optical Materials.

GACIAC SOAR 87-01 JANUARY 1989

State of the Art Review
OPTICAL PROCESSING

J. S. Harris

Published by GACIAC
IIT Research Institute
10 West 35th Street
Chicago, Illinois 60616-3799

**Approved for public release:
Distribution unlimited**

GACIAC -- A Dod Information Analysis Center
Operated by IIT Research Institute, 10 West 35th Street, Chicago, IL 60616-3799
DoD Technical Sponsor -- U.S. Army Missile Command, Redstoned Arsenal, AL 35898-5243

PREFACE

Optical processing is a rapidly developing technology that is based on the ability of an optical system to perform mathematically complex operations on light passing through it and to do so in a straightforward manner. Thus, this technology uses light waves to do the work performed in electronic systems by electrons. Digital electronics surpasses other technologies in performing complex mathematical operations that can be broken down into many simple arithmetical steps performed in sequence. Optical signal processing offers very different capabilities: its greatest strength is in performing complex mathematical transformations on large masses of data in parallel. While simple subtraction is much easier to implement electronically than optically, the powerful Fourier transform, which presents a difficult challenge for electronic processors, is a simple operation for an optical processor.

These unusual properties of optical processors have made them invaluable for solving otherwise intractable signal processing problems such as those encountered in radar processing, video image enhancement, spectrum analysis of the frequencies in a radio signal, object recognition, and robotic vision. Since optical processors by their very nature handle many simultaneous inputs, they are able to perform mathematical operations involving large sets of matrices much more rapidly than digital computers. These capabilities of optical processors have made them very important to existing and future weapons systems for guidance and electronic warfare. Although the emphasis since the beginning of research in optical processing has been on military applications, recent advances in optical system components, system architecture, and pattern-recognition algorithms have spurred a real advance in research for commercial applications in computers, artificial intelligence, robotics, and product inspection.

This document offers the reader both a concise definition of optical processing theory and a description of the state of the art in optical processing applications.

CONTENTS

CONTENTS (continued)

TABLES

FIGURES

FIGURES (continued)

FIGURES (continued)

STATE-OF-THE-ART REVIEW OF OPTICAL PROCESSING

J. S. Harris

ABSTRACT

The unusual properties of optical processors have made them invaluable for solving otherwise intractable signal processing problems such as those encountered in radar processing, video image enhancement, spectrum analysis of the frequencies in a radio signal, object recognition, and robotic vision. Since optical processors by their very nature handle many simultaneous inputs, they are able to perform mathematical operations involving large sets of matrices much more rapidly than digital computers. These capabilities of optical processors have made them very important to existing and future weapons systems for guidance and electronic warfare. Although the emphasis since the beginning of research in optical processing has been on military applications, recent advances in optical system components, system architecture, and pattern-recognition algorithms have spurred a real advance in research for commercial applications in computers, artificial intelligence, robotics, and product inspection.

This document offers the reader both a concise definition of optical processing theory and a description of the state of the art in optical processing applications.

LIST OF ABBREVIATIONS AND ACRONYMS

AO	acousto-optic
BSO	bismuth silicon oxide
CCD	charge coupled device
CdS	cadmium sulfide
CGH	computer-generated hologram
FK	Fukunaya-Koontz
FS	Foley-Sammon
GaAs	gallium arsenide
GaAs/AlGaAs	gallium arsenide/aluminum gallium arsenide
GaInAs	gallium indium arsenide
GaInAs/AlInAs	gallium indium arsenide/aluminum indium arsenide
GS	Gram-Schmidt
HeNe	helium-neon
KL	Karhunen-Loeve
LCLV	liquid crystal light valve
LCT	liquid crystal television
LED	light-emitting diode
LPE	liquid phase epitaxy
MBE	molecular beam epitaxy
MO-CVD	metal organic chemical vapor deposition
MOSLM	magneto-optic spatial light modulator
MQW	multiple quantum well
MSF	matched spatial filter
OMVM	optical matrix-vector multiplier
QCSE	quantum-confined Stark effect
RF	radio frequency
SDF	synthetic discriminant function
SEED	self-electro-optic effect device
SLM	spatial light modulator
TB	time-bandwidth product
TCNQ	tetracyanoquinodimethane
TNAP	11,11,12,12-tetracyano-2,6-napthoquinodimethane
WRD	wedge-ring detector

1. THEORY OF OPTICAL PROCESSING

An understanding of optical processing[1-3] is best obtained from an understanding of the natures of coherence, diffraction, and image formation, and how they are involved in the operation of simple optical systems.

If we consider an object plane, an electro-optic system, and an image plane (see Figure 1), the problem of forming the image of an extended object is dominated by the question of coherence. Every element of the object will emit waves, and the amplitudes of the waves will build a diffraction pattern in the image plane; the manner of combining these diffraction patterns depends entirely on the knowledge of the degree of coherence between the various elements of the object. We must consider three cases for all electro-optic systems because of the relative degree of coherence in the object illumination. These three cases are:

- coherent illumination
- incoherent illumination
- partially coherent illumination

1.1 COHERENT ILLUMINATION

In the case of coherent illumination, the various elements of the object have to be illuminated by a single very small source, S (see Figure 2). Consider a small object element in the central part of the object, O. The image of the small (i.e., not resolved by the system) object element will produce in the image plane a diffraction pattern whose amplitude distribution is shown in Figure 3. This diffraction pattern is represented by the law of repartition of amplitudes, E(M'), where M' is the position in the image plane of O'. If we displace the small source in the object by a quantity characterized by M of the same amount, its image is characterized by E(M'-M). If the object is extended where the repartition of amplitudes is A(M), every element of the

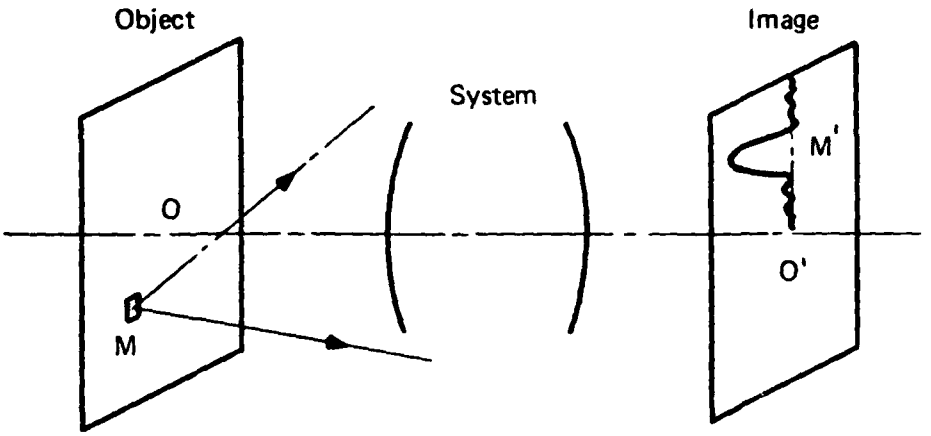


Figure 1. Concept of image formation with an electro-optic system.

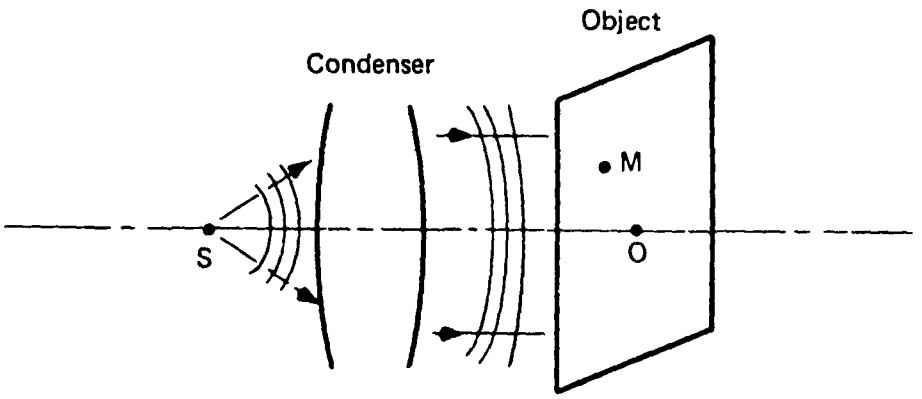


Figure 2. Coherent object illumination.

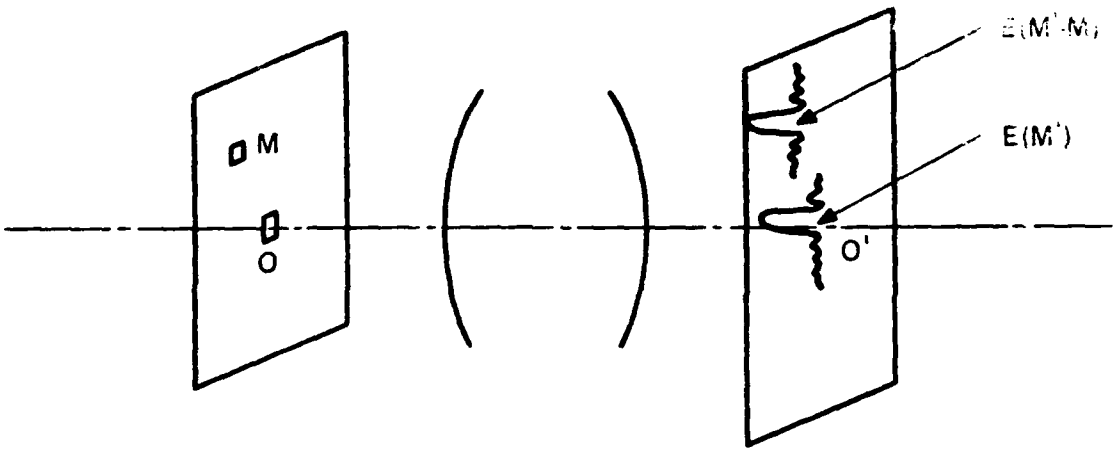


Figure 3. Diffraction image pattern E (M') of O at O' and its invariance over the image plane E (M'-M).

object will be responsible for a contribution A(M) E(M'-M), and the image will be represented by the function

$$A'(M') = \int A(M)\ E(M'-M)\ dM \qquad (1)$$

Equation 1 is a mathematical relation describing a convolution, and we see that the repartition of amplitudes A' in the image plane is obtained by the convolution of the repartition of amplitude in the object plane A and repartition of amplitude in the image of a point E and

$$A' = A * E \qquad (2)$$

The diffraction pattern E(M') associated with the image of an unresolved object O is the amplitude point spread function of the system. This function completely characterizes the image-forming properties of a system for coherent illumination, since from the Huygens-Fresnel principle of light propagation we know that the image plane distribution E(M') is the Fourier transform of the light's amplitude distribution in the exit pupil of the electro-optical system.

Ernst Abbe's experimental work in the 1870s on improving the performance of microscope objectives laid the foundation of the approach used today in considering coherent imaging. In experiments with periodic specimens, Abbe showed that the influence of a large aperture is associated with diffraction at the object. He demonstrated, as shown in Figure 4, how the diffraction maxima formed in the back focal plane of a lens contribute to the formation of an image, the higher orders (higher spatial frequencies) controlling the fine detail in the image. Abbe introduced the wave theory of image formation, which led to the modern concept that such systems act as spatial filters. This concept has made possible the development of current research in optical processing. Abbe's description of image formation in coherent light was interpreted in terms of Fourier series by A. B. Porter in 1906. Porter demonstrated the effects on the image of periodic objects when the various diffracted orders are prevented from contributing to the image. This work was the first to show the effect of spatially filtering for manipulation of image contrast.

1.2 INCOHERENT ILLUMINATION

The case of incoherent illumination is the exact opposite of coherent illumination. In this case, the various amplitudes from the object are assumed to be incoherent: the energies (not the amplitudes) have to be summed in order to know what the repartition of energy in the image plane is, and the mechanism of convolution will be applied to energies rather than to amplitudes. If O(M) is the repartition of luminance in the object, and if D (M') is the repartition of illumination in the image of an isolated central point, the illumination in the image is obtained by the convolution in the form

$$I(M') = \int D(M'-M)\ O(M)\ dM \qquad (3)$$

and as in Equation 1 we see that

$$I(M') = O * D \qquad (4)$$

so that the situation is very similar to the case of coherent illumination except that the convolution is applied to energies rather than to amplitudes.

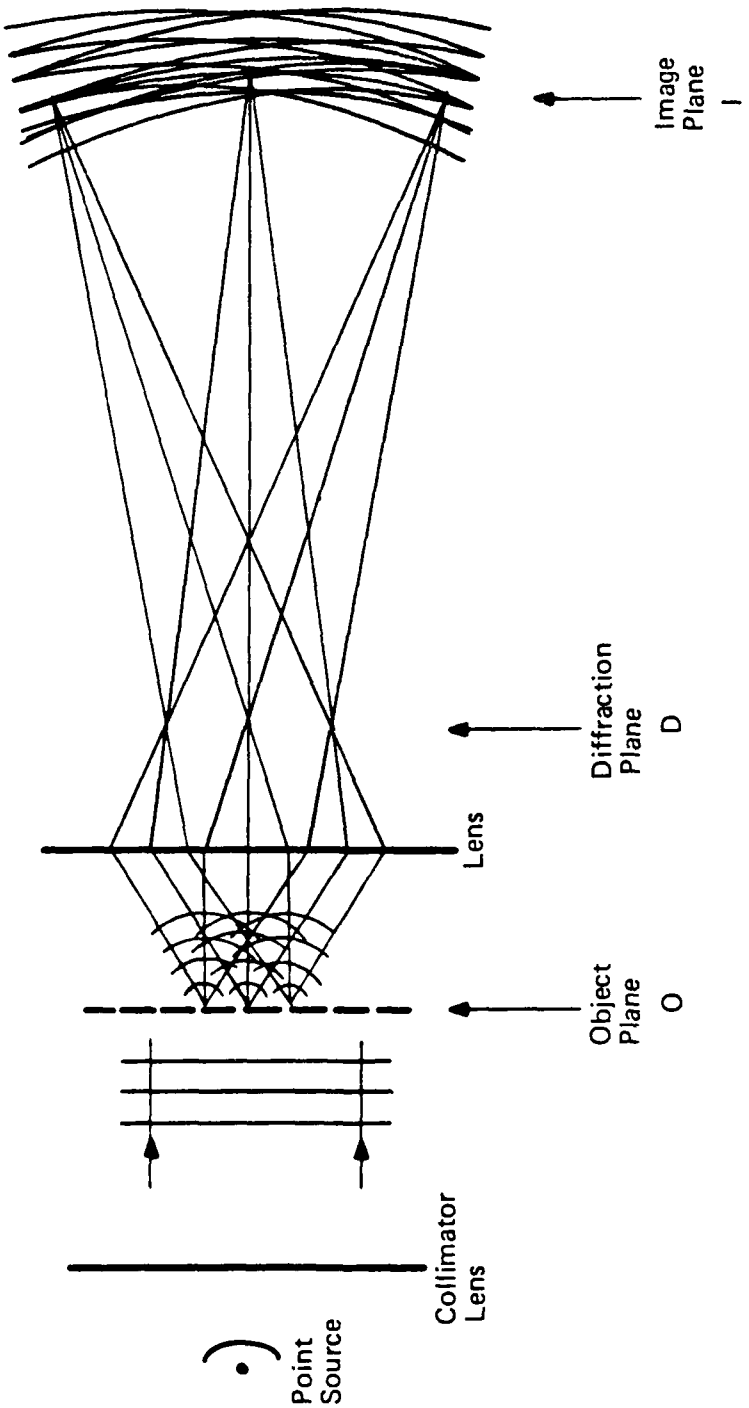


Figure 4. Abbe's demonstration of image formation in a microscope.

1.3 PARTIALLY COHERENT ILLUMINATION

In some cases, the conditions of full coherence or full incoherence are not satisfied, and the illumination in the image I (M') becomes

$$I(M') = \iint A(M_1)\, E(M'-M)\, A^*(M_2)\, E^*(M'-M_2)\, \gamma(M_2-M_1)\, dM_1\, dM_2 \qquad (5)$$

A^*, E^* are complex conjugates, and γ is a function that shows how the degree of coherence between the points of interest varies over the object plane.

In all three cases above, we see that the imaging properties of the optical system are described in terms of the diffraction pattern image of an unresolved object point, i.e., the system's point spread function. It is important to note that

$$D(M'-M) = E\,(M'-M)\, E^*(M'-M) \qquad (6)$$

Since most of the work in optical processing is based on coherent illumination, we will concentrate on this case and on how the work of Abbe led to the modern concept that such systems act as spatial filters, modifying the spatial frequency spectrum of the wave amplitudes transmitted by the object and thus producing a spatially filtered image. Consider a one-dimensional, multiple-slit transmission grating as an object being imaged by a lens (see Figure 4). Wavefronts constituting the various diffraction maxima are brought to separate foci in the back focal plane of the lens, and it is the light that passes through these foci in the diffraction plane that combines in the image plane to produce an optical reconstruction of the object. Alone, any pair of the foci produce a set of sinusoidal bands in the image plane. In this sense, image formation can be thought of as a double process of diffraction--an approach put forth by Fritz Zernike in 1935, when he described his invention of the phase-contrast microscope. Figure 5 shows how an image is built up in this way, where the pairs of beams diffracted by the object grating in the nth order combine in the image plane to interfere and produce a harmonic distribution of illumination, with period D'_n given by

$$D'_n \sin \theta'_n = \lambda \qquad (7)$$

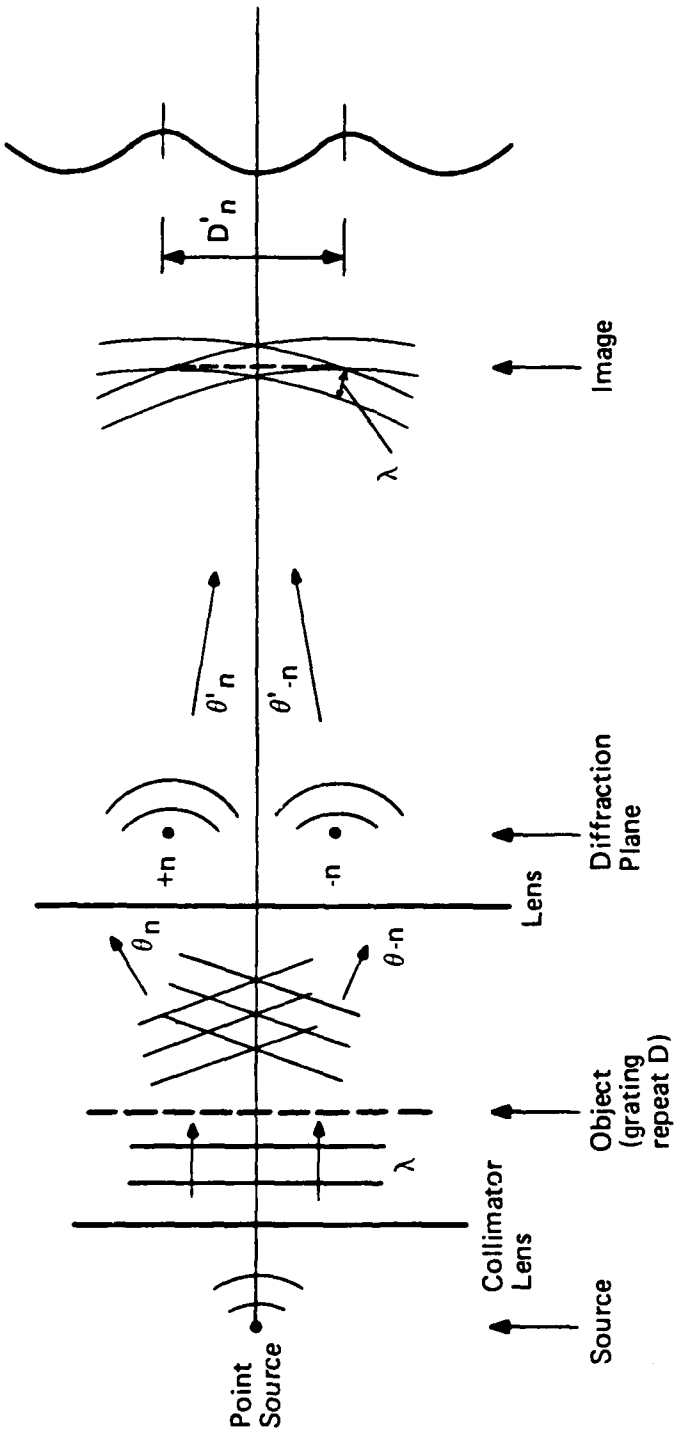


Figure 5. Image formation: Diffraction and recombinations.

The condition for the formation of the nth order diffraction maxima at the object is

$$D \sin \theta_n = N\lambda \tag{8}$$

where

$$D'_n = \frac{D}{N} \left(\frac{\sin \theta_n}{\sin \theta'_n}\right) \tag{9}$$

and $\sin \theta_n / \sin \theta'_n$ is related to the magnifying power of the lens.

The first-order pair of maxima from the object interfere in the image plane to give a simple-harmonic variation of illumination that corresponds to the basic period of the object grating; this period is the bare minimum of information about the object, with no fine detail of its optical structure. Each pair of successively higher-order maxima adds a harmonic of successively shorter period to the total illumination that forms the image, until the full detail of the image is built up by what is clearly recognizable as Fourier synthesis. In the normal illumination of a periodic object (i.e., grating) by a coherent plane wave, the diffraction maxima in the various diffracted orders comprise the Fourier analysis of the object grating, and the diffraction plane on which the various diffracted orders lie is referred to as the Fourier plane. Thus the image formation process in Figure 5 can be regarded as a double Fourier process, with the diffraction pattern representing a Fourier analysis of the object grating, and the image representing a Fourier synthesis of the Fourier analysis. This concept of a double Fourier description is in accord with the double diffraction interpretation and is both very important and responsible for the development of the many directions in modern optical processing research.

For perfect imaging, an infinite Fourier synthesis would be required, which would require the generation of an infinite number of diffraction orders and the transmission of all of them through the optical system to the image plane. This is not possible. Equation 8 shows how the values of D and $\sin \theta_n$ impose a limit to the number of possible diffraction maxima that can be produced by any lens with a finite aperture ($\sin \theta < 1$).

The previous description of periodic objects can be extended to nonperiodic objects because discrete orders of diffraction are not a prerequisite, and a nonperiodic object can be regarded as an aperture of a grating. In this case the diffraction pattern, instead of being a Fourier series, becomes a Fourier transform. This means that the formation of a diffraction image is expressed by a Fourier transform. Thus, if we assume the geometric optical description of a lens with an entrance pupil, exit pupil, and front and rear focal points, the repartition of amplitude of an object A will produce a Fourier transform T(A) in the entrance pupil. The exit pupil is the image of the entrance pupil formed by the lens. For a perfect lens, neglecting aberrations and magnification effects, the amplitude distributions in the entrance and exit pupils are equal. In the same way, the exit pupil plane is the Fourier transform of the image, and the correspondence between pupil and image is also a Fourier transform. This is represented in Figure 6. However, since any actual lens will not perfectly image a point to a point, the Fourier transform in the exit pupil T(A') is not equal to T(A), and

$$T(A') = T(E)\ T(A) \tag{10}$$

where T(E) is the transform of the image, E, of an isolated central point. For an isolated central point object, the amplitudes on the pupil become generally a constant inside the aperture of the pupil, and the physical significance of Equation 10 is very simple; in this case, the Fourier transform of the object located on the pupil plane is filtered by the pupil itself in order to produce the Fourier transform of the image. This case gives a simple law of transmission of object spatial frequencies. This law is a rectangular law (see Figure 7) having a limiting frequency α_0 corresponding to the amplitude located at the edge of the pupil where α_0 is the angular aperture of the lens's entrance pupil, and $\alpha_0 = \lambda/D$.

From the preceding description of coherent image formation it is possible to readily understand the nature of optical processing systems. The fundamental operation performed in a coherent optical processor (or optical computer) is the Fourier transform. We have just seen how diffraction in an optical system is a Fourier-related process; understanding this concept is vital to developing real insight into the technology of optical processing.

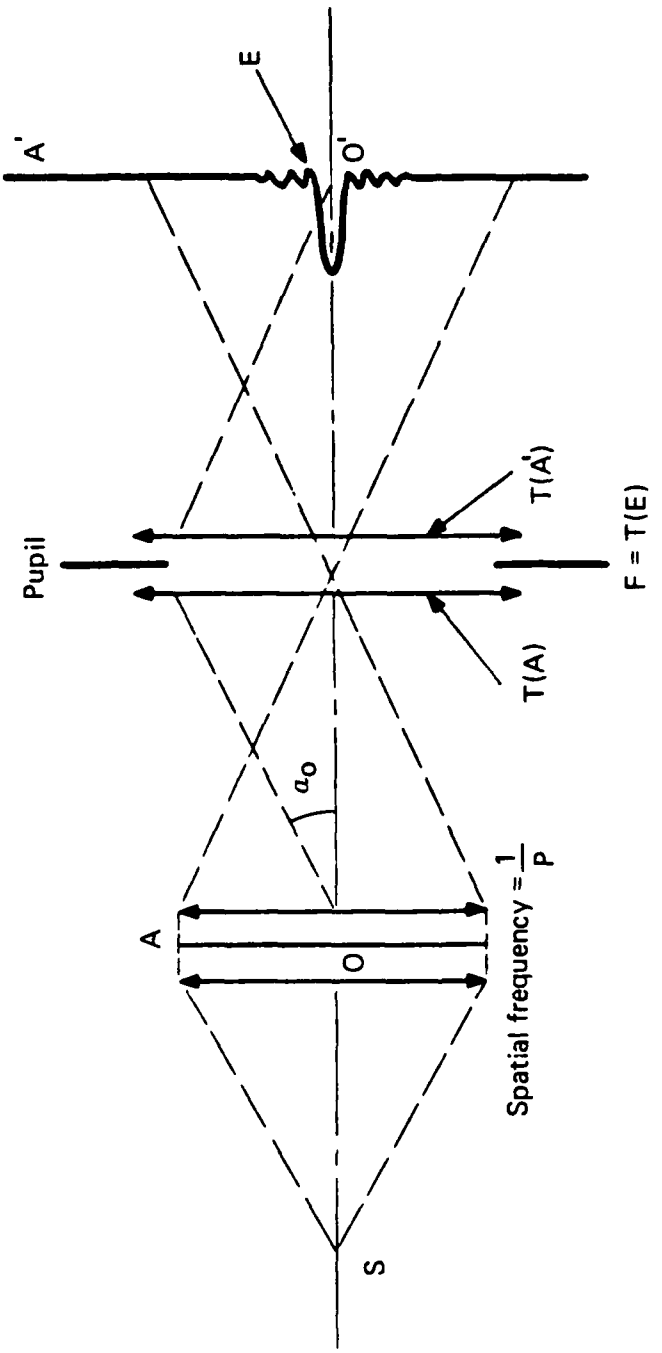


Figure 6. Fourier transform description of image formation.

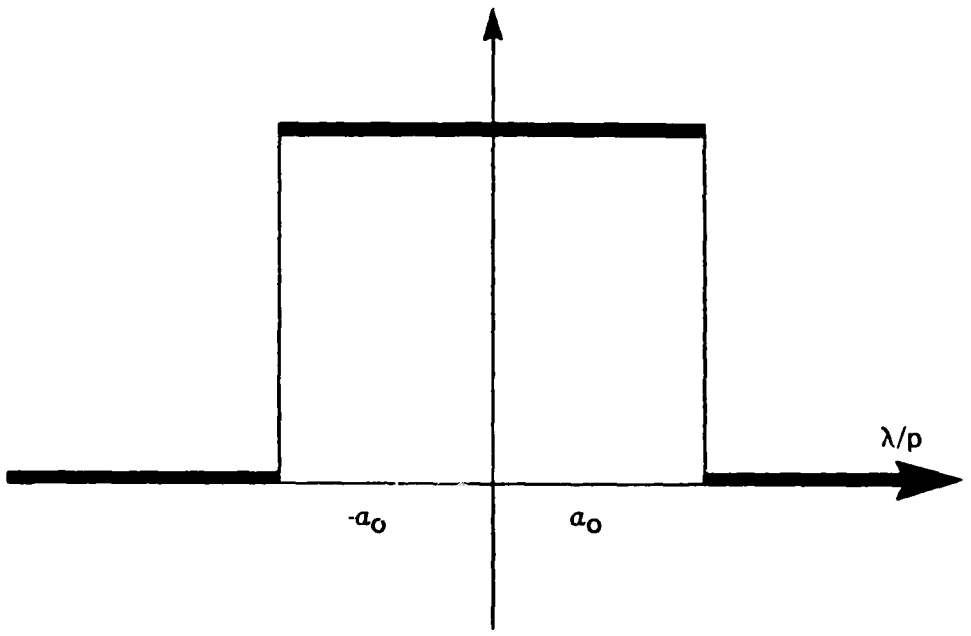


Figure 7. Rectangular band-pass of transmitted spatial frequencies.

From the discussion of coherence and the Fourier transform relationship between the pupil plane and image plane of any optical system (e.g., a lens), we are able to expand our explanation to describe a basic optical processing system. Consider Figure 8, where the plane P_1 is the input plane. The input plane lies one focal plane in front of lens L_1 and has an amplitude transmittance $g(x_1,y_1)$. If P_1 is illuminated by a uniform coherent and monochromatic plane wave of amplitude U_0, the amplitude distribution in the rear focal plane P_2 of the lens will be the complex two-dimensional Fourier transform G of g. This Fourier relationship is written as

$$G(u,v) = G\left(\frac{x_2}{\lambda f_1}, \frac{y_2}{\lambda f_1}\right) = \frac{U_0}{\lambda f_1} \iint_{\infty}^{\infty} g(x_1,y_1) \exp\left[-2\pi i(ux_1+vy_1)\right] dx_1\, dy_1 \qquad (11)$$

where input spatial functions are denoted by lower-case variables and their Fourier transforms (FT) by their corresponding upper-case letters. The incident amplitude distribution $U_0(x,y)$ is a complex quantity equal to $a_0(x_1,y_1)$ $\exp\left[i\phi_0\ (x_1,y_1)\right]$. Convention calls for the input wave to have a uniform

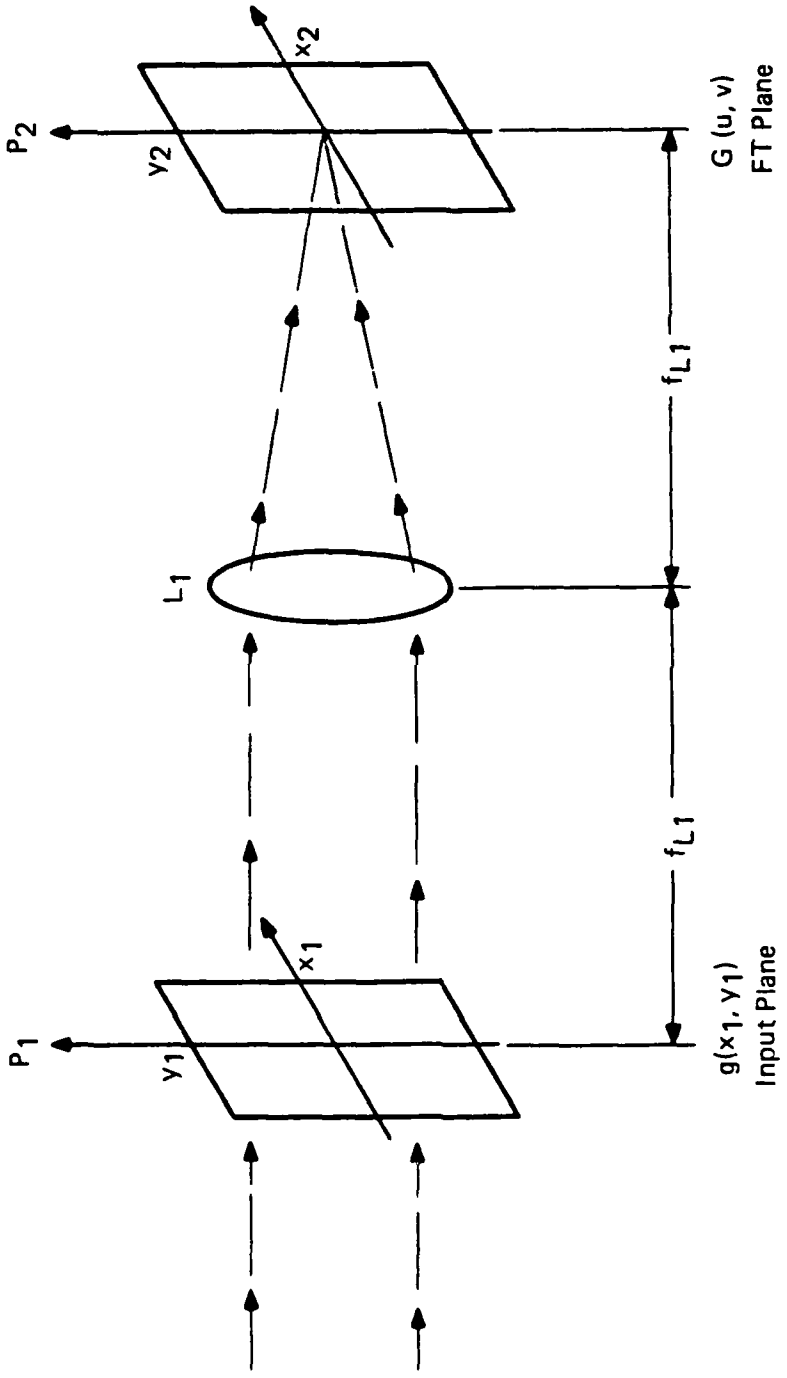


Figure 8. Optical Fourier system.

amplitude and phase, so U_0 is a constant, and after the incident plane wave passes through plane P_1 the light amplitude distribution is gU_0. Thus, passing one amplitude distribution through another produces a multiplication of the two distributions.

Although the amplitude of the light distribution incident on P_2 is a Fourier transform, the detectors used for recording in plane P_2 are only capable of recording the intensity or absolute values of Equation 11. In addition, for most applications in optical processing, the transmittance of P_1 will generally be both real and positive. Thus, the complex input data for plane P_1 must be recorded by quadrature carrier modulation and on a bias, i.e., a hologram or interferogram.

In these types of optical systems the variables used to describe the various functions are spatial instead of temporal. This is only logical, since any recording in the x,y plane of Figure 8 is a snapshot in the time history of a given signal. The spatial coordinates u,v of P_2 have units of cycles divided by length (i.e., cycles/mm), and are related to the coordinates x_2,y_2 of P_2 by

$$u = \frac{x_2}{\lambda f_1}$$

$$v = \frac{y_2}{\lambda f_1} \qquad (12)$$

where λ is the wavelength of the input monochromatic plane wave. From the study of diffraction and propagation of light we know that the propagation of a wave $U_0(x,y)$ from a plane (x,y) through a distance d to a plane with coordinates (x_1,y_1) results in a wave $U_1(x_1,y_1)$ at plane x_1,y_1 that is related to $U_0(x,y)$ by

$$U_1(x_1,y_1) = \iint U_0(x,y) \exp\{i(2\pi/\lambda d)[(x-x_1)^2+(y-y_1)]\}dxdy \qquad (13)$$

From this discussion it is evident that the optical Fourier system of Figure 8 consists of lenses, empty space, and intensity modulators. Typical intensity modulators include film transparencies, light-emitting diode (LED) arrays, liquid crystal displays, and light valves. Propagation of the input wave U_0 through P_1 can be described by a multiplication of U_0 by g; propagation from P_1 to L_1 is described by Equation 13; propagation through L_1 is equivalent to a multiplication by the transfer function T_2 of lens L_1, where

$$t_L = \exp\left[-2\pi i\,(x^2+y^2)/\lambda f_1\right] \tag{14}$$

and propagation from the lens L_1 to plane P_2 is given again by Equation 13. Thus

$$U_2(x_2,y_2) = \frac{U_0}{\lambda f_1} \iint g(x_1,y_1) \exp\left[-2\pi i(x_1x_2+y_1y_2)/\lambda f_1\right] dx_1 dy_1$$

or

$$U_2(x_2,y_2) = \frac{U_0}{\lambda f_1} G\left(\frac{x_2}{\lambda f_1},\frac{y_2}{\lambda f_1}\right) = \frac{U_0}{\lambda f_1} G(u,v) \tag{15}$$

In the application of the optical Fourier system to optical processing or optical computing, the basic properties of Fourier transform relationships and their simple implementation require description. The light distributed in the Fourier transform plane P_2 of Figure 8 is an ensemble of points of light, each of which indicates the presence of a particular spatial frequency in the input. The distance of a point of light from the origin is proportional to the input spatial frequency, with higher spatial frequencies lying further from the origin, as shown by Equation 12. The angle at which the point of light lies with respect to the x_2 or y_2 axes shows the orientation of the input data with respect to the x_1 and y_1 axes. The relative amplitudes of the points of light indicate the relative distribution of the spatial frequency content of the input pattern. The Fourier transform pairs that are implemented in optical processing systems include the common ones where

$$g(x) \Rightarrow G(u) \tag{16}$$

$$g(ax) \Rightarrow \frac{1}{|a|} G\left(\frac{u}{a}\right) \tag{17}$$

$$\delta(x-a) \Rightarrow \exp(-2\pi iua) \tag{18}$$

$$g(x-a) \Rightarrow G(u) \exp(-2\pi iua) \tag{19}$$

$$\exp(2\pi iU_0 x)\, g(x) \Rightarrow G(u-u_0) \tag{20}$$

$$g^*(x) \Rightarrow G^*(-u) \tag{21}$$

and the inverse transform functions

$$G(u) \Rightarrow g(x) \tag{22}$$

$$\exp(2\pi ua) \Rightarrow \delta(u+a) \tag{23}$$

$$G(u)\, H(u) \Rightarrow g(x) * h(x) \quad \text{(convolution)} \tag{24}$$

$$G(u)\, H^*(u) \Rightarrow g(x) \otimes h(x) \quad \text{(correlation)} \tag{25}$$

Equations 16 and 22 are the definitions of the basic Fourier transform. Equations 17 and 19 show that the position of the input function in the input plane is encoded as a linear phase term in the Fourier transform plane. The Fourier transform G itself is complex, and because of this the phase of the Fourier transform must be preserved. The Fourier transform always appears centered on the axis at P_2; this is very important in many practical applications, as will be seen later in the discussions of various applications of optical processing. Equation 20 shows that the Fourier transform of the function g(x) modulating a complex spatial frequency carrier u_0 is the Fourier transform G of the function centered at the spatial frequency coordinate u_0 in the Fourier transform plane P_2. Equation 21 shows the relationship of the Fourier transform of the complex conjugate G^* of g and the Fourier transform of g. Equations 22 through 25 are illustrated in Figure 9. In that figure,

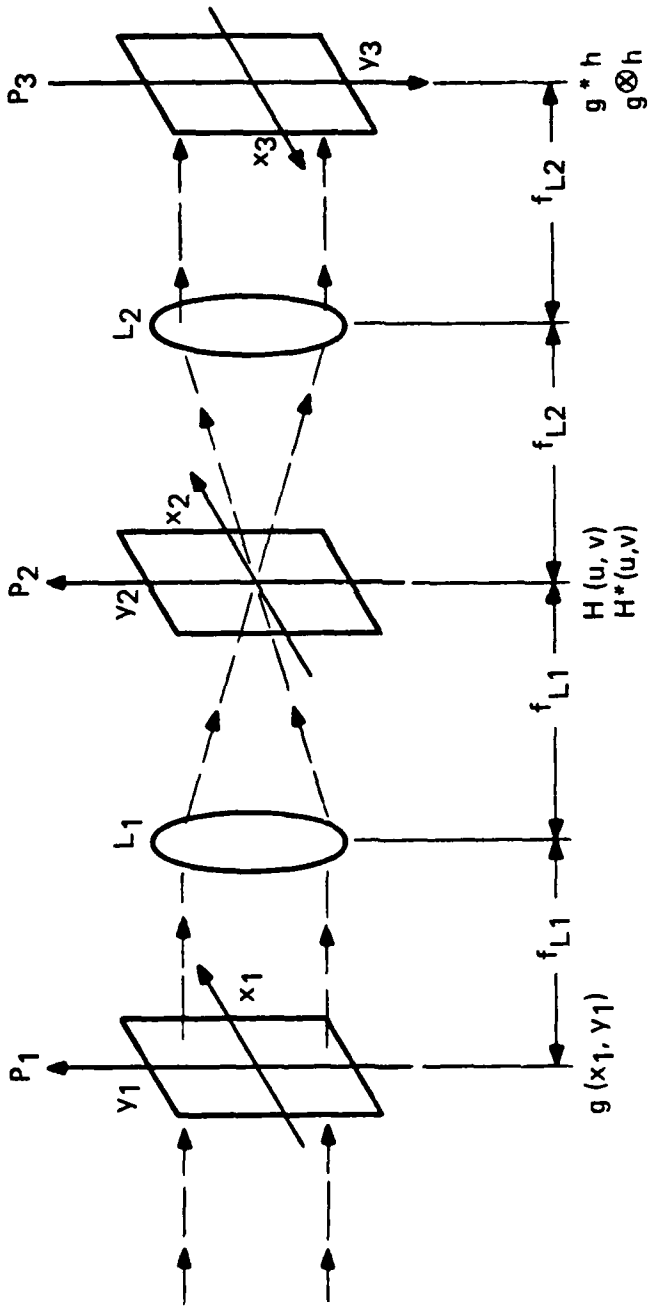


Figure 9. Simple optical data processing system.

lens L_1 produces the Fourier transform G of g at P_2 and lens L_2 produces the Fourier transform of the Fourier transform of g. At P_3 we find g(-x,-y), which is a reversed and inverted image of the input g(x,y), and shows that an optical system can only perform forward Fourier transforms. The inverse transform of G is

$$g(x) = F^{-1}\,[G(u)] \tag{26}$$

The input functions to be transformed in Equations 22 through 25 correspond to the light distributions transmitted through plane P_2, and the coordinates of this plane are spatial frequencies. The output function g(x) in Equation 26 is the output of the output plane P_3. By the choice of the inverted x,y coordinate of Figure 9 in plane x_3,y_3 it is possible to realize an optical inverse Fourier transform, and in this way the lens L_2 can be defined to perform an inverse transform. This assumption is only a pretext to allow the equations of Fourier transform theory to be used without the need to change variables. Equations 22 and 23 show the reversibility of previous Fourier transform pairs; Equation 24 shows that an optical processor can be used to perform a convolution operation, and Equation 25 shows that a correlation operation can also be performed, i.e., the use of H(u,v) or $H^*(u,v)$.

The use of these basic properties for linear optical systems has produced a rapid growth in research and development of optical processing technologies since the late 1950s. This work has involved the development not only of new applications, but also of the components and systems required for implementation of the various applications. The applications developed have included those in image, signal, and numerical processors. Image processing applications have involved image enhancement, restoration, preprocessing, feature extraction, and pattern recognition. Signal processing refers to spectrum analysis: extracting range information; extracting Doppler frequency shift information; extracting angle- or direction-of-arrival and time-of-arrival information from received signals; and synchronization and message demodulation of communications and spread spectrum data. Numerical processing applications include discrete processing applications for matrix-vector multiplication.

Recent work in developing an alternative to analog matrix-vector multiplication has been directed toward digital numerical processing techniques using nonlinear optical devices. This work has involved the development of nonlinear bistable devices for implementation of binary and residue digital logic.

2. SIGNAL PROCESSING APPLICATIONS

The first class of problems to be solved using optical processing technologies involved signal processing. Although the Fourier transform property of a lens offers some unique advantages over electronic systems in two-dimensional problems, parallel computations, high-speed data rates, and simple performance of the Fourier transform operation, the Fourier transform property of a lens suffers from the limitations of being an analog system, an inability to make decisions, and difficulty in programming. In order to overcome the limitations of optical systems in solving other than image problems, hybrid optical-electronic processors were developed to solve problems such as radar signal processing.

Hybrid systems have been designed to perform many different tasks, and because of this diversity they often bear little resemblance to each other. Figure 10 is a block diagram of a generalized hybrid system that includes most of the elements of practical systems now in use. In a hybrid optical-electronic system the interface between the optical and electronic systems is a major design problem. Three interface subsystems are shown in Figure 10. The input interface is required to convert the raw input signal (a one-dimensional electrical signal) into a two-dimensional optical signal suitable as input to the optical processor. The output interface converts the two-dimensional optical output of the optical processor into the desired output format. This format could be as simple as displaying the optical output on a screen, or could require converting the optical output into an electrical signal. The control interface is provided for controlling the optical processor, which makes it possible to program the processor automatically. The control interface will depend significantly on the specific processor used in the hybrid system. The three interfaces (i.e., input, control, and output) often are connected to a central controller, usually a digital electronic computer because of its flexibility and ease of programming. The operator programs the digital computer to control the input, analyze the output, and change the operation of the optical processor. The last element of the hybrid

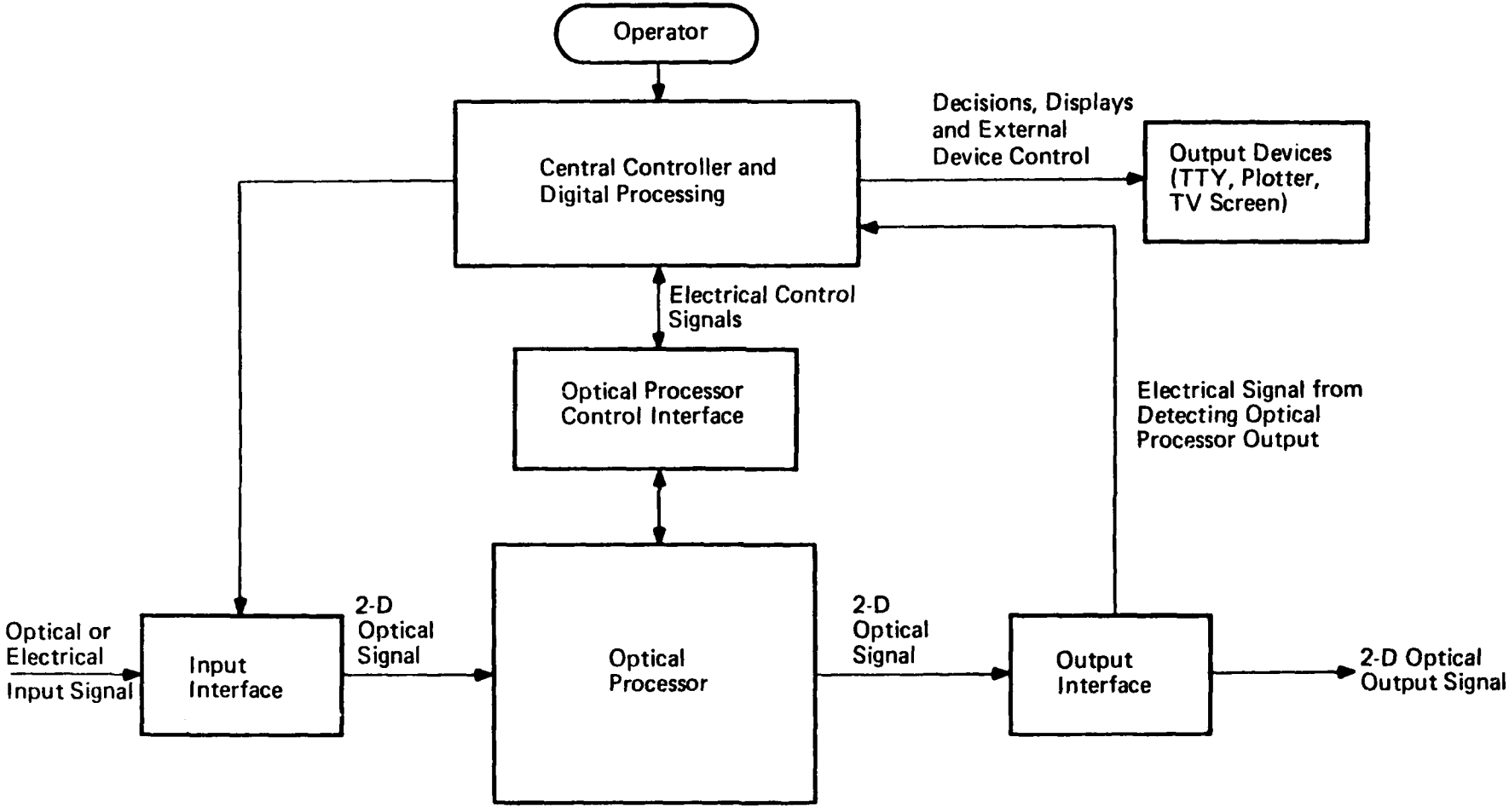


Figure 10. Generalized hybrid system.

system is the optical processor itself, which is used to perform the desired mathematical operation on the two-dimensional optical signal. The Fourier-optical processor configuration has been applied to a diverse family of radar signal processing functions including detection of radar returns; estimating radar target range, radar range gate, and target angular position for phased array radars; and mapping target scattering distributions for synthetic-aperture radar. Fourier-optical processors are successful as radar signal processors because they supply a flexible linear filter implementation that can be adapted to the particular processing needs of radar.

2.1 SYNTHETIC-APERTURE RADAR

The first significant technique developed for practical applications of optical processing was developed in the 1950s to process data gathered by a new type of radar. This system was developed because of the need for an all-weather, day-or-night imaging system. In order to penetrate clouds, the system employed radiation that is not attenuated or dispersed by water vapor, i.e., microwave radiation with a wavelength between 1 and 30 cm. To perform high-resolution mapping from an aircraft, however, would require a radar antenna too large to fit on the aircraft. The limiting resolution is proportional to the ratio between the wavelength of the radiation source and the size of the radar antenna. As an example, a 6000-m antenna would be required to provide the resolution of a 10-cm aperture camera lens in the visible. To overcome the limited resolution of a conventional circular-scan radar, the approach selected was to simulate a large antenna by continually moving a smaller one. This was accomplished by mounting a fixed 5-m antenna on the belly of an aircraft and collecting the returns of transmitted radar signals from an aircraft in flight. The return radar signals are integrated over time to obtain the results corresponding to those from a stationary large antenna. The problem of this synthetic-aperture radar was that it required complex processing to convert this huge collection of radar return data into a useful map. It was while working on this problem that Emmett Leith and a group at the University of Michigan's Willow Run Laboratories developed an optical technique for processing the return data of the synthetic-aperture radar. In their system, the return radar signal controlled the intensity of an electron-beam spot sweeping across a cathode ray tube. Each sweep (corresponding to a

single radar pulse) writes a line across a strip of photographic film, and the film moves lengthwise between sweeps. The pattern recorded on the film contains all the range and reflectivity data the radar has collected. The reconstruction process involves taking the inverse transform of the recorded diffraction pattern to produce a map of the radar-scanned terrain. Figure 11 shows how the radar signals are transmitted and received from a radar on an aircraft if flight. Figure 12 shows how the received radar signal is used to modulate a cathode ray tube display, and how this image is recorded on a moving strip of film. Figure 13 shows an expanded view of the recorded radar sweeps and the corresponding slice through a circular diffraction pattern (Fresnel zone plate). Figure 14 shows one means of using a coherent light processor to reconstruct the image of the terrain.[4] In the reconstruction process, a collimated beam of light is used to illuminate the moving data film, and after diffraction of the light by the data film and imaging by a cylindrical lens, the radar image is formed on a moving image film. In this way, one generates a continuous picture of the terrain for the entire length of the aircraft's flight. In operation, a conical lens is placed in front of the data film of Figure 14 to compensate for the reconstructed image's tilted image plane. This effect is produced by differences in the scattered wavefronts from terrain as a function of their distance from the moving aircraft. Although extremely complex, this solution for synthetic-aperture radar imagery continues to be one of the most important applications for optical processing.

2.2 ONE-DIMENSIONAL ACOUSTO-OPTIC MODULATORS FOR RADAR OPTICAL SIGNAL PROCESSING

In many applications of optical signal processing the input/output is a rapidly changing and one-dimensional signal. For this important class of problem that includes outputs such as range, velocity, and acceleration, a one-dimensional Bragg cell acousto-optic modulator is usually used as the electronic/optic-to-optical transducer. The major attributes of these devices are their ability to accommodate high carrier frequencies (1 GHz) and signal bandwidths (300 MHz).

In an acousto-optic cell, an ultrasound generator sends a high-frequency acoustic wave through a block of material, producing a periodic variation in

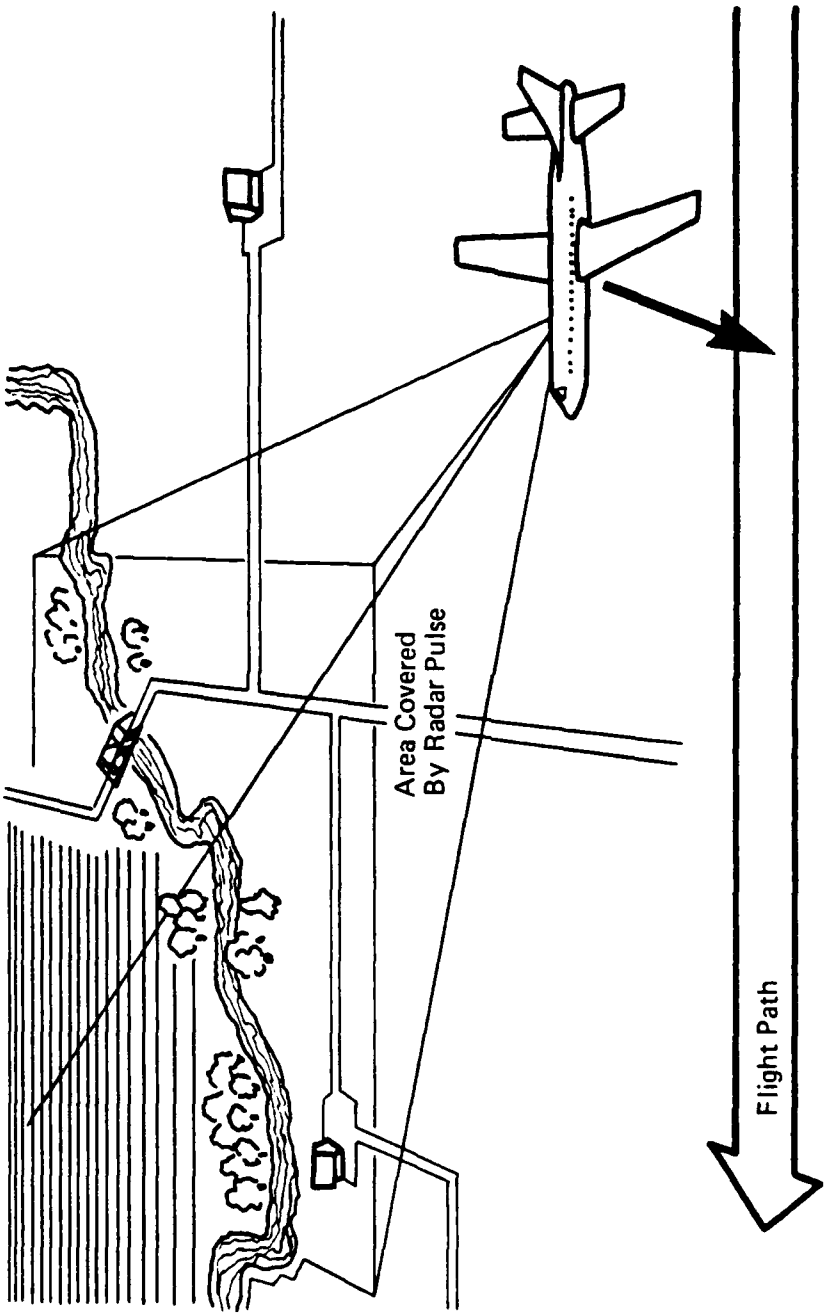


Figure 11. Transmission and collection of radar pulses from a moving aircraft.

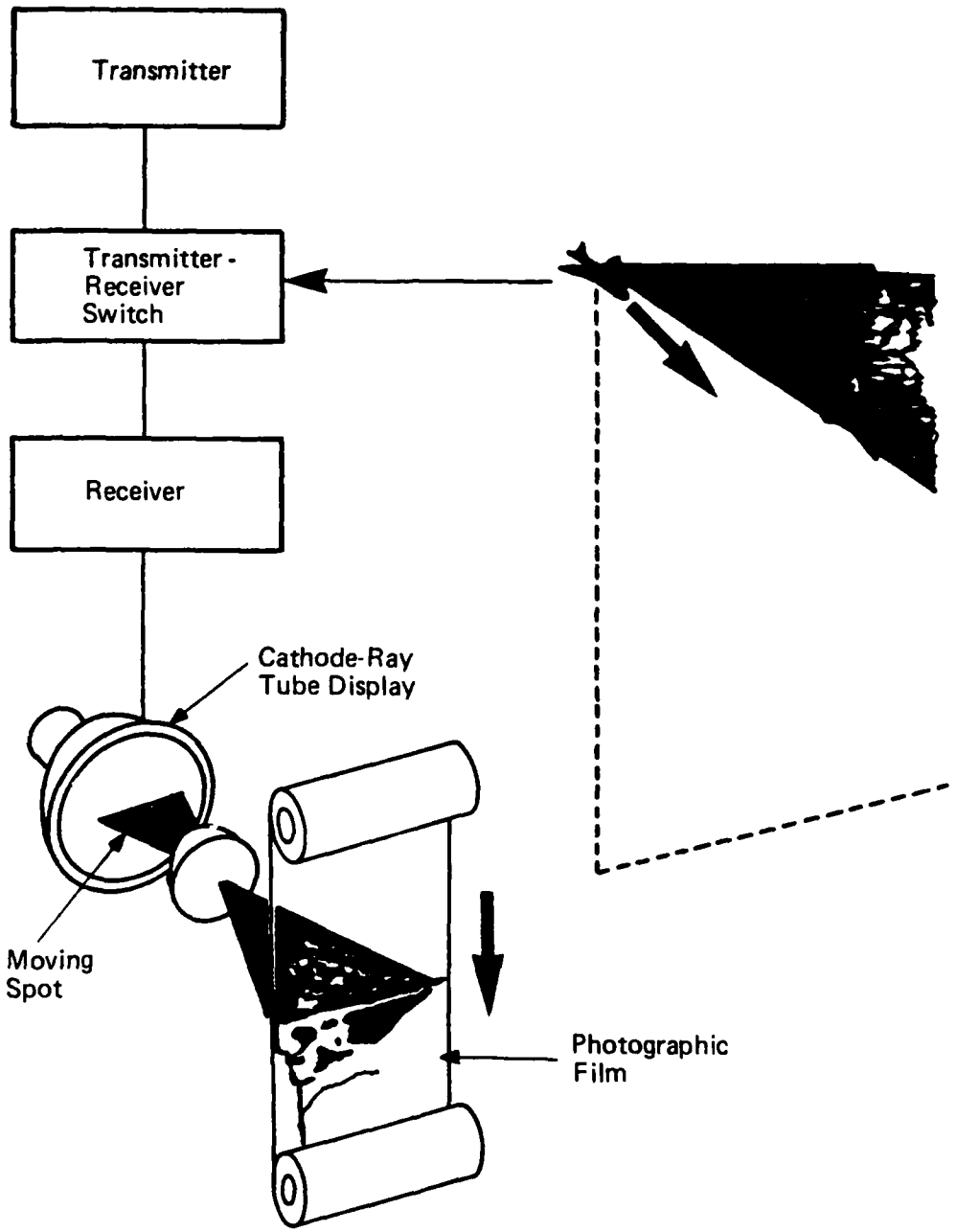


Figure 12. Concept for transmitting and receiving radar signals and recording of images on a cathode-ray tube display.

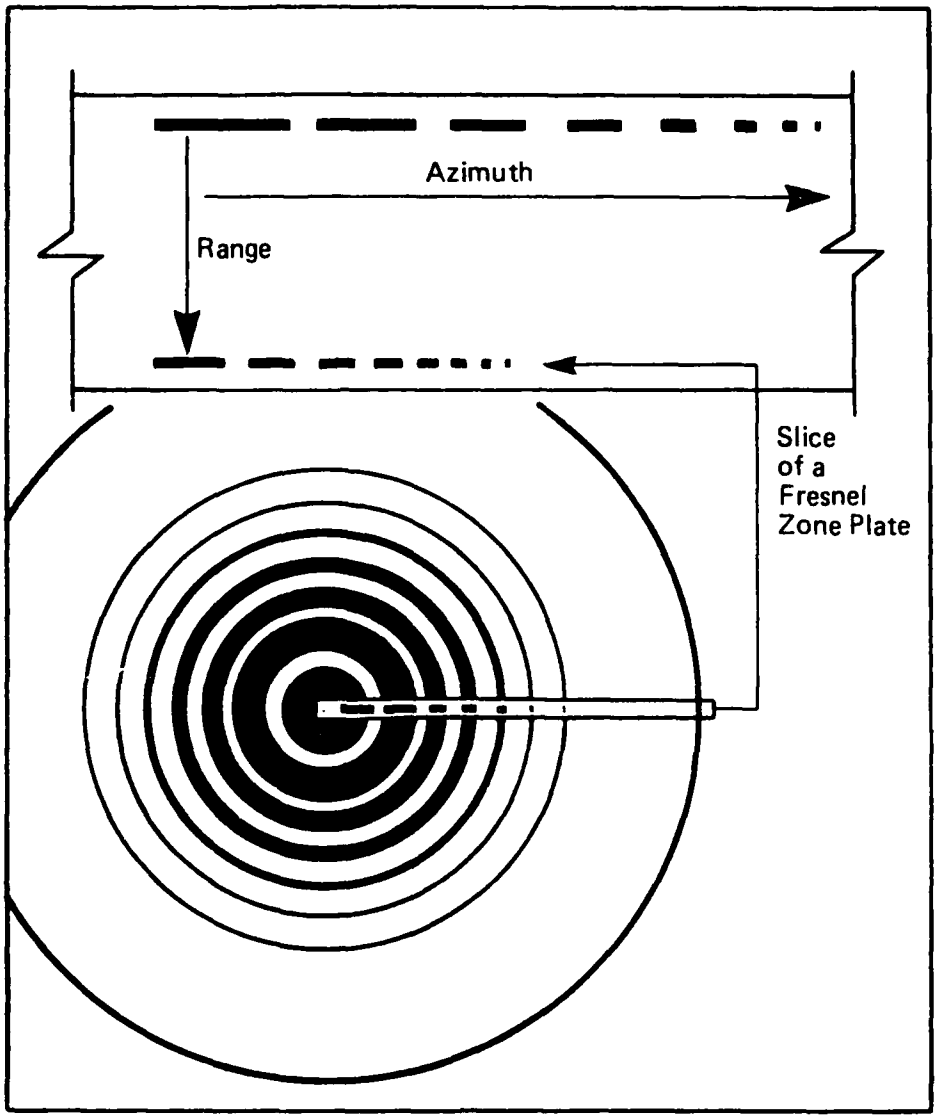


Figure 13. Recorded radar sweep and corresponding slice of Fresnel zone diffraction pattern.

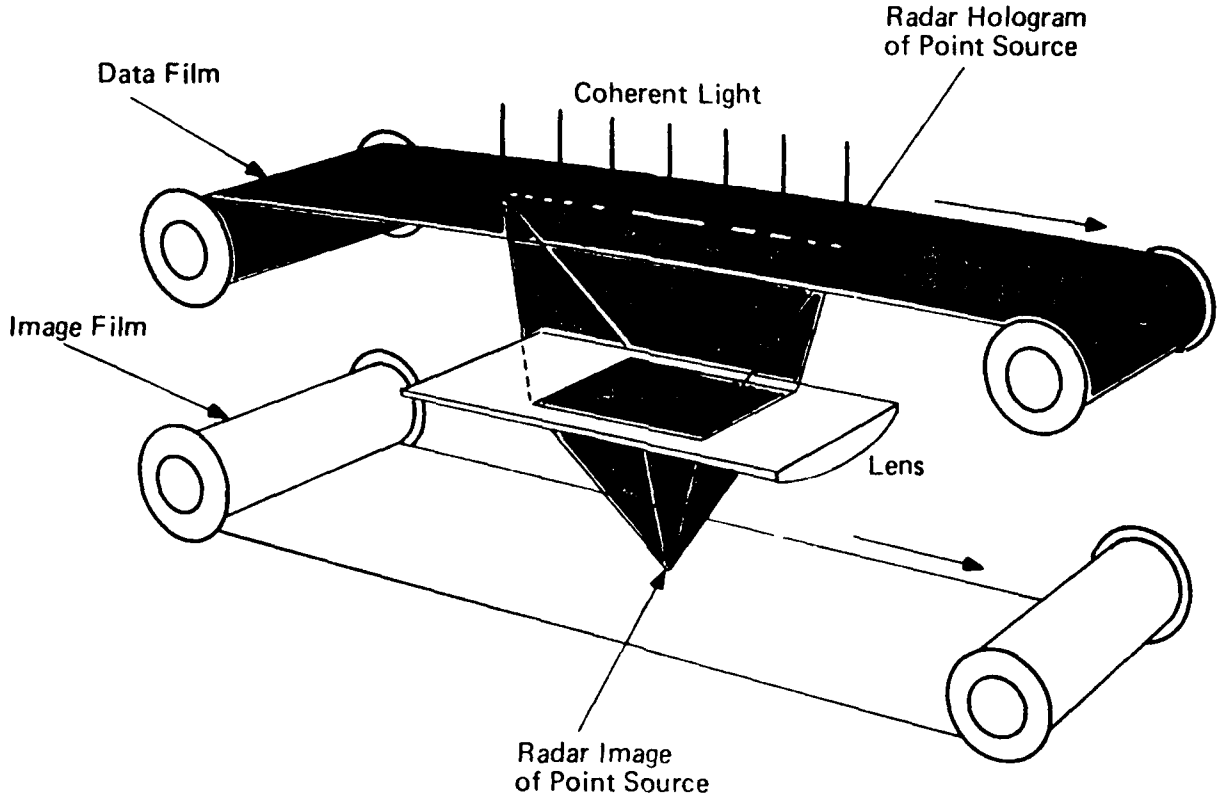


Figure 14. Optical processor concept for terrain image forming for a synthetic aperture radar.

the material's density and hence in its refractive index. A beam of light traveling perpendicular to the motion of the acoustic wave is deflected at an angle that varies according to the spacing of the dense regions in the cell; this spacing, in turn, depends on the frequency of the acoustic wave. Thus, different acoustic frequencies scatter light at different angles.

The electronic signal to be analyzed drives the ultrasound generator, which creates a refractive-index wave pattern in the cell. The cell is fabricated from materials, such as quartz or tellurium dioxide, that provide a high acoustic response. Figure 15 shows the basic structure and operation of an acousto-optic cell. An input signal of amplitude A and temporal frequency ω_1 is fed to a transducer bonded to the cell. This converts the input electrical signal to an acoustic wave that propagates across the cell. The acoustic wave can be described as a sine-wave grating with spatial frequency proportional to ω_1 and modulation proportional to A. When light enters the

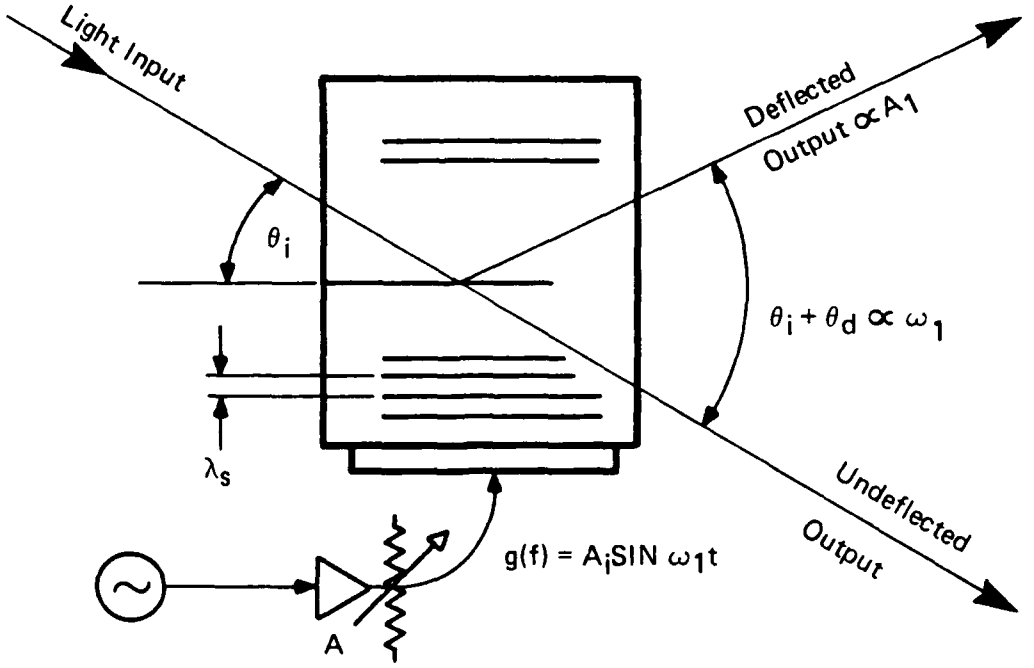


Figure 15. Description of acousto-optic modulator operation.

cell at the Bragg angle θ_B, it will be diffracted. The angle of diffraction is proportional to ω_1, and the amount of diffracted light is proportional to A. When multiple signals of different amplitudes A_n and frequencies ω_n are present in the input data, multiple light waves are diffracted at angles proportional to ω_n and with amplitudes proportional to A_n. A lens placed behind the cell will focus these different light waves at different spatial locations in the rear focal plane of the lens. Such a system forms the Fourier transform of a complex input signal: multiple input signals are separated according to their temporal frequencies.

Shifting the input frequency linearly causes the output beam to scan left and right or up and down; in this way, the modulator becomes a beam deflector or scanner.

In another mode of operation, where both temporal and spatial effects are considered, the input signal is represented by g(t) and the transmittance of the modulator is described by $g(t-\tau)$. The shift variable τ equals $\frac{x}{v}$, where x is the distance along the modulator and v is the velocity of propagation of the acoustic wave. Both the temporal and spatial transmittance variables of the cell and their coupling are used so that the spatial transmittance of the modulator is controlled as a function of time.

In the operation of an acousto-optic modulator, the light diffracted consists of three beams (orders) corresponding to the 0, ±1 order beams. In use, only one of the first-order diffracted beams is required. Depending on the cell type, the acoustical wave frequency is typically in the 20 MHz to 2 GHz range, and the diffraction angles will be less than 10 degrees. To avoid nonlinear interactions that result in additional diffraction components, the Bragg cell is operated at relatively low diffraction efficiencies when multiple frequencies and associated gratings are present.

Rhodes and Guilfoyle[5] describe acousto-optic signal processing architectures, and include an excellent review of the fundamental limitations on acousto-optic (AO) Bragg cell operation. They stress the importance of understanding the fundamental limitations of AO cell operation in order to determine the limitations of the devices in specific processing configurations. They first examine beam modulation capability. In Figure 16, each input beam illuminates a segment of acoustic wave signal of duration T_B (equal

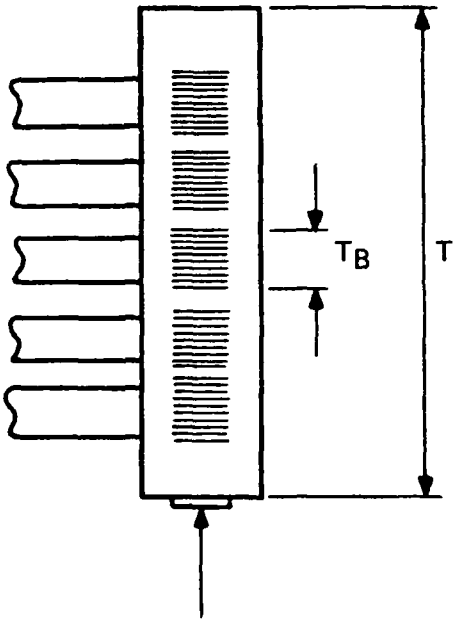


Figure 16. Fundamental limitations on multiple input-output beam modulator. Acoustic wave traverses beam width in time T_B, entire cell in time T.

to the width of the illuminating beam divided by the acoustic wave velocity). Since the signal driving the Bragg cell cannot change at a rate exceeding B, the bandwidth of the cell (usually limited by the attenuation of acoustic waves in the cell that are above and below material-dependent cutoff frequencies), T_B must satisfy the condition

$$T_B > \frac{1}{B} \tag{26}$$

if the intensity of each input beam is to be modulated essentially independently. If the acoustic wave transit time for the entire cell is T seconds, then the maximum number of inputs and outputs, N, is restricted by

$$N = \frac{T}{T_B} < TB \tag{27}$$

where TB is the time-bandwidth product of the cell. This number typically is between 200 and 2000. To assure independence of beam modulation from one beam to another, a practical upper limit on the number of interconnections is lower

than that--perhaps between 100 and 500. Practical packaging difficulties suggest a number toward the lower end of that range.

Next, Rhodes and Guilfoyle establish fundamental limitations on the beam deflector approach. Figure 17a assumes that there is one input beam. The configuration shown is essentially that of an AO spectrum analyzer,[6] with performance limited by diffraction. Light from the single input beam can be directed, in parallel and with individually controlled weighting, to any combination of N outputs, so long as N is less than the time-bandwidth product of the cell, i.e., $N < TB$. This upper limit is determined by the inverse relationship between the size of the illuminating beam and the diffraction spreading of the focused spots in the output plane. If N exceeds TB, the amount of light sent to each detector cannot be controlled independently, and crosstalk will result. In Figure 17b, there are two input beams. Because only half the cell is used for a given input beam, diffraction spreading is twice as great, and the number of resolved output detectors is reduced by a factor of two. In general, as shown in Figure 17c, M inputs can be coupled to no more than TB/M outputs if the connection weights are to be reasonably independent. TB/M is thus the ideal limit; practical limitations, however, dictate a number perhaps two to three times smaller. Any additional source of spreading of the detector-plane light points--for example, poorly collimated input beams or excessive wavelength spread (diffraction angles are proportional to the reciprocal of the wavelength of light)--further reduces the number of interconnections achievable.

The characteristics and fundamental limitations of the two modes of operation are summarized in Table 1, which shows both the maximum number of point-to-point connections and the maximum number of inputs and outputs. Two significant differences between the two modes stand out. First, the beam deflector mode allows for one-to-many interconnections--essentially global in nature--whereas the beam modulator mode allows only for one-to-one, or local interconnections. On the other hand, the number of inputs and outputs both equal TB for the beam modulator case, whereas for the beam deflector case the product of the number of inputs times the number of outputs is limited to TB. Assuming $TB = 1000$ and $M = N$, no more than 31 inputs can be connected on a one-to-many basis to 31 outputs.

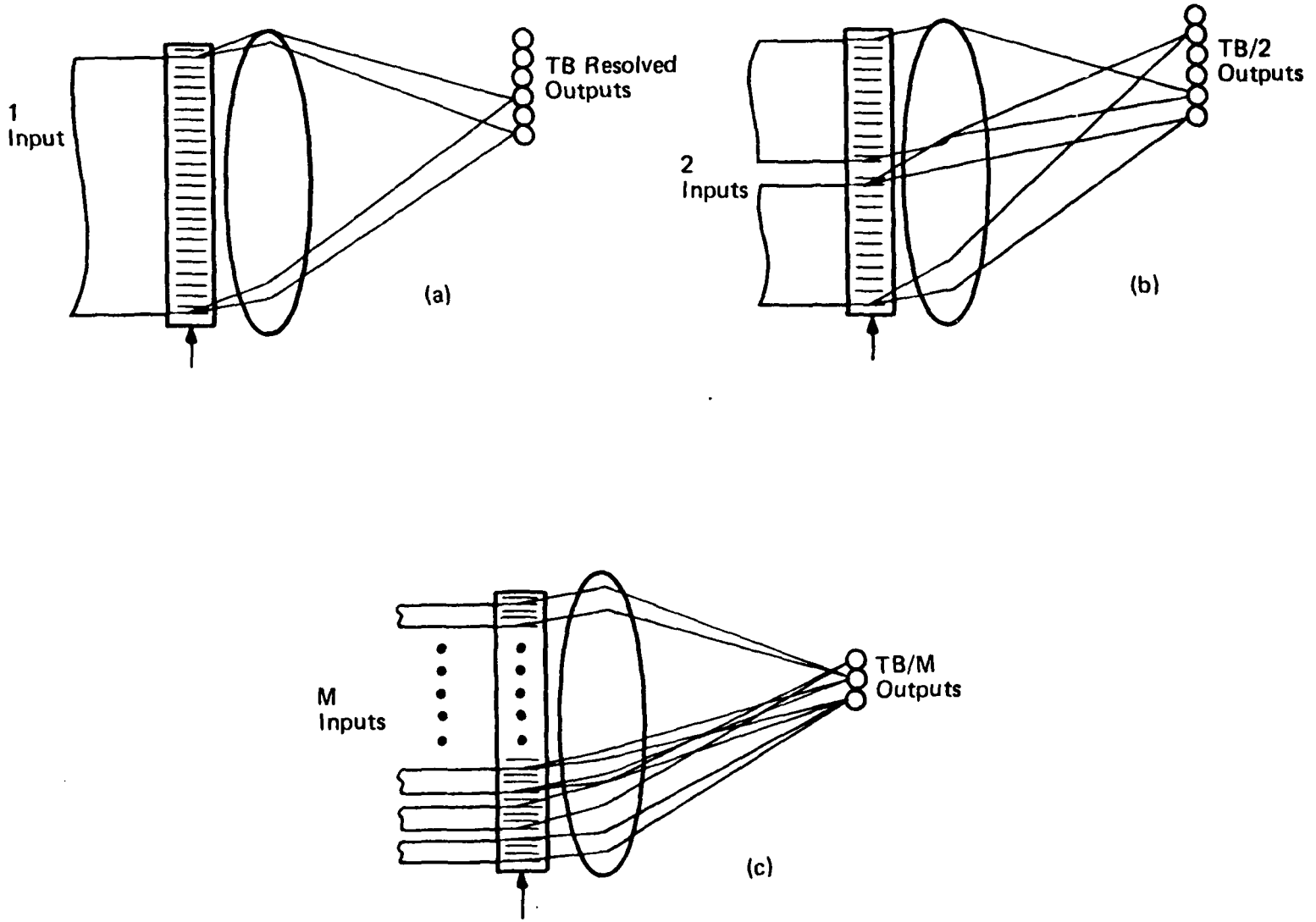


Figure 17. Fundamental limitations on multiple input-output beam deflector. Diffraction will limit to TB/M the number of detectors that can be addressed by M inputs, where TB is the cell time-bandwidth product.

TABLE 1. SUMMARY OF CHARACTERISTICS AND LIMITATIONS OF BRAGG CELLS FOR AO PROCESSORS

	Beam-Modulator Mode	Beam-Deflector Mode
Interconnect	one-to-one	one-to-many
Type	(local)	(global)
Number of possible point-to-point connections	< TB	< TB
Number of inputs and outputs	N inputs N outputs N < TB	M inputs N outputs MN < TB

TB = Bragg cell time-bandwidth product.

In the beam-modulation mode of operation, relatively high diffraction efficiencies can be achieved without nonlinearities (which are inherent in the AO diffraction process) seriously affecting the accuracy of the processor. In the beam-deflector mode of operation, on the other hand, it is quite difficult to control the individual intensities of the different diffracted beams if high diffraction efficiency is desired: nonlinear coupling and harmonic components in the grating transmittance distribution introduce too much crosstalk. Multifrequency beam deflector devices are therefore generally used at low diffraction efficiencies.

2.3 APPLICATIONS OF OPTICAL SIGNAL PROCESSING TO ONE-DIMENSIONAL SIGNAL ANALYSIS

In electronic warfare, spectrum analysis can tell which radio frequencies the enemy is using. This knowledge would allow battlefield commanders to monitor or jam enemy communications and to alert potential targets to radar illumination. Since a spectrum analyzer can also filter out unwanted frequencies, such a device could serve as an input filter in spread-spectrum communications--a popular military communications tactic in which the signal is dispersed over a range of carrier frequencies. An acousto-optic spectrum analyzer is shown in Figure 18, in which three different radio frequencies are input to an AO modulator, and each electrical frequency diffracts the single-frequency laser light at a different angle. The intensity of the light

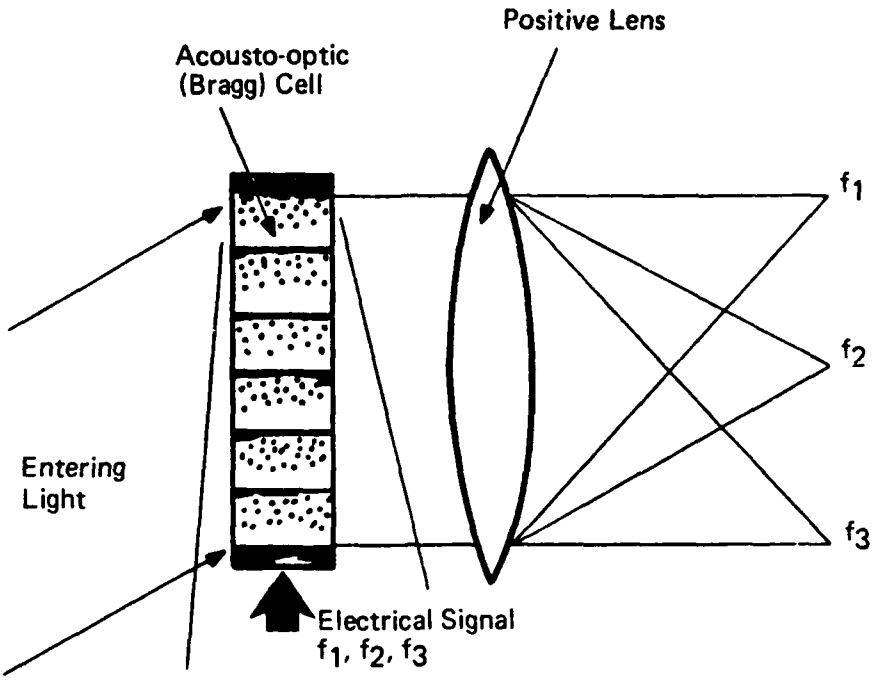


Figure 18. Acousto-optic spectrum analyzer for radio frequency signals f_1, f_2, f_3.

diffracted at each angle is proportional to the energy of the corresponding electrical frequency.

Another important operation that acousto-optic processors can perform is to measure the correlation or degree of similarity of two images or signals. Correlators are powerful tools, and could perform tasks such as robotic vision and identification of such complex patterns as fingerprints, signatures, body cells, and military targets. In such systems, a single input pattern might be matched simultaneously to hundreds or thousands of candidates. The most important use of signal correlation has been in radar and sonar, in which a system that "knows" the signal pattern of a target, such as an enemy aircraft, can search through a maze of signals to find any that correlate. An ideal correlator would automatically warn of an enemy aircraft and pinpoint its location, while ignoring friendly aircraft. The goal of the military is to develop high quality, low cost correlators for such applications as one-time use on "fire and forget" weapons programmed to home in on a target specified

by a particular correlation function. Other applications include the ability to separate signals from noise in spread spectrum communications.

Signals that vary in time or in only one spatial dimension are usually correlated with acousto-optic processors. Acousto-optic correlators have developed along two different paths. The first type, known as a space-integrating correlator, resembles a spectrum analyzer in that the electronic signal is applied to an acousto-optic cell, where the resulting sound-induced wave diffracts a light beam. The diffracted beam passes through a pair of lenses and a spatial filter to remove extraneous light, and then through a fixed mask whose transmission corresponds to the reference pattern (the transform of the reference object). The greater the correlation between the object being viewed and the reference pattern, the more light the mask will transmit. Figure 19 shows a simple implementation of the space-integrating correlator. This system was the first acousto-optic correlator developed. The received signal g(t) is fed to the acousto-optic cell at P_1, which is illuminated at the correct angle. Lenses L_1 and L_2 image plane P_{1a} onto plane P_{1b}. A slit filter at P_2 performs the necessary single-sideband modulation of the data. The signal incident on P_{1b} (effective transmittance of P_{1a} for the term of interest) is described by $g(x-vt) = g(t-\tau)$. Consequently, the wavefront incident on P_{1b} is proportional to the complex-valued signal

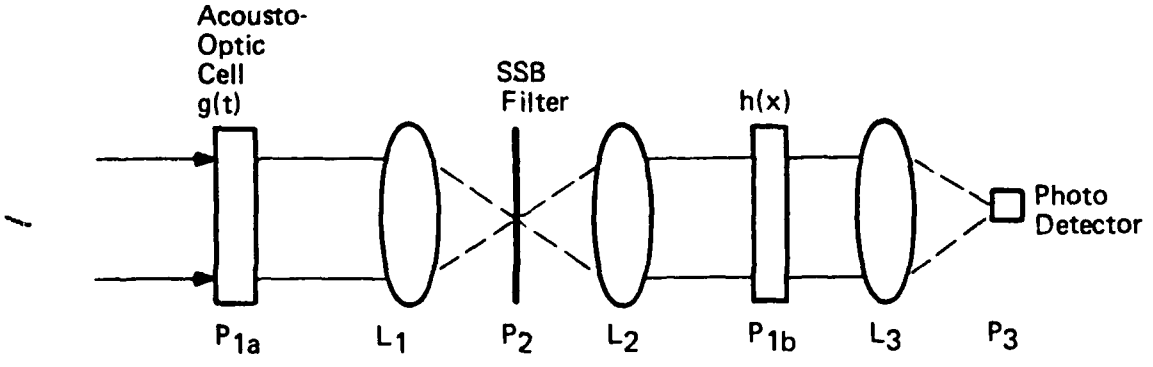


Figure 19. Space-integrating acousto-optic correlator.

$g(x-\tau)$. The mask at P_{1b} has stored the reference transmitted signal code $h(x)$, and the light distribution leaving P_{1b} will then be $h(x)\ g(x-\tau)$. The Fourier transform of this signal product is formed by L_3 at the output plane, where

$$U_2(u,t) = \int g(x-\tau)\ h(x)\ \exp(-i2\tau ux)dx \qquad (28)$$

When evaluated by an on-axis photodetector at $u = 0$, Equation 28 becomes

$$U_2(t) = \int h(x)\ g(x-\tau)\ dx - h \circledast g \qquad (29)$$

or the correlation of g and h. The integration in Equation 29 is performed over the spatial coordinate x. The output correlation variable is time, since the time output from the simple on-axis photodetector is the correlation pattern. Hence the name space-integrating correlator is given to this architecture.

Figure 20 shows the second type of acousto-optic correlator, known as a time-integrating correlator. Here, the integration is performed in time on the output detector. In this system, the signal to be correlated modulates a light source. The modulated light beam is spread out and passed through an acousto-optic cell to which the reference signal is applied. The intensity of the light emerging from the cell is the product of the reference and the

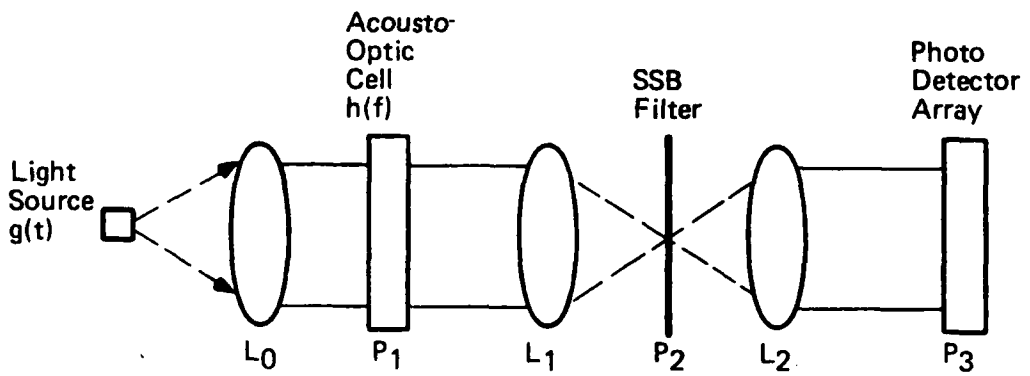


Figure 20. Time-integrating acousto-optic correlator.

sample signals. To obtain the correlation function, this product is integrated over time--a task peformed by an array of photodetectors. In the time-integrating acousto-optic correlator of Figure 20, the output correlation appears as a function of distance across the output detector array. An input light source such as a light-emitting diode or diode laser can be modulated with the received signal g(t). Lens L_0 collimates this output, and an acousto-optic cell at P_1 is uniformly illuminated with the time-varying light distribution g(t). The transmittance of the cell is now described by $h(t-\tau)$, where $\tau = \frac{x}{v}$. The light distribution leaving the cell is thus $g(t)h(t-\tau)$. Lenses L_1 and L_2 image P_1 onto P_3. Any spatial filtering required for improved system performance can be done at plane P_2, and time integration on a linear detector array is done at plane P_3. The light distribution at P_3 after time integration is

$$U_3(x) = \int h(t-\tau)\, g\,(t)dt = g \circledast h \qquad (30)$$

and we see that $U_3(x)$ is the correlation of the received and reference signals. In this case the integration is performed in time, and the correlation is displayed in space.

In practice, space integration performs better in searching large regions, and time integration performs better in correlation of long waveforms. The space-integrating system can accommodate large range-delay searches between the received and reference signals. In the space-integrating system, the signal integration time and signal time-bandwidth product (TB) that these systems can handle is small. The typical dwell time is 40 μs, and the typical acousto-optic cell TBWP is 1000. In the time-integrating system, the signals must be time-aligned, and only a much smaller range-delay search window is possible, typically 40 μs. However, the time-integrating processor allows longer integration times and the associated correlation of longer TB signals.

3. IMAGE PROCESSING APPLICATIONS

In Section 2 we demonstrated that one-dimensional correlation is required for recognition of one-dimensional signals. In this section we will demonstrate that two-dimensional correlation is required for image processing.

Image processing is concerned with image enhancement, restoration, preprocessing, feature extraction, and pattern recognition.[8,9] The outstanding features of this technology are its real-time and parallel processing features and its ability to generate the Fourier transform of two-dimensional input data, to generate linear system features of coherent optical systems, and to perform correlations on two-dimensional data.

The architecture of the coherent optical processor in its simplest form is that shown in Figure 21. The light distribution in the output plane P_2 is the Fourier transform of the amplitude transmittance of the input plane P_1. This light distribution is written $F[g(x_1,y_1)] = G(x_2,y_2)$, where lower-case letters represent the spatial coordinates of the input plane and the corresponding upper-case variables represent their Fourier transform. The coordinates of P_2 and the spatial frequencies (u,v) in the input plane P_1 are related by $(x_2,y_2) = (\lambda f_L u, \lambda f_L v)$. The ability to optically produce this two-dimensional Fourier transform in parallel and real-time is one of the major advantages of a coherent optical system. The two-dimensional Fourier transform distribution in P_2 will be observed as an ensemble of points of light. From the results of Section 1 we find that the magnitude of the Fourier transform is shift invariant (i.e., translations of the input image do not change the amplitude of the Fourier transform), higher input spatial frequencies correspond to peaks of light in P_2 that lie further from the origin in P_2, and, as the input rotates, the light distribution in the Fourier transform plane also rotates.

The space-bandwidth products in the input plane (i.e., the square of the product of the maximum input spatial frequency and the physical size of the input plane in one dimension) and in the Fourier transform plane (i.e., the number of spatial frequencies present) are equal. For even modest imagery

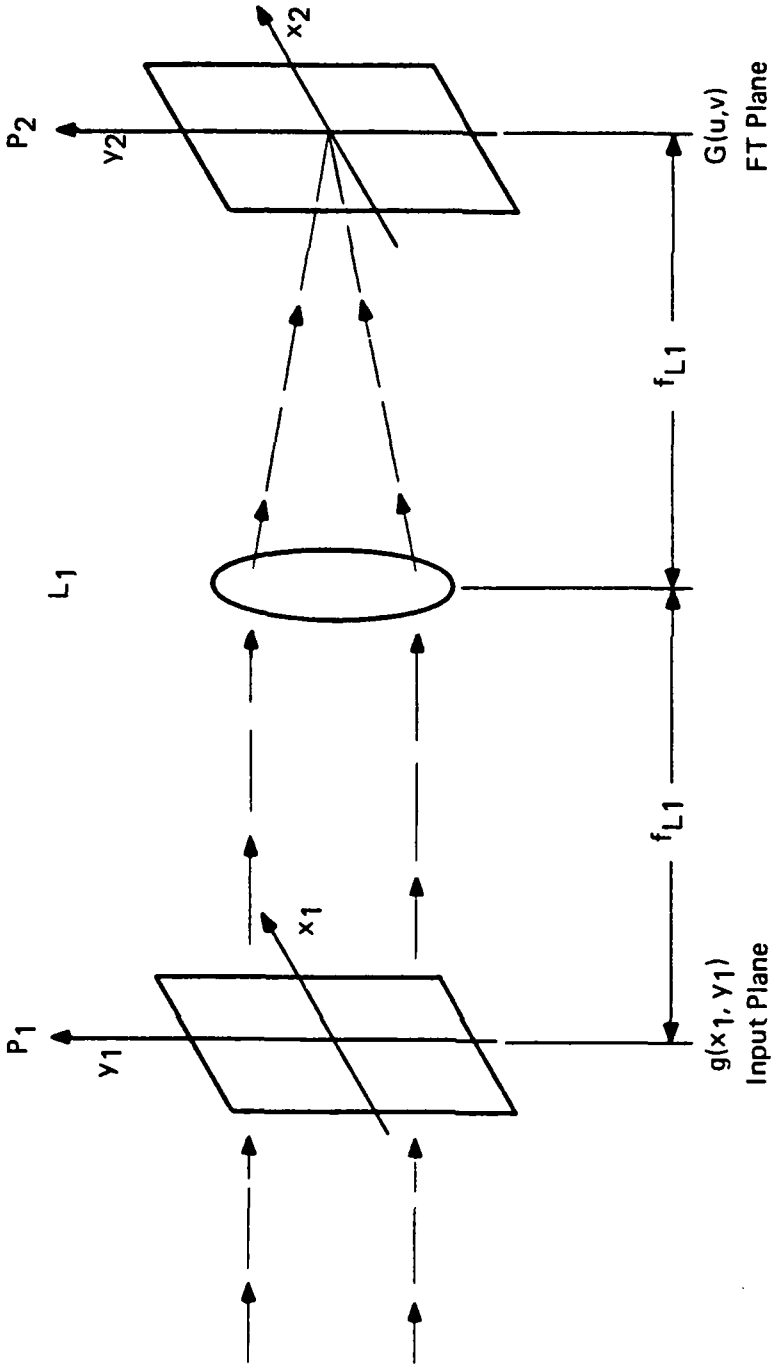


Figure 21. Coherent optical Fourier transform system.

there is a large amount of data, and the three properties of Fourier transforms mentioned above can be used to achieve a significant data compression. If the Fourier transform plane is sampled by a detector with wedge-shaped elements in one-half of the Fourier transform plane, and by a detector with annular ring-shaped elements in the other half, information on the scale and orientation of the input object can be obtained with significant data compression. The imagery is real and positive; thus, the Fourier transform is symmetrical, and no information is lost in separating the Fourier transform plane into symmetrical halves. From the above three properties of Fourier transforms, the wedge data provide object orientation information that is scale invariant, and the ring data provide object scale information that is rotation invariant. Figure 22 shows a schematic representation of a wedge-ring detector (WRD) developed for use in optical processing. One device reported[8] had 32 wedge and 32 ring elements (64 WRD).

One of the simplest optical processors is a Fourier coefficient generation and analysis system using the architecture of Figure 21 with the output detection of Figure 22. Such a system will experience a rapid degradation in

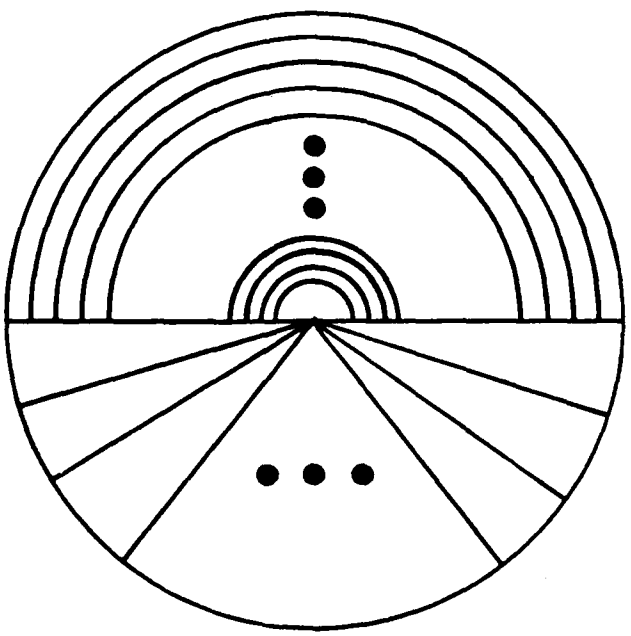


Figure 22. Simplified wedge-ring detector.

performance as input noise is introduced, when multiple input objects are present, or when the input objects to be separated are very complicated. The reason for this is that a loss of data occurs when the input space-bandwidth product is reduced to 64 WRD readings. However, in many situations these systems are adequate.[8] Stark[3] has described successful applications of the WRD architecture in Fourier power spectrum sampling that include aerial views of agricultural scenes with one- and two-dimensional periodicity, aerial views of railroad cars, and handwriting analysis.

The optical correlation of two two-dimensional functions is achieved in parallel with the architecture of Figure 23. The operation of this system in the linear theory of optical imagery was presented in Section 1. In this system, the Fourier transform G of the input scene g is multiplied at P_2 by the conjugate Fourier transform H^* of a reference object h. The light distribution leaving P_2 is thus the product GH^* of two Fourier transforms, and the output plane P_3 is the Fourier transform plane of GH^*, i.e., $F(GH^*)$. The Fourier transform of the product of two Fourier transforms is a convolution function of the two spatial functions h and g. Thus, $F(GH^*) = g \circledast h$.

The system illustrated in Figure 23 requires two spatial functions for its operation. The first is the input function in plane P_1 of $g(x_1,y_1)$, which is an input transparency of complex transmittance $g(x_1,y_1)$ for a system not operating in real-time. For real-time operations, a real-time spatial light modulator is required. The second spatial function is located in plane P_2, where the spectrum of the input is physically accessible and therefore can be manipulated simply by the placement of masks or optical filters. Simple beam blocks can be used in plane P_2 for spatial filtering of features and noise, and complex spatial filters can be used for such functions as object recognition through signal extraction. Such a complex filter is usually known as a matched filter or a Vander Lugt filter, after its originator. An optical filter matched to the signal g(x,y) will have a transfer function proportional to the complex conjugate of the signal spectrum:

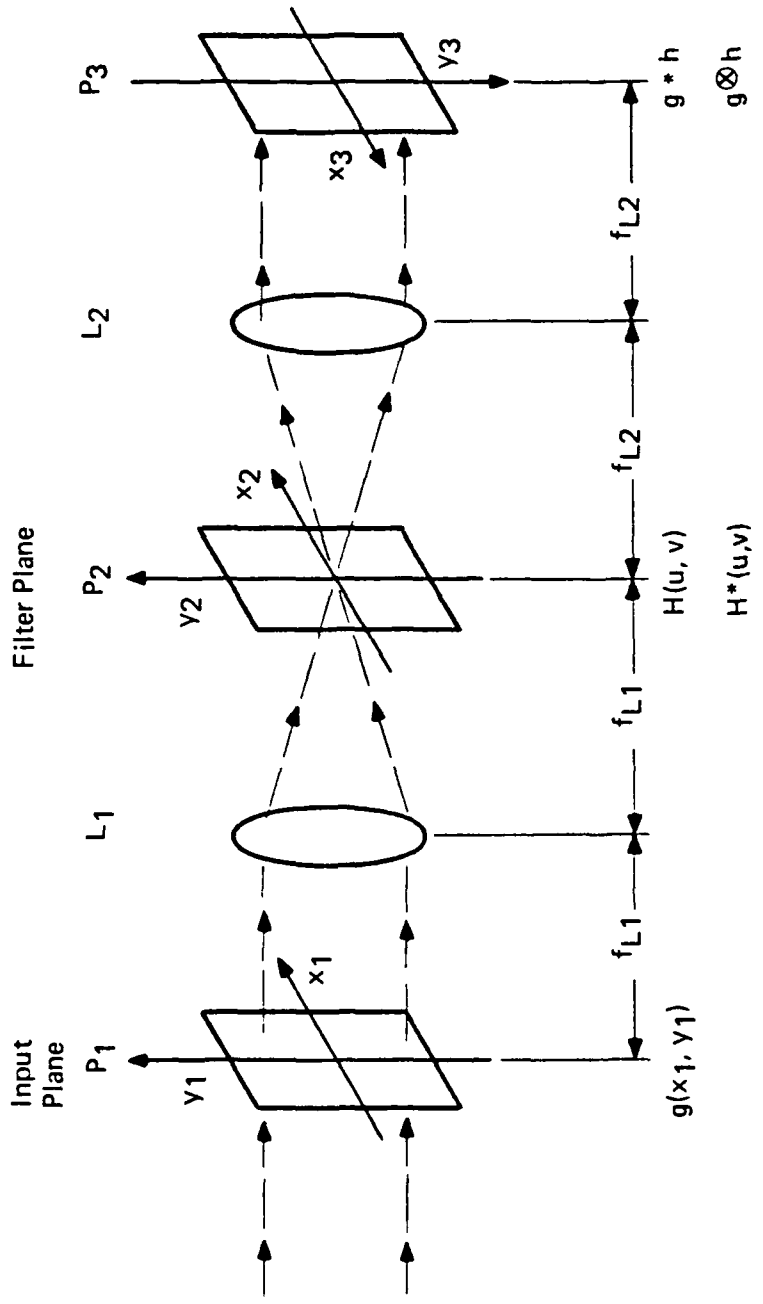


Figure 23. Two-dimensional coherent optical frequency plane correlator.

$$H(u,v) = \frac{k\, G^{*}(u,v)}{|N(u,v)|^{2}} \tag{31}$$

where

$$H(u,v) = F[h(x,y)] \tag{32}$$

and

$$G^{*}(u,v) = F[g^{*}(-x,-y)] \tag{33}$$

Where the noise spectral density can be assumed constant, the filter transfer function and impulse response become

$$H(u,v) = K'G^{*}(u,v) \tag{34}$$

and

$$h(x,y) = k'g^{*}(-x,-y) \tag{35}$$

The filter described by Equations 34 and 35 is either recorded by holographic means or computer-generated. A linearly recorded Fourier hologram of the signal g(x,y) will reconstruct, in one of the sidebands, a bright spot proportional to the point-spread function $g^{*}(-x,-y)$.

3.1 SPATIAL LIGHT MODULATORS FOR REAL-TIME SIGNAL PROCESSING

Parallel processing and real-time operation are two significant advantages of optical processing. However, as just described, the input and filter plane data must be presented as transparencies. Thus, the data introduced in these two planes must be available as transparencies and be capable of changing in real-time. The materials and devices capable of this type of performance are referred to as spatial light modulators (SLM).[10,11]

Spatial light modulators are optical image transducers, electro-optic devices that are characterized by a transmission that can be changed point-by-point in response to an applied electric field or an incident light intensity. In these devices light is incident on an electrically biased photosensitive semiconductor. The charge distribution in the semiconductor is distributed in accordance with the incident light intensity or the applied electric field.

The charge distribution affects the electro-optical properties of another semiconductor layer, a liquid crystal cell, or a membrane.

Optically addressed modulators require a photosensitive medium. The approaches used include:

- the use of photoconductive layers or individual phototransistors that, when activated, control the voltages to an electro-optic or liquid crystal material or to a mechanical structure such as a membrane
- the use of a photocathode in conjunction with a microchannel plate that results in an electron beam that impinges on an electro-optic plate or a membrane
- the use of a material whose properties change due to the heating effect of the light beam
- the use of photorefractive materials whose refractive index changes with exposure to light
- the use of nonlinear optical arrays

The electronically addressed modulators use either serial addressing by means of electron beams or some form of addressing by means of microcircuitry, which can be serial, parallel, or a combination, and which often involves some form of matrix addressing. The electrical signals then act on many of the same materials as mentioned for the optically addressed devices, as well as materials such as magneto-optic films that can only be addressed electrically.

In the operation of a spatial light modulator, the desired data is first recorded on the SLM as a charge pattern (write mode); then a laser beam is passed through the device (read mode) to emerge spatially modulated by the recorded data. The contents of the SLM are erased after each read mode, and a new cycle is begun.

At present these two-dimensional modulators can achieve the following performance, although not all in the same device:

- at least 100 x 100 resolution elements
- data rate of at least 10 frames/s
- storage rate of at least 1 frame/s
- sensitivity of less than 50 $\mu J/cm^2$
- dynamic range greater than five levels
- spatial nonuniformity less than 10 percent
- optical quality less than five wavelengths flatness

It has been reported[12] that at least 45 types of SLM have been designed and built, of which eight are commercially available. Table 2 lists commercially available SLMs, and includes data on SLM type, the company producing it, modulating materials, addressing medium, resolution, optical sensitivity, writing speed, erasure time, and storage time.

Because of applications of the SLM to areas other than two-dimensional image processing the following should be noted. The SLM is used to multiply an input two-dimensional pattern on a beam of light by the two-dimensional data on the SLM (see Figure 24); it can perform many functions on an image, such as amplification, inversion, thresholding, wavelength conversion, and conversion from incoherent light to a coherent replica (see Figure 25).

Figure 24 illustrates how an SLM modulates the amplitude or phase of a "readout" light beam as a function of the intensity of a controlling "write" light beam. Many SLMs, as illustrated, have a reflective structure in which the controlling write beam is incident on one side and the readout beam is reflected from the other side, with an effective reflectivity proportional to the intensity of the write beam.

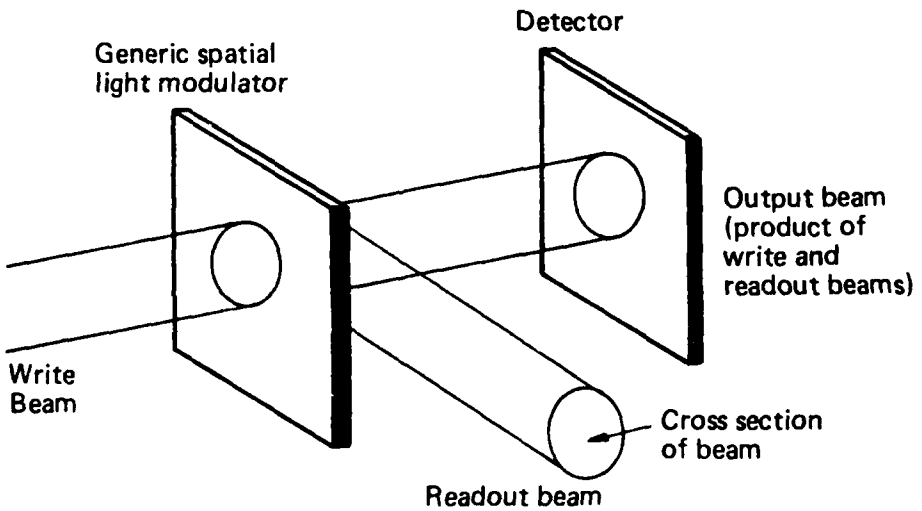


Figure 24. Basic operation of a spatial light modulator.

TABLE 2. COMMERCIALLY AVAILABLE SPATIAL LIGHT MODULATORS

Name or Type	Developer	Modulating Material	Addressing Medium	Resolution,[a] Line Pairs Per mm^2	Optical Sensitivity, $\mu J/cm^2$	Time Response[b] Write, ms	Erase, ms	Store
			Optically addressed devices					
Liquid-crystal light valve	Hughes[c]	Twisted nematic liquid crystal	Cadmium sulfide	30	6	10	15	15 ms
Microchannel spatial light modulator	MIT, Hamamatsu,[c] Optron[c]	Electro-optic crystal (lithium niobate)	Photocathode and microchannel plate	10	3×10^{-6}	10	20	Days to months
Pockels readout optical modulator	Itek, USC, Sumitomo[c]	Bismuth silicon oxide or bismuth germanium oxide	Bismuth silicon oxide or bismuth germanium oxide	6	5	<0.1	<0.1	<2 hr
PRIZ	USSR	Bismuth silicon oxide or bismuth germanium oxide	Bismuth silicon oxide or bismuth germanium oxide	10	5	<0.1	<0.1	<2 hr
Librascope	Hewlett-Packard, Singer[c]	Smectic liquid crystal	Liquid crystal (heat absorption)	40	10^4	0.005	0.001	Months
Thermoplastic spatial light modulator	Harris, NRC,[c] CBS, ERIM, Fujinon, Honeywell, Kalle-Hoechst AG[c]	Heat-sensitive plastic	PVK:TNF	1400	5	10	100	Years
			Electronically addressed devices					
LightMod	Litton,[c] Semetex[c]	Yttrium iron garnet (magnetooptic)	Matrix of electrodes	128 x 128	NA	0.001	0.001	Years
Talaria	Switzerland, GE[c]	Oil film	Scanning electron beam	50	NA	15	15	300 ms

[a]For the LightMod, the resolution is stated in terms of pixels (picture elements).
[b]Time response is given for a full frame except for the smectic liquid crystal modulator and the thermoplastic spatial light modulator, where the address is given for a single pixel; for those devices, the time to address the whole array depends on the addressing mechanism.
[c]Developer is also marketing the indicated spatial light modulator.

Abbreviations: NA = not applicable; CBS = CBS Research Laboratories; ERIM = Environmental Research Institute of Michigan; GE = General Electric; MIT = Massachusetts Institute of Technology; NRC = Newport Research Corp.; PVK:TNF = Poly-n-vinylcarbazole 2,4,7 trinitrofluorenon; USC = University of Southern California.

Source: Arthur Fisher, Naval Research Laboratory, Washington, D.C. Note: the spatial light modulators listed are only the ones that are commercially available at this time; there are at least 37 others, mostly laboratory prototypes--see "A Review of Spatial Light Modulators," A. Fisner, Technical Digest, Topical Meeting on Optical Computing, Lake Tahoe, Nev., March 1985, Optical Society of America, pp. TUC 1-1 to T-4.

Figure 25 illustrates how the construction of a specific SLM and the nature of the readout beam will determine the actual function of the device. For example, if the reflectivity of the surface is directly proportional to the intensity of the write beam and the readout beam is very strong, the SLM will function as an amplifier (Figure 25a). If the reflectivity of the surface is inversely proportional to the intensity of the write beam, the SLM will function as an inverter (Figure 25b). If the output modulation is a threshold version of the write beam, the SLM can function as one step in an analog-to-digital converter (Figure 25c), and if the readout beam is coherent laser light, the SLM can convert an incoming incoherent image to a replica of the incoming wavefront (Figure 25d). If the readout beam is a different wavelength from the write beam, the SLM can convert the input to a different wavelength (Figure 25e). If data are encoded on the readout beam as well as on the write beam, the SLM can perform mathematical multiplications on patterns such as two-dimensional matrices (Figure 25f).

It is this last feature that has greatly increased interest in the SLM. As we will demonstrate in Section 4, if patterns or intensities of light beams can be coded to represent numerical values, the SLM can be used as a rapid analog numerical processors.

3.2 TWO-DIMENSIONAL IMAGE PROCESSING APPLICATIONS

The most common type of two-dimensional signal is an image, and the two major operations in optical image processing are (1) frequency plane blocking for restoring and enhancing degraded images and (2) image pattern recognition. Although image enhancement is of obvious value, pattern recognition has received the greatest amount of interest and research.

Work in pattern recognition has developed along two approaches: correlation and feature extraction or classifier techniques. Correlation attempts to recognize the whole object at once using the technique of matched filtering. An input reference object is used to interferometrically produce the matched filter as described earlier. When the light from the input scene transparency (Figure 23) is Fourier transformed and imaged onto the matched filter, the matched filter diffracts this pattern of light and produces a second diffraction pattern, which is then refocused by a second lens onto the output plane.

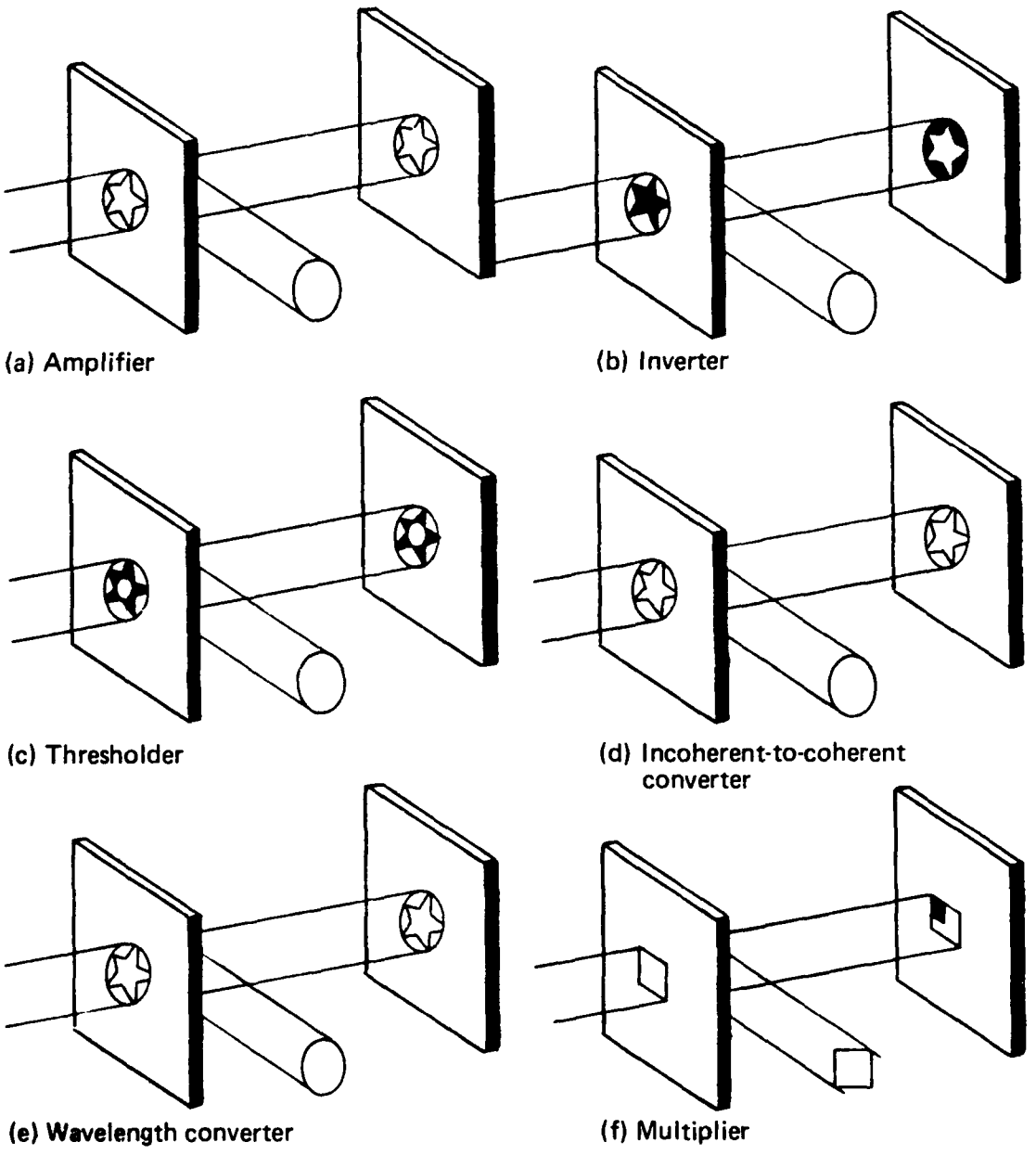


Figure 25. Illustration of various functions of a spatial light modulator.

If the object used in producing the matched filter is in the field of view, a peak of light in the output indicates its location in the field of view of the scene.

Feature extraction or classifier techniques attempt to describe an object by identifying its parts. They take advantage of various features of an object that do not change with different orientations of the object and therefore are better descriptions for an automated recognition system than would be an image of the object itself.

The advantage of classifier techniques is that a single filter works for different objects, however, these techniques do not work well if several objects are in the field of view or if the input scene is noisy. Correlator techniques, on the other hand, are not bothered by multiple objects or noise, but they do not work if the object is oriented (rotated) differently from the reference object used to record the matched filter. Current research is directed toward minimizing the limitations of both these techniques.

3.2.1 Classifiers

Figure 26 is a block diagram of the operations performed by a classifier: feature extraction, dimensionality reduction, and classification. The feature extractor organizes a set of descriptors that together represent an object. The feature extractor must be able to perform this function for various distortions in the input plane, i.e., scale, rotation, focus, and blur. Although feature extractors usually reduce the dimensionality of an image, the information they generate is typically of high dimensionality ($N > 50$). This is particularly true for optical feature extractors. Thus, it is essential to reduce the information to a manageable dimension ($N < 5$). Classification

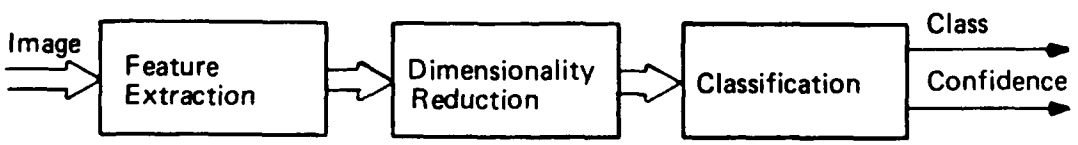


Figure 26. Schematic of classic pattern recognition system.

usually takes the form of comparing the reduced information with test data to determine the identify of the object. Since classification is never perfect, a confidence level can be used by both a user and an expert system to gauge the reliability of the classification.

Optical processors are used primarily as feature extractors in classifiers. The remainder of this section describes various optical feature extractors and presents an overview of dimensionality reduction algorithms and classification procedures.

3.2.1.1 Feature Extraction

The task of a feature extractor is to assemble a set of descriptors that uniquely distinguish one object from any others. Possible descriptors include color, curvature, elongation, geometric moment, length, width, and area. It is useful if the feature extractor organizes the descriptors in a vector form: a feature-vector. The components of a feature-vector depend on the objects to be distinguished and should include only discriminating characteristics. For example, a feature-vector designed to discriminate circles of different radii would not include elongation because elongation does not vary with radius, and a feature-vector that depended on the scale of the input object would not be useful for trying to classify objects regardless of scale. To distinguish between rectangles of different dimensions, the feature extractor may measure the two descriptors' length and width and organize them into a feature-vector $\vec{F} = (a_1, a_2)^T$, where a_1 is the measured length and a_2 is the measured width.

The feature-vector method is useful because it transforms an image into a point in a feature-space. Continuing the rectangle example, the rectangle feature-vector $\vec{F}$ describes a point in a feature-space with basis vectors (i.e., axes) width and length, as shown in Figure 27. Every rectangle with a different width and length is mapped to a distinct point in feature-space. To distinguish between rectangles of different dimensions, feature-space is partitioned into different regions by discriminant functions (Figure 27). Thus, by identifying the region into which the feature-vector $\vec{F}$ maps, rectangle discrimination is achieved.

The primary optical feature-vector generator used is the Fourier transform, because a spherical lens can readily perform a two-dimensional

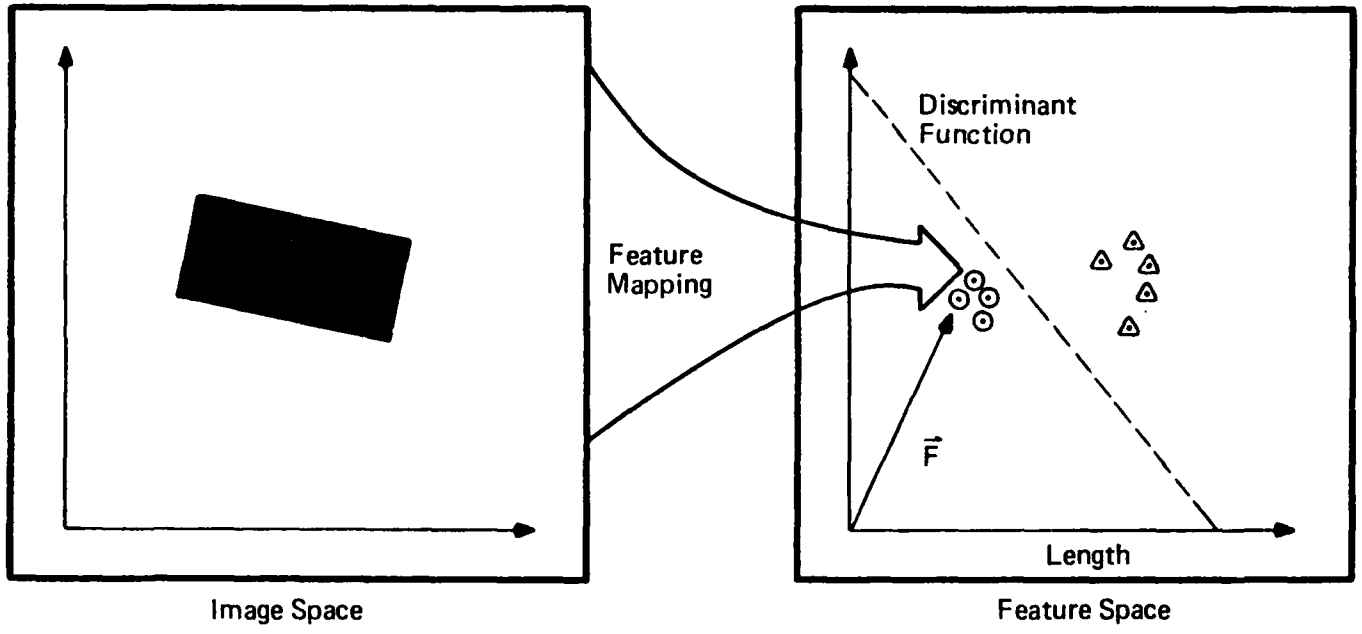


Figure 27. Illustration of mapping an image into a feature-space and using a discriminant function to partition the feature-space into class regions.

Fourier transformation of an input image. The Fourier transform has several properties that make it useful as a feature extractor: (1) the modules of the Fourier transform are invariant to positional shifts in the input plane, and because of this an object can be anywhere in the input plane and the intensity of the Fourier transform will be the same; (2) Fourier transforms are unique in that every image has a different Fourier transform; (3) the Fourier transform facilitates data compression because an image can be adequately described by a few Fourier coefficients; and (4) a rotation in the input plane corresponds to an equal rotation in the Fourier plane.

A convenient method for sampling the Fourier spectrum is the previously described wedge-ring detector (WRD). A WRD samples the Fourier plane with 32 wedge-shaped detectors in one-half of a circular array of detectors; 32 concentric ring detectors occupy the other half. Since an image is real and positive, its Fourier transform is radially symmetrical and a WRD can record the entire Fourier transform. WRD sampling has these convenient properties: the ring outputs are invariant to in-plane rotation changes in the input plane, and the wedge outputs are invariant to scale changes in the input plane. Furthermore, WRDs provide dimensionality reduction: a 512-square element image can be reduced to a feature-vector of 64 elements. Nevertheless, a WRD-produced feature-vector is not completely invariant to changes in scale and in-plane rotation. The wedge components undergo a translation for an in-plane rotation change, and the ring components undergo a scaling for a scale change.

An alternative transformation that can be used to generate a feature-vector is the Mellin transform. The Mellin transform has the property of being invariant to scale changes. For example, let the Mellin transform of $f(x)$ be $M[f(x)] = m(w)$, where M is the Mellin transform operator. The Mellin transform of $f(ax)$ is then $M[f(ax)] = a^{iw}m(w)$, and thus

$$|M[f(x)]| = |M[f(ax)]| \tag{36}$$

The Mellin transform is defined in one dimension as

$$m(u) = \int_0^\infty f(x)\, x^{-iu^{-1}}\, dx \tag{37}$$

If the substitution $\rho = \ln(x)$ is made, then

$$M(u) = \int_{-\infty}^{\infty} f(e^{\rho})\ 1^{-iu\rho}\ d\rho \qquad (38)$$

Since M(u) of Equation 38 is the Fourier transform of $f(e^{\rho})$, the Mellin transform can easily be generated optically. Casasent[2] has developed many systems for implementing the Mellin transform.

3.2.1.1.1 Fourier-Mellin Feature-Vectors

The Mellin transform can be used to develop a scale-invariant form of pattern recognition. Let G(u,v) be the Fourier transform of the image t(x,y); that is, G(u,v) = F[t(x,y)], where F is the Fourier transform operator. The effects on G(u,v) of in-plane rotation and scaling of the image can be separated by the polar coordinate transformation $G(u,v) \rightarrow G(r,\theta)$. In polar coordinates, an in-plane rotation in the input plane by an angle ϕ corresponds to a translation in the angular component, i.e., $(r,\theta) \rightarrow (r,\theta-\phi)$, and a scaling in the input plane corresponds to a scaling in the radial component, i.e., $(r,\theta) \rightarrow (ar,\theta)$. If a Mellin transform is performed on the radial component of $G(r,\theta)$, a completely scale-invariant feature-vector would result: a Fourier-Mellin feature-vector. Figure 28 shows that the difference between the Fourier-Mellin feature-vector of a rectangle and that of a scaled and in-plane rotated version of the same rectangle is a translation by the rotated angle.

In the optical system of Figure 29, the Fourier-Mellin transform is used to generate a feature-vector that is completely invariant to translation and scale, and in-plane rotation variant with only a translation. An input SLM such as a liquid crystal light valve (LCLV) or liquid crystal television (LCT) with transmission t(x,y) is uniformly illuminated by an argon laser. A Fourier transform is performed using lens L_3 and is projected onto a bismuth silicon oxide (BSO) photorefractive crystal (or onto an LCLV). The resulting index of refraction change of the BSO crystal thus corresponds to the modulus of the Fourier transform. A polarized readout beam from a HeNe laser also illuminates the BSO crystal and is passed through an analyzing polarizer to modulate the beam. A computer-generated hologram (CGH) is used to perform a Cartesian-to-polar coordinate transformation along with a logorithmic scaling

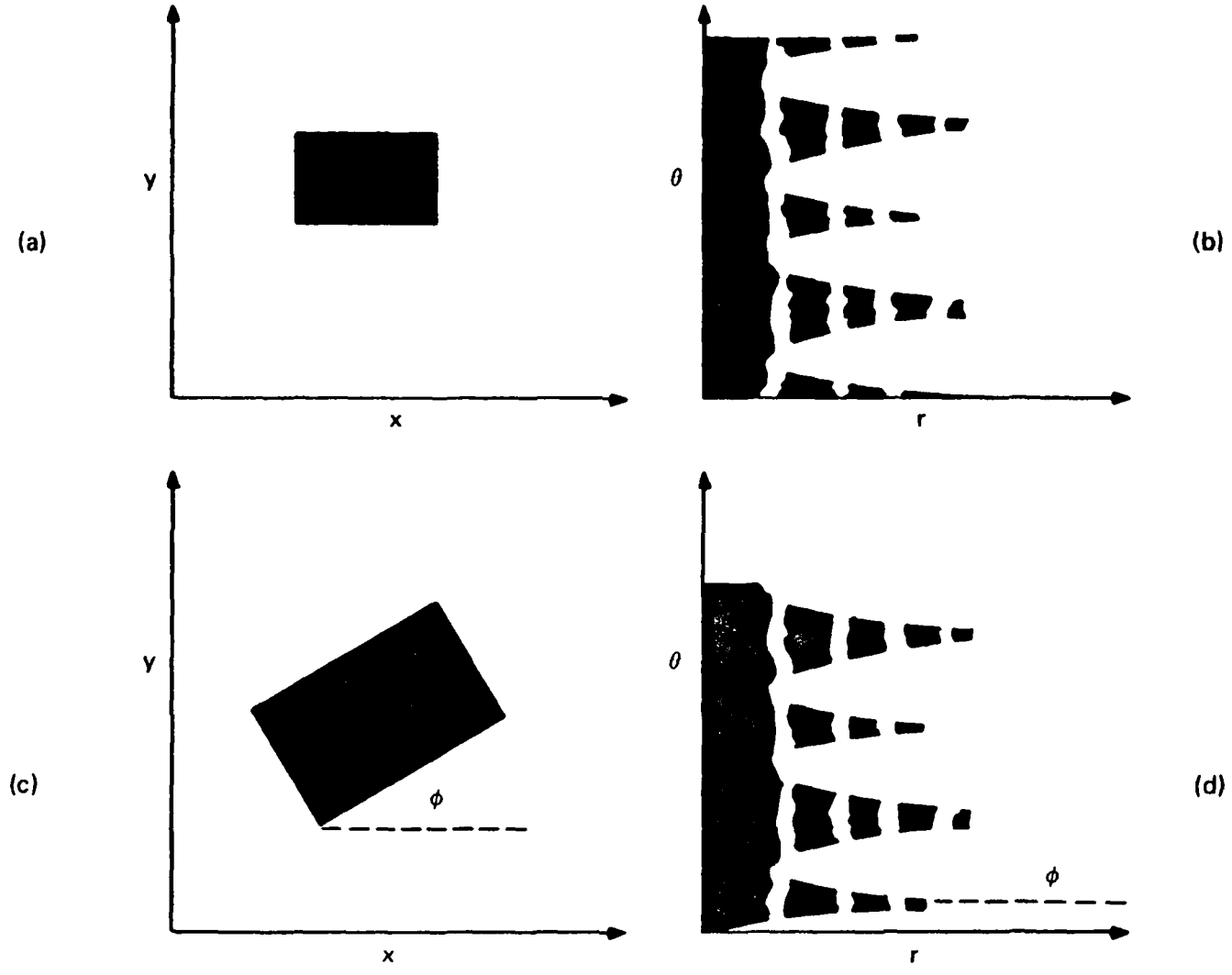


Figure 28. Effects of scaling and rotation on the Fourier-Mellin transform:
(a) Input image;
(b) Fourier-Mellin transform of (a) in polar coordinates;
(c) Scaled and rotated version of (a); and
(d) Fourier-Mellin transform of (c) in polar coordinates.

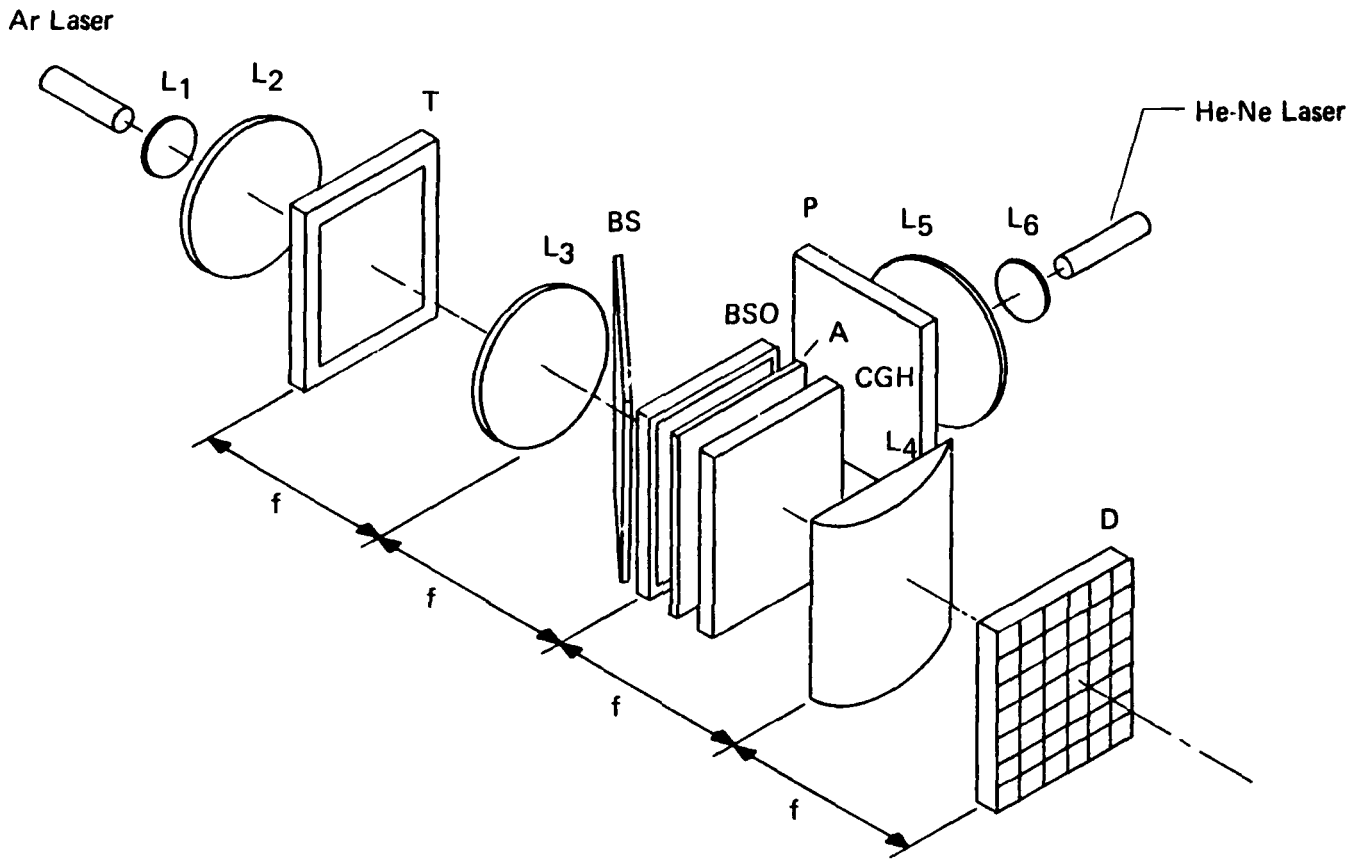


Figure 29. Fourier-Mellin transform feature extractor to generate invariant feature-vectors:
L_1, L_2, L_5, and L_6 = Expanding and collimating lenses;
T = Input SLM;
L_3 = Fourier transform lens;
BS = Beam splitter;
BSO = Bismuth silicon oxide crystal;
P and A = Polarizers;
CGH = Computer-generated hologram;
L_4 = Cylindrical Fourier transform lens;
D = Two-dimensional array of detectors.

in the radial direction.[13] A cylindrical lens L_4 performs a Fourier-Mellin transform on the radial component. A rectangular detector array is used to record the resulting feature-vector. Thus, in operation, the modulus of the Fourier transform of t(x,y) is detected by the BSO crystal in the form

$$|F[t(x,y)]| = |G(u,v)| \tag{39}$$

where t(x,y) is the transmission of the input transparency. The conversion from Cartesian to polar coordinates,

$$|G(u,v)| \rightarrow |F(r,\theta)| \tag{40}$$

is then accomplished using a CGH. The radial component is also scaled by the CGH by $\rho = \ln(r)$

$$|F(r,\theta)| \rightarrow |F(e^{\rho},\theta)| \tag{41}$$

The Fourier-Mellin transform is then performed

$$M(\omega_\rho,\theta) = \int |F(e^{\rho},\theta)| e^{(-\omega_\rho \rho)}\, d\rho \tag{42}$$

to generate the invariant feature-vector.

The Fourier-Mellin feature-vector is invariant to scale and position changes and variant to in-plane rotation changes. But as with WRDs, an in-plane rotation corresponds to a translation. The Fourier-Mellin feature-vector facilitates dimensionality reduction by means of a Fourier transform and by varying the resolution of the detector array. For instance, if a 512-by-512 element image is input, an n-by-m detector array can be used, where n <512 and m <512. Since rectangular detector arrays of different sizes can be easily obtained, the dimensionality reduction properties of the processor can be easily varied, allowing the resolution of the feature-vector to be variable to match the needs of the classifier. This is an advantage over WRDs, which do not permit the resolution to be varied. Fourier-Mellin transforms can also be used in other feature-vector architectures. Yatagai et al.[14]

describe a pattern classification system that accomplished an optical Mellin transform by a logarithmic scaling of both spatial coordinates using a CGH, followed by a Fourier transform. The Fourier-Mellin transform was sampled using a circular array of detectors to generate a feature-vector. Casasent et al.[15] developed a digital simulation of a Fourier-Mellin feature extractor using multiple linear discriminant function to analyze large-dimensional feature-space, and this system has demonstrated greater than 90 percent correct identification for various classes of ships with in-plane rotations.

3.2.1.1.2 Geometric Moments

Another approach to feature extraction for a more universal operation is that of geometric moments. The geometric moments of an image $t(x,y)$ are defined as

$$m_{pq} = \iint t(x,y)\ x^p y^q dxdy \qquad (43)$$

The basic motivation behind the use of geometric moments in feature extraction is that it is possible to form a nonlinear combination of moments that is invariant to translation, scale, and in-plane rotation distortion.[16] Several architectures have been developed to calculate the moments: Casasent et al.[17] employed a multiplex method to generate spatially separated signals corresponding to the moments; Teague[18] described a method that shows how to calculate the geometric moments from the intensity of the Fourier transform of an input image; and Blodgett et al.[19] developed and used a spatial-frequency multiplexing scheme for computing the moments.

3.2.1.1.3 Chord Distribution

This linear discriminant function is defined as follows: Consider an edge-enhanced image $t(x,y)$ such that $t(x,y) = 1$ at the edges and $t(x,y) = 0$ elsewhere. Each point on the edge can be connected by a chord with length r and angle θ if

$$t(x,y)t(x+r\cos\theta,\ y+r\sin\theta) = 1 \qquad (44)$$

The chord distribution is the distribution of r's and θ's as given by the integral

$$h(r,\theta) = \iint t(x,y)\ t(x+r\cos\theta,\ y+r\sin\theta)\ dxdy \qquad (45)$$

The substitution of $(\alpha,\beta) = (r\cos\theta, r\sin\theta)$ into Equation 45 gives

$$h(r,\theta) = \iint t(x,y)\ t(x+\alpha,y+\beta)\ dxdy \qquad (46)$$

which is the autocorrelation of $t(x,y)$ in polar coordinates. An autocorrelation can be easily performed using an optical correlator; the chord distribution is useful for feature extraction because the chord lengths are invariant to in-plane rotation, and the chord angles are invariant to scale changes.

Casasent and Chang[20] have demonstrated an optical chord distribution feature extractor for classifying ships with out-of-plane rotations, and they were able to show greater than 90 percent correct results.

3.2.1.2 Dimensionality Reduction

In all of the approaches to image processing discussed in Section 3.2 except for geometric moments, the feature vectors generated have a large dimensionality. Large-dimensional problems are difficult and cumbersome to deal with, and it is important to reduce the dimensionality of the problem. The data contained in an optically generated feature-vector of N dimensions must be distilled down to one, two, or three dimensions. In this way the problem is reduced to the point where it is easily solved using a computer. It is also important that minimal information be lost in the dimensionality reduction so that object discrimination is still feasible. This criterion is usually gauged by how well the reduced feature-vectors of the objects to be distinguished are separated in feature-space. For instance, an algorithm that maps two objects to be identified to the same region in feature-space would not be useful, but an algorithm that maps two objects to be identified to different quadrants of feature-space would be useful.

The algorithms that have been successfully used in the past for dimensionality reduction are the Karhunen-Loeve (KL) expansion,[21] the

Fukunaya-Koontz (FK) expansion,[22] the Foley-Sammon (FS) transform,[23] and the Gram-Schmidt (GS) expansion.[24] The first three have been successfully applied to wedge-ring detector (WRD) classifiers.

With the KL expansion, the eigenvalues and eigenvectors of the autocorrelation matrix for a set of training images that represent an object class or type are calculated for each class to be discriminated. Essentially, the set of training images is expanded in terms of its eigenvectors. Then only the dominant eigenvectors (those with the largest eigenvalues) for each class are retained (ϕ_1, ϕ_2, ...). In practice, it is possible to retain more than one dominant eigenvector per class; thus, a set of dominant eigenvectors can be formed (ϕ_1^1, ϕ_1^2, ..., ϕ_1^n; ϕ_2^1, ϕ_2^2, ..., ϕ_2^n; ...; ϕ_m^1, ϕ_m^2, ..., ϕ_m^n). The KL expansion reduces the dimensionality of the feature-space to nm dimensions, where n is the number of eigenvectors retained per class and m is the number of classes. Each eigenvector then serves as a basis vector in a reduced feature-space. However, there is no assurance that the KL expansion will select the important features necessary to discriminate the classes. It is conceivable that two classes would have similar features, so corresponding dominant eigenvalues and eigenvectors would be comparable. Thus, the usefulness of the KL expansion is restricted to intraclass discrimination.

The FK expansion was developed to overcome this limitation of the KL expansion. The FK expansion is similar to the KL expansion in that the eigenvectors and eigenvalues of the autocorrelation matrices are calculated, but the FK expansion contains an additional constraint when selecting the eigenvectors to be retained that ensures that the dominant eigenvectors between classes are different. The important features of class 1 become the least important features of class 2, and the important features of class 2 become the least important features of class 1. The FK expansion has the drawback of being limited to two-class discrimination, although multiclass classification can be accomplished by the sequential pairing of classes. Furthermore, the FK expansion does not optimally define a feature-space that conveniently separates classes linearly.

With the FS transform it is even easier to separate two classes. To find the best basis vectors to define a feature-space, the FS transform maximizes the interclass distance and minimizes the intraclass distance for a set of training images.

The GS expansion produces an orthonormal set of basis vectors from a set of training images by expanding the training set as a linear combination of Gram-Schmidt vectors $\vec{\Psi}_s$, s = 1, 2, ..., n. By the selection of a set of $\vec{\Psi}_s$ such that n < N, where N is the dimension of the training set, dimensionality is reduced. Using the GS expansion to find a set of basis vectors may be easier than calculating eigenvalues and eigenvectors. The eigenvector method, on the other hand, provides a convenient ordering of the eigenvectors by eigenvalues, making optimal vector selection easier.

3.2.1.3 Classification

The final operation performed by a pattern classifier is classification. The main criterion for classification is the minimization of misrecognition. Figure 30 is a schematic example of a feature-space for a three-class problem.

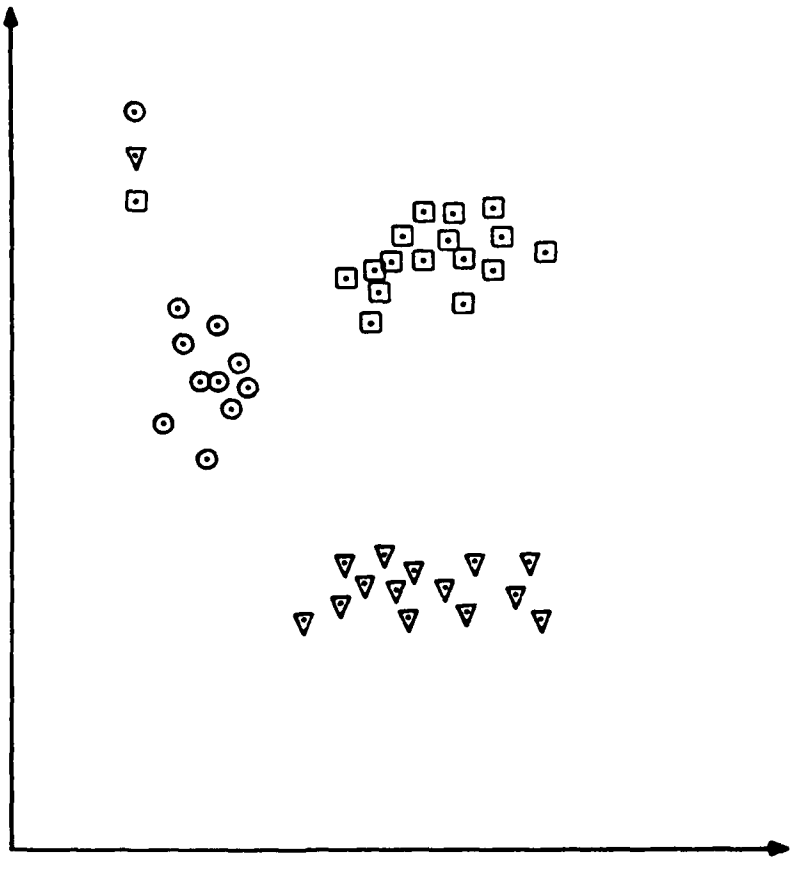


Figure 30. Example of clustering in feature-space for a three-class problem.

The main task of the classifier is to partition such a feature-space into regions corresponding to the various classes. This procedure is straightforward for classes that do not overlap. The task becomes more involved if the classes overlap and they are not linearly separable. Algorithms have been developed to deal with the partitioning of feature space. Most of them require a set of training images and a mathematical model for the discrimination function. For instance, discriminant functions have been modeled as linear, quadratic, polynomial, and piecewise linear.[25] In addition, statistical approaches, such as the Bayesian and stochastic, have been used to analyze training sets.[26] Furthermore, there are classification techniques in which the classifier learns to discriminate classes unsupervised, by analyzing the clustering in feature-space.[26-28]

Purely statistical classifiers and optical feature extractors have limitations. To accommodate multiobject scenes, preprocessing is necessary in the form of image segmentation. Statistical classifiers have an additional limitation in scene analysis: contextual information is not included in the classification algorithm.[29] For example, a statistical classifier would not use the information that cars are usually found on roads and boats are usually found on water to discriminate cars from boats. It is anticipated that the inclusion of contextual information with statistical classifiers within an expert system will minimize misrecognition and improve time efficiency in object identification. Since contextual information is inherently symbolic, artificial intelligence machines and languages are probably best suited for such scene understanding.

3.2.2 Correlators

The critical components of a matched or Vander Lugt filter optical correlator are shown in Figure 31. These components are an SLM to input the viewed scene, a Fourier transform lens, a matched spatial filter (MSF), a correlating lens, and an output plane detector array. The MSF is the heart of the correlator. If the complex conjugate of the Fourier transform of a reference image is placed in the Fourier plane, the resulting output of the processor is the correlation of the input scene with the reference image. This is a convenient method of locating objects in the viewed scene, since the correlation function dramatically increases wherever the input scene matches

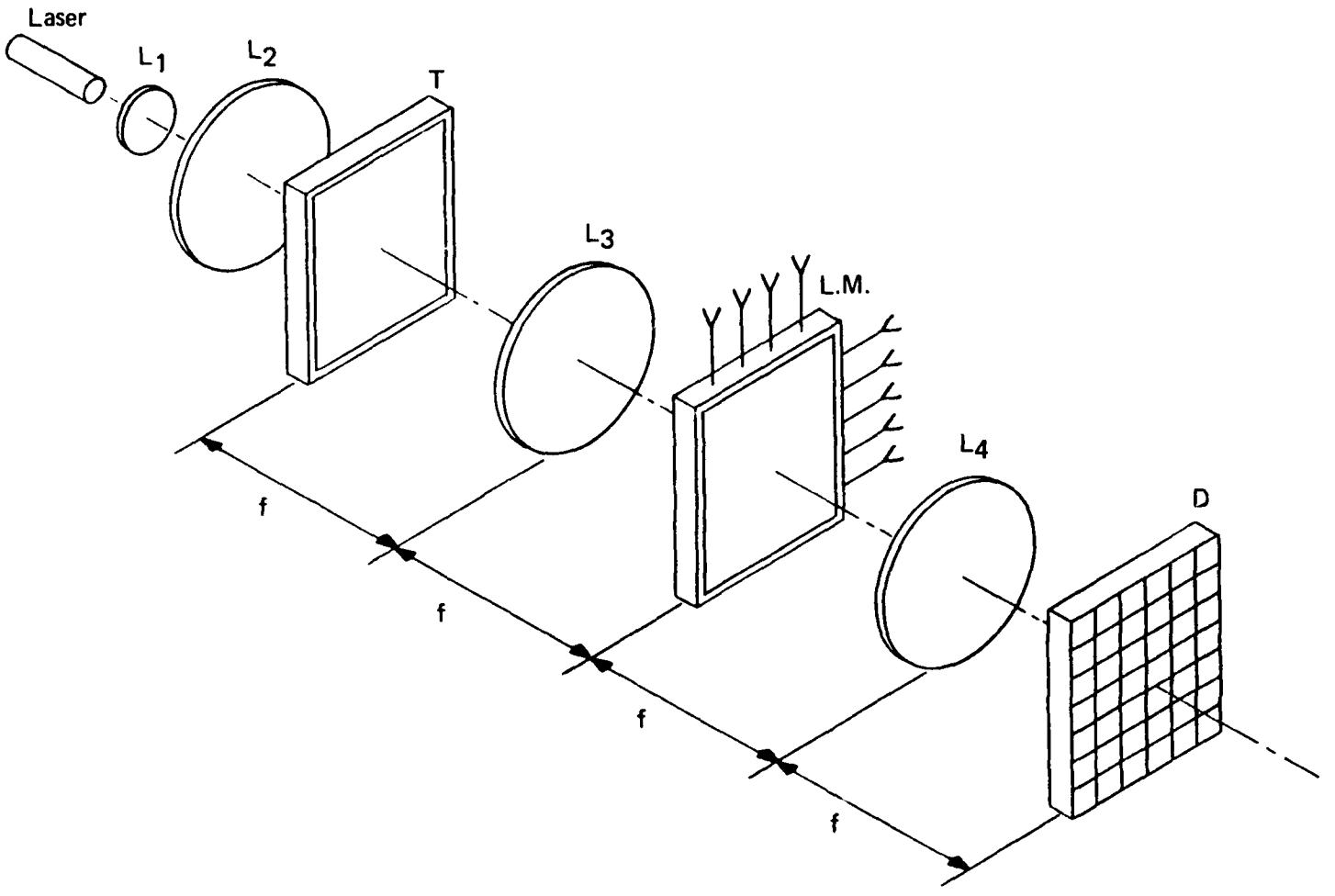


Figure 31. Optical correlator:
L_1, L_2 = Expanding and collimating lenses;
T = Input SLM;
L_3 = Fourier transform lens;
L.M. = SLM in Fourier plane;
L_4 = Correlating lens;
D = Two-dimensional array of detectors.

the reference image. A serious limitation, however, is that the locating performance of this type of correlator degrades rapidly for small distortions between the input scene and the reference image. For instance, a decrease in signal-to-noise ratio of a factor of 500 results from only a 2% change in scale or a 3.5-degree change in rotation. Techniques such as the Fourier-Mellin correlator have been developed to surmount these difficulties, but they have not been the answer to all problems. The Fourier-Mellin correlator is scale- and rotation-invariant, but it is not able to provide position information.

An alternative approach to distortion-invariant optical correlation involves synthetic discriminant functions (SDFs). The advantage of this technique is its versatility in accommodating a variety of pattern recognition problems. Optical correlators employing SDFs can be made variant or invariant to the distortions of scale, translation, in-plane rotation, out-of-plane rotation, and class. Casasent has described the theory of SDFs.[30] The concepts behind this technique are as follows: A set of centered training images tn(x,y) with the desired distortion-invariant features is used to construct an SDF s(x,y). For example, if the correlator is to locate a rectangle of a given aspect ratio regardless of size, the training set will contain images of the rectangles at various scales. The main restriction placed on the SDF s(x,y) is that the correlation peaks (at the origin) between s(x,y) and all the training images t_n, n = 1, 2, ..., L, must be a constant,

$$t_n \circledast s = c_n = \text{constant} \tag{47}$$

where $\circledast$ denotes correlation. It is possible to expand s as a linear combination of the training set,

$$s = \sum_{n=1}^{L} a_n t_n \tag{48}$$

The coefficients a_n must now be determined. The substitution of Equation 48 into Equation 47 gives

$$t_n \circledast s = t_n \circledast \sum_{m=1}^{L} a_m t_m = \sum_{m=1}^{L} a_m r_{nm} = c_n \tag{49}$$

where r_{nm} are the elements of the autocorrelation matrix for the data set t_n. In matrix form, Equation 49 is

$$R\,\vec{a} = \vec{c} \tag{50}$$

where R is the correlation matrix and c is a vector of constants. Thus, the coefficients a_n are given by

$$\vec{a} = R^{-1}\,\vec{c} \tag{51}$$

To construct the spatial filter to be used in the correlator, the coefficients found from evaluating Equation 51 are substituted into Equation 48 to generate the SDF s(x,y). A Fourier transform is then performed on s(x,y) to generate the SDF-MSF. The method described below is optimized for systems with Gaussian noise. Kumar[31] has developed a method of fabricating optimized SDF-MSFs for any noise source.

In an SDF correlator, the training set of images t_n and the values used in the constant vector c determine the function of the processor. For instance, for intraclass discrimination, t_n would contain the various distorted images, while c would contain all ones (ones are used for simplicity, but any constant could be used). For interclass discrimination, t_n would be broken down into sets of subvectors $t_n = \{t_{m_1}, t_{m_2}, \ldots, t_{m_\ell}\}$, where the subvector set t_{m_ℓ} contains the various distorted images within a class and where $\vec{c}$ would have a different constant corresponding to the different subvectors, $\vec{c} = \{1, \ldots; 1, 2, \ldots; 2, \ldots\ell; \ldots.\ell\}$. As can be seen, SDF-MSFs are computationally intensive regardless of the specific application, but all of the required calculations can be done off-line, thereby allowing high-speed correlation.

Ennis and Jared[9] have described an SDF correlator utilizing the architecture of Figure 31 using a magneto-optic spatial light modulator (MOSLM) in the Fourier plane that can scan a library of SDF-MSFs. The MOSLM is a binary SLM that modulates light by means of the Faraday effect. The rotation of the plane of polarization of the incident light beam depends on the direction of the magnetization of the material. The intensity of the beam can be modulated by passing the beam through an analyzing polarizer. Devices with 128-by-128

pixels are available at present, and larger devices are expected in the future. The advantage of the MOSLM over other SLMs is its ability to be electronically addressed very quickly. With fast electronics, frame rates of up to 1000 Hz appear possible with MOSLMs. These high frame rates compare to 4 Hz for photorefractive crystals and 30 Hz for CdS LCLVs. Although the MOSLM is a binary device and photorefractive crystals and LCLVs are gray-scale devices, MOSLMs perform well in optical correlators. Psaltis et al.[32,33] have shown that thresholding does not seriously degrade the performance of the correlator, and in some cases actually enhances it. Furthermore, computer-generated MSFs have traditionally been binary due to the simplification of the calculations involved. Psaltis et al. used a MOSLM in the Fourier plane of a Vander Lugt correlator, but only a classic MSF with minor enhancements was displayed. Nevertheless, a MOSLM is very capable of encoding SDF-MSFs.

By utilizing both SDF-MSFs and the high frame rate of a MOSLM device, the correlator of Figure 31 can be used as a fast and flexible distortion-invariant pattern recognizer. For example, the MOSLM can be driven by a microcomputer system that contains a library of SDF-MSFs. Each SDF-MSF would correspond to a set of distorted images for a given object. By scanning the library and sampling the correlation plane, all of the objects in the input scene could be identified and located. Since the MOSLM has a frame rate of 1000 Hz, the scanning procedure is not too time costly. Furthermore, it is anticipated that intelligent construction (via expert system control) of the hierarchical ordering of the SDF-MSF library will improve time efficiency. An input/output time bottleneck can exist for this correlator design because of the requisite processing of the 1000 frames per second of two-dimensional data from the output plane detector array.

Since many SDF-MSFs can be used in the identification process, it is not necessary to synthesize a single SDF-MSF from a large collection of training images that are multiclass and multidistortion, a process that would be computationally prohibitive. If the SDF-MSFs contained in the library were synthesized from only a few training images, the computational overhead would be greatly reduced. However, the scanning ability of the MOSLM would not compromise the multiclass and multidistortion property of the SDF-MSFs. In fact, a large library of SDF-MSFs would greatly enhance the applications and flexibility of the correlator.

3.2.3 Learned Pattern Recognition Using SDFs

A major area of research in artificial intelligence is the encoding of learned knowledge into a data base so that the knowledge can be applied to future decision making. For example, in a visual expert system, it is necessary to encode images so that they can be recalled to perform object classification. If the expert system is to learn to discriminate circles from squares, the salient aspects of roundness and squareness must be represented in such a fashion that inferences about future circles and squares can be made. An extremely powerful application of the correlator system shown in Figure 31 is in image encoding for a visual expert system.

The learning procedure is based on the fact that SDF-MSFs are generated from a set of training images. The training set represents a collection of object images that the system can positively identify. Learning is facilitated by continually synthesizing new SDF-MSFs as the training set of images increases for a given object. As mentioned above, the correlator system intelligently scans the library of SDF-MSFs to identify an object. If a positive correlation is not made, the system queries the operator (this approach needs a method of unsupervised learning) for an identification. Once information on the object is provided by the operator, the system then searches its library to see if an SDF-MSF already exists for that object. If one does exist, implying that the input object is an unrecorded/unrecognized distortion, the input image is added to the training set for that object and a new SDF-MSF is constructed including the input image. If the object does not already exist in the library, a new SDF-MSF is generated and added to the library. Thereafter, the system would be able to identify the object.

It is possible to configure the system for unsupervised learning by providing the system with a feature-space to map images into. In a good feature-space, objects within the same class will cluster into distinct regions. By identifying the boundaries of clusters in feature-space, the system can set up a framework within which it can organize its knowledge. For example, in Figure 30 the mapping of 40 images into feature space is shown. As can be seen, three easily identifiable clusters formed. The system would generate an SDF-MSF for each cluster. The training images used would be those images that mapped into a particular cluster.

3.2.4 Optical Pattern Recognition: Summary

Table 3 summarizes the distortion invariances of each optical pattern recognition system discussed. A space-bound pattern recognition system needs to be completely distortion invariant. To some degree, all of the systems exhibit some invariance to translation, scale, and in-plane rotation. Most feature extractors can accommodate out-of-plane rotation only with classification algorithms. However, SDF-MSFs are by far the most distortion invariant. SDF-MSFs can provide invariance to translation, scale, in-plane rotation, out-of-plane rotation, and class. Furthermore, SDF-MSFs can be used on multi-object scenes with a minimal amount of preprocessing (i.e., edge enhancement), whereas feature extractors can be used only on one-object images.

TABLE 3. SUMMARY OF OPTICAL PATTERN RECOGNITION SYSTEMS

Processor	Dimensionality	Distortion			
		Translation	Scale	In-Plane Rotation	Out-of-Plane Rotation
Fourier transform with rectangular detector	n x m (see note a)	invariant	variant, scaling in radial component	variant, translation in angular component	variant
Fourier transform with WRD	64	invariant	variant, scaling of rings	variant, translation of wedges	variant
Fourier-Mellin transform	n x m (see note a)	invariant	invariant	variant, translation in angular component	variant
Chord distribution	64	invariant	variant, scaling in radial component	variant, translation in angular component	variant
Moments	p x q (see note b)	invariant	invariant	invariant	variant
SDF-MSF correlator	n x m (see note a)	invariant/variant selectable	invariant/variant selectable	invariant/variant selectable	invariant/variant selectable

[a]Dimensionality depends on the size of the detector array used.

[b]Dimensionality depends on the number of moments calculated.

4. ALGEBRAIC PROCESSING APPLICATIONS

The operations performed by optical systems are described in terms of simple mathematics: convolution, multiplication, integration, etc. It requires only a minor change in approach to recognize that optics can be used to perform mathematical operations. This view has been stated by many researchers since the earliest days of optical processing. In the 1960s, Cutrona[34] described the application of optical systems to the evaluation of general superposition integrals and to the multiplication of a vector by a matrix, and numerous other researchers recognized the potential of optical systems for performing a variety of mathematical operations. During the past 10 years, a great deal of work has been done in developing different applications of optics to mathematical operations that are numerical and algebraic in nature. This work, while not directed toward or thought of as leading to a general-purpose optical computer, will lead to general optical-array processors. These processors will be used as adjuncts to digital computers to perform specific algebraic computations at very high speeds. Designs are currently under consideration for high-speed optical processors to evaluate polynomials, matrix-vector products, matrix-matrix products, and solutions of sets of linear equations.

As described earlier the acousto-optic (AO) Bragg cell is used to perform signal convolution, correlation, and spectrum analysis. During the past several years, new applications based on algebraically oriented operations such as matrix-vector and matrix-matrix multiplication have been developed for this versatile device. The processors under development are capable of presenting significant competition to alternative all-electronic approaches such as the CRAY-1 signal-processing computer, which operates at an average rate of 30 x 10^6 floating-point multiplications-additions per second (180 x 10^6 burst rate) with 64-bit word length.

4.1 ALGEBRAIC SIGNAL PROCESSING OPERATIONS

The operation and basic principles of an AO Bragg cell and AO signal processing were presented in Section 2, and the two capabilities of a Bragg cell for intensity modulation and frequency-dependent beam deflection are exploited in AO algebraic processors.

A number of algebraic, matrix-oriented operations are important to modern signal-processing applications such as control, pattern recognition, adaptive beam forming, direction finding, and spectral analysis. Particularly important operations are matrix-vector and matrix-matrix multiplication, Gram-Schmidt orthogonalization, solutions of sets of linear equations, the determination of eigenvectors and eigenvalues of matrices, singular value decomposition of matrices, and least-squares estimates of solutions of sets of linear equations. Of these, the first two, i.e., matrix-vector and matrix-matrix multiplication, are the most fundamental, and they often form an integral part of the other operations. Because of this, there has been considerable emphasis on developing accurate, high-speed, versatile processors for these two operations. A subsequent task will be to determine how such processors can best be configured in larger systems to perform the higher order algebraic operations.

4.2 STANFORD OPTICAL MATRIX-VECTOR MULTIPLIER

The first of these new optical array processors was the invention by Goodman[35] of the Stanford optical matrix-vector multiplier (OMVM). This device, illustrated in Figure 32, is capable of multiplying a 100-component vector by a 100 x 100 matrix in about 20 ns. Components of the input vector $\vec{x}$ are input via a linear array of LEDs or laser diodes. The light from each source is spread out horizontally by cylindrical lenses, optical fibers, or planar lightguides to illuminate a two-dimensional mask that represents the matrix A. Light from the mask that has been reduced in intensity by local variations in the mask transmittance function is collected column-by-column and directed to discrete, horizontally arrayed detectors.

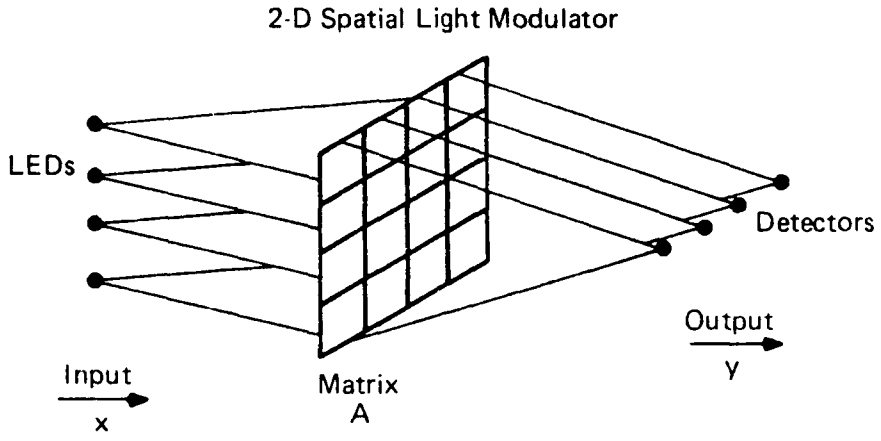


Figure 32. The Stanford matrix-vector multiplier. Not shown in the figure are light-spreading and collecting optics.

The outputs from these detectors represent the components of output vector $\vec{y}$, where $\vec{y}$ is given by the matrix-vector product $\vec{y} = A\,\vec{x}$:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_M \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1N} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{M1} & a_{M2} & \cdot & \cdot & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} \qquad (52)$$

The Stanford matrix-vector multiplier architecture is fully parallel: input and output as well as the computation itself are handled in parallel. In principle, therefore, this architecture is as fast as any currently conceivable processor can be, and is in that sense an ideal optical architecture.

As originally conceived, the Stanford OMVM suffers from several serious limitations:

- Accuracy is limited by the accuracy with which the source intensities can be controlled and the output intensities read.
- Dynamic range is source and/or detector limited.
- Rapid updating of the matrix A requires the use of a high-quality two-dimensional read-write transparency--a spatial light modulator (SLM)--whose optical transmittance pattern can be changed rapidly.

Because the speed of operation of the OMVM optical processor is far greater than any existing system for the input and output of data, research has continued toward developing a more compatible interface of optical processor and surrounding electronic system. Research[5,36] by Casasent, Caulfield, Goodman, and Rhodes resulted in one solution to this problem in which the OMVM is used for iterative algorithms, and the processor output is directed in analog form back to the input to circumvent the data input/output time limitation.

4.3 SYSTOLIC-ARRAY PROCESSING

The next significant development in algebraic processing applications was the systolic-array processor. Systolic-array processing was developed principally by H. T. Kung at Carnegie-Mellon University and S. Y. Kung at the University of Southern California, and is an algorithmic and architectural approach to overcoming limitations of very large scale integrated (VLSI) electronics in implementing high-speed signal-processing operations. Systolic processors are characterized by regular arrays of identical processing cells, primarily local interconnections between cells, and regular data flow. This concept is illustrated in Figure 33a, where an input x, representing one element of an input vector, is passed to the block from the left. An input a, representing one element of the input matrix, is passed to the block from above. An input y, representing a partially accumulated value of one element of the output vector is passed to the block from the right. The block produces two outputs, one simply a duplication of x passed to the right, and the second the output y+ax passed to the left. The joining of several such blocks, as shown in Figure 33b, and a flow of input vector elements from the left and matrix elements from above, all input with the proper timing, results in the sequential output of the components of the vector representing the product of the matrix with the input vector. In Figure 33, any element represented by a heavy black dot does not affect the output components of interest, but a time slot must be present for such elements to assure proper timing. The systolic-array algorithms and architectures are readily implemented in optical versions because of the regular data-flow characteristics of optical devices like AO cells and CCD detector arrays, and the ease of implementing regular interconnect patterns optically. The first optical approach

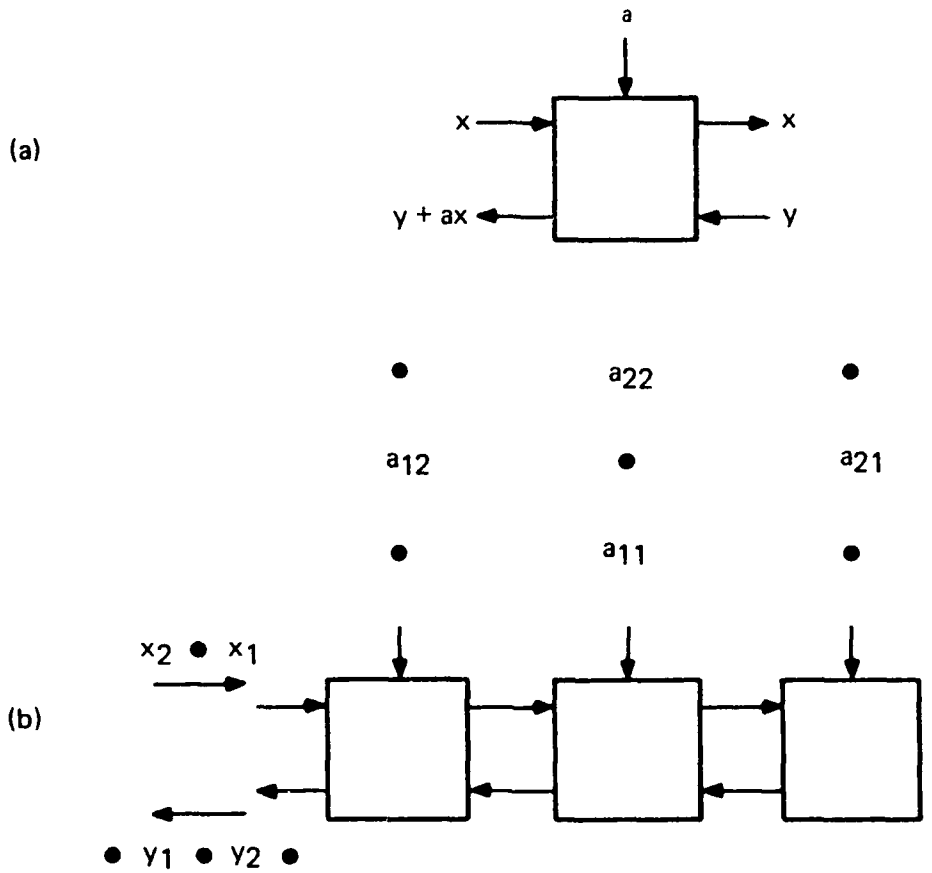


Figure 33. Basic concept of a systolic processor:
(a) Basic building block of a systolic processor ;
(b) Three processors interconnected in a systolic array.

was proposed by Caulfield, Rhodes, Foster, and Horvitz.[37] With their architecture, a single-transducer acousto-optic cell is used to perform matrix-vector multiplication, and for the case of a 2 x 2 matrix-vector where

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and

$$y_1 = a_{11}\, x_1 + a_{12}\, x_2$$

$$y_2 = a_{21}\, x_1 + a_{22}\, x_2 \qquad (53)$$

implementation is based on the beam-modulator mode of operation of the AO cell. Figure 34 shows a system configured for the multiplication of a two-component vector by a 2 x 2 matrix. The processor consists of an input LED or laser diode source array, a collimation lens for each source, an AO cell, an imaging system with focal plane stop, and a linear array of integrating detectors that store charge in proportion to their exposure. The first input to the Bragg cell driver, vector component x_1, produces a short diffraction grating with diffraction efficiency proportional to x_1 that moves across the cell. When that grating segment is in front of LED 1, as shown in Figure 34b, the LED is pulsed with light intensity proportional to matrix coefficient a_{11}, and integrating detector 1 is illuminated with light intensity in proportion to the product $a_{11}x_1$. The next critical moment occurs when the x_1 grating sement is in front of LED 2 and a second grating segment, with diffraction efficiency in proportion to vector component x_2, has moved in front of LED 1, as shown in Figure 34c. At that moment, LED 1 is pulsed with light intensity proportional to a_{12}, and LED 2 is pulsed with light intensity proportional to a_{21}. The integrated output of detector 1 is now proportional to $(a_{11}x_1 + a_{12}x_2)$, which is the output vector component y_1. The integrated output of detector 2 is $a_{21}x_1$ at this stage. The final critical moment in the computation, shown in Figure 34d, occurs after grating segment x_2 has moved in front of LED 2. A final pulse from LED 2, in proportion to a_{22}, yields at the output of detector 2 a voltage in proportion to $(a_{21}x_1 + a_{22}x_2)$, the second component y_2 of the output vector. The computation is now complete.

The dimensionality of the vector entering into the product, N, is straightforwardly increased from two by adding more modulated beam sources. Vector dimensionality is theoretically limited to a maximum value of TB, as noted in Section 2. The evaluation of a matrix-vector product by this processor takes T seconds (the AO cell time window) before the first y_i comes out. T is thus the latency of the processor. It takes another T seconds to complete the entire matrix-vector product, yielding a total of 2T processing time. The maximum number of operations (multiply-adds) performed in that time is N^2, where $N < T$. Thus the theoretical limit on processing rate is given by

$$\text{processing rate} < \frac{B^2 T}{2} \text{ operations/s}$$

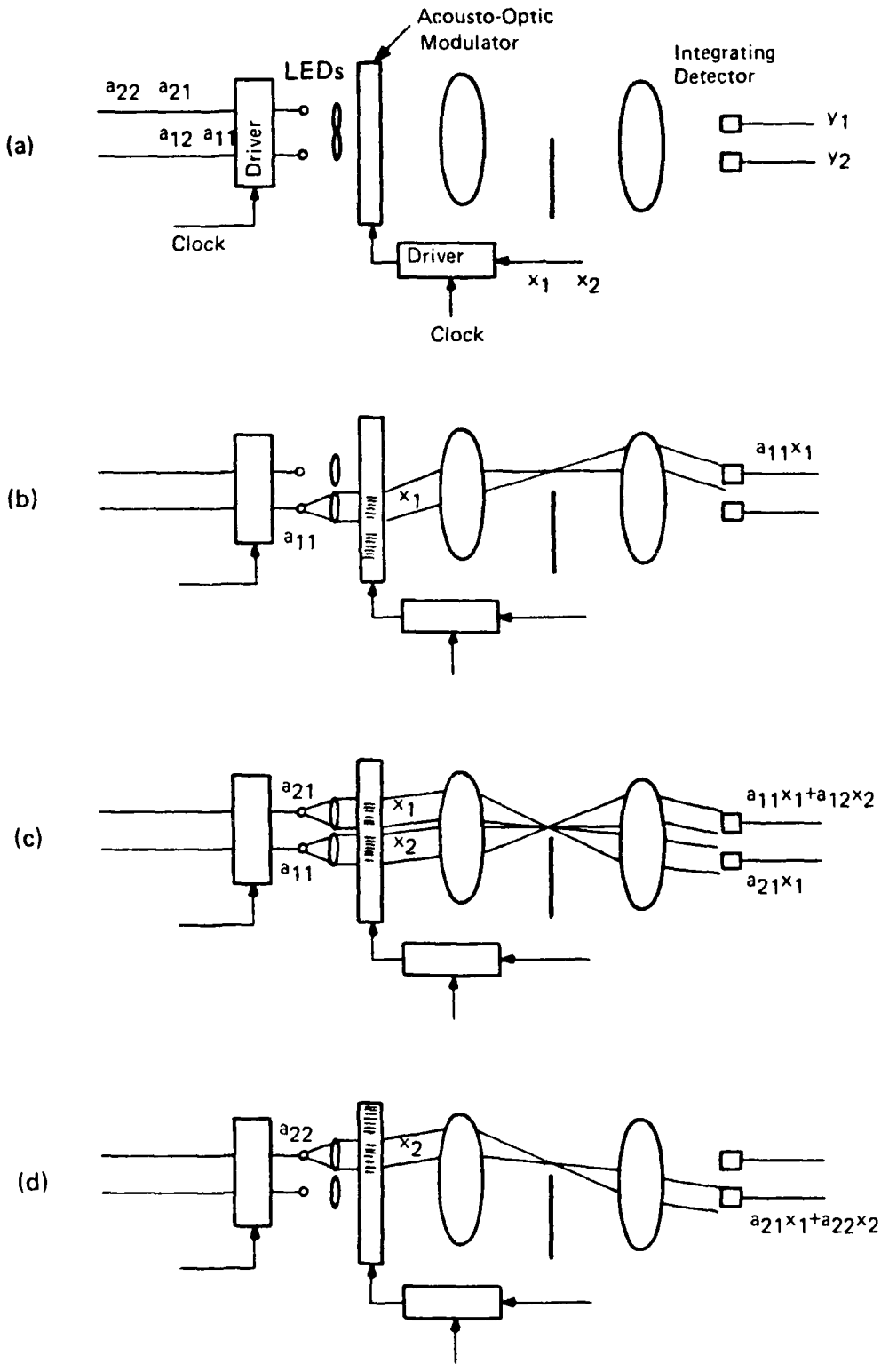


Figure 34. Systolic architecture matrix-vector multiplier, 2 x 2 example:
(a) General system;
(b) First critical moment in operation;
(c) and (d), Subsequent critical moments in system operation.

which evaluates to 5×10^{10} operations/s assuming a 100 MHz bandwidth cell with a 10 μs time window--numbers typical of many off-the-shelf Bragg cells. An additional factor of 10 can be achieved with available higher bandwidth cells.

Following the development of the optical systolic matrix-vector multiplier, two other important advances took place: the invention of optical matrix-matrix multipliers and the achievement of digital accuracy with optical algebraic processors.

4.4 MATRIX-MATRIX MULTIPLICATION SYSTEMS

The operation of a matrix-vector processor is easily extended to accommodate matrix-matrix multiplication, since a matrix-matrix product can be evaluated as a succession of matrix-vector products, i.e., the matrix-matrix product AB = C given in the 3 x 3 case by

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \tag{54}$$

can be written as

$$A\,[b_1 b_2 b_3] = [c_1 c_2 c_3] \tag{55}$$

where A is the 3 x 3 matrix of coefficients a_{ij} above and

$$b_1 = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \qquad b_2 = \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} \qquad c_3 = \begin{bmatrix} c_{31} \\ c_{32} \\ c_{33} \end{bmatrix} \tag{56}$$

Equation 56 suggests that matrix C is calculated by evaluating column vectors c_1, c_2, and c_3 in sequence. The data flow is important to these operations. For this example, the calculation of the partial sums of c_2 can commence prior to completion of the calculations of c_1, and so forth. The processor latency T thus applies only to computation of column vector c_1. With matrices 10 x 10 or larger, this latency can be ignored relative to the overall processing time, giving a processing rate for the matrix-matrix product case of B^2T operations per second, or 10^{11} multiply-adds per second for the 3 x 3 case.

The processors discussed in this section operate with light intensities, which are always non-negative. Thus, if bipolar or complex-valued vectors and matrices are to be multiplied, multiplexing or coding schemes must be used. Techniques available include two-component representations for real numbers and biased real-imaginary representations as well as three- and four-component representations for complex numbers. All of these schemes result in some reduction in system throughput (by a factor of two or three) and an increase in system complexity.

Figure 35 shows a system proposed by Casasent et al.[38] that uses both AO beam modulation and beam deflection for vector-matrix and matrix-matrix multiplication. In the vector-matrix multiplication mode, the initial operation requires that all laser diode sources are off while the Bragg cell is loaded with a sequence of composite grating segments, each of which can diffract light from a given input beam to any combination of output detectors with arbitrary weighting. When the composite gratings are in the correct positions, the laser diodes are strobed with intensities proportional to vector components x_1, x_2, etc., as shown. As an example, assume an output vector component y_1 is given by $y_1 = 3x_1 + 4x_2 + 2x_4$. When the sources are strobed, beam 1, with intensity proportional to x_1, has part of its energy diffracted to detector 1 with diffraction efficiency 3k, k being some proportionality constant. Simultaneously, beam 2, with intensity proportional to x_2, is diffracted to the same detector with diffraction efficiency 4k. Beam 4, with intensity proportional to x_4, is diffracted to detector 1 with efficiency 2k. The result is an output at detector 1 proportional to $k\ (3x_1 + 4x_2 + 2x_4)$, i.e., proportional to y_1. As this is happening, light is also being diffracted in correct amounts to the other detectors to calculate y_2, y_3, etc.

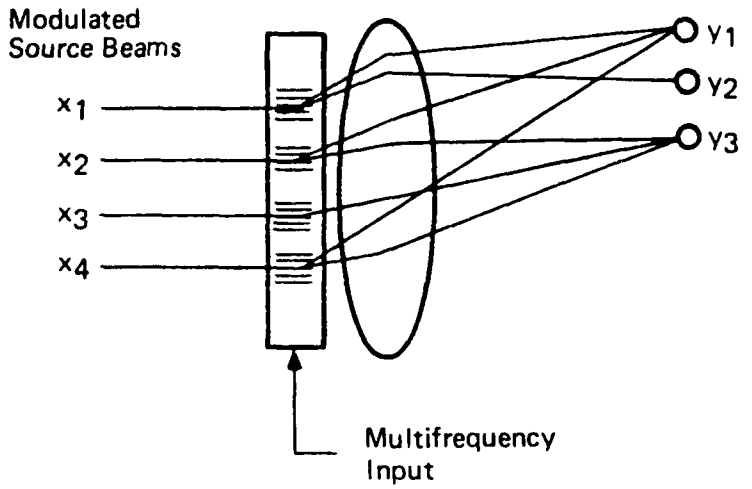


Figure 35. Beam deflector based matrix-vector multiplier.

The time it takes for a single N-component matrix-vector product to be evaluated is determined almost entirely by the fill time T for the Bragg cell, the strobe time being negligible by comparison. N^2 analog multiply-adds are performed, where N must satisfy the constraint ($M = N$ for the vector-matrix product case) $N < (BT)^{1/2}$. The number of operations performed per second thus cannot exceed the cell bandwidth B, which may range from 10^7 to 10^9 Hz. The numbers again assume a Bragg cell bandwidth of 100 MHz and a time window T of 10 μs, which is a factor of 10 less than that now achievable with available Bragg cells.

Because the matrix size N is limited to $(BT)^{1/2}$, the beam-deflector approach is suitable only for relatively small matrices (although larger matrices can be accommodated by partitioning methods), and it is clearly slower for single matrix-vector multiplication than the beam-modulator architecture. However, the processing rate increases when matrix-matrix products are evaluated as a succession of vector-matrix products, for it is not necessary to refill the Bragg cell prior to calculating the next vector-matrix product.

Figure 36 shows how matrix-matrix multiplication is performed using this basic architecture. The approach is philosophically the same as for the vector-matrix multiplication. The matrix C, given by the matrix-matrix product AB, is calculated vector by vector as in Equation 55. At the instant

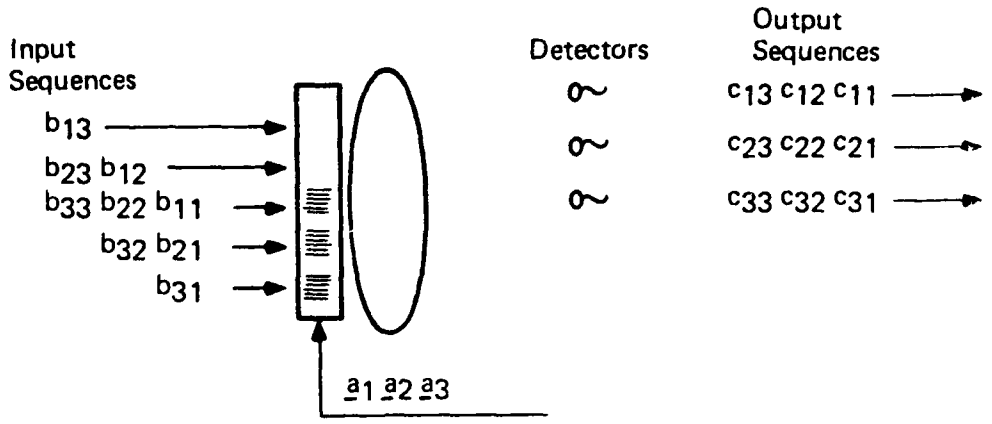


Figure 36. Matrix-matrix multiplication with the system of Figure 35. Vectors a_i denote columns of Input A matrix.

depicted in Figure 36, vector c_1, represented by $[c_{11}c_{21}c_{31}]^t$ (where t denotes transpose), is calculated by flashing the laser diodes in proportion to b_{11}, b_{21}, and b_{31}, as shown. T/N seconds later, the grating segments have moved up to the next position, where $c_2 = [c_{12}c_{22}c_{32}]^t$ can be evaluated, and so forth. For the matrix-matrix product evaluation, there is a T-second latency (assuming T is still the transit time of the entire Bragg cell), T-second additional processing time, and a total of N^3 multiply-adds performed, where $N < (BT)^{1/2}$, for an overall processing rate of $(BT)^{3/2}/Ts^{-1}$. A typical rate with B = 100 MHz, T = 10 μs is 4×10^8 operations per second.

It would appear from the above that the beam-modulator approach would have a clear advantage over the beam-deflector approach for both vector-matrix and matrix-matrix multiplication. However, the rates calculated are based on fundamental theoretical limitations, and in practice the advantage is not so clear. For example, the processing rates for the beam-modulator architecture were calculated assuming that the number of input beams were given by the time-bandwidth product of the cell. The time-bandwidth product is typically about 1000, and in practice it is very difficult to achieve so many individually modulated input beams. For the beam-deflector approach, the corresponding upper limit of about 30 input and 30 output beams is not unrealistic to expect. Another important consideration is the requirements imposed on the support electronic systems. To achieve maximum allowable throughput from the

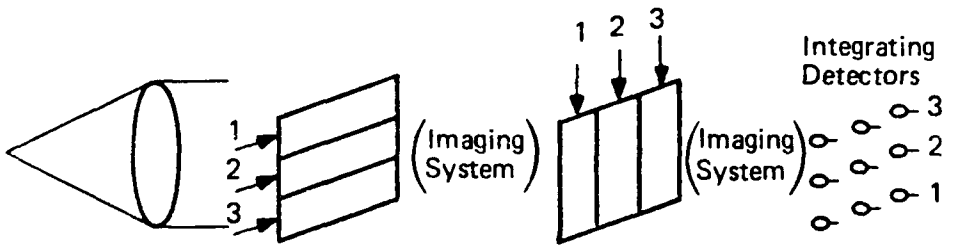
beam-modulator architecture, each source must operate as fast as 100 MHz. Conditions are greatly reduced for the beam-deflector architecture, which requires input beam source modulation at only 1/30 that rate. Also, with significantly fewer sources and detectors, the number of electronic support components is greatly reduced. Finally, the processing rate of the beam-deflector architecture can be increased to essentially that of the beam-modulator architecture, if the sources are modulated at significantly higher rates, for then the entire calculation can be performed without requiring any pipeline motion of the grating segments in the Bragg cell.

4.4.1 Multitransducer Processor Architectures

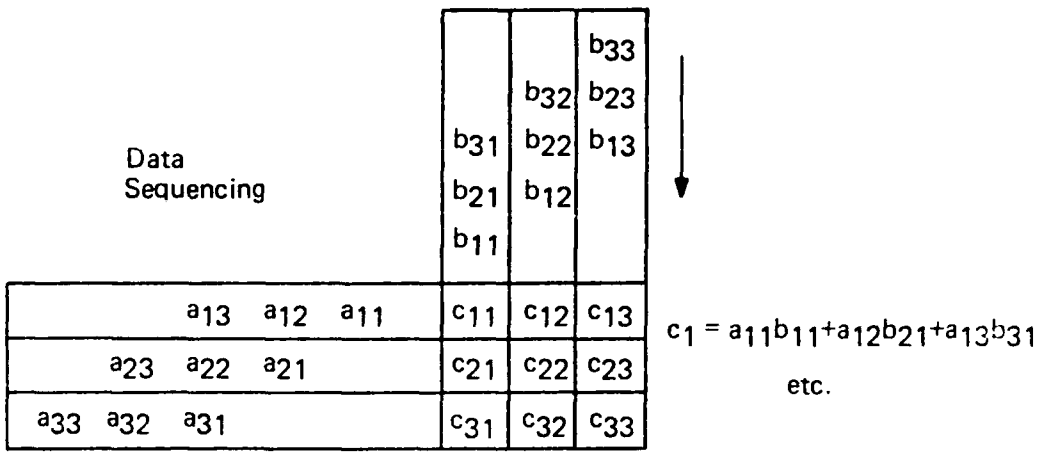
The systems discussed so far have used Bragg cells with single transducers where only a single acoustic beam is present for AO interaction. Some of the most recent developments in AO signal processing have been based on multitransducer cell architectures, e.g., high-quality Bragg cells have been fabricated with as many as 100 transducers. In these multitransducer Bragg cells, each transducer produces its own acoustic beam, and cells with 10 to 30 transducers are now readily available.[39]

Figure 37 shows one example of a multitransducer cell arthitecture for doing matrix-matrix multiplication using AO beam-modulating methods.[37] The systems consist of two three-transducer Bragg cells, imaged (with appropriate stops to block undiffracted light) onto one another and then onto a 3 x 3 array of detectors. Illumination is spatially uniform and pulsed in time. Because of optical stops in the imaging system, only light diffracted by both Bragg cells arrives in the detector plane. Thus, if row transducer 1 and column transducer 2 are the only two to receive signals, only the detector at location (1,2) will be illuminated.

For matrix-matrix multiplication, the components of the input matrices are sequenced into the two orthogonal cells as suggested by Figure 37b. The coefficients aij are input horizontally, a_{11} first, then a_{12} and a_{21}, and so forth. Simultaneously, the coefficients bij are input to the vertical cell transducers, b_{11} first, then b_{21} and b_{12}, and so forth. As they move, the grating segments representing these numbers effectively cross each other in space, causing light to be diffracted to detectors in corresponding spatial locations. The first significant event occurs when grating segments b_{11} and



(a)



(b)

Figure 37. Multitransducer Bragg cell architecture for matrix-matrix multiplication:
(a) System;
(b) Output end data sequence.

a_{11} are imaged onto each other. At that instant, the common source is pulsed, and doubly diffracted light energy in proportion to the product $a_{11}b_{11}$ is sent to the integrating detector at position (1,1). A short time later, aftermovement of the grating segments through one beam width, light intensity in proportion to the product $a_{12}b_{21}$ is sent to the same detector, and so on, until the entire sum $k(a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31})$ (k constant), proportional to c_{11}, has been integrated. Similarly, at other integrating detectors, other partial sums are being evaluated to calculate output matrix coefficients c_{12}, c_{13}, c_{21}, etc. With this architecture as in previous examples, all numbers must be positive, and multiplexing must be used to implement full real or complex arithmetic, and extension to larger matrices is straightforward.

Another approach to multitransducer architecture is the optical outer product calculator.[40] Figure 38 illustrates the system schematically. In this approach, the individual sound columns in the multitransducer Bragg cell are short and are used as point modulators for light passing through them. Light from the vertical laser diode array is spread out and recollected by optics (not shown) so as to illuminate a square array of detectors in the output plane. The intensity of each horizontal row is proportional to the intensity of the corresponding laser diode, and the intensity of each column of the output array is determined by the diffraction efficiency of a given Bragg cell sound wave. With this architecture it is possible to calculate outer products, i.e., matrix-matrix products of the type

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} [y_1\ y_2\ y_3] = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{bmatrix} \qquad (57)$$

and a row vector when left-multiplied by a column vector produces a two-dimensional rank-1 matrix; this operation is commonly called an outer product between the two vectors. These calculations are integral to the calculation of covariant matrices, which are very important in algorithms for linear algebra, image processing, and signal processing. Algorithms that can be implemented optically using outer product concepts include matrix multiplication, convolution/correlation, matrix decomposition, and binary arithmetic operations.

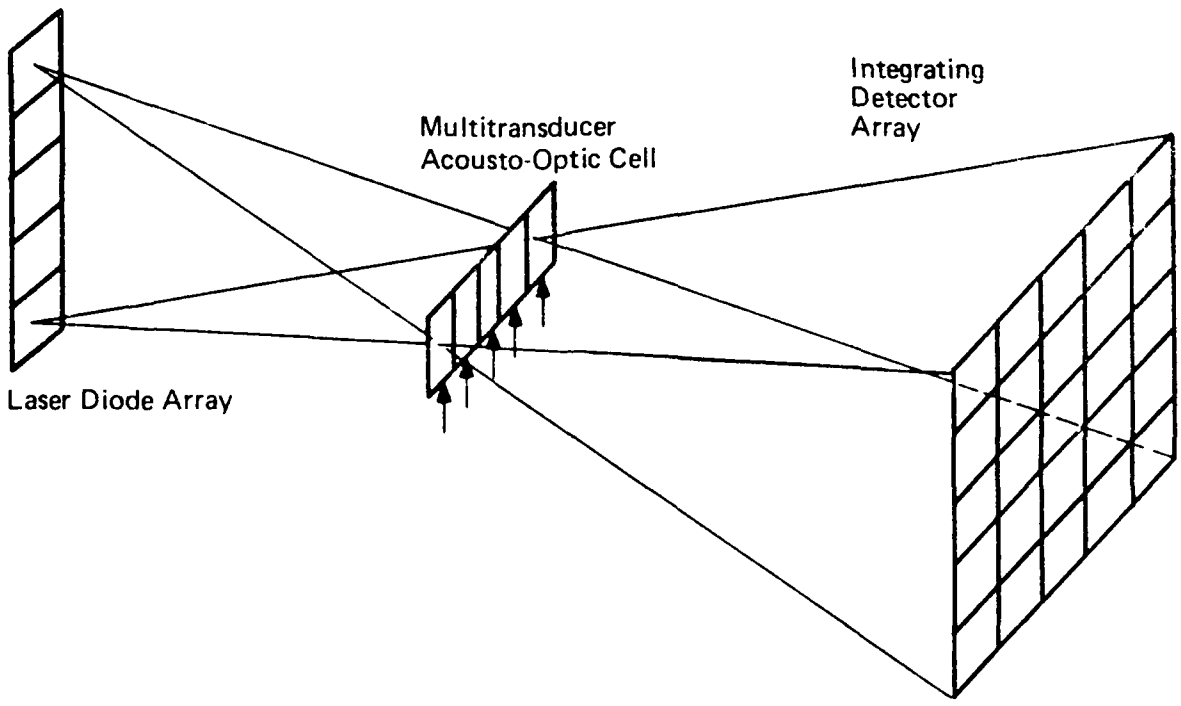


Figure 38. Illustration of an optical outer-product calculator using multitransducer Bragg cell architecture: Optical components not shown spread light and image it in appropriate directions.

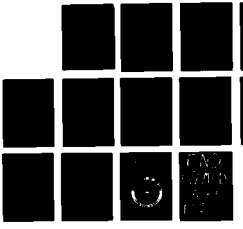
A succession of outer products can be used to calculate arbitrary matrix-matrix products by appropriate decomposition and summing, as suggested by the following equation:

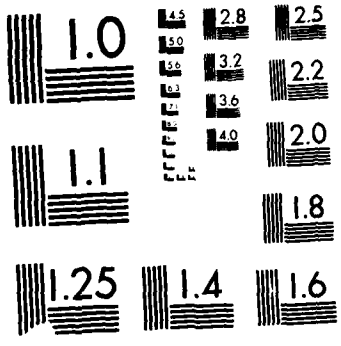
$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} [b_{11}\ b_{12}\ b_{13}]$$

$$+ \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} [b_{21}\ b_{22}\ b_{23}] + \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} [b_{31}\ b_{32}\ b_{33}] \qquad (58)$$

AD-A207 524 STATE OF THE ART REVIEW: OPTICAL PROCESSING(U) TACTICAL 2/2
WEAPONS GUIDANCE AND CONTROL INFORMATION ANALYSIS
CENTER CHICAGO IL J S HARRIS JAN 89 GACIAC-SOAR-87-01
UNCLASSIFIED DLA900-86-C-0022 F/G 20/6 NL





UTION TEST CHART

AO devices are not the only modulators that can be used for this operation, which requires only some kind of multitransducer linear array modulator. However, because of the extent of development in AO devices, they are very attractive candidates for the task.

4.4.2 Digital Accuracy in Matrix-Matrix Multiplication

Since the optical matrix multipliers discussed earlier are analog processors, the accuracy of the final computational results is typically limited to 8 to 10 bits by the dynamic range of the input light sources, output detector array, and the spatial light modulator's transmission. For applications such as algebraic signal processing, where considerably greater accuracy is required, this limitation can be circumvented by combining a binary matrix representation with the algorithm of binary multiplication by analog convolution.[41] This advance takes advantage of the fact that the multiplication of two binary numbers can be viewed as a convolution between the two binary bit systems involved, followed by a carry propagation operation that converts the set of nonbinary partial products obtained to a binary representation.

As an example of this algorithm, consider the multiplication of 23 by 25. If the binary multiplicands {10111} and {11001} are viewed solely as sequences of ones and zeros, their discrete convolution will give the output number sequency {1000111111}, which is the mixed binary representation of the product 575.

A discrete convolution can be performed acousto-optically, as shown in Figure 39 for the binary sequence {10111} and {11001} just described. The sequences are input as square wave modulation on RF carriers to a pair of Bragg cells, which are constructed so that the signals propagate in opposite directions. (In practice, the two cells might be imaged onto each other.) Light that is diffracted by both cells is collected by the lens and brought to a focus at the detector. The analog output signal from the detector is shown as in inset in Figure 39, and conveys the mixed binary sequence {11123111} as the heights of triangular pulses. To obtain the full binary digital result, the triangular peaks of the analog output signal are digitized, shifted, and added electronically to base 2^5.

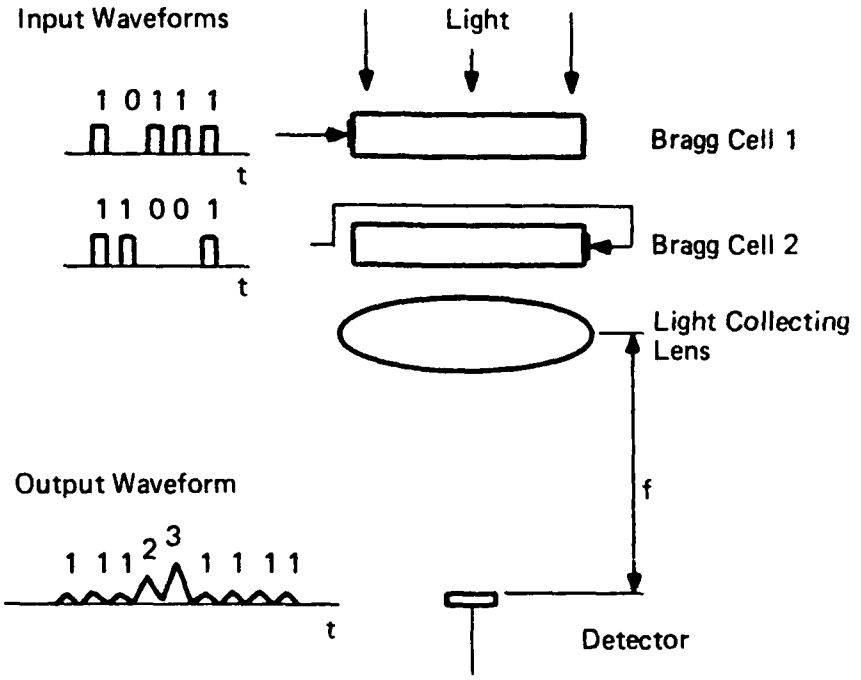


Figure 39. AO implementation of discrete convolution: Inset triangular waveform conveys mixed-binary results.

Since only ones and zeros are represented at the input, the Bragg cells can be operated at peak diffraction efficiency without concern for nonlinear response. Furthermore, the AO convolver is only required to have sufficient accuracy to allow a small number of levels to be distinguished at the output. For 5-bit inputs (the example 23,25) the triangular peaks of the analog output signal will range after quantitization from zero to a maximum of 5. In general, N-bit inputs require that N levels be correctly distinguished at the output. Since binary representations are used, negative numbers can be accommodated using 2^5 complement arithmetic[43] or similar methods.

The strengths of optical processing in convolutions have led to innovative proposals for optical processors capable of achieving numerical accuracies comparable to those of electronic digital computers. One of the most important in this area is the systolic acousto-optic binary convolver proposed by Guilfoyle,[42] whose approach uses a multitransducer AO technology that can operate with at least 32 bits of accuracy at processing rates of 10 multiplications-adds per second. The key to the approach is a combining of the systolic AO vector-matrix multiplier concept shown in Figure 34 with a

method for digital multiplication via discrete convolution (serial product). Guilfoyle's system was a single collimated light source for illumination, a pair of multitransducer Bragg cells (one with a low acoustic wave velocity, the other with a high velocity), a linear array of detectors, and imaging and light collecting lenses. The two Bragg cells are imaged onto each another, and subsequent optics are adjusted so that only light diffracted by both cells reaches the detector plane.

The operation of the processor is described with the help of Figure 40, which shows the two multitransducer Bragg cells with their respective inputs. The system configuration is appropriate for the case of a 32-component vector-matrix product where the components of the input vector and matrix are represented by 10-bit binary numbers. The cell with 32 transducers is loaded bit-serially with the binary sequences representing the components aij of the matrix A. The least significant bits of the 10-bit numbers are loaded first. The time sequence of the components is the same as for the systolic vector-matrix processor of Figure 34. The other multitransducer Bragg cell is loaded bit-parallel with the bit representations of vector components b_1, b_2, ..., b_{32}. The most significant bit is loaded into the bottom channel, the least significant bit at the top. Within the Bragg cells, the bits are represented by short acoustic grating segments that diffract the incident light.

The two Bragg cells are imaged onto each other, and only light diffracted by both cells reaches the detector plane. A convenient way to visualize system operation is to think of the grating segments as being small holes in an opaque sheet. The holes move with the velocity of sound, and light reaches the detector plane only when one hole passes over another. The light is summed in the vertical direction by lens components, i.e., all light transmitted by a particular column reaches one specific detector.

Cell 1, carrying the bits of the matrix A components, has an acoustic velocity 10 times greater than that of cell 2. This means, for example, that the 10-bit signal stream corresponding to, matrix coefficient a_{11} moves through and entirely past the bit pattern representing vector component x_1 before the latter pattern has had an opportunity to move a significant distance. This relative movement of the two bit patterns is fundamental to the discrete convolution operation. To complete the convolution, the doubly

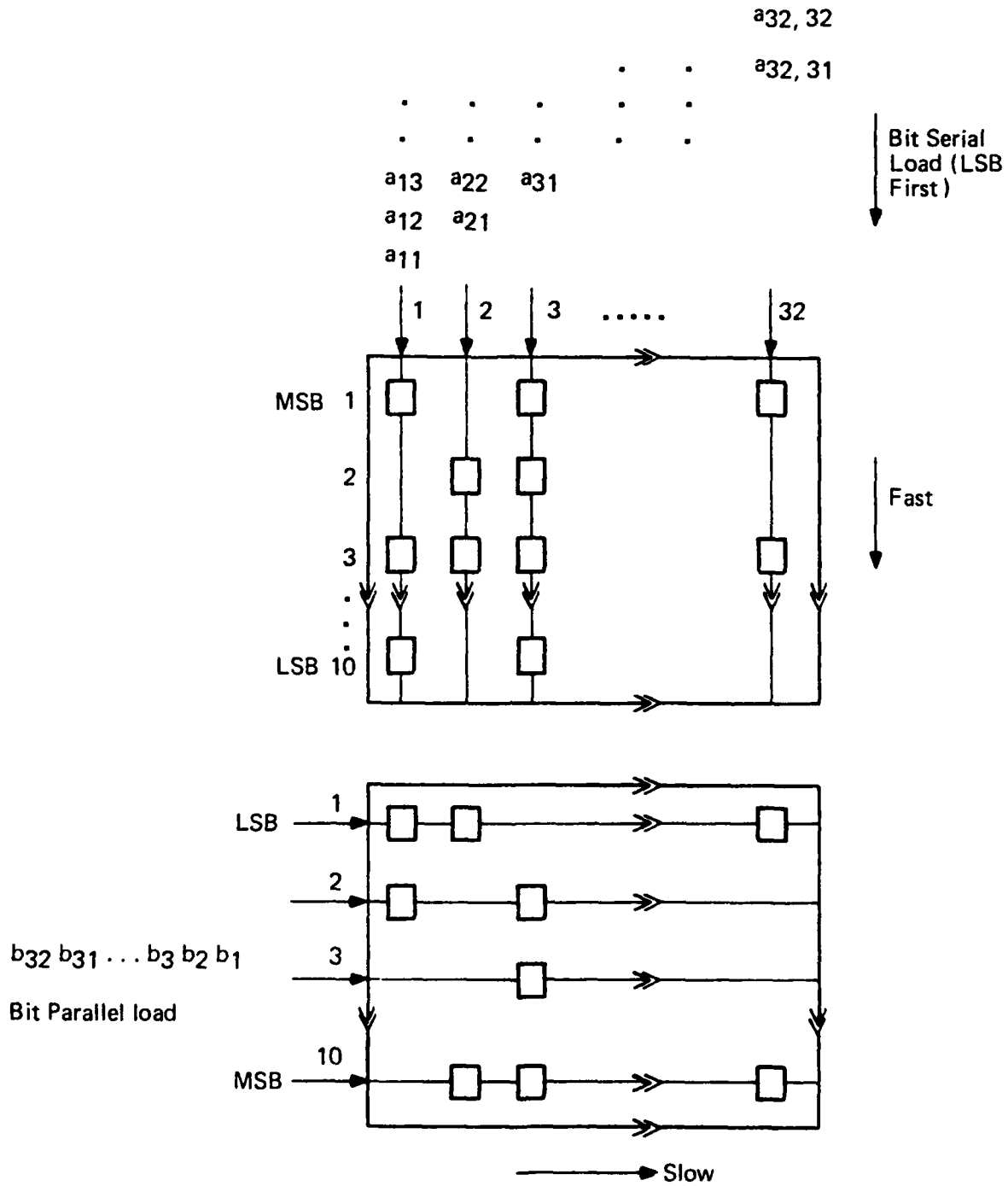


Figure 40. Sequencing of data into two multitransducer Bragg cells for multiplication of 32-component Vector b by 32 x 32 Matrix A: All components specified by 10-bit binary numbers.

diffracted light in a particular column (corresponding to a given aij) is collected by a cylindrical-spherical lens combination and focused to the appropriate detector. With time, the output plane detectors thus output triangular waveforms representing products of the general form aijxj. These waveforms are converted into the full binary representations and then summed to produce full binary representations of the output vector components yi.

5. BISTABLE OPTICAL DEVICES FOR OPTICAL SIGNAL PROCESSING APPLICATIONS

Bistable optical devices[44-46] are nonlinear optical devices that can be switched between two stable conditions with different optical characteristics. Applications of interest to optical processing include the ability to build optical logic elements analogous to electronic logic components, differential gain in which the device allows one beam to control a more intense beam, and power limiters to provide an almost constant transmittance as the incident intensity is varied. In these and most other applications, the bistable optical device is operated as a three-port device in much the same way as an optical transistor.

5.1 BISTABLE OPTICAL DEVICES

The primary functions of a bistable optical device are to determine when an incident optical signal exceeds a set threshold at a particular spatial location, to use the detected optical signal to spatially modify a local material property in such a fashion that distinct stable states correspond to regions illuminated below and above threshold, and to encode the amplitude and/or phase of a readout illumination beam with material state-dependent information.

Applications of bistable optical elements in optical processors and computers include the implementation of logic functions (AND, OR, NOR, etc.), level restoration, level amplification, bistable switching with momentary contact for simultaneous inputs, bistable latching for serial inputs, and variable thresholding. The use of bistable optical devices for these applications will parallel the use of semiconductor switching elements in integrated electronic circuits. The fundamental difference and potential advantage of the optical device is the availability of two-dimensional arrays of bistable optical elements that can be accessed in parallel by static or dynamically programmable optical interconnections.[47] "Optical interconnections" refers to the ability of optics to provide multiple independent paths between different computational components, i.e., an array of light

emitting diodes or diode lasers can be simultaneously imaged through a spatial light modulator or bistable array to a readout device in an optical processor.

The parallel-access feature is crucial to the eventual incorporation of bistable optical elements in optical processing and optical computing systems. If the only capability offered by a bistable array were the contruction of two-dimensional arrays of switching elements, the interest in them would not be so great because of the massive investment required to develop this and the required associated technology. The success of the rapidly developing very large scale and wafer-scale integrated circuits in providing densely packed two-dimensional arrays of switching elements would greatly limit the interest in funding this effort. If, however, reconfigurable interconnections can be implemented in the context of a given switching technology, distinct advantages accrue for the solution of extremely complex classes of computational and processing problems.[47] In this case, the distinctions between the interconnections and the processor as separate functional elements fade, and one can see the potential of the combined entity as a reconfigurable machine. It is the goal of these efforts to develop an array of all-optical switches in order to meet the long-term goal:[45] a clocking speed of tens (or even hundreds) of gigahertz through the system.

5.1.1 Nonlinear Fabry-Perot Etalon

The work in a number of fundamental physical effects has been used to produce bistable optical devices.[44,48-51] A large number of these devices depend for their operations on intensity-dependent modification of the effective optical path lengths of a Fabry-Perot etalon. The essential operating principle of this type of bistability is illustrated in Figure 41, which shows an optical material with intensity-dependent index of refraction arranged between two partially transmitting mirrors to form a resonant cavity (left side of figure). Incident illumination results in a buildup of intensity within the cavity by means of multiple reflections between the partially reflective mirrors. The transmitted and reflected intensities from the device depend on the optical path length within the cavity, and the difference in optical path determines whether constructive or destructive interference occurs.

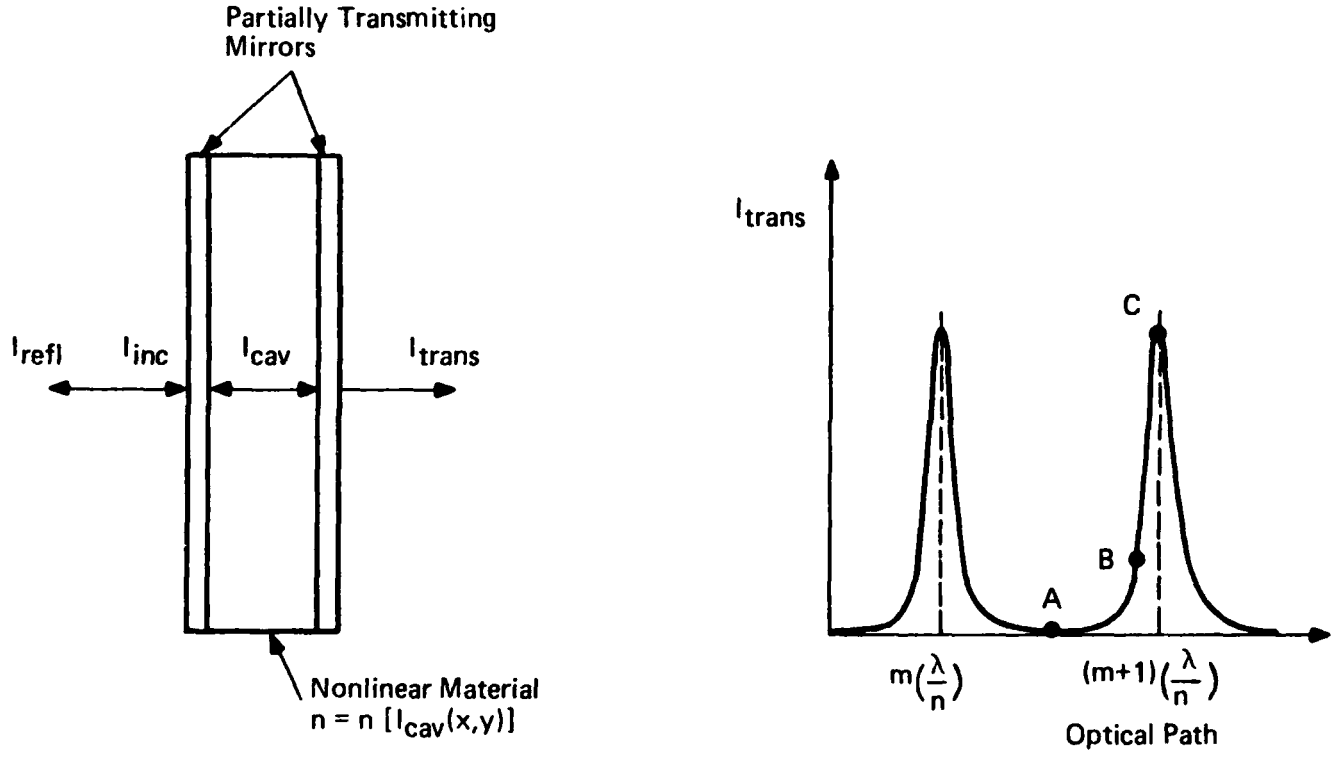


Figure 41. Fundamental principles of optical bistability as observed in a nonlinear medium placed in a Fabry-Perot etalon.

The dependence of the transmitted intensity on path length (plotted in units of the optical wavelength divided by the cavity's index of refraction) is shown at the right of the figure. Resonant behavior is observed such that peaks in the transmitted intensity occur whenever the path length is equal to an integral number of wavelengths. For normal incidence as shown in Figure 41, the optical path length is 2nd where n is the medium's index of refraction and d is the cavity thickness.

The most important consideration for an optical bistable device is that it exhibit, over a certain range of input intensities, two distinct output levels for each input level, i.e., two different output levels may be achieved, depending on the input intensity. This is illustrated by Figure 42, and is obtained as follows for the Fabry-Perot etalon of Figure 41: Consider the initial optical thickness of the cavity to be such that the system intensity transmitted by the etalon is at the point labeled A. This point also corresponds to the origin of the plot of transmitted intensity as a function of incident intensity as shown in Figure 42. As the incident intensity is increased, the intensity in the resonant cavity increases and, as seen from Figure 41, the transmitted intensity increases. Since it is assumed that the cavity medium is a nonlinear material whose index of refraction is a function of the intensity within the cavity, the increase in intensity within the cavity alters the resonance condition. If the index of refraction is increased (the resonance condition of an optical path difference equal to some multiple of the wavelength λ divided by the index of refraction n), the resonance will occur at a lower value of the path length. This is the equivalent of moving from A toward B in Figure 41.

Above a critical threshold in incident intensity, this situation is unstable. The intensity-induced shift in the resonance condition lowers the reflected intensity and increases both the transmitted intensity and the intracavity intensity. The latter increase tends to further increase the refractive index of the material, which creates positive feedback and drives the system to a stable equilibrium condition as represented by point C in Figure 41. Further increases in the incident intensity increase the cavity intensity, increasing the refractive index, the result of which is to decrease the transmitted intensity. This results in negative feedback, which stabilizes the transmitted intensity in the saturation region of Figure 42.

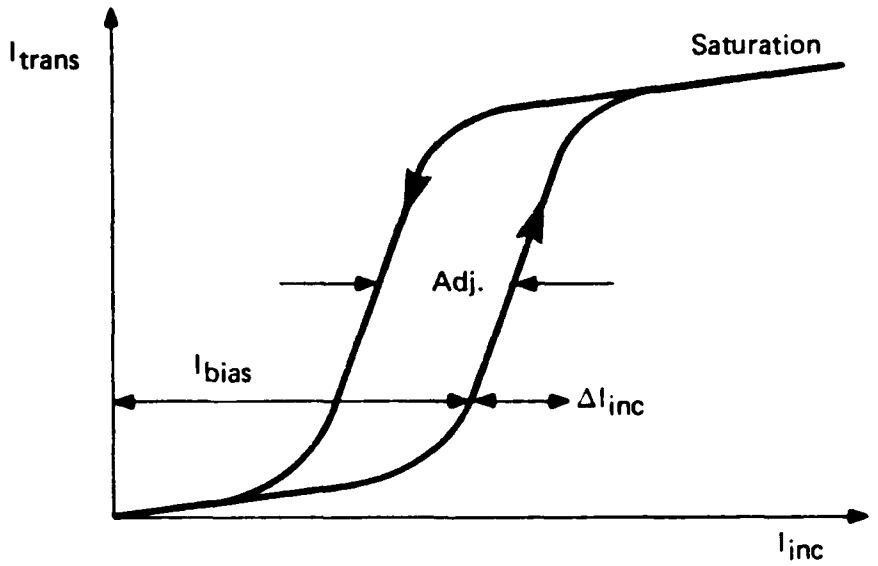


Figure 42. Illustration of the hysteresis and saturation effects characteristic of optical bistability.

Hysteresis occurs as a result of the fact that a relatively large decrease in the incident intensity is now required to significantly detune the cavity from the resonance condition and switch back to a region below the threshold for positive feedback. The existence of two distinct values of the transmitted intensity for a given value of the incident intensity is a clear indication of the presence of a bistable mechanism. As indicated in Figure 42, the width of the bistable hysteresis characteristic is adjustable, and depends on the initial relationship between the actual path length of the cavity and the optical path length at zero input intensity.

The addition of a bias intensity level allows a number of different functions to be performed. For example, if the bias is set as indicated in Figure 42, the device will amplify small additional signal intensities as represented by the steep slope of the characteristic curve. If binary input intensities are used, the bias provides a significant increase in the optical sensitivity of the switch. The bias level can be set so that two units of coincident signal intensity are required to switch into saturation, which implements the Boolean logical AND function. Other logical operations can be performed with

appropriate use of the bias input or of the phase delay between the bias input pulse and the signal inputs.[52]

5.1.2 Self-Electro-Optic Effect Devices

The basis of operation of the self-electro-optic effect device (SEED), which was developed and patented by David Miller[53] at AT&T Bell Laboratories, is a multiple quantum well (MQW) material. MQW structures are a special case of the general class of periodic multilayer materials, as shown schematically in Figure 43. The layers are grown by molecular beam epitaxy (MBE), metal organic chemical vapor deposition (MO-CVD), or liquid phase epitaxy (LPE) techniques. Layer thicknesses range from atomic monolayers of only a few angstroms to more than a thousand angstroms. Depending on the nature of the

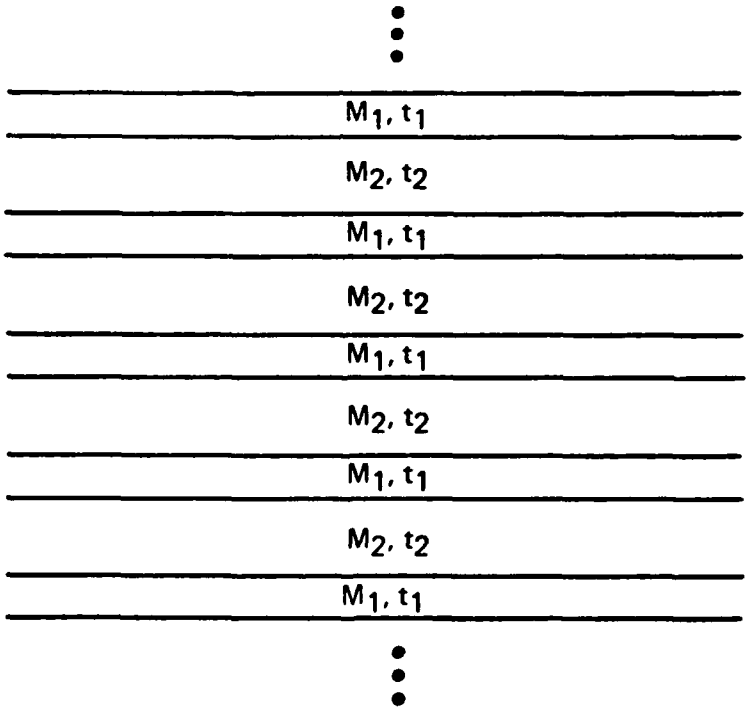


Figure 43. Generalized diagram of a periodic multilayer multiple quantum well structure/superlattice: Grown by MBE, MO-CVD, and LPE techniques.

structure, the total number of layers can range from five to 500. MQW structures show some interesting effects when electric fields are applied to the material. One such novel effect is called the quantum-confined Stark effect (QCSE),[54] which is both large and rapid. One important aspect of the effect is that it appears to rely on the confinement of the carriers within the thin semiconductor layers; thus it is truly a quantum well effect and will not exist in bulk semiconductors at any temperature. The existence of excitons (sharp absorption features near the optical band edge) at room temperature is what makes MQW structures so interesting and useful.

At low temperatures in bulk semiconductors the absorption near the fundamental edge is governed by excitonic effects. Excitons are electron hole pairs forming a bound state analogous to the hydrogen atom. They produce very sharp resonance peaks just below the band gap, where a large oscillator strength is concentrated in a narrow spectral domain. These resonances correspond to the creation of excitons, not the excitation of existing particles; a good analogy is the creation of positronium atoms in a vacuum by a photon. The excitonic resonances have been extensively investigated in linear and nonlinear optics, but so far they have not been used because of the low temperatures at which they are usually observed. The physical mechanism that prevents observation of exciton resonances at room temperature is that, in polar semiconductors, longitudinal optic vibrations produce strong electric fields that ionize the weakly bound excitons.

In recent years, modern techniques of crystal growth have enabled the fabrication of semiconductor heterojunctions that are smooth down to one atomic monolayer, with almost perfectly controlled composition. Using pairs of semiconductors with very specific physical and chemical compatibility, it is possible to grow alternate ultrathin layers of each compound to form MQW structures using III-V and II-VI semiconductors.

Because of the band gap differences between the two components, the electron hole pairs are confined in the low gap layers. The motion of the carriers has to be quantitized both along the normal to the layers (z) and in the plane (x,y). For ultrathin layers with thickness $L_z \approx 100$ Å, the confinement along z produces a series of discrete states, whereas the particles are free to move in the x-y plane (see Figure 44).

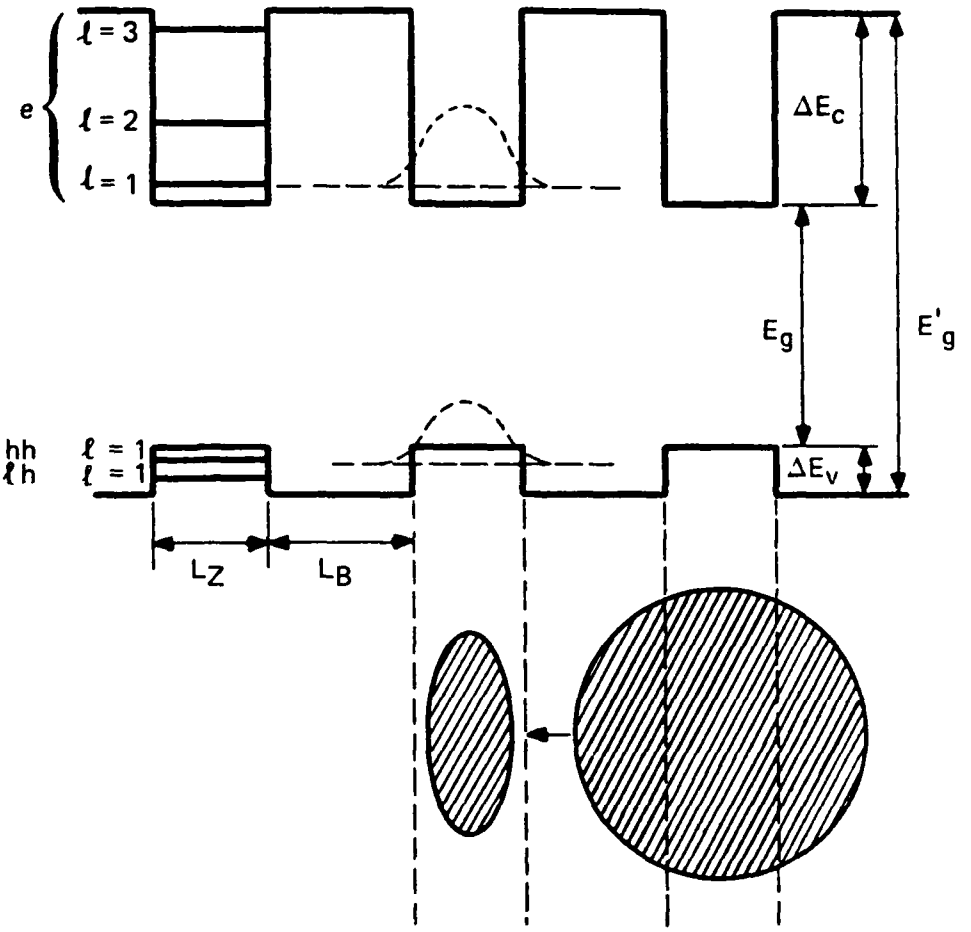


Figure 44. Schematic of the band structure of a multiple quantum well structure along the normal to the layers (z) in real space:
(1) Dashed line represents the electron and the hole wave function along z;
(2) Striped circle and ellipse illustrate how the exciton is compressed by the confinement.

The effect of the confinement on the density of states that describe the optical transitions is to transform the usual parabolic edge into a series of steps. It also raises the degeneracy of the upper valence band of III-V semiconductors by introducing a splitting between the heavy and the light holes. The electron and holes still interact through the Coulomb interaction, but in MQW structures the electron hole bound states are flattened, the average distance between the two particles is reduced, and the exciton binding energy

is increased, resulting in enhanced excitonic effects. In fact, because MQW structures have two valence subbands at each absorption edge, two excitons can be seen involving the heavy and light holes, respectively. However, although the exciton binding energy is increased in the MQW structure, the interaction with the phonons is almost unaffected for two reasons. First, the pair of compounds that form an MQW structure usually have very similar phonon spectra; in addition, the quasi two-dimensional excitons are mostly localized in the low gap layers and thus do not significantly probe the other compound. This is sufficient to produce sharp excitonic resonances at room temperature in the absorption spectra of high quality GaAs/AlGaAs and GaInAs/AlInAs MQW structures, as shown in Figure 45. At room temperature, the excitons live just

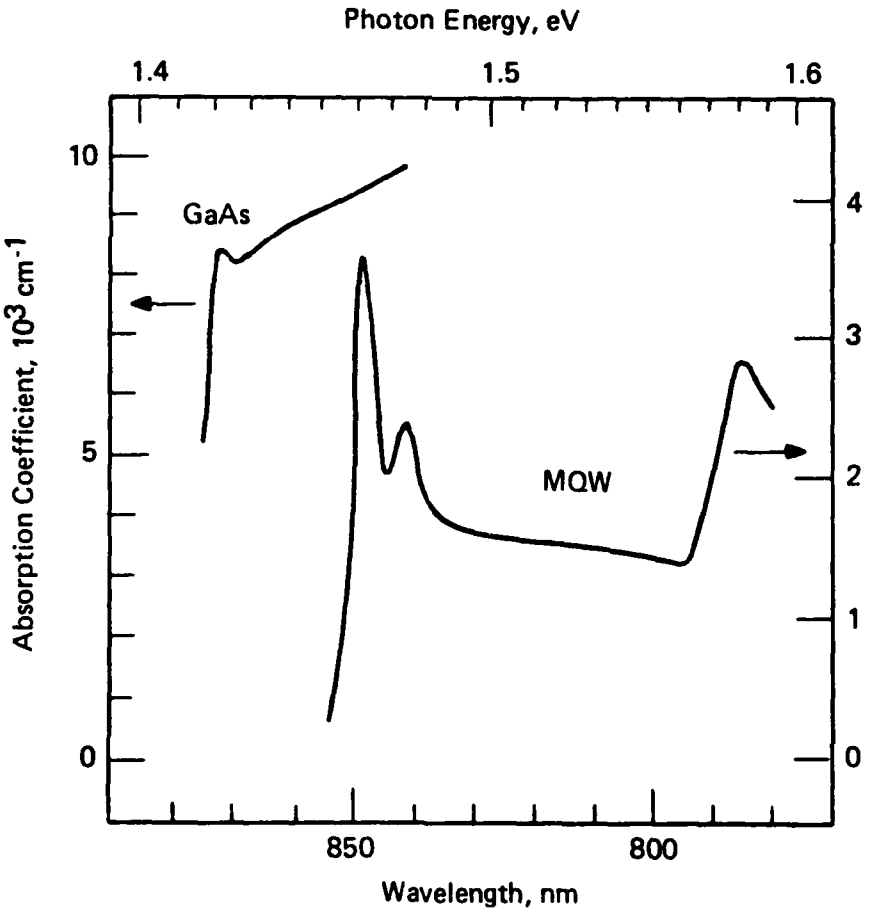


Figure 45. Absorption of a GaAs sample (3.2 μm thick) with an MQW structure consisting of 77 periods of 102 Å GaAs layers alternating with 207 Å AlGaAs layers.

long enough to produce these resonances; because the energy of the longitudinal optic vibration phonons is much greater than the binding energy, the excitons are promptly ionized. Line shape studies as a function of temperature show that the mean time for thermal phonon ionization is 0.4 ps for GaAs MQW structures and 0.24 ps for GaInAs MQW structures.

When free electron hole pairs are generated, they induce changes in the absorption coefficient and in the refractive indices. The physical mechanisms involved are the screening of the Coulomb interaction and phase space filling. Screening by free electron hole pairs is much more efficient than that due to excitons. In both cases it produces a real shift of the band gap, whereas the absolute energy of the exciton is not significantly changed due to its electric neutrality. As the band gap diminishes, the exciton binding energy reduces, loses oscillator strength, and eventually disappears.

The preceding discussion was concerned with the linear absorption in MQW structures and the nonlinear effects of optically adding electrons and holes and/or excitons. As mentioned at the beginning of this section, MQW structures also show some interesting absorption effects near the band edge when electric fields are applied to the material. One of these effects is the quantum-confined Stark effect (QCSE), which is rapid and large. One important aspect of the effect is that it appears to rely on the confinement of the carriers within the thin semi-conductor layers; consequently, it is truly a quantum well effect and will not exist in bulk semiconductors at any temperature.

Conventional semiconductors show an effect known as the Franz-Keldysh effect when electric fields are applied. The application of electric fields results in a slight shift of the absorption edge to lower photon energies, but the predominant consequence is really a broadening of the edge. In conventional semiconductors, exciton resonances are not normally directly relevant in such devices at room temperature because they are not resolvable. However, if the semiconductor is cooled so that the excitons can be seen, the broadening becomes particularly apparent on the exciton resonances themselves. This broadening can be understood as field ionization of the excitons; no sooner is the exciton created than it is ripped apart by the strong electric field. The field ionization therefore shortens the life of the exciton,

thereby broadening the absorption line. The exciton starts to show some shift to lower energies, and this shift is analogous to the Stark shift of the ground state of a hydrogen atom. However, this shift is limited to about 10 percent of the binding energy of the hydrogenic system at a few times the classical ionization field. After this point the field ionization is so rapid that the hydrogenic system no longer exists as a quasi-bound system, and the resonance is no longer resolvable.

When electric fields are applied parallel to the quantum well layers, the effects observed are similar to those expected in low-temperature conventional semiconductors, except that now they are observed at room temperature. The exciton progressively broadens and disappears as the electric field is increased.

However, when the electric field is applied perpendicular to the layers, a qualitatively different phenomenon is seen. In this case, the excitons shift to lower photon energies, and there is little or no broadening even with very high electric fields. A typical set of absorption spectra with increasing electric fields is shown in Figure 46.

The explanation of this effect is essentially that the potential barriers inhibit the field ionization of the exciton, hence inhibiting the broadening mechanism. Consequently, field strengths can be applied that would normally completely destroy the exciton resonance, but because the particle continues to exist, the Stark shift continues up to much larger values than are possible for conventional hydrogenic systems. The shift has been measured[54] to be 2.5 times the binding energy at 50 times the classical ionization field.

There are two different ways in which the potential barriers inhibit the field ionization, both of which are important for the quantum-confined Stark effect. First, the walls of the potential wells inhibit the electrons and holes from tunneling totally away from one another. Second, because the wells are narrow (100 Å) compared to the conventional exciton diameter (300 Å), even when the electron and hole are pulled to opposite sides of the well the Coulomb interaction between them is strong, and the exciton remains strongly bound. The resulting effect is in fact still a Stark effect, but the quantum confinement has qualitatively changed the behavior, which is reflected in the name for this effect: quantum-confined Stark effect.

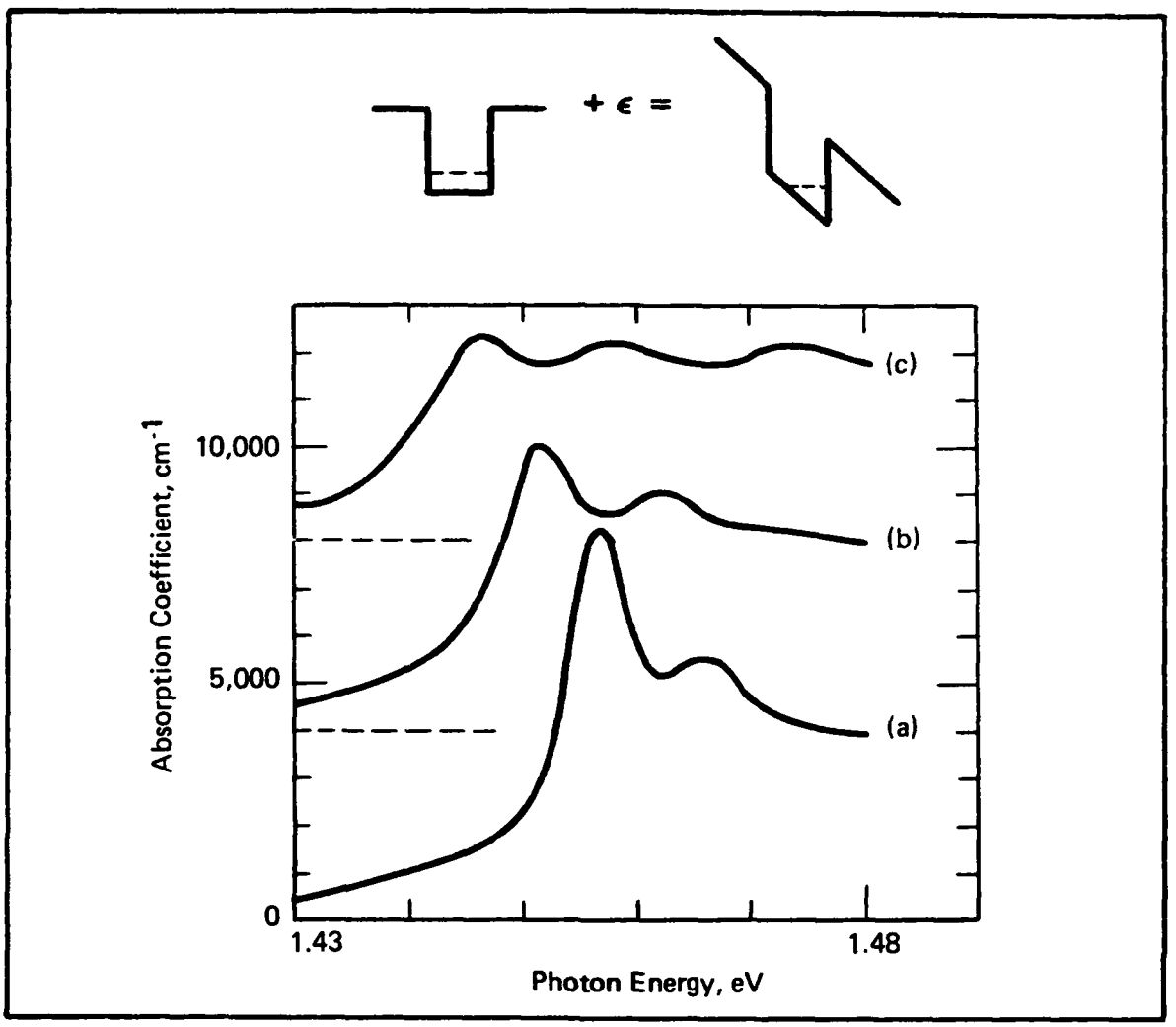


Figure 46. Absorption spectra for various electric fields perpendicular to the quantum well layers:
(1) 1×10^4 V/cm;
(2) 4.7×10^4 V/cm;
(3) 7.3×10^4 V/cm.
Zeros of the spectrum are displaced for clarity.
Inset illustrates the effect of a static field on the potential seen by the carriers.

Although the ultimate speed of the QCSE has not been measured yet, it has been tested down to 100 ps.[55] The practical limits on response times so far have been the RC time constants of the package; the fundamental limit appears to be the speed at which the quantum mechanical wave function can respond, which is limited by the uncertainty principle to times less than or of the order of 1 ps. The speed is not limited by carrier life.

The MQW structure is used between two transparent p and n doped regions to form a p-i-n diode in the fabrication of self-electro-optic effect devices (SEED).[53,54] The field is applied perpendicular to the quantum well layers by reverse-biasing the diode. In this mode, a field of the order of 10 kV/cm can be applied with about one volt of bias.

This device has been shown to operate both as a modulator and photodetector. The modulation operation occurs because of the shift of optical absorption with the applied field. For example, it is possible to choose to operate at a photon energy just below the band edge at zero field where the material is substantially transparent; then, by turning on the electric field the absorption is shifted down to the operating photon energy. Because the shifted absorption is so large (~5 x 10^3 to 10^4 cm^{-1}), substantial modulation is achieved with only microns of material thickness. When the photodetector operation occurs, it has been found that for every photon absorbed in the quantum wells a carrier can flow round an external electrical circuit, i.e., the internal quantum efficiency is unity, as would be expected for absorption within the depletion region of a diode.

The importance of these results is that the same device can operate simultaneously as a modulator and a photodetector. Thus, by the appropriate choice of external electrical circuit, it is possible to construct an opto-electronic feedback loop; it is this principle that has been used to develop the self-electro-optic effect device. SEED applications will change with the selection of a positive or negative feedback. With positive feedback a low energy optical bistable switch is obtained, and with negative feedback the SEED functions as an optical level shifter.[54] Although SEEDs are hybrid devices in that they use both optics and electronics for their operation, the ability to have the detector and modulator in the same integrated structure means that the electronics can be minimal (e.g., a resistor or a photodetector and a power supply), and the extremely low energy requirements of the QCSE modulator can give these devices exceptionally low operating energies for devices with both optical inputs and outputs.

Figure 47 shows a schematic illustration of a bistable SEED. For bistable operation, the operating wavelength is chosen near the main exciton peak position for zero fields so that for increasing field (i.e., increasing

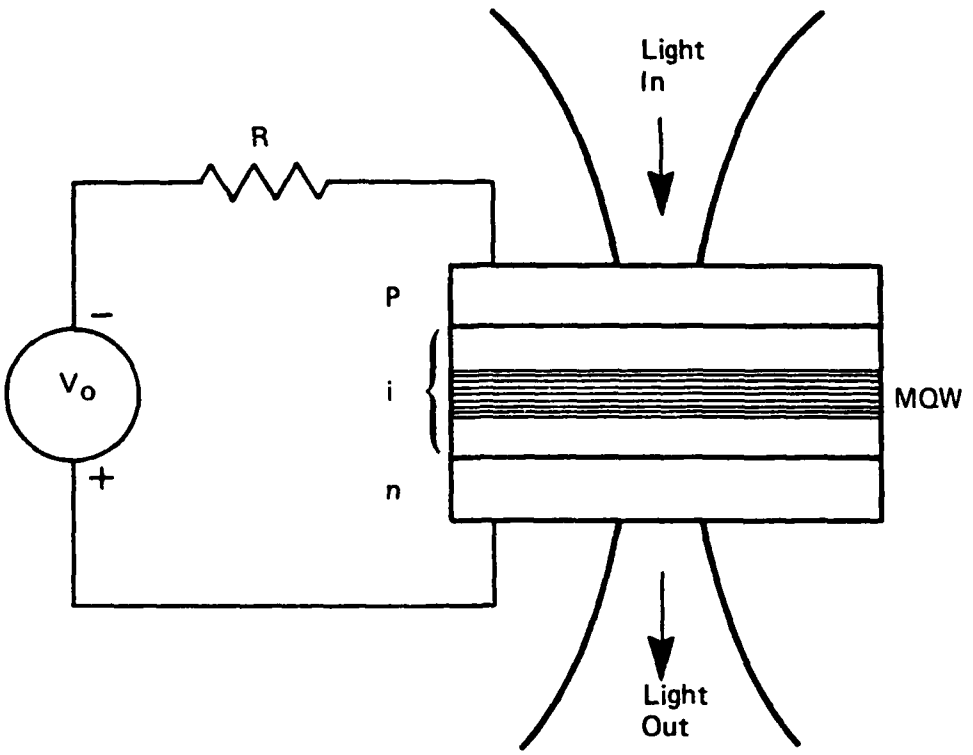


Figure 47. Schematic of an optical bistable SEED device.

reverse bias on the diode) the absorption decreases as the exciton moves to lower energy. When no light is shining on the device there is negligible current, and the full supply voltage reverse-biases the device so that the absorption is relatively low. With increasing incident light, the resulting photocurrent causes a voltage drop across the resistor, and the voltage across the diode decreases, resulting in increased absorption and hence increased photocurrent. Thus, a positive feedback is established, and under the right conditions this can lead to switching into a high absorption state with the exciton shifted back to the operating wavelength. This leads to the bistable optical input/output characteristic shown in Figure 48.[53] This bistability belongs to a recently identified general class of optical bistability without mirrors. Figure 48 shows the theoretical curve and measured transmission of the type of device shown in Figure 47.

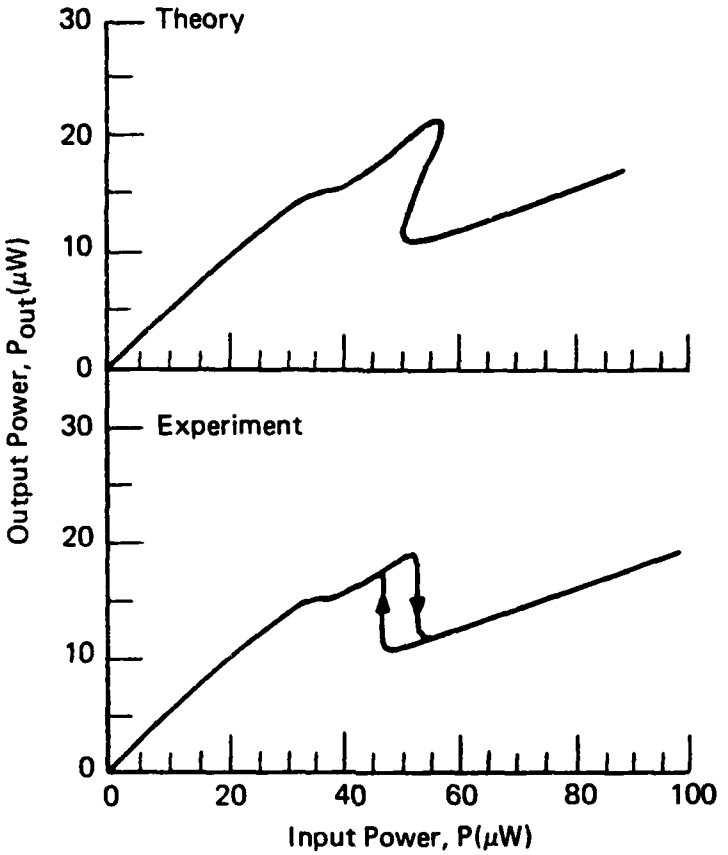


Figure 48. Optical bistable operation of a SEED device.

The device of Figures 47 and 48 is of interest for two reasons. First, it operates under reasonably practical conditions; it runs at room temperature, requires no cavity or other external feedback, is compatible with laser diode wavelengths and powers, operates over a wide range of time scales and powers (powers as low as 650 nW have been demonstrated with speeds as fast as 400 ns, and faster operation will be possible with smaller devices[54]), and can operate with incoherent light. Second, it offers extremely low switching energy per unit area (20 $fJ/\mu m^2$), and this is a factor of six smaller than any other device operating at a comparable wavelength; this fact is of special interest because the device uses no resonant cavity to operate at reduced switching energy.

5.2 NONLINEAR OPTICAL MATERIALS

If the work in the area of nonlinear optical devices for high-speed parallel processing architectures is to advance beyond the present research phase to the development of commercially viable products, significant advancements must occur in nonlinear optical materials.

Table 4 lists the characteristics that will be required in these new nonlinear optical materials. The first three have been discussed in this section: (1) the material must have a large nonlinear optical response so that a large bias does not have to be applied to the device to hold it near its nonlinear operating region; (2) the energy needed, in addition to the

TABLE 4. CHARACTERISTICS OF AN IDEAL NONLINEAR OPTICAL MATERIAL

1.	Large nonlinear optical response
2.	Low switching energy
3.	Rapid switching times
4.	Nondispersive
5.	Mechanically tough and formable
6.	High damage thresholds
7.	Formable into thin films and coatings
8.	Useful at high and low temperatures
9.	Immune to corrosive and oxidation environments

bias energy, to move the operations into the nonlinear region and thereby implement switching must also be low; and (3) the more rapid the switching time, the better for optical computing. Other characteristics are also required to develop practical devices. The material must be nondispersive and preserve the narrow bandwidth of the laser light; otherwise, pulses of light representing bits of information will tend to get broader and overlap with other pulses. This overlap will limit the throughput rate. The material must be able to withstand the rigors of device fabrication and have a damage threshold sufficient to withstand the power levels necessary to drive as many as one million resolution elements. For example, if each resolution element or channel were to require one microwatt of power, the material would have to tolerate

one watt of power over the operating aperture. The final three characteristics are desirable but not necessary qualities. Some device configurations require the fabrication of thin films; likewise, to avoid the expense and space requirements needed to produce special environments, a material should be both useful over a wide temperature range and environmentally stable.

The effect of the interaction of incident photons and the atoms of a material determine the nonlinear effects. The photon imparts its energy to those electrons that are loosely coupled to the nuclei, causing them to separate from their atoms, producing a charge separation. If the separation of the charges can be maintained momentarily, the resulting electric field leads to a nonlinear response of the material that is related to the degree of separation. The nonlinear materials currently receiving the most attention fall into three categories: inorganic insulators, large-molecule organic materials, and inorganic superlattices.

In inorganic insulators, the charge separation is primarily the result of the free electrons (created by the incoming photons) being trapped at other sites in the material. Once an electron gains enough energy to escape from its atom, it is attracted to a nearby site containing an atom capable of adding the free electron to its electron structure. The three leading materials in the inorganic insulator category are strontium barium niobate, bismuth silicon oxide, and barium titanate. Although the state of the art of these materials is much more advanced than that of the other two categories, their response times are only in milliseconds.

Research into large-molecule organic materials[56] indicates a possibility of achieving not only a greater degree of nonlinearity than in the inorganic insulators, but also a much shorter response time. The charge can be attained easily at low power levels owing to the existence of pi electrons that have low binding energies with their associated atoms. Also, the organic molecular structures are considerably larger than the inorganic ones, leading to the capability of sweeping free electrons down long molecular chains, thereby creating a relatively large separation between the centers of positive and negative charge. Table 5 lists the organic materials receiving the majority of attention at present. The major disadvantage of these materials is their relative environmental instability compared to the inorganics.

TABLE 5. CANDIDATE ORGANIC NONLINEAR OPTICAL COMPOUNDS

1.	Substituted and disubstituted acetylenes and diacetylenes
2.	Anthracines and derivatives
3.	Dyes
4.	Macrocyclics
5.	Polybenzimidazole
6.	Polybenzobisthiazole and polybenzobisoxazole
7.	Polyester and polyesteramids
8.	Polyetherketone
9.	Polyquinoxalines
10.	Porphyrins and metal-porphyrin complexes
11.	Metal complexes of TCNQ and TNAP
12.	Urea

The third category of materials are those with structures known as superlattices (also known as MQW structures). These materials are built up of alternating thin films of two different semi-conductors. The two semiconductors must have different band gaps where different levels of photon energy are required to free their electrons. In these materials, incident photons create pairs of electrons and holes in the wider band gap semiconductor. These electrons are rapidly swept into the neighboring layers because they represent a lower energy state. These layers act as potential wells that trap the electrons, thus creating the desired charge separation. A great deal of materials engineering is possible with these structures; for example, one can change the layer thickness on the potential well depths to create devices with very different operating characteristics. Most superlattices to date have been fabricated with alternating layers of GaAs and AlGaAs for room temperature operation. Future efforts should include not only other semiconductor materials but also organic nonlinear optical compounds.

REFERENCES

1. J. Goodman, Introduction to Fourier Optics, McGraw-Hill Co., New York, 1968.

2. D. Casasent, Optical Data Processing, Springer-Verlag, New York, 1978.

3. H. Stark, Applications of Optical Fourier Transforms, Academic Press, New York, 1982.

4. H. Jensen, L. Graham, L. Porcello, and E. Leith, "Side-looking Airborne Radar," Scientific American, 237, No. 4, (1977).

5. W. Rhodes and P. Guilfoyle, "Acousto Optic Algebraic Processing Architectures," Proc. of the IEEE, 72, No. 7 (1984).

6. T. Turpin, "Spectrum Analysis Using Optical Processing," Proc. IEEE, 69, 79-92 (1981).

7. D. Hecht, Proceedings of the SPIE, 90, 148 (1976).

8. D. Casasent, "Coherent Optical Pattern Recognition: A Review," Opt. Eng., 24, 26 (1985).

9. D. Ennis and D. Jared, "Optical Processing and Space Station Automation," Opt. Eng., 25, 808 (1986).

10. A. Tanguay, Optical Computing, J. Neff, ed., Proc. SPIE, 456, 129 (1984).

11. J. Lee, "Devices for Optical Signal Processing," Optics News, 11, 10 (1985).

12. T. Bell, "Optical Computing a Field in Flux," IEEE Spectrum, August 1986.

13. Y. Saito, S. Komatsu, and H. Ohzu, "Scale and Rotation Invariant Real-Time Optical Correlation Using Computer Generated Holograms," Opt. Commun., 47, 1 (1983).

14. T. Yatagai, K. Choji, and H. Saito, "Pattern Classification Using Optical Mellin Transforms and Circular Photodiode Array," Opt. Commun., 38, 3, 162 (1981).

15. D. Casasent and H. Okuyama, "A High-Dimensionality Pattern Recognition Feature Space," Proc. SPIE, 579, 245 (1985).

16. M. K. Hu, "Visual Pattern Recognition by Moment Invariants," IRE Trans. Inf. Theory, IT-8, 179 (1962).

17. D. Casasent, L. Cheatham, and D. Fetterly, "Optical Systems to Compute Intensity Moments: Design," Appl. Opt., 21, 18, 3292 (1982).

18. M. Teague, "Optical Calculations of Irradiance Moments," Appl. Opt., 19, 8, 1353 (1980).

19. J. Blodgett, R. Athale, C. Giles, and H. Szu, "Multiplexed Coherent Optical Processor for Calculating Generalized Moments," Opt. Lett., 7, 1, 7 (1982).

20. D. Casasent and W. Chang, "Generalized Chord Transformation for Distortion-Invariant Optical Pattern Recognition," Appl. Opt. 22, 14, 2087 (1983).

21. S. Watanabue, 4th Prague Conf. on Information Theory, (1965).

22. F. Fukunaya and W. Koontz, "Application of the Karhunen-Lowe Expansion to Feature Selection and Ordering," IEEE Trans. Comp., C-19, V, 311 (1970).

23. D. Foley and J. Sammon, "An Optimal Set of Discriminant-Vectors," IEEE Trans. Comp., C-24, 3, 281 (1975).

24. E. Patrick, Fundamentals of Pattern Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1972.

25. J. R. Ullmann, Pattern Recognition Techniques, Russak and Company, New York (1973).

26. K. Fu, Editor, Pattern Recognition and Machine Learning, Plenum Press, New York, 1971.

27. N. Nilsson, Learning Machines: Foundations of Trainable Pattern - Classifying Systems, McGraw-Hill, New York, 1965.

28. J. Mendel and K. Fu, Adoptive Learning and Pattern Recognition Systems, Academic Press, New York, 1970.

29. K. Fu and T. Yu, Statisticsl Pattern Classification Using Contextual Information, Research Studies Press, New York, 1980.

30. D. Casasent, "Unified Synthetic Discriminant Function Computational Formulation," Appl. Opt., 23, 10, 1620 (1984).

31. B. Kumar, "Optimability of Projection Synthetic Discriminant Functions," Proc. SPIE, 579, 86 (1985).

32. D. Psaltis, E. Park, and S. Venkatesh, "Optical Image Correlation with a Binary Spatial Light Modulator," Opt. Eng., 22, 4, 485 (1983).

33. D. Psaltis, F. Mok, and E. Paek, "On the Use of the Magneto-Optic Device in Optical Processors," Proc. SPIE, 465, 29 (1984).

34. L. J. Cutrona et al., "On the Application of Coherent Optical Processing Techniques to Synthetic-Aperture Radar," Proc. IEEE, 54, 1026 (1966)

35. J. W. Goodman, A. R. Dias, and L. M. Woody, "Fully Parallel, High-Speed, Incoherent Optical Method for Performing Discrete Fourier Transforms," Opt. Lett., 2, 1-3 (1978).

36. H. J. Caulfield, J. A. Neff, and W. T. Rhodes, "Optical Computing: The Coming Revolution in Optical Signal Processing," Laser Focus, 100-110, November 1983.

37. H. J. Caulfield, W. T. Rhodes, M. J. Foster, and S. Horwitz, "Optical Implementation of Systolic Array Processing," Opt. Commun., 40, 86-90 (1981).

38. D. Casasent, J. Jackson, and C. Neuman, "Frequency - Multiplexed and Pipelined Iterative Optical Systolic Array Processors," Appl. Opt., 22, 115-124 (1983).

39. A. B. Vander Lugt, G. S. More, and S. S. Mathe, "Multichannel Bragg Cells: Compensation for Acoustic Spreading," Appl. Opt., 22, 3906-3912 (1983).

40. R. A. Athale and W. C. Collins, "Optical Matrix-Matrix Multiplier Based on Outer Product Decomposition," Appl. Opt., 21, 2089-2092 (1982).

41. H. J. Whitehouse and J. M. Speiser, Aspects of Signal Processing Part 2, D. Reidel, Boston (1976).

42. P. S. Guilfoyle, "Systolic Acousto-Optic Binary Convolver," Opt. Eng., 23, 20-25 (1984).

43. R. P. Bocker, S. R. Clayton, and K. Bromley, "Electro-Optical Matrix Multiplication Using 2's Complement Arithmetic for Improved Accuracy," Appl. Opt., 22, 2019-2021 (1983).

44. H. M. Gibbs, Optical Bistability: Controlling Light with Light, Academic Press, New York (1985).

45. J. A. Neff, "Major Initiatives for Optical Computing," Opt. Eng., 26, 2-9 (1987).

46. A. R. Tanguay, "Materials Requirements for Optical Processing and Computing Devices," Opt. Eng., 24, 2-18 (1985).

47. J. W. Goodman, F. J. Leonberger, S. Y. Kung, and R. A. Athale, "Optical Interconnections for VLSI Systems," Proc. IEEE, 72, 850 (1984).

48. H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesan, "Optical Bistable Devices: The Basic Components of All - Optical Systems," Opt. Eng., 19, 463 (1980).

49. P. W. Smith and W. J. Tomlinson, "Bistable Optical Devices Promise Subpicosecond Switching," IEEE Spectrum, June 1981.

50. C. M. Bowden, M. Ciftan, and H. R. Robl, Optical Bistability, Plenum Press, New York (1981).

51. E. Abraham, C. T. Seaton, and S. D. Smith, "The Optical Computer," Sci. Am., 248, No. 2, 85 (1983).

52. J. L. Jewell, M. C. Rushford, and H. M. Gibbs, "Use of a Single Nonlinear Fabry Perot Etalon as Optical Logic Gates," Appl. Phys. Lett., 44, 172 (1984).

53. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, "Novel Hybrid Optically Bistable Switch: The Quantum Well Self Electro Optical Effect Device," Appl. Phys. Lett., 45, 13 (1984).

54. D. S. Chemla, D. A. B. Miller, and P. W. Smith, "Nonlinear Optical Properties of GaAs/GaAlAs Multiple Quantum Well Material: Phenomena and Applications," Opt. Eng., 24 , 556-564 (1985).

55. T. H. Wood, C. A. Burrus, D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, and W. Wiegmann, IEEE J. Quantum Electron., QE-21, 117 (1985).

56. A. F. Garito and K. D. Singer, "Organic Crystals and Polymers - a New Class of Nonlinear Optical Material," Laser Focus, 18, No. 2, 59-64 (1982).

THE TACTICAL WEAPON GUIDANCE AND CONTROL INFORMATION ANALYSIS CENTER (GACIAC)

GACIAC is a DoD Information Analysis Center operated by IIT Research Institute under the technical sponsorship of the Joint Service Guidance and Control Committee with members for OUSDRE, Army, Navy, Air Force, and DARPA. The U.S. Army Missile Command provides the Contracting Officer's Technical Representative. Its mission is to assist the tactical weapon guidance and control community by encouraging and facilitating the exchange and dissemination of technical data and information for the purpose of effecting coordination of research, exploratory development, and advanced technology demonstrations. To accomplish this, GACIAC's functions are to:

1. Develop a machine-readable bibliographic data base -- currently containing over 36,000 entries;

2. Collect, review, and store pertinent documents in its field of interest -- the library contains over 11,000 reports;

3. Analyze, appraise and summarize information and data on selected subjects;

4. Disseminate information through the GACIAC Bulletin, bibliographies, state-of-art summaries, technology assessments, handbooks, special reports, and conferences;

5. Respond to technical inquiries related to tactical weapon guidance and control; and

6. Provide technical and administrative support to the Joint Service Guidance and Control Committee (JSGCC).

The products and services of GACIAC are available to qualified industrial users through a subscription plan or individual sales. Government personnel are eligible for products and services under block funding provided by the Army, Navy, Air Force and DARPA. A written request on government stationery is required to receive all the products as a government subscriber.

Further information regarding GACIAC services, products, participation plan, or additional copies of this State-of-the-art Review may be obtained by writing or calling: GACIAC, IIT Research Institute, 10 West 35th Street, Chicago, Illinois 60616-3799, Area Code 312, 567-4519 or 567-4526.

JSGCC

JOINT SERVICE GUIDANCE AND CONTROL COMMITTEE

END

FILMED

6-89

DTIC