

AD-A220 332

# OPTICAL SIGNAL PROCESSING

U.S. Army Research Office  
DAAL03-87-K-0011

FINAL REPORT  
24782-F

A. VanderLugt  
Principal Investigator  
28 February 1990

North Carolina State University  
Raleigh, North Carolina 27695



**S** DTIC  
ELECTE  
APR 10 1990  
**D**  
B

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

The view, opinion, and/or finding contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION <b>Unclassified</b>		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT  Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)  24782-F		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Electrical and Computer Engineering Department	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION  U. S. Army Research Office	
6c. ADDRESS (City, State, and ZIP Code) North Carolina State University PO Box 7911 Raleigh, NC 27695		7b. ADDRESS (City, State, and ZIP Code)  P. O. Box 12211 Research Triangle Park, NC 27709-2211	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U. S. Army Research Office	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code)  P. O. Box 12211 Research Triangle Park, NC 27709-2211		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO.	PROJECT NO.
		TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Optical Signal Processing (Unclassified)			
12. PERSONAL AUTHOR(S) A. VanderLugt			
13a. TYPE OF REPORT Final Report	13b. TIME COVERED FROM 20 Nov86 to 31 Jan90	14. DATE OF REPORT (Year, Month, Day) 1990 February 28	15. PAGE COUNT 103
16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	Optical signal processing; heterodyne spectrum analyzers; angle of arrival; adaptive optical processing; Fresnel trans- form; optical switches; acousto-optics; cross spectrum analysis	
		18, cont. decimated arrays.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) We studied optical processing techniques to detect and track frequency hopped radio signals in a dense electro-magnetic environment by using cross-spectrum of two signals obtained from antenna elements spaced by one-half an RF wavelength. We decimate the array by retaining only every M <sup>th</sup> element and scan the cross-spectrum past the decimated array. We therefore reduce the circuit complexity, but suffer some loss in system performance because we require more photodetector bandwidth to accommodate the scanning action. An extension of the decimated array concept is to decimate the reference waveform in the heterodyne spectrum analyzer. In this case, we generate only 64 optical probes in the Fourier domain instead of the 2048 that are normally required. The advantage is that all of the optical power in the reference beam can be concentrated into the reduced number of optical probes, thereby offsetting some of the intrinsic loss in performance experienced by the cross-spectrum analyzer. <i>keywords:</i>			
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

We analyzed the connection between the temporal and spatial frequencies in interferometric analyzers, using both Fourier and space plane analyses. The mixed transform is used to gain an accurate picture of the frequency content for cw signals, short pulse signals, and evolving pulse signals - - the most difficult ones to analyze.

We performed an extensive analysis on a free space propagating optical switch, based on acousto-optic cell technology and carried out computer analyses of the Fresnel transform for the most likely configurations. We also performed analyses and conducted experiments to verify insertion loss, worst-case optical crosstalk, and acousto-optic cell reconfiguration time predictions for a 1x4 Fourier domain switch in a multimode fiber-optic system. Insertion loss for the switch is approximately 2-4 dB, worst-case signal-to-crosstalk ratio is better than 25 dB, and the reconfiguration time is 880 nsec. These measured values are in good agreement with the theory, and support our claims concerning the high performance level of our acousto-optic architecture.

We recognized the opportunity to process signals in domains other than the time or frequency domains arises naturally in optical Fourier transform systems. We showed that  $N$  samples are sufficient to sample a signal in any Fresnel plane, provided that a specified nonuniform sampling distribution is followed. We showed that the highest spatial frequencies in any Fresnel transform is concentrated near the optical axis so that the samples must be most closely spaced in this region. We showed that the highest possible spatial frequency may occur in one or more of four planes in a generalized imaging system: the direct plane, the Fourier plane, the second crossover plane, or the image plane.

We studied a new method for detecting short pulses using direction of arrival information derived through finding the time difference of arrival at a dual antenna receiver. This technique has broader application such as to time-domain reflectometry. The key idea is to use a Fresnel transform to disperse a short pulse and its time delayed replica into longer time duration signals. These two signals interfere after dispersion to produce a Fresnel diffraction pattern having a strong sinusoidal component whose frequency is directly related to the time difference of arrival. A photodetector element detects the light at specially chosen positions and the output is directed to a spectrum analyzer that displays the frequency content of the new signal. The time-of-arrival and frequency content are now directly related.

---

TABLE OF CONTENTS

1.0 INTRODUCTION.....3  
2.0 SPECTRUM ANALYSIS.....3  
3.0 ACOUSTO-OPTIC CROSSBAR ..... 5  
4.0 SAMPLING OF FRESNEL TRANSFORMS.....6  
5.0 RECENT STUDIES.....7  
6.0 MISCELLANY.....8  
7.0 REFERENCES.....9

LIST OF APPENDICES

- APPENDIX A: USE OF DECIMATED PHOTODETECTOR ARRAYS IN  
SPECTRUM ANALYSIS  
APPENDIX B: DECIMATED ARRAYS FOR SPECTRUM ANALYSIS  
APPENDIX C: TEMPORAL FREQUENCIES OF SHORT AND EVOLVING PULSES  
IN INTERFEROMETRIC SPECTRUM ANALYZERS  
APPENDIX D: ROLE OF PHOTODETECTORS IN OPTICAL SIGNAL PROCESSING  
APPENDIX E: ACOUSTO-OPTIC PHOTONIC SWITCH  
APPENDIX F: ACOUSTO-OPTIC PHOTONIC SWITCH: AN OPTICAL CROSSBAR  
ARCHITECTURE  
APPENDIX G: ACOUSTO-OPTIC PHOTONIC SWITCH  
APPENDIX H: OPTIMUM SAMPLING OF FRESNEL TRANSFORMS

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



---

**OPTICAL SIGNAL PROCESSING**

---

**1.0 Introduction**

As the bandwidth of signals increase and as the electromagnetic environment becomes increasingly dense, processing operations such as convolution, spectrum analysis, correlation, ambiguity function generation and adaptive filtering become computationally intensive operations. Optical processing provides high-speed, parallel computations so that digital post-processing techniques can be used for lower-speed, serial computations. One objective of this study is to modify the basic architecture of an interferometric spectrum analyzer to include time signal modulation in either one or both branches of the interferometer so that operations such as detecting frequency hopped signals, calculating cross-spectral densities, or estimating angle-of-arrival can be implemented; techniques will be developed for reducing the complexity of the photodetector array. A second objective is to extend the use of Fresnel transforms for wideband analog signal protection, and to study how Fresnel transforms can be used to detect the angle-of-arrival of a short pulse or to detect the pulse repetition interval of a short pulse train. The third objective is to investigate the use of Fresnel transforms in computing systems. In particular, we shall investigate improved methods for implementing a dynamic crossbar switch that can be rapidly reconfigured.

The research effort described in this report has resulted in several innovative optical processing techniques for improved performance. The research covers the period from 20 November 1986 to 31 January 1990. The major accomplishment can be divided into three major areas: (1) studies relating to improvements in spectrum analyzers, including a study of the improvements needed of photodetectors in optical signal processing, (2) optical crossbar switches that are non-blocking, and (3) the optimum sampling of Fresnel transforms. In the following paragraphs, we summarize the key results in each area: further details can be found in the referenced journal articles that have been published and included in the Appendices.

**2.0 Spectrum Analysis**

Spectrum analysis is probably the more widely used tool in signal processing, independently of how it is implemented. Optical techniques for spectrum analysis has steadily evolved during the past decade, with some notable breakthroughs such as the heterodyne spectrum analyzer developed on a previous grant (DAAG29-80-0149). Improvements to this basic concept were developed on a subsequent grant (DAAG29-83-C0033), in the form of better understanding the fundamental requirements of the distributed local oscillator provided by the reference beam in an

---

interferometric spectrum analyzer. Under the current grant (DAAL03-87-K-0011), A. VanderLugt has studied optical processing techniques to detect and track frequency hopped radio signals in a dense electro-magnetic environment. One method is to form the cross-spectrum of two signals obtained from antenna elements spaced by one-half an RF wavelength. The phase difference between these two signals provides angle of arrival information which is an important sorting parameter. Previous methods obtained the phase through spatial frequency heterodyning but the photodetector array is extremely complicated, and the dynamic range is limited. If we use a temporal heterodyne technique, however, we can use discrete photodetectors that have much higher dynamic ranges, but more complex readout circuitry. In the proposed approach, we decimate the array by retaining only every  $M^{\text{th}}$  element and scan the cross-spectrum past the decimated array. We therefore reduce the circuit complexity, but suffer some loss in system performance because we require more photodetector bandwidth to accommodate the scanning action. In some applications, the tradeoff may be a good one; we can reduce the number of photodetectors from 2048 to 64 with only a 6dB loss in system performance (either need more laser power or tolerate less dynamic range). The performance is referenced to the normal mode of operation in which the spectrum is not scanned, we call this the staring mode.

An extension of the decimated array concept is to decimate the reference waveform in the heterodyne spectrum analyzer. In this case, we generate only 64 optical probes in the Fourier domain instead of the 2048 that are normally required. The advantage is that all of the optical power in the reference beam can be concentrated into the reduced number of optical probes, thereby offsetting some of the intrinsic loss in performance experienced by the cross-spectrum analyzer. If the system is thermal noise limited in the staring mode due, say, to insufficient laser power, we can actually improve the system performance by using this decimation approach. At the other extreme, if we are shot noise limited in the staring mode, the performance may be reduced somewhat because the photodetector bandwidth is increased. The performance between these extremes is application dependent. Two papers have been published on this research effort,<sup>1,2</sup> and are included in Appendix A and Appendix B.

P. H. Wisseman analyzed the connection between the temporal and spatial frequencies in interferometric analyzers, using both Fourier and space plane analyses. This problem is important because such analyzers provide more dynamic range than power spectrum analyzers due to the heterodyne action, and the signal that gets through the bandpass filter is highly dependent on its temporal frequency content. He used the mixed transform to gain an accurate picture of the frequency content for cw signals, short pulse signals, and evolving pulse signals -- the most difficult ones to analyze. He set up an interferometric spectrum analyzer to test the coupling of temporal and spatial frequencies for CW signal, short pulse signals, and evolving pulse signal. He tested the coupling of temporal and spatial frequencies. The experimental

results validate the theoretical calculations for all three signal types. He completed work on this topic for his Master's Thesis, and published the results in *Applied Optics*<sup>3</sup> as shown in Appendix C.

Photodetector arrays have been, and continue to be, designed primarily for imaging applications. Optical signal processing applications place new and more stringent requirements on these devices. We need less blooming, faster recovery, less electrical readout noise, more video output lines, segmented arrays, and so forth. The major improvements, however, must be in dynamic range and in techniques for reducing the output data rates. A paper on this topic was published as a result of this Army Research Office Palantir study.<sup>4</sup> The participants in the study consisted of the authors of the paper included in Appendix D.

### 3.0 Acousto-optic Crossbar Switch

As the need increases for switching systems that serve large numbers of high capacity terminals, the limitations of electronic switching techniques become more evident. Large electronic switching systems have always required a high degree of complexity, and, with the advent of fiber-optic communication, these systems are now becoming information flow bottlenecks. Fortunately, optical technologies are well suited to switching applications. They exhibit excellent information processing and carrying capabilities that alleviate some of the problems encountered with the presently used electronic systems. In this paper, we propose a powerful architecture for space-division switching that is based on acousto-optic technology.

Through the years, three basic techniques for circuit switching have emerged. These include space-division switching, in which reconfigurable dedicated physical paths are used, time-division switching, which relies on rearrangement of time-division multiplexed source information, and frequency-division switching, where each source signal is frequency shifted and discrimination is provided by destination filters. Most of the work with photonic switching has been in the space division switch area; this is because the necessary optical devices for the other techniques, such as volatile optical memories for time division switches and broadly tunable optical sources for frequency division switches, are either unavailable or impractical.<sup>5</sup>

Aside from the complexity constraints of crosspoint networks, electronic switching has some limitations that are more fundamental. Signal bandwidth is severely limited by RC time constants, and signals within the switch are vulnerable to electromagnetic interference. Also, broadcasting is difficult with multistage networks, because electronic crosspoints do not provide signal splitting capability. Finally, lightwave technology is used extensively in high-speed network applications; in some instances, it may be more cost-effective to switch without regenerating the optical signals within these networks. If switching can only be performed



electronically, nonregenerative switched optical networks will not be possible.

D.O.Harris performed an extensive analysis on a free space propagating optical switch, based on acousto-optic cell technology. He carried out some computer analyses of the Fresnel transform for the most likely configurations and used these results to predict the level of crosstalk expected. One critical issue is to evaluate possible methods to control the RF frequency of the channels of the cell. The best approach based on current technology is to use frequency synthesizers that can be set to within  $\pm 1$ Hz of the desired frequency with an settling time of about 100 nanoseconds. Both figures exceed the requirements established for the switch. He also performed analyses and conducted experiments to verify insertion loss, worst-case optical crosstalk, and acousto-optic cell reconfiguration time predictions for a 1x4 Fourier domain switch in a multimode fiber-optic system. Insertion loss for the switch is approximately 2-4 dB, worst-case signal-to-crosstalk ratio is better than 25 dB, and the reconfiguration time is 880 nsec. These measured values are in good agreement with the theory, and support our claims concerning the high performance level of our acousto-optic architecture. The results of this research effort has been published<sup>6-8</sup> and are included in Appendices E, F and G, the latter being a short summary article written for Optics News.

#### 4.0 Sampling of Fresnel Transforms

A. VanderLugt has recognized the opportunity to process signals in domains other than the time or frequency domains arises naturally in optical Fourier transform systems. It is easy to show that the Fourier plane can be added to the list, since all the information must pass thorough a restricted aperture if the object is bandlimited. In this study, we showed that N samples are sufficient to sample a signal in any Fresnel plane as well, provided that a specified nonuniform sampling distribution is followed. We showed that the highest spatial frequencies in any Fresnel transform is concentrated near the optical axis so that the samples must be most closely spaced in this region. We showed that the highest possible spatial frequency may occur in one or more of four planes in a generalized imaging system: the object plane, the Fourier plane, the second crossover plane, or the image plane.

This nonuniform sampling is similar to the the visual system in which the region of greatest acuity for the eye is at the optical axis. It may have application to showing how to properly dilute arrays in other wavelength region such as discrete element, phased arrays in the microwave region. The optimum sampling procedure reduces the required number of samples by up to a factor of four, for the two-dimensional case, leading to less computations in applications such as image restoration. This work has been submitted for publication in Applied Optics and is included as Appendix H.<sup>9</sup>

A. VanderLugt has studied a new method for detecting short pulses. Isolated pulses of short duration are particularly difficult to detect. The energy per pulse is typically low and, because the pulse duration is short, the energy is spread over a wide band of frequencies in the Fourier domain so that direct spectrum analysis often is useless. There may be many pulses received per unit time from different sources, each with a unique signature. The issue is how to detect these short pulses and related low probability of intercept signals. One way to help detect short pulses is to find the direction of arrival through finding the time difference of arrival at a dual antenna receiver. Once the methodology is established, it is clear that the technique has broader application such as to time-domain reflectometry. The key idea is to use a Fresnel transform to disperse a short pulse and its time delayed replica into longer time duration signals. These two signals interfere after dispersion to produce a Fresnel diffraction pattern having a strong sinusoidal component whose frequency is directly related to the time difference of arrival. A photodetector element detects the light at specially chosen positions and the output is directed to a spectrum analyzer that displays the frequency content of the new signal. The time-of-arrival and frequency content are now directly related. One of the nice features of this approach is that the pulse shape is not very important and that the time-of-arrival can be accurately measured even if there are no easily identifying characteristics on the pulse.

### 5.0 Recent Studies

Closed loop adaptive processing at optical wavelengths is a difficult task, although significant progress was made under grant DAAG29-83-C0033. In this research effort, we began investigations into application that can use optical processing architectures in a quasi-closed loop mode, which we call open loop adaptive processing. The main concept is that there are important applications in which the signal processing environment can be monitored and appropriate processing algorithms introduced quickly enough to provide a performance level essentially the same as that for a closed loop system.

C. J. Anderson has begun an analysis of using optical methods for digital radio channel equalization. These microwave radio channels are characterized by a four-parameter model. The parameters can be measured from the power spectrum of the transmitted signal. Anderson is studying how appropriate spatial filters can be constructed to equalize the channels. Both amplitude and phase conjugation is required to correct for the channel distortion.

T.P. Karnowski is investigating methods to construct a general spatial filtering operation. Since the impulse response for processing time signals is real we will use area modulation to produce the required amplitude and a half-wave plate to produce the required negative amplitudes. He has begun to set up the required interferometric optical system and will begin to

---

process known signals with known impulse responses to calibrate the system. The longer range goal is to use a computer controlled liquid crystal television spatial light modulator to produce quasi-realtime adaptive response.

R. S. Ward is investigating a different form of a quasi-realtime adaptive system for signal excision. Previous efforts demonstrated that adaptive notch filters could be implemented optically, but that the system performance is quite sensitive to vibrations, thermal effects, and so forth. Ward is investigating the use of the auxiliary beam from a signal excision system to establish the jammer environment and to use this information to set the notches. This work is still in its early stages.

### 6.0 Miscellany

A. VanderLugt attended a workshop on "Acousto-optical Signal Processing", sponsored by the Army Research Office, held on May 19-20, 1987, in Columbia, MD. The purpose of the workshop was to identify areas in which further research is needed so that the engineering development of these systems is not impeded. He gave a plenary talk describing my perception of the University/Industry roles in these development efforts.

Dr. R. Hartman from MICOM visited in May, 1987, to discuss their development efforts on a matched spatial filtering system for terminal guidance. VanderLugt described our research over the past few years; one area of special interest is the scrambling of wideband analog signals such as those that produced by the imaging systems on RPV's.

A. VanderLugt presented a paper on "Optical Signal Processing" at the Fifth International Workshop on Integrated Electronics and Photonics in Communications, Research Triangle Park, NC, on October 23, 1987 and summarized the research activity reported here at a Workshop sponsored by the Center for Communication and Signal Processing at North Carolina State University on October 28, 1987.

A. VanderLugt presented a paper on "Decimated Arrays for Spectrum Analysis" at the Technical Symposium on Optics, Electro-Optics, and Sensors held in Orlando, FL, 6-8 April, 1988 .

D.O. Harris presented a paper on "Acousto-Optic Photonic Switch: A Optical Crossbar Architecture," at the SPIE Symposium on Optoelectronic & Fiber Optic Devices & Applications held in Boston, MA, 5-8 September, 1989

The participating personnel on this research study were Dr. A. VanderLugt, Principal Investigator, D.O. Harris, P.H. Wisseman, P.J.W. Melsa, T.P. Karnowski, R.S. Ward, and C.S. Anderson. P.H. Wisseman received his Masters of Science degree in December, 1988, with a thesis entitled "Spatial and Temporal Frequencies of Short and Evolving Pulses in Optical

---

---

Interferometric Spectrum Analyzer" partially fulfilling the requirements for the MSc degree.

### 7.0 References

1. A. VanderLugt, " Use of Decimated Photodetector Arrays in Spectrum Analysis," *Appl. Opt.*, Vol 27, p. 2061 (1988)
  2. A. VanderLugt, "Decimated Arrays for Spectrum Analysis," *Proc. SPIE*, Vol. 936, p. 221-228, 1988
  3. P.H. Wissman and A. VanderLugt, "Temporal Frequencies of Short and Evolving pulses in Interferometric Spectrum Analyzers," *Appl. Opt.*, Vol. 28, pp. 3800-3809, (1989)
  4. G.W. Anderson, B.D. Guenther, J.A. Hyncek, R.J. Keyes, and A. VanderLugt, "Role of Photodetectors in Optical Signal Processing," *Appl. Opt.*, Vol. 27. p.2871, (1988)
  5. S. F. Su, L. Jou, and J. Lenart, "A review on classification of optical switching systems", *IEEE Comm. Mag.*, vol. 24, no. 5, pp. 50-55, May 1986.
  6. D.O. Harris and A. VanderLugt, "Acousto-Optic Photonic Switch," *Optics Letters*, Vol. 14, pp. 1177-1179, (1989)
  7. D.O. Harris and A. VanderLugt "Acousto-Optic Photonic Switch: An Optical Crossbar Architecture" *Proc. SPIE*, Vol. 1178, (1989)
  8. D.O. Harris and A. VanderLugt, "Acousto-Optic Photonic Switch," *Optics News*, Vol. 14, pp. 49-50, (1989)
  9. A. VanderLugt, "Optimum Sampling of Fresnel Transforms," Submitted to *Applied Optics*.
-

---

APPENDIX A

USE OF DECIMATED PHOTODETECTOR ARRAYS IN SPECTRUM ANALYSIS

Reprinted from Applied Optics

Volume 27, Pages 2061-2070, 15 May 1988

---

# Use of decimated photodetector arrays in spectrum analysis

A. VanderLugt

Temporal heterodyne spectrum analysis requires the use of discrete photodetector arrays that have a large number of elements. Each element is generally followed by an amplifier, a bandpass filter, a demodulator, and nonlinear devices to handle the large dynamic range. When the number of elements in the array is of the order of 1000–2000, the readout hardware is difficult to implement. We consider decimating the array so that a much smaller number of elements are used. The spectrum is scanned across this array so that each element reads out a set of spatial frequencies in a time division multiplexing fashion. In some cases there is no penalty in dynamic range; in others, the penalty is more strongly related to the reduction in the number of photodetectors. Similar techniques are applied to a cross-spectrum analyzer that uses temporal heterodyning to derive angle of arrival information from wideband signals.

## 1. Introduction

The application of acousto-optic techniques to the spectral analysis of wideband signals is well established. A wideband signal drives an acousto-optic cell, generally operated in the Bragg mode, to produce a space-time signal. The optical system produces the instantaneous Fourier transform, or spectrum, of the signal resident in the cell. A 1-D photodetector array then samples the instantaneous power spectrum at a rate of no more than once per time interval  $T$ , where  $T$  is the fill time of the Bragg cell.

A variation on the instantaneous power spectrum analyzer is the radiometer. The optical systems are structurally the same; the key difference is that a radiometer integrates the detected output samples for a time duration much longer than  $T$ . The radiometer, therefore, applies time integration to data produced by a space integrating spectrum analyzer. This mixed integration, or hybrid approach, allows for the detection of weak cw signals along with those whose instantaneous SNRs are high.

Another variation is the use of heterodyne techniques to increase the dynamic range by detecting spectral amplitudes instead of spectral powers. The

heterodyne action is produced by adding a reference function to the spectrum. The reference waveform may be a plane wave<sup>1</sup> or an array of distributed local oscillators.<sup>2</sup> In the latter case, the temporal beat frequency is fixed at all spatial frequency locations so that identical discrete photodetectors and postdetection circuitry can be used.

In yet other variations, multichannel Bragg cells have been used to determine the angle of arrival of a signal. If, for example, we use two antenna elements to receive the signal from a common source, we can determine the angle that the source makes relative to the receiver geometry; the measurement technique may rely on a comparison of the amplitudes or the phases of the spectral components of the two signals. In the latter case, the phase can be obtained by forming the cross spectrum of the signals received by two antenna elements. The relative phase of the cross spectrum provides a measure of the angle of arrival of cw emitters at each frequency.

These applications place severe demands on the photodetector array and its associated circuitry. Although 1-D photodetector arrays typically have multiple video readout lines, the composite bandwidth usually does not allow a readout every  $T$  seconds. Instead, the light is integrated for a time interval longer than  $T$  seconds, and the resulting system is, therefore, not information preserving. For example, if we wish to detect short pulse signals, we read the array at least once while the pulse is in the cell to avoid a significant loss in SNR or, even worse, miss a pulse altogether. Another example of signals whose envelopes are time dependent are frequency hopped radios in which a

The author is with North Carolina State University, Raleigh, North Carolina 27695.

Received 19 October 1987.

0003-6935/88/102061-10\$02.00/0.

© 1988 Optical Society of America.

narrowband signal may propagate at randomly chosen frequencies within a wide available band of frequencies. Sometimes the dwell time at each frequency is also randomly varied. Hop rates are generally of the order of 100 hops/s; the trend is toward rates of 1000 hops/s to as high, perhaps, as 10,000 hops/s. The number of photodetector elements required in the array is application dependent; it ranges from 500 to 2000 elements.

We describe spectrum analyzers that support the sampling rate required to avoid missing signals and yet operate with a significantly reduced number of photodetector elements. The basic idea is to decimate an  $N$ -element photodetector array by retaining only every  $M$ th element. We time share the remaining elements by scanning the spectrum across the decimated array. Each photodetector therefore produces, as a time sequence, the spectral content of the received signal over a small frequency range. In Sec. II we establish the background necessary to understand the basic approach. In Sec. III we discuss the photodetector geometry and detection scheme and, in Sec. IV, the required reference and scanning functions. In Sec. V we analyze the effects of decimation on SNR and dynamic range. In Sec. VI, we introduce a reference waveform that significantly improves the dynamic range of a heterodyne spectrum analyzer. The decimation technique can also be used to extend the performance parameters of heterodyne cross-spectrum analyzers as we show in Secs. VII-IX.

## II. Background for the Heterodyne Spectrum Analyzer

The heterodyne spectrum analyzer consists of a conventional spectrum analyzer, modified to include a reference function in the Fourier plane. Our description is brief; a detailed account is given in Ref. 2. In Fig. 1 we show a Mach-Zehnder interferometer in which the lower branch contains a Bragg cell in plane  $P_1$  driven by the signal  $f(t)$  that we want to analyze. The Fourier transform of this signal, created in plane  $P_2$ , is given by

$$F(\alpha, t) = \int_{-\infty}^{\infty} a(x)f(t - x/v - T/2) \exp(j2\pi\alpha x) dx, \quad (1)$$

where  $a(x)$  is the aperture weighting function that accounts for the illumination profile, acoustic attenuation, and truncation effects caused by the Bragg cell or lenses,  $\alpha$  is a spatial frequency, and  $v$  is the velocity of the acoustic wave. We have ignored an exponential function that describes the Bragg angle, and we assume that the rf signal is centered at  $f_c$ , the midpoint of the bandpass of the Bragg cell. In a similar fashion, the Fourier transform of the reference signal  $r(t)$  that drives the Bragg cell in plane  $P_3$  in the upper branch is

$$R(\alpha, t) = \int_{-\infty}^{\infty} a(x)r(t - x/v - T/2) \exp(j2\pi\alpha x) dx. \quad (2)$$

The reference waveform can be shifted geometrically by a spatial frequency  $\alpha_d$  to provide a fixed temporal offset frequency  $f_d$  at each spatial frequency position.<sup>2</sup> An alternative method is to frequency shift the entire

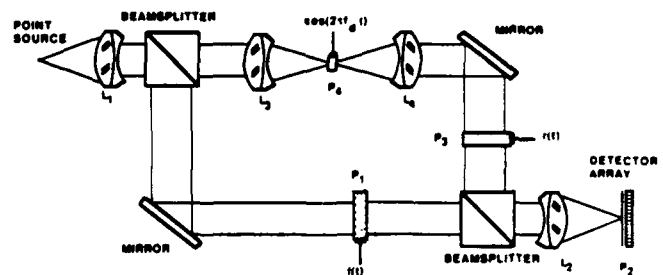


Fig. 1. Heterodyne spectrum analyzer.

reference waveform by  $f_d$ , using the acousto-optic point modulator in the upper branch. This method for achieving the offset frequency provides more flexibility and can be used directly in the cross-spectrum analysis application in which a geometric shift cannot be tolerated. The intensity at plane  $P_2$  is given by

$$I(\alpha, t) = |F(\alpha, t) + R(\alpha, t) \exp(j2\pi f_d t)|^2. \quad (3)$$

We substitute Eqs. (1) and (2) into Eq. (3) and expand the result to find that two intensity terms  $I_1(\alpha, t) = |F(\alpha, t)|^2$  and  $I_2(\alpha, t) = |R(\alpha, t)|^2$  have frequencies at baseband. For the condition that  $R(\alpha, t)$  is a narrowband signal relative to  $f_d$ , the cross-product term does not overlap the baseband terms and becomes

$$I_3(\alpha, t) = 2|F(\alpha, t)| |R(\alpha, t)| \cos(2\pi f_d t + \phi), \quad (4)$$

which is centered at  $f_d$ , where  $\phi$  is the phase difference between  $R(\alpha, t)$  and  $F(\alpha, t)$ . The spectral information is measured by a photodetector array consisting of  $N$  discrete elements, each followed by a bandpass filter. For a spectrum analyzer,  $N$  must be greater than or equal to  $2TW$ , where  $T$  is the effective duration of the signals in the Bragg cells and  $W$  is the signal bandwidth.

The photodetector current is given by the integral of  $I(\alpha, t)$  over the photodetector area. Since the baseband terms have bandwidths no greater than  $1/T$ , we can remove them by a bandpass filter centered at  $f_d$ , where  $f_d \gg 1/T$ . The response of a detector element centered at a spatial frequency  $\alpha_k$  to the intensity  $I_3(\alpha, t)$  is then

$$i_3(\alpha_k, t) = A_0 \cos(2\pi f_d t + \phi) \int_{-\infty}^{\infty} |F(\alpha, t)| |R(\alpha, t)| H(\alpha) d\alpha, \quad (5)$$

where  $A_0$  is a scaling constant,  $H(\alpha) = S \text{rect}\{(\alpha - \alpha_k)/\alpha_e\}$  is the photodetector response,  $S$  is the sensitivity of the photodetector,  $\alpha_e$  is the width of the detector element expressed as a spatial frequency, and  $\text{rect}(x) = 1$  for  $|x| \leq 1/2$ . We see that the current given by Eq. (5) is a measure of the magnitude of  $F(\alpha, t)$ , in sharp contrast to a power spectrum analyzer where the current is a measure of the square of the magnitude of  $F(\alpha, t)$ . As a result of the heterodyne action provided by  $R(\alpha, t)$ , the dynamic range is increased significantly. To measure the spectrum accurately, we must carefully select an appropriate reference signal  $r(t)$ ; we return to this point in Sec. IV.

### III. Photodetector Geometry and Detection Scheme

The number of photodetector elements required in a typical spectrum analysis application may be large. For example, we may need to measure energy in each 25-KHz channel over a 40–50-MHz frequency band; this requires 1600–2000 detector elements. Figure 2 shows the photodetector geometry relative to the spectrum. The spectrum is centered at a spatial frequency  $\alpha_c$  corresponding to  $f_c$ ; each cw signal produces a response  $A(\alpha)$ , which is the Fourier transform of the aperture weighting function  $a(x)$ . Suppose that we retain only every  $M$ th element of the array so that the total number of elements is  $Q = N/M$ . We now arrange to scan the spectrum across this decimated array so that each element detects, during a period of time  $T_s$ , the  $M$  frequency components within its sector. The sampling rate of the spatial frequencies is then  $R = M/T_s$  samples per second for each element in the photodetector array.

In Fig. 3 we show the side and top views of a spectrum analyzer that incorporates the required scanning action. The interferometer is created by using a dual channel Bragg cell in plane  $P_3$  with its electrodes stacked in the vertical direction. This cell is illuminated by two parallel beams of light; the diffracted beams are combined by a prism so that the sum of the spectra occurs at plane  $P_4$ . Since the interferometer is compact, the effects of vibrations and thermal variations in the Bragg cell are minimized relative to those typically present in a Mach-Zehnder interferometer.

In a recent paper, Koontz described a heterodyne spectrum analyzer in which a beam splitting prism, similar to that used to combine the beam, was placed in plane  $P_2$  to generate the desired illumination pattern.<sup>3</sup> Since we require that one beam be offset by a frequency  $f_d$ , an alternative way to generate the two beams is to use an acousto-optic modulator in plane  $P_2$ . When the modulator is operated with 50% diffraction efficiency, the two beams have equal intensities, but one of them is frequency shifted by  $f_d$ . If the beam separation at the Bragg cell is not adequate, we can use a prism, similar to the beam combining prism, to split the light. An acoustooptic or electrooptic modulator with suitable spatial filtering can then be inserted into one beam to provide the offset frequency  $f_d$ . Yet another means to achieve the offset is to use  $f_c + f_d$  as the center frequency for one of the signal channels; we then require an optical element, such as a thin prism, in that channel after the Bragg cell to cause this spectrum to geometrically overlap that from the channel whose center frequency is  $f_c$ .

The physical size of the individual photodetector elements is governed by the frequency spacing  $f_0$  and the physical parameters of the optical system. We want the detectors to be large enough to collect photons efficiently but small enough to resolve the spatial frequencies. The procedure is to find an aperture function  $a(x)$  whose transform  $A(\alpha)$  is adequate to control the sidelobe levels while still providing the required dip between two frequencies that are  $f_0$  apart. For a typical case where  $a(x)$  is a Gaussian distribution

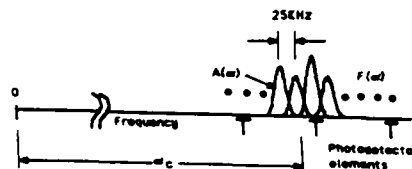


Fig. 2. Spatial frequency plane and photodetector geometry.

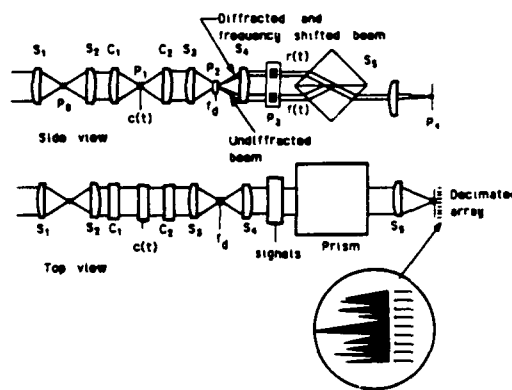


Fig. 3. Practical implementation of heterodyne spectrum analyzer.

truncated by the Bragg cell at the  $1/e^4$  intensity levels, the photodetector size  $\alpha_e$  can be made equal to the full width of  $A(\alpha)$  at the half-power points, while maintaining a 1-dB dip between two equal amplitude cw signals.

Scanning the spectrum can be most easily accomplished by scanning the source in plane  $P_2$ . As shown in the top view, a Bragg cell placed in plane  $P_1$  is driven by a repetitive function  $c(t)$ , such as a linear FM signal, whose period is  $T_s$ . If the value of  $M$  is small, we might consider driving the Bragg cell by a set of  $M$  discrete frequencies to provide a staircase scanning pattern. In either case, a set of cylindrical and spherical lenses ( $C_1$ ,  $C_2$ ,  $S_2$ , and  $S_3$ ) image the primary source from plane  $P_0$  into the secondary source in plane  $P_2$  while providing the requisite illumination for the Bragg cell. The time-bandwidth product of this scanning cell must be equal to  $M$ ; the bounds are  $M = 1$ , when no scanning is invoked, to  $M = N$ , when only a single photodetector is used.

The modulator in plane  $P_2$  must accommodate the change in the spatial position of the secondary source produced by the scanning action of the first Bragg cell. As arranged, the two cells are orthogonal so that the modulator must have a interaction length of  $Md_0$ , where  $d_0$  is the diameter of the secondary source in plane  $P_2$ . The phase responses of the scanning Bragg cell and the modulator are not important because they affect both branches of the interferometer equally. Amplitude changes must as usual be compensated either by a feedback technique or by scaling the measured data.

### IV. Reference and Scanning Functions

The reference signal is obtained by driving the upper channel of the Bragg cell in plane  $P_3$  with a signal



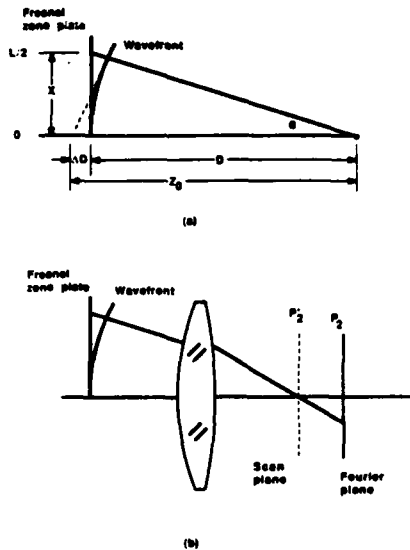


Fig. 4. Reference waveform for scanning spectrum: (a) convergent wavefront produced by a Fresnel zone plate; (b) lens acting on a chirp waveform.

$$r(t) = \sum_{n=N_1}^{N_2} \cos(2\pi n f_0 t + \phi_n), \quad (6)$$

where  $N_2 - N_1 + 1 = N$  is equal to the number of photodetectors required and  $f_0$  is the spacing of the frequency components of  $r(t)$ . If  $N_2$  and  $N_1$  are both positive integers, the signal represented by Eq. (6) is a bandpass signal whose center frequency is  $f_c = f_0(N_2 + N_1)/2$ ; this is the signal we use in practice. It is clear that  $r(t)$  is the sum of  $N$  equal-amplitude discrete frequencies spaced  $f_0$  apart. Such a signal is an ideal distributed local oscillator because it causes the term  $|R(\alpha, t)|$  in the integrand of Eq. (5) to have a fixed amplitude at each sample point of the spectrum  $F(\alpha, t)$ ; the spectrum is, therefore, accurately measured.

The scanning action is obtained by driving the Bragg cell in plane  $P_1$  with a similar signal

$$c(t) = \sum_{m=M_1}^{M_2} \cos(2\pi m f_s t + \phi_m), \quad (7)$$

where  $M = M_2 - M_1 + 1$  is the decimation rate. It is clear that  $c(t)$  has period  $T_s = 1/f_s$ , which is the time allowed between successive samples of the same spatial frequency; we call  $T_s$  the revisit interval. We obtain the required scanning action by setting

$$\phi_m = \frac{d\pi m^2}{M}, \quad (8)$$

which causes  $c(t)$  to be an infinite train of chirp signals having a duty cycle  $0 \leq d \leq 1$ .<sup>4</sup> In general, the light is most efficiently used when  $d = 1$  because the Bragg cell is then completely filled at all times. We now show the conditions for which  $c(t)$  produces such a chirp function.

Consider the convergent wave front produced by an on-axis Fresnel zone function as shown in Fig. 4(a). This wave front represents the baseband version of  $c(t)$ ; the basic arguments are the same for the more

general case, except for an angular tilt in the diffracted light caused by the carrier frequency. The unit amplitude wave front can be expressed as

$$c(x) = \exp(j\pi x^2/\lambda D), \quad (9)$$

so that the spatial frequency  $\alpha = x/\lambda d$  is a linear function of  $x$ . We might, therefore, expect that at any instant in time the function  $c(t - x/v - T/2)$  within the Bragg cell consists of  $M$  plane waves, which combine to produce the converging wave front.

Bardos<sup>5</sup> suggested an interesting way to find the focal point of the traveling function  $c(t - x/v - T/2)$  and show what  $\phi_m$  should be in the bargain. Figure 4(a) shows that a tangent to the wave front intercepts the  $z$  axis a distance  $\Delta D$  from the central ordinate; this distance, multiplied by  $k = 2\pi/\lambda$ , is the phase advance of the wave front relative to the phase at  $x = 0$ . We see that  $Z_0 = D/\cos\theta$  so that the distance  $\Delta D = Z_0 - D = D[(1/\cos\theta) - 1]$ . For small  $\theta$ , we find that  $\Delta D = (D/2)\theta^2 = (D/2)(x^2/D^2) = x^2/2D$ . The phase is obtained by multiplying  $\Delta D$  by  $2\pi/\lambda$  so that  $\phi = \pi x^2/\lambda D$ , in agreement with Eq. (9).

We express the phase of the  $m$ th plane wave in the same way:

$$\phi_m = (2\pi/\lambda)(D/2)m^2\theta_0^2, \quad (10)$$

Since  $\theta_0 = \lambda/L$  is the smallest resolvable angular change for any signal of total physical length  $L$ , we have

$$\phi_m = \frac{\pi}{\lambda} D m^2 \left( \frac{\lambda^2}{L^2} \right) = \frac{\pi D \lambda m^2}{L^2}. \quad (11)$$

The highest spatial frequency of a chirp waveform is

$$\alpha_{\max} = \frac{1}{2\pi} \frac{\partial}{\partial x} \left( \frac{\pi x^2}{\lambda D} \right)_{x=L/2} = \frac{L}{\lambda D}. \quad (12)$$

The space-bandwidth product of the signal is  $L\alpha_{\max}$ , which must be equal to  $M$ , the number of discrete frequencies scanned within the revisit interval  $T_s$ . Thus we find that  $\phi_m = \pi m^2/M$  for  $d = 1$  as stated in Eq. (8).

The evolution of the function  $c(t - x/v - T/2)$  resident in the Bragg cell to the focused spot as depicted in Fig. 4(a) can be obtained by expressing the amplitude of the light leaving the Bragg cell as

$$c(x, t) = \text{Re} \left\{ \sum_m \exp[j2\pi m f_s (t - x/v - T/2) + j\phi_m] \right\}, \quad (13)$$

where we have ignored unimportant factors. We now ignore the temporal frequency and consider the propagation of these plane waves; we define  $k_{mx}$  and  $k_{mz}$  as the  $x$  and  $z$  components of the normal to the  $m$ th plane wave. The propagation into free space is then represented by

$$c(x, z) = \text{Re} \left\{ \sum_m \exp[j(k_{mx}x + k_{mz}z + \phi_m)] \right\}, \quad (14)$$

where  $k_{mx} = k \sin(m\alpha_0)$  and  $k_{mz} = k \cos(m\alpha_0)$ ;  $\alpha_0$  is the incremental spatial frequency. We use small angle approximations to find that

$$c(x,z) = \text{Re} \sum_m \exp(jkm\alpha_0 x) \exp(jkz) \times \exp(-jkm^2\alpha_0^2 z/2) \exp(j\pi m^2/M). \quad (15)$$

The focal point occurs when the summation is maximized, which in turn occurs when the exponent terms in  $m$  are set equal to zero:

$$km\alpha_0 x - km^2\alpha_0^2 z/2 + \pi m^2/M = 0,$$

from which we get

$$\frac{2x}{L} - \frac{m\lambda z}{L^2} + \frac{m}{M} = 0. \quad (16)$$

To satisfy this equation for all  $m$ , we set both the constant and the term linear in  $m$  to zero. We find that the focus occurs at  $x = 0$  and  $z = D$ , where

$$\frac{L^2}{\lambda D} = M. \quad (17)$$

The focal point is, therefore, located at  $x = 0$  in the plane  $z = D$ .

The chirp rate is a function of the revisit interval  $T_s$  and the decimation ratio  $M$ ; we need to scan across  $M$  distinct spatial frequencies, each separated by  $\alpha_0$  in the time interval  $T_s$ . We can express the chirp rate in terms of temporal frequencies by noting that  $\alpha_0 v = f_0$ . The chirp rate is then  $CR = Mf_0/T_s = W/QT_s$ .

If the Bragg cells in planes  $P_1$  and  $P_3$  in Fig. 3 have time apertures of the order of  $T$ , only a fraction  $T/T_s$  of a given chirp waveform is in the cell at any instant. The focused spot created by lens  $S_3$  will be well formed except during the transition from one chirp period to the next. For the most usual situation in which  $T_s \gg T$ , we can blank the output during this transition. If  $T_s \approx T$ , we require a Bragg cell having twice the chirp length if the data must be continuously collected. The latter implementation results in a factor of two penalty in usable light source power.

The chirp function travels through the Bragg cell at a velocity which may be too fast to cover  $M$  spots in  $T_s$  seconds. We can control the scan velocity by choosing the focal length of  $S_3$ , given the value of  $D$  from Eq. (17). From Fig. 4(b) and by some straightforward ray tracing, we find that the spot scanning velocity  $v_s$  in plane  $P_2$  is  $v_s = v(F_3/D)$ , where  $F_3$  is the focal length of lens  $S_3$ . Plane  $P_2$  is not, however, the Fourier plane of lens  $S_3$ ; the focusing action of the chirp causes the scan plane to be at plane  $P'_2$ , short of the Fourier plane by a distance  $F_3^2/(D + F_3)$ .

The time-bandwidth product of the chirp Bragg cell is a factor of  $T/T_s$  less than the signal processing cell due to the lower bandwidth required. As  $T \rightarrow T_s$ , the required time-bandwidth product doubles because the time aperture must double. From these considerations, as well as those to follow based on SNR and dynamic range analyses, we conclude that this scanning technique is most useful for spectrum analysis when  $T_s \gg T$ .

We can represent the Fourier transform of the scanning chirp function at plane  $P_2$  as

$$C(\alpha, t) = \text{rect}(\alpha/\alpha_0) \sum_{n=-\infty}^{\infty} \delta[\alpha - (n + t/T_s)\alpha_0], \quad (18)$$

where  $\alpha_s$  is the total spatial frequency interval scanned in the revisit interval  $T_s$ . Without loss of generality, we can set the magnification between planes  $P_2$  and  $P_4$  equal to unity. The spectrum in plane  $P_4$  is expressed as the convolution of  $C(\alpha, t)$  with the spectra  $F(\alpha, t)$  and  $R(\alpha, t)$ . After detection and bandpass filtering, we express the photocurrent output of the  $k$ th photodetector element as

$$i_k(t) = 2A_0 \sum_{n=-\infty}^{\infty} \text{rect}\{[\alpha_k - (n + t/T_s)\alpha_s]/\alpha_s\} \times F[(n + t/T_s)\alpha_s] R^*[(n + t/T_s)\alpha_s] \times \cos[2\pi f_d t + \phi[(n + t/T_s)\alpha_s]]. \quad (19)$$

The rect function reveals how the scanning action takes place. For any index number  $n$ ,  $i_k(t)$  has a value for  $(-1/2 + k - n)T_s < t < (1/2 + k - n)T_s$ . When  $t$  exceeds the upper bound, the index increases by one so that the scanning revisits the lower bound of the frequency interval sampled by the  $k$ th photodetector element. We must increase the bandwidth of the Bragg cell in plane  $P_4$  to accommodate the mismatch in the Bragg angle caused by the scanning action. The increased bandwidth is  $W_1 = W(1 + M/N)$  so that  $M/N$  represents the fractional increase in  $W$ . Generally the increase in bandwidth will be  $< 10\%$ .

#### V. Signal-to-Noise Ratio and Dynamic Range

The number of samples that each photodetector must take in each revisit interval is  $R = M/T_s$ . The revisit interval is application dependent and strongly affects the required postdetection bandwidth, which in turn affects the SNR and dynamic range. Frequency hopped signals typically dwell at a randomly selected frequency for  $T_h \gg T$  seconds. In some cases, we may need to detect simply the angle of arrival of the hopping signals in the time interval  $T_s \approx T_h$ . In other cases, we may need to detect the angle of arrival of the hopper in a time interval much less than  $T_h$  to implement a replica jammer or to demodulate the signal; in this case  $T_s < T_h$ , but  $T_s$  may still be long relative to  $T$ .

To find the effect of  $T_s$  on the dynamic range, we define the SNR at the output of the photodetector as the signal power divided by the sum of the shot noise and thermal noise terms:

$$\text{SNR} = \frac{\langle i_k^2 \rangle R_L}{2eB(\bar{i}_d + \bar{i}_k)R_L + 4kTB} \quad (20)$$

where  $i_k(t)$  is the photocurrent produced by the  $k$ th photodetector element,  $R_L$  is the load resistor,  $e$  is the charge on an electron,  $B$  is the postdetection bandwidth,  $\bar{i}_d$  is the dark current  $\bar{i}_k = \langle i_k(t) \rangle$  is the average signal current due to the two baseband terms,  $k$  is Boltzmann's constant, and  $T$  is the temperature in degrees kelvin.

The value of the load resistance is determined by the highest frequency required to pass the signal. This

frequency is  $f_d + B/2$ , which for  $B \ll f_d$  is approximately  $f_d$ . The load resistance is then

$$R_L = \frac{1}{2\pi c_d f_d} \quad (21)$$

where  $c_d$  is the photodetector capacitance. The value of  $f_d$  is generally determined by the desire to keep  $i_3(\alpha, t)$  as given by Eq. (5) a narrowband function.

We substitute Eq. (21) into Eq. (20) and rearrange terms to obtain

$$\text{SNR} = \frac{\langle i_k^2 \rangle}{2eB(\bar{i}_d + \bar{i}_k) + 8\pi k T B c_d f_d} \quad (22)$$

A few comments regarding Eq. (22) are in order so that we understand how the SNR is to be calculated under a wide range of photodetector applications. The performance of the system is highly dependent on whether thermal noise or shot noise dominates. The average current  $\bar{i}_k$  always dominates the dark current  $\bar{i}_d$  in a heterodyne application, but its value may not be large enough so that shot noise dominates thermal noise when  $f_d$  is large. For these conditions, the system is thermal noise limited due to limited available laser power.

Consider the photodetector current at a particular spatial frequency of the form

$$i_k = \delta(1 - \gamma)i_s + (1 - \delta)\gamma i_r + 2\sqrt{\delta(1 - \delta)\gamma(1 - \gamma)i_s i_r} \cos(2\pi f_d + \phi_k) \quad (23)$$

where  $i_s$  is proportional to the signal intensity,  $i_r$  is proportional to the reference intensity, and  $\delta$  and  $\gamma$  are the fractional reflection coefficients of the beam splitter and beam combiner. After bandpass filtering, the power in the detected signal is  $\delta(1 - \delta)\gamma(1 - \gamma)i_s i_r$ . When the thermal noise dominates the shot noise, we have

$$\text{SNR} = \frac{\delta(1 - \delta)\gamma(1 - \gamma)i_s i_r}{8\pi k T B c_d f_d} \quad (24)$$

The maximum input signal level  $i_s$  is generally limited by intermodulation products produced by the Bragg cell.

Our primary interest is to maximize the dynamic range, not necessarily the SNR. The two measures are connected, however, because the minimum detectable signal level occurs when the SNR is equal to one and establishes the dynamic range as

$$\text{DR} = 20 \log \left( \frac{i_{\max}}{i_{\min}} \right) \quad (25)$$

The dynamic range is, therefore, a function of the minimum value of  $i_s$ , which is in turn determined by the SNR.

We now consider how the dynamic range is affected by the decimation rate. The reference function  $R(\alpha, t)$  contains  $N$  discrete frequency terms, called beads, distributed over the analysis bandwidth  $W$ . These beads interfere with the signal spectrum  $F(\alpha, t)$  to produce an output that has a nominal frequency  $f_d$ . The frequency difference between the beads is of the order of  $f_0 = 1/T$ . Since a received signal can have any frequency in

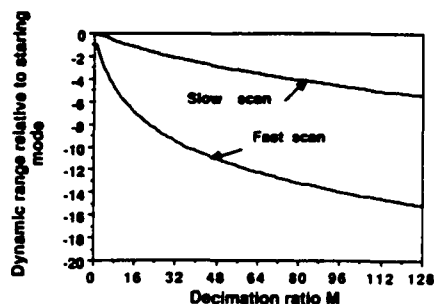


Fig. 5. Dynamic range for heterodyne spectrum analyzer using a decimated array.

the range  $\pm f_0/2$  relative to that of the nearest bead, the postdetection bandpass filter must respond from  $f_d \pm f_0/2$  to measure accurately the spectrum; the postdetection bandwidth in the nonscanning, or staring, mode is, therefore,  $B_s = f_0 = 1/T$ .

In the scanning mode, the instantaneous carrier frequency  $f_i$  within the band  $f_d \pm f_0/2$  is amplitude modulated by the detected spectrum. The bandwidth of the amplitude modulation is  $B_{am} = M/2T_s$ , based on the Nyquist sampling rate. The postdetection bandwidth required is now increased to  $B = B_s + B_{am}$ . The loss in performance when using the decimated array may not be too severe if  $B_{am} \ll B_s$ . As a base line system, consider a spectrum analyzer whose analysis bandwidth is  $W = 50$  MHz and whose effective processing time is  $T = 40 \mu\text{s}$ . We characterize the performance of the system for a slow scan rate wherein each frequency is revisited in a time interval  $T_s = 25T$ ; the slow scan might be used to detect frequency hopped signals whose dwell time is  $T_h \approx 1$  ms. We also consider a fast scan rate wherein  $T_s \approx 2T$ ; the fast scan might be used to track rapidly changing signals or to detect short pulses. In Fig. 5 we plot the dynamic range, referenced to the staring mode for which  $B = 1/T = 25$  kHz, vs the decimation ratio  $M$  for the fast and slow scan modes.

The fast scan results show that the dynamic range loss is 4.8 dB for  $M = 8$  and 6.9 dB for  $M = 16$ . We note that the loss in performance falls less rapidly than  $1/M$  because the postdetection bandwidth  $B = B_s + B_{am}$  does not increase linearly with  $M$ . The slow scan performance is no better than that of the fast scan mode for small values of  $M$  because the postdetection bandwidth cannot be reduced below  $B = B_s$  due to the frequency uncertainty. A side benefit of this general performance level is that the dynamic range does not deteriorate rapidly as a function of the decimation ratio. For example, when  $M = 32$ , the loss in dynamic range referenced to the staring mode is only 2.1 dB; the use of the decimated technique is, therefore, an attractive trade.

## VI. Improved Reference Waveform

From the thermal noise limited result given in Eq. (24), we see that an increase in the reference signal  $i_r$  increases the SNR and, therefore, the dynamic range. In the heterodyne spectrum analyzer we can alter the

reference waveform  $r(t)$  to more efficiently use the light in the Fourier domain to increase  $i_r$ . The geometry of the photodetector elements is the same as that shown in Fig. 2 except that the photodetector elements are now  $M$  times wider than before so that they are abutting. In this case, we do not scan the signal beam relative to the set of  $M$  enlarged photodetector elements. Instead, we make use of the fact that the cross-product term as given, for example, by Eq. (5), exists only if the two beams overlap. To produce a controlled overlap, we generate a reference waveform that produces  $Q = N/M$  optical probes or beads at plane  $P_2$ . Each bead overlaps the spectrum  $F(\alpha, t)$  for a spectral range that is one part in  $M$  of the photodetector width. Then, instead of scanning the joint spectra across the fixed photodetector array as before, we scan only the reference beam across the spectrum  $F(\alpha, t)$  and photodetector array, which are both fixed in space. Each reference beam probe scans the  $M$  spatial frequencies associated with its assigned photodetector element before the scan is reset.

The appropriate reference waveform is a chirp function similar to that given by Eq. (7) where the fundamental frequency of the waveform  $f_s$  is now given by the full analysis bandwidth  $W$  divided by the number of required probes:  $f_s = W/Q$ . From this relationship we immediately deduce that  $T_s = T/M$  so that, within the Bragg cell, there are exactly  $M$  chirp waveforms, each of time duration  $T_s$ . Note that the general form of  $c(t)$  as given by Eq. (7) provides the desired waveform, although it is quite different from the one used in plane  $P_1$  of Fig. 3. Here there are  $M$  chirp waveforms within the reference Bragg cell, whereas before the chirp waveform was typically much longer than the scanning Bragg cell ( $T_s \gg T$ ).

The reference beam geometry for this case is shown in Fig. 6 along with the detail associated with the Fourier domain. Although the chirp waveforms in plane  $P_2$  are traveling at the acoustic velocity  $v$ , the optical probes are both focused and stationary in plane  $P_2$ . The claim may seem strange at first, especially when compared to our previous result as illustrated in Fig. 4. The difference is that we now have  $M$  chirp waveforms in the cell at one time. It is true that they focus individually at some plane short of the Fourier plane and that, in this plane, the focused spots move with velocity  $v_s = v(F_2/D)$  as before. But since the chirp length is now  $L/M$  instead of  $L$ , the distance  $D$  is reduced a factor of  $M^2$ . As a result, these focused and scanning spots become secondary sources that completely overlap at the Fourier plane and mutually interfere to produce the required  $Q$  stationary probes.

We cause the probe ensemble to scan with a controlled velocity by chirping the center frequency of the reference signal. The bandwidth of the Bragg cell is increased to  $W_1 = W(1 + 1/Q)$  independently of the scanning rate; this increase is generally  $<10\%$  or so. The scanning rate merely determines the chirp rate, not the highest frequency in the chirp waveform.

The major advantage of using this modified reference waveform is that it increases the dynamic range

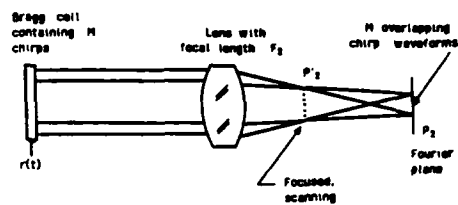


Fig. 6. Optical subsystem to produce a decimated reference beam.

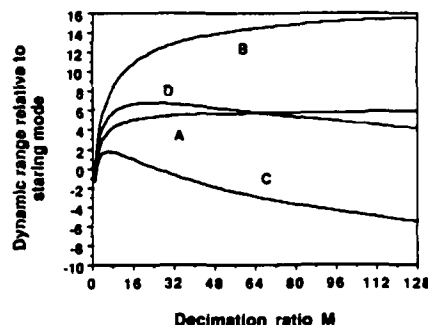


Fig. 7. Dynamic range with improved reference function: curve A, fast scan mode; curve B, slow scan mode; curve C, fast scan mode with shot noise an order of magnitude less than the thermal noise; curve D, slow scan mode with shot noise an order of magnitude less than the thermal noise.

when the system would otherwise be thermal noise limited. The reference current  $i_r$  is increased by a factor of  $M$ , which improves the SNR and dynamic range. Figure 7 shows that the dynamic range increases relative to the staring mode by 6 dB for the fast scan mode (curve A) and by 16 dB for the slow scan mode (curve B). The improvement is due to the fact that the numerator of Eq. (24) increases more rapidly than the denominator, the increase being more pronounced for the slow scan mode. For large  $M$ , the numerator and denominator of Eq. (24) are both proportional to  $M$  so that the dynamic range approaches an asymptotic value in both cases.

From Eq. (22) we see that the system eventually becomes shot noise limited because the local oscillator current also increases according to  $M$ . Consider the SNR when we use Eq. (23) in Eq. (22) so that  $i_r$  dominates  $i_d$ , and, furthermore, the shot noise is comparable with the thermal noise:

$$\text{SNR} = \frac{\delta(1-\delta)\gamma(1-\gamma)i_r}{2eB\gamma(1-\delta)i_r + 8\pi kTBc_d f_d} \quad (26)$$

From Eq. (26) we see that a significant increase in the local oscillator current  $i_r$  will cause the shot noise to dominate the thermal noise; further increases in  $i_r$  will not result in an improvement in the SNR or the dynamic range. In Fig. 7 we show the performance of the system when the shot noise is 10 dB less than the thermal noise in the staring mode. As the decimation rate increases, the dynamic range increases for small values of  $M$ , as shown in curve C, for the reasons discussed relative to the thermal noise limited case. At  $M = 6$  in the fast scan mode and at  $M = 22$  in the

slow scan mode, the shot noise is equal to the thermal noise; the dynamic range then begins to decrease as  $M$  increases because the signal power is proportional to  $M$ , whereas the shot noise, which is now dominant, is proportional to  $M^2$ .

The results of Fig. 7 show that a significant reduction in the number of photodetectors can be obtained without a large penalty in dynamic range. For example, from curve  $C$  we see that the loss in dynamic range for  $M = 32$  is only 0.7 dB. In all other cases the performance is more favorable.

## VI. Background for the Cross-Spectrum Analyzer

Suppose that we want to detect a burst of energy associated with a narrowband radio signal of bandwidth  $B_h$  located somewhere in a total rf bandwidth  $W$ ; the minimum dwell time of the burst signal is  $T_h$ . We want to measure the angle of arrival of these signals to assist in a sorting process. The receiver consists of two antenna elements as shown in Fig. 8. The distance between the two elements is  $D_{ab}$ , and the received signals are  $s_a(t)$  and  $s_b(t)$ . Suppose that a source is situated at an angle  $\theta$  with respect to the boresight of the array. The rf energy arrives at antenna element  $b$  with time delay  $\tau = (D_{ab}/c) \sin\theta$  with respect to that at antenna element  $a$  so that  $s_b(t) = s_a(t - \tau)$ .

One way to measure the angle of arrival is to feed these received signals to a two-channel Bragg cell spectrum analyzer as shown in Fig. 9. In the top view we show the Bragg cell illuminated by the laser via lens  $L_1$ , the lens  $L_2$  that performs the Fourier transformation, and the Fourier transform centered at  $\alpha_c$  in plane  $P_2$ . The side views shows the two channels driven by  $s_a(t)$  and  $s_b(t)$ . Lens  $L_2$  also produces the Fourier transform of the light distribution created by the Bragg cell in the vertical direction.

In the horizontal direction we describe the Fourier transform of the first input signal as

$$S_a(\alpha, t) = \int_{-\infty}^{\infty} a(x) s_a(t - x/v - T/2) \exp(j2\pi\alpha x) dx, \quad (27)$$

similar to Eq. (1). To obtain the full 2-D Fourier transforms from the upper channel, we account for the light distribution in the vertical direction by the integral

$$S_a(\beta) = \int_{-\infty}^{\infty} \text{rect}\left(\frac{y - H/2}{h}\right) \exp(j2\pi\beta y) dy \quad (28)$$

$$= \text{sinc}(\beta h) \exp(j\pi\beta H), \quad (29)$$

where  $h$  is the height of the transducer and  $H$  is the distance between the two channels. Therefore,

$$\begin{aligned} S_a(\alpha, \beta, t) &= S_a(\alpha, t) S_a(\beta) \\ &= S_a(\alpha, t) \text{sinc}(\beta h) \exp(j\pi\beta H). \end{aligned} \quad (30)$$

For  $s_b(t) = s_a(t - \tau)$ , we have

$$S_b(\alpha, t) = S_a(\alpha, t) \exp(j2\pi\alpha v\tau), \quad (31)$$

so that the spectrum  $S_b(\alpha, t)$  is identical to  $S_a(\alpha, t)$  save for a phase term  $\phi(\alpha) = 2\pi\alpha v\tau$ , which contains the

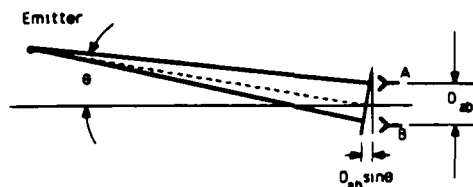


Fig. 8. Dual antenna geometry to determine angle of arrival.

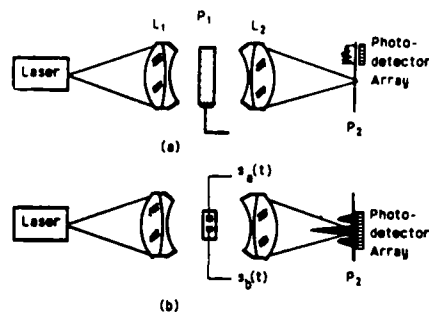


Fig. 9. Dual channel Bragg cell processor: (a) top view; (b) side view.

angle-of-arrival information. By a similar calculation, we find that  $S_b(\beta) = S_a^*(\beta)$  so that

$$S_b(\alpha, \beta, t) = S_a(\alpha, t) \exp(j2\pi\alpha v\tau) \text{sinc}(\beta h) \exp(-j\pi\beta H). \quad (32)$$

The total spectrum in plane  $P_2$  is the sum of  $S_a(\alpha, \beta, t)$  and  $S_b(\alpha, \beta, t)$  as given by Eqs. (30) and (32). The intensity is the magnitude squared of this sum:

$$I(\alpha, \beta, t) = |S_a(\alpha, t)|^2 \text{sinc}^2(\beta h) \{1 + \cos[2\pi H\beta + \phi(\alpha)]\}. \quad (33)$$

This result shows that, in the  $\alpha$ -direction, we have the same output as from an instantaneous power spectrum analyzer. In the vertical, or  $\beta$ -direction, the resultant intensity reveals a  $\text{sinc}(\beta h)$  envelope function due to the height of the transducers that modulates a spatial fringe structure whose frequency is established by the distance between the transducers. As we see from Eq. (33), the fringe pattern is encoded by the phase  $\phi(\alpha)$  as a function of the spatial frequency.

One way to detect  $\phi(\alpha)$  is to use a 2-D photodetector array that has  $N$  elements in the horizontal direction to sample the spatial frequencies and  $M$  elements in the vertical direction to sample the phase. Another method is to demodulate the phase by multiplying  $I(\alpha, \beta, t)$  by a function  $D_c(\beta) = 1 + \cos(h\beta)$  and by integrating the product over all values of  $\beta$ ; another function  $D_s(\beta) = 1 + \sin(h\beta)$  provides the quadrature component. Variations on these two basic detection schemes are possible. It is clear that we need at least two  $N$ -element detector arrays to implement the latter scheme and at least one  $M \times N$ -element array to implement the former scheme. The available light is not used efficiently, and, for large  $N$ , the arrays cannot be sampled often enough to ensure that signals are not missed. We now consider a cross-spectrum analyzer that uses a temporal, instead of spatial, heterodyne technique to measure the phase. This approach efficiently uses the available light and can do so with fewer photodetector elements.

### VIII. Cross-Spectrum Analyzer with Temporal Heterodyning

The cross-spectrum analyzer is a special case of the interferometric spectrum analyzer described earlier. If we associate  $S_a(\alpha, t)$  and  $S_b(\alpha, t)$  with  $F(\alpha, t)$  and  $R(\alpha, t)$ , the interferometric spectrum analyzer architecture shown in Fig. 3 is exactly what we need to generate the desired cross-spectral products. The intensity at plane  $P_2$  is

$$I(\alpha, t) = |S_a(\alpha, t) + S_b(\alpha, t) \exp(j2\pi f_d t)|^2, \quad (34)$$

where  $S_a(\alpha, t)$  is given by Eq. (27). The exponential in Eq. (34) shows that the light in the upper branch has been upshifted by  $f_d$ ; it could equally well be downshifted without affecting the results. We do not show an explicit functional dependence of the intensity on  $\beta$  because both Bragg cells are located on the same vertical axis. We expand the intensity function to get

$$I(\alpha, t) = |S_a(\alpha, t)|^2 + |S_b(\alpha, t)|^2 + 2|S_a(\alpha, t)S_b^*(\alpha, t) \cos[2\pi f_d t + \phi(\alpha, t)]|. \quad (35)$$

For  $s_b(t) = s_a(t - \tau)$ , we have

$$I(\alpha, t) = |S_a(\alpha, t)|^2 [1 + \cos[2\pi f_d t + \phi(\alpha, t)]], \quad (36)$$

where  $\phi(\alpha, t) = 2\pi\alpha v\tau$  is the phase we seek to measure. The phase term is now associated with a sinusoidal function of time instead of space. Therefore, we need to demodulate  $I(\alpha, t)$  in time instead of demodulating  $I(\alpha, \beta, t)$  in space as described in Sec. VII. As a consequence, we must now use discrete element photodetectors that have sufficient bandwidth to pass the desired signal whose energy is centered at  $f_d$ .

As before, we identify three modes of operation: the staring mode, the fast scan mode, and the slow scan mode. To illustrate the phase detection technique for all three modes, consider the simple case of an emitter at angle  $\theta_j$  with respect to the boresight of the antenna array having frequency  $f_j$  and amplitude  $c_j$ . We assume that the spectral amplitude does not change significantly over the revisit time  $T_s$ . The signals from the two antenna elements then become

$$s_a(t) = c_j \cos(2\pi f_j t + \varphi_k), \quad (37)$$

$$s_b(t) = c_j \cos[2\pi f_j(t - \tau) + \varphi_k]. \quad (38)$$

We can let  $\varphi_k = 0$  without loss of generality. From Eqs. (27) and (34) we find that the cross-product term can be expressed as

$$g_1(t) = c_j^2 \cos(2\pi f_d t + \phi_j) \int_{-\infty}^{\infty} |A(\alpha - \alpha_j + \alpha_c/2)|^2 H(\alpha) d\alpha, \quad (39)$$

where  $H(\alpha)$  is the photodetector response defined before. The integral is an amplitude scaling factor that can be combined with several others. For example, we need to account for the optical efficiency of the system, the Bragg cell diffraction efficiency, the beam splitter and beam combiner ratios, the laser power. The output term can then be conveniently written as

$$g_1(t) = A_0 c_j^2 \cos(2\pi f_d t + \phi_j), \quad (40)$$

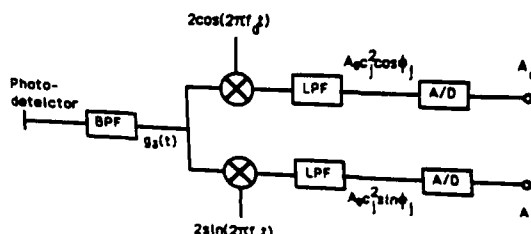


Fig. 10. Postdetection processing.

where  $A_0$  accounts for all the amplitude scaling factors.

The problem of extracting the phase and amplitude information from the temporal carrier  $f_d$  is similar to that of demodulating an angle modulated signal in a communication system. One way to demodulate the signal is through quadrature detection which recovers the phase angle and removes the sign ambiguity on  $\phi_j$  as shown in Fig. 10. We first eliminate the signals  $g_1(t)$  and  $g_2(t)$  produced by the intensity terms at baseband by a bandpass filter centered at  $f_d$ . We then multiply  $g_3(t)$  by  $2 \cos(2\pi f_d t)$  and  $2 \sin(2\pi f_d t)$ , which are derived from the same signal source used to drive the temporal modulator in plane  $P_2$  of Fig. 3. Thus we guarantee that any instabilities or drifts in the frequency  $f_d$  are automatically compensated. The results are then low pass filtered to provide the signals  $A_0 c_j^2 \cos \phi_j$  and  $A_0 c_j^2 \sin \phi_j$ . After an analog-to-digital conversion, the arctangent operation is applied to obtain the phase  $\phi_j$ , and, through evaluating  $A_0$ , the amplitude  $c_j$  can be obtained by forming the square root of the sum of the squares of the output signals  $A_s$  and  $A_c$ . Finally, the angle of arrival  $\theta$  can be computed from the relationship that  $\phi_j = 2\pi f_j \tau$  and  $\tau = (D_{ab}/c) \sin \theta$ .

There are several ways to implement the circuitry in hardware to extract the phase information including quadrature detectors, limiter/discriminators, or phase lock loops. Although these circuits are available in small packages, the challenge of implementing the electronics for 1000–2000 discrete detectors is formidable. The use of a decimated array coupled with a scanning of the spectrum similar to that described before can significantly reduce the hardware complexity.

### IX. Signal-to-Noise Ratio and Dynamic Range for the Cross-Spectrum Analyzer

There are two significant differences between the cross-spectrum analyzer and the heterodyne spectrum analyzer. The first major difference is that  $R(\alpha, t)$  serves as a local oscillator that produces a current  $i_r$  that dominates the dark current  $i_d$ . For the cross-spectrum application we consider the modified shot noise term

$$SN = 2eB(\bar{i}_d + \bar{i}_k). \quad (41)$$

The current  $\bar{i}_k$  is the average value of the bias terms resulting from the square-law detection of the sum of the outputs from the two branches of the interferometer. For the cross-spectrum analyzer case, we have

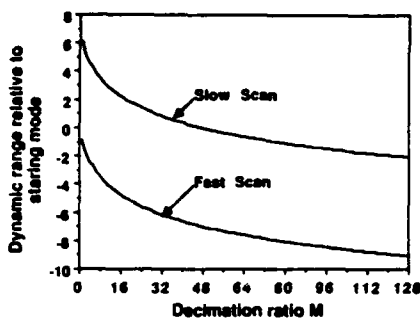


Fig. 11. Dynamic range performance for cross-spectrum analyzer.

$$\begin{aligned} \bar{i}_k &= \langle i_k \rangle \propto |S_o(\alpha, t) + S_b(\alpha, t)|^2 \\ &\propto |S_o(\alpha, t)|^2 + |S_b(\alpha, t)|^2 \\ &\propto 2|S_o(\alpha, t)|^2. \end{aligned} \quad (42)$$

We progress from the first to the second line because the cross-product term as given by Eq. (35) has zero average value; we progress from the second to the third line because  $s_b(\alpha, t)$  differs from  $s_o(\alpha, t)$  by only a phase factor.

A SNR of unity is obtained when the input signal level drops sufficiently. From Eq. (22) we find that, for almost any practical value of  $f_d$ , the shot noise term drops below the thermal noise term when.

$$2e(\bar{i}_d + \bar{i}_k) < 8\pi k T c_d f_d.$$

Since  $\bar{i}_k \propto 2|S_o(\alpha, t)|^2$  in the cross-spectrum analyzer case, its value eventually approaches  $\bar{i}_d$  as the input signal is reduced;  $\bar{i}_d$  then dominates the shot noise term. Furthermore, the system is now always thermal noise limited, since there is no local oscillator available to help obtain the more favorable shot-noise-limited case. We expect, therefore, to find that the performance resembles that shown in Fig. 5 for the heterodyne spectrum analyzer. There are, however, some significant differences as shown in Fig. 11. First, we note that the dynamic range decreases at a slower rate for the fast scan mode. For any value of  $M$ , the loss in dynamic range is a factor of 2 less in decibels for the cross-spectrum analyzer vs the heterodyne spectrum analyzer. The reason is that the system loss due to scanning is allocated between the two input signals in the cross-spectrum analyzer so that the system loss of 6 dB results in a dynamic range loss of 3 dB. For the heterodyne spectrum analyzer, the full system loss is allocated to the received signal because the reference signal has a fixed amplitude. As a result, the cross-spectrum analyzer performance deteriorates less quickly, as a function of the decimation ratio, than the heterodyne spectrum analyzer performance. It should be remembered, however, that the dynamic range for the heterodyne spectrum analyzer is nearly double that the cross-spectrum analyzer; the loss of dynamic range is, therefore, not as severe as it seems at first.

Finally, the slow scan mode results are 7 dB better than the fast scan mode because the postdetection

bandwidth can be reduced by virtue of the heterodyne frequency being exactly at  $f_d$  at all spatial frequencies. These results show that rather large savings can be made in the photodetector circuitry for more modest penalties in performance. For  $M = 32$ , we need only 64 photodetector elements instead of 2000. We can equal the staring mode performance if the laser power is increased by 6 dB; of course, the dynamic range will be reduced by 6 dB if we are already laser power limited. A 6-dB loss in dynamic range may, however, be a good trade for a thirty-two fold reduction in the complexity of the readout circuitry.

## X. Conclusions

The use of a decimated array in heterodyne spectrum analysis significantly reduces the complexity of the photodetector readout electronics with a reduction in the dynamic range that is often tolerable. We described an improved reference beam that concentrates light at only the probe positions, thereby increasing the dynamic range when the system is thermal noise limited. When the system is shot noise limited, the dynamic range is reduced by the decimation rate.

We also introduce temporal heterodyning techniques to determine the angle of arrival of signals received by two antenna elements. In this case, the phase information contained in the cross spectrum of the two signals, after suitable demodulation, is a measure of the angle of arrival. The decimated array technique is applicable here as well, noting that the dynamic range drops only half as fast as the decimation rate because the loss is shared by both signals.

We would expect to use decimation rates that provide a significant reduction in the photodetector complexity, say  $M = 8$  to  $M = 32$ . The loss in dynamic range relative to the staring mode is then generally  $< 4$  dB, and this approach is particularly attractive when used in the slow scan mode where the performance is even better. In cases where the system is thermal noise limited, no penalty is extracted; there may actually be an increase in the dynamic range as shown in Figure 7.

This work was performed with support from the U.S. Army Research Office.

## References

1. M. C. King, W. R. Bennett, L. B. Lambert, and M. Arm, "Real-Time Electrooptical Signal Processors with Coherent Detection," *Appl. Opt.*, **6**, 1367 (1967).
2. A. VanderLugt, "Interferometric Spectrum Analyzer," *Appl. Opt.*, **20**, 2770 (1981).
3. M. D. Koontz, "Miniature Interferometric Spectrum Analyzer," *Proc. Soc. Photo-Opt. Instrum. Eng.* **639**, 126 (1986).
4. A. VanderLugt and A. M. Bardos, "Spatial and Temporal Spectra of Periodic Functions for Spectrum Analysis," *Appl. Opt.*, **23**, 4269 (1984).
5. A. M. Bardos, Harris Corp.

---

**APPENDIX B**

**DECIMATED ARRAYS FOR SPECTRUM ANALYSIS**

**Reprinted from Proceedings of the SPIE**

**Volume 936, Pages 221-228, 1988**

---



## Decimated Arrays for Spectrum Analysis

A. VanderLugt

Electrical and Computer Engineering Department

North Carolina State University

Raleigh, NC 27695

### ABSTRACT

We describe heterodyne spectrum analyzers that operate with a significantly reduced number of photodetector elements. The technique is to decimate an N-element photodetector array by retaining only every M<sup>th</sup> element. We time-share the remaining elements by scanning the spectrum across the decimated array. Each photodetector therefore produces, as a time sequence of samples, the spectral content of the received signal over its associated spatial frequency range.

### 1. BACKGROUND FOR THE HETERODYNE SPECTRUM ANALYZER

The heterodyne spectrum analyzer consists of a conventional spectrum analyzer, modified to include a reference function in the Fourier plane.<sup>1</sup> Figure 1 shows a Mach-Zehnder interferometer in which the lower branch contains a Bragg

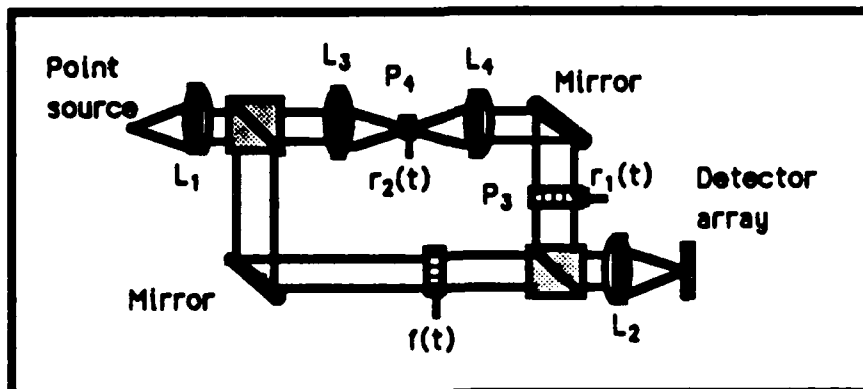


Figure 1. Heterodyne spectrum analyzer

cell in plane P<sub>1</sub> driven by the signal f(t) that we want to analyze. The Fourier transform of this signal, created in plane P<sub>2</sub>, is given by

$$F(\alpha, t) = \int_{-\infty}^{\infty} a(x) f(t - T/2 - x/v) e^{j2\pi\alpha x} dx, \quad (1)$$

where  $a(x)$  is the aperture weighting function that accounts for the illumination profile, acoustic attenuation, and truncation effects caused by the Bragg cell or lenses. In a similar fashion, the Fourier transform of the reference signal  $r(t)$  that drives the Bragg cell in plane  $P_3$  in the upper branch is

$$R(\alpha, t) = \int_{-\infty}^{\infty} a(x)r(t-T/2-x/v)e^{j2\pi\alpha x} dx. \quad (2)$$

The reference waveform can be temporally or geometrically shifted by a spatial frequency  $\alpha_d$  to provide a fixed temporal offset frequency  $f_d$  at each spatial frequency position.<sup>2</sup> In Figure 1,  $r_2(t)$  provides the required offset frequency shift. The intensity at plane  $P_2$  is then given by

$$I(\alpha, t) = |F(\alpha, t) + R(\alpha, t)e^{j2\pi f_d t}|^2. \quad (3)$$

We substitute (1) and (2) into (3) and expand the result to find that two intensity terms  $I_1(\alpha, t) = |F(\alpha, t)|^2$  and  $I_2(\alpha, t) = |R(\alpha, t)|^2$  are at baseband. The photodetector current is given by the integral of  $I(\alpha, t)$  over the photodetector area. Since the baseband terms have bandwidths no greater than  $1/T$ , we can remove them by a bandpass filter centered at  $f_d$ , where  $f_d \gg 1/T$ . The detector response to the cross product term  $I_3(\alpha, t)$  is then

$$i_3(\alpha, t) = A_0 \cos(2\pi f_d t + \phi) \int_{-\infty}^{\infty} |F(\alpha, t)| |R(\alpha, t)| H(\alpha) d\alpha, \quad (4)$$

where  $A_0$  is a scaling constant,  $\phi$  is the phase difference between  $F(\alpha, t)$  and  $R(\alpha, t)$ , and  $H(\alpha)$  is the photodetector response.

## 2. PHOTODETECTOR GEOMETRY AND DETECTION SCHEME

Figure 2 shows the photodetector geometry relative to the spectrum. The spectrum is centered at a spatial frequency  $\alpha_c$ , corresponding to the center frequency  $f_c$ ; each CW signal produces a response  $A(\alpha)$ . Suppose that we retain only every  $M^{\text{th}}$  element of the array scan the spectrum so that each element detects, over a period of time  $T_s$ , the  $M$  frequency components in its sector. Scanning the spectrum can be most easily accomplished by scanning the point source by means of an auxiliary Bragg cell (not shown) driven by a repetitive function  $c(t)$ , such as a linear FM signal, whose period is  $T_s$ . The time-bandwidth product of this scanning cell must be equal to  $M$ . The scanning action can be

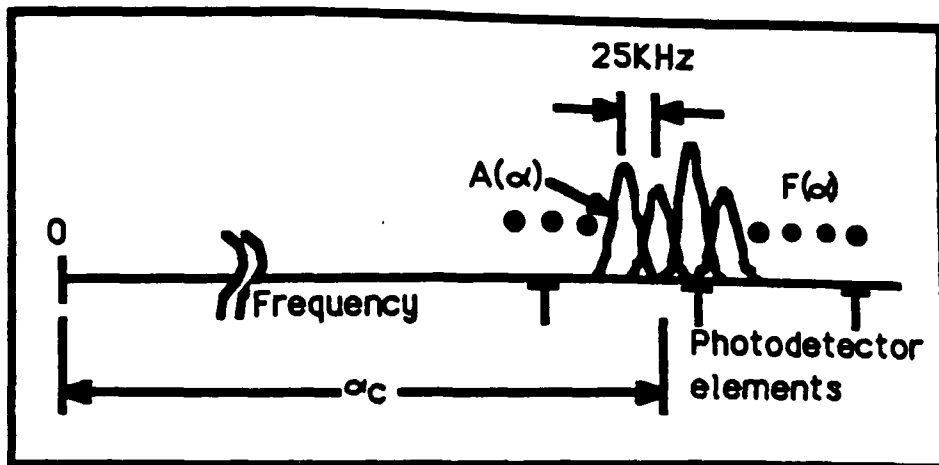


Figure 2. Photodetector array geometry

obtained by driving the Bragg cell in plane  $P_1$  with a chirp signal

$$c(t) = \sum_{m=M_1}^{M_2} \cos(2\pi m f_s t + \phi_m), \quad (5)$$

where  $M = M_2 - M_1 + 1$  is the decimation rate.<sup>3</sup> It is clear that  $c(t)$  has period  $T_s = 1/f_s$ , which is the time allowed between successive samples of the same spatial frequency; we call  $T_s$  the revisit interval.

The number of samples that each photodetector must take during each revisit interval is  $R = M/T_s$ . The revisit interval is application dependent and strongly affects the required post-detection bandwidth which, in turn, affects the signal-to-noise ratio and dynamic range. Frequency hopped signals typically dwell at a randomly selected frequency for  $T_h \gg T$  seconds. In some cases, we may need simply to detect the angle of arrival of the hopping signals in the time interval  $T_s \approx T_h$ . In other cases, we may need to detect the angle of arrival of the hopper in a time interval much less than  $T_h$  to implement a replica jammer or to demodulate the signal; in this case  $T_s < T_h$ , but  $T_s$  may still be long relative to  $T$ .

To find the effect of  $T_s$  on the dynamic range, we define the signal-to-noise ratio at the output of the photodetector as the signal power divided by the sum of the shot noise and thermal noise terms:

$$\text{SNR} = \frac{\langle i_k^2 \rangle R_L}{2eB(\bar{i}_d + \bar{i}_k)R_L + 4kTB}, \quad (6)$$

where  $i_k(t)$  is the photocurrent produced by the  $k^{\text{th}}$  photodetector element,  $R_L$  is the load resistor,  $e$  is the charge on an electron,  $B$  is the post-detection bandwidth,  $\bar{i}_d$  is the dark current,  $\bar{i}_k = \langle i_k(t) \rangle$  is the average signal current due to the two baseband terms,  $K$  is Boltzmann's constant, and  $T$  is the temperature in degrees Kelvin. Our primary interest is to maximize the dynamic range; the minimum detectable signal level occurs when the signal-to-noise ratio is equal to one and establishes the dynamic range as

$$DR = 20 \log \left[ \frac{i_s |_{\max}}{i_s |_{\min}} \right]. \quad (7)$$

We now consider how the dynamic range is affected by the decimation rate. The reference function  $R(\alpha, t)$  contains  $N$  discrete frequency terms, called probes, distributed over the analysis bandwidth  $W$ . These probes interfere with the signal spectrum  $F(\alpha, t)$  to produce an output that has a nominal frequency  $f_d$ . The frequency difference between the probes is of the order of  $f_0 = 1/T$ . Since a received signal can have any frequency in the range  $\pm f_0/2$  relative to that of the nearest probe, the post-detection bandpass filter must respond from  $f_d \pm f_0/2$  to accurately measure the spectrum; the post-detection bandwidth in the non-scanning, or staring, mode is therefore  $B_s = f_0 = 1/T$ .

In the scanning mode, the instantaneous carrier frequency  $f_i$  within the band  $f_d \pm f_0/2$  is amplitude modulated by the detected spectrum. The bandwidth of the amplitude modulation is  $B_{am} = M/2T_s$ , as based on the Nyquist sampling rate. The post-detection bandwidth required is therefore increased to  $B = B_s + B_{am}$ . Consider a spectrum analyzer whose effective processing time is  $T = 40 \mu\text{sec}$ . We characterize the performance of the system for a slow scan rate wherein each frequency is revisited in a time interval  $T_s = 25T$ ; the slow scan might be used to detect frequency hopped signals whose dwell time is  $T_h \approx 1 \text{ms}$ . We also consider a fast scan rate wherein  $T_s \approx 2T$ ; the fast scan might be used to track rapidly changing signals or to detect short pulses. In Figure 3 we plot the dynamic range, referenced to the staring mode for which  $B = 1/T = 25 \text{KHz}$ , versus the decimation ratio  $M$ . The fast scan results show that the dynamic range loss is 4.8dB for  $M=8$  and 6.9dB for  $M=16$ . We note that the loss in performance falls less rapidly than  $1/M$  because the post-detection bandwidth  $B = B_s + B_{am}$  does not increase linearly

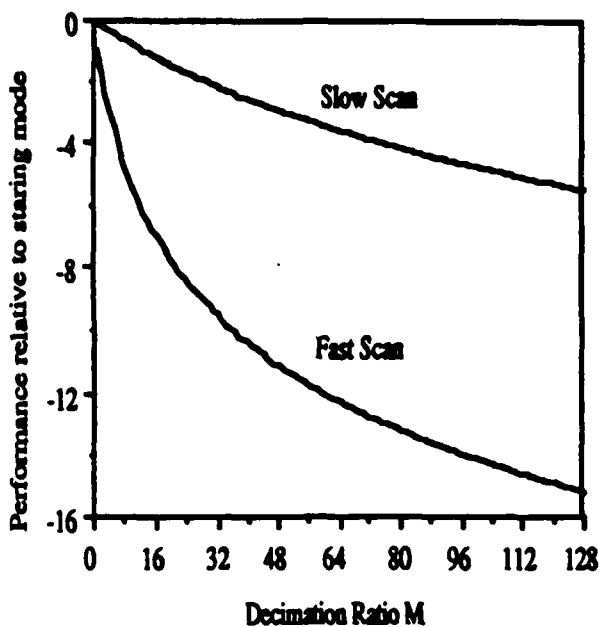


Figure 3. Relative performance with normal reference beam

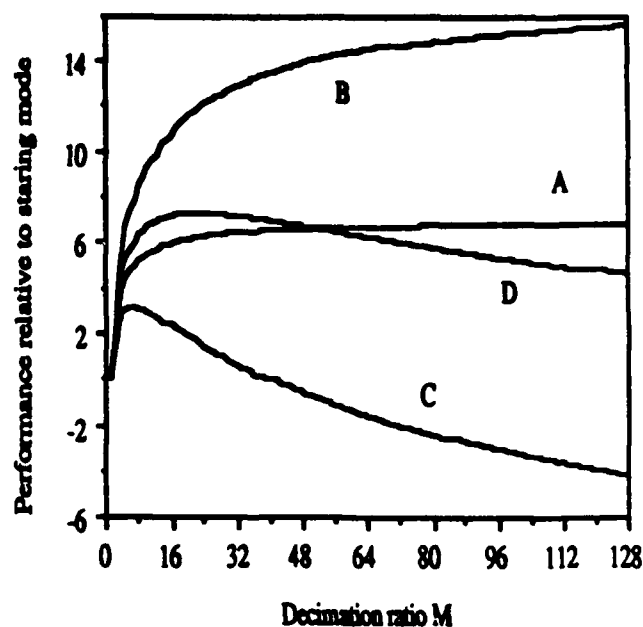


Figure 4. Relative performance with improved reference beam

with  $M$ . The slow scan performance is no better than that of the fast scan mode for small values of  $M$  because the post-detection bandwidth cannot be reduced below  $B = B_s$  due to the frequency uncertainty. A side benefit of this general performance level is that the dynamic range does not deteriorate rapidly as a function of the decimation ratio. For example, when  $M = 32$ , the loss in dynamic range referenced to the staring mode is only 2.1dB, the use of the decimated technique is therefore an attractive trade.

### 3. IMPROVED REFERENCE BEAM WAVEFORM

In the heterodyne spectrum analyzer we can alter the reference waveform  $r(t)$  to more efficiently use the light in the Fourier domain to increase the local oscillator current  $i_r$ . The geometry of the photodetector elements is the same as that shown in Figure 2 except that the photodetector elements are now  $M$  times wider than before so that they are abutting. In this case, we do not scan the signal beam relative to the set of  $M$  enlarged photodetector elements. Instead, we make use of the fact that the cross-product term exists only if the two beams overlap. To produce a controlled overlap, we generate a reference waveform that

produces only  $Q = N/M$  optical probes at plane  $P_2$ . Each probe overlaps the spectrum  $F(\alpha, t)$  for a spectral range that is one part in  $M$  of the photodetector width. Then, instead of scanning the joint spectra across the fixed photodetector array as before, we scan only the reference beam across the spectrum  $F(\alpha, t)$  and photodetector array which are both fixed in space. Each reference beam probe scans the  $M$  spatial frequencies associated with its assigned photodetector element before the scan is reset. We cause the probe ensemble to scan with a controlled velocity by chirping the center frequency of the reference signal.

The major advantage of using this modified reference waveform is that it increases the dynamic range when the system would otherwise be thermal noise limited. The local oscillator current  $i_r$  is increased by a factor of  $M$ , which improves the signal to noise ratio and the dynamic range. Figure 4 shows that the dynamic range increases by 6dB for the fast scan mode (curve A) and by 16dB for the slow scan mode (curve B). For large  $M$ , the signal and noise are proportional to  $M$  so that the dynamic range approaches an asymptotic value in both cases.

In Figure 4 we also show the performance of the system when the shot noise is 10dB less than the thermal noise in the staring mode. As the decimation rate increases, the dynamic range increases for small values of  $M$ , as shown in curve C, for the reasons discussed relative to the thermal noise limited case. At  $M = 8$  in the fast scan mode and at  $M = 22$  in the slow scan mode, the shot noise is equal to the thermal noise; the dynamic range then begins to decrease as  $M$  increases because the signal power is proportional to  $M$  whereas the shot noise, which is now dominant, is proportional to  $M^2$ . The results of Figure 4 show that a significant reduction in the number of photodetectors can be obtained without a large penalty in dynamic range. For example, from curve C we see that the loss in dynamic range for  $M = 32$  is only 0.7dB. In all other cases the performance is more favorable.

#### 4. CROSS-SPECTRUM ANALYSIS WITH HETERODYNE DETECTION

A cross-spectrum analyzer, used to obtain the angle of arrival of a signal  $s(t)$  detected by a dual-antenna system, is a special case of the interferometric spectrum analyzer described earlier. If we associate the Fourier transforms

$S_a(\alpha,t)$  and  $S_b(\alpha,t)$  of the two received signals with  $F(\alpha,t)$  and  $R(\alpha,t)$ , the interferometric spectrum analyzer architecture shown in Figure 1 is basically what we need to generate the desired cross-spectral products. There are two significant differences between the cross-spectrum analyzer and the heterodyne spectrum analyzer. The first major difference is that  $R(\alpha,t)$  serves as a local oscillator that produces a current  $i_r$  that dominates the dark current  $i_d$ . For the cross-spectrum application, the system is always thermal noise limited since there is no local oscillator available to help obtain the more favorable shot noise limited case. We expect, therefore, to find that the performance resembles that shown in Figure 3 for the heterodyne spectrum analyzer. There are, however, some significant differences as shown in Figure 5. First, we note that the dynamic

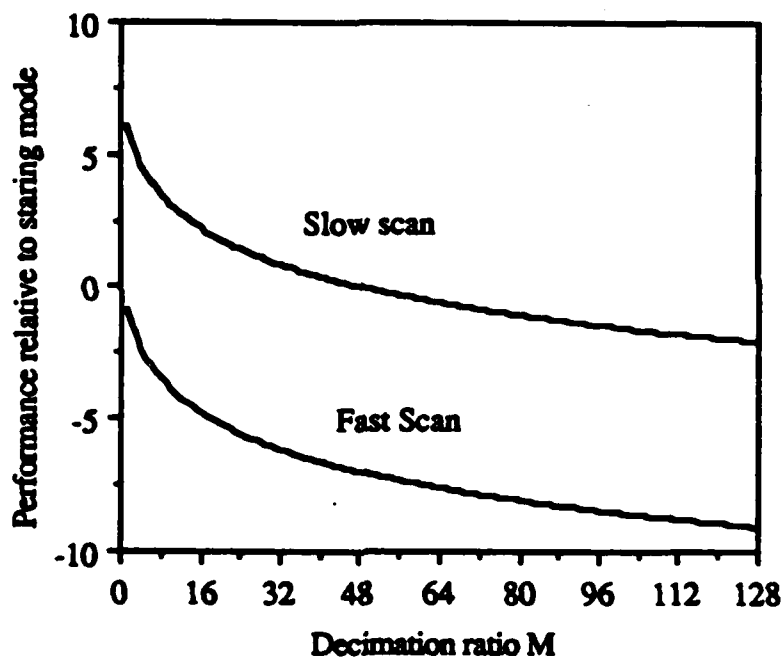


Figure 5. Performance relative to staring mode for cross-spectrum analyzer

range decreases at a slower rate for the fast scan mode. The loss in dynamic range is a factor of two less in decibels for the cross-spectrum analyzer versus the heterodyne spectrum analyzer. The reason is that the system loss due to scanning is allocated between the two input signals in the cross-spectrum analyzer so that a total system loss of 6dB results in a dynamic range loss of only 3dB.

Finally, the slow scan mode results are 7dB better than the fast scan mode because

the post detection bandwidth can be reduced by virtue of the heterodyne frequency being exactly at  $f_d$  at all spatial frequencies. These results show that rather large savings can be made in the photodetector circuitry for more modest penalties in performance. For  $M=32$ , we need only 64 photodetector elements instead of 2000. We can equal the staring mode performance if the laser power is increased by 6dB; of course, the dynamic range will be reduced by 6dB if we are already laser power limited. A 6dB loss in dynamic range may, however, be a good trade for a 32-fold reduction in the complexity of the readout circuitry.

## 5. CONCLUSIONS

The use of a decimated array in heterodyne spectrum analysis reduces the complexity of the photodetector readout electronics with a reduction in the dynamic range that is often tolerable. We introduced the use temporal heterodyning techniques to determine the angle of arrival of signals received by two antenna elements. The decimated array technique is applicable here as well, noting that the dynamic range drops only half as fast as the decimation rate because the loss is shared by both signals.

## 6. ACKNOWLEDGEMENT

This work was performed with support from the U.S. Army Research Office.

## 7. REFERENCES

1. A. VanderLugt, "Interferometric Spectrum Analyzer," Appl. Opt., Vol. 20, p. 2770 (1981)
2. A. VanderLugt and A.M. Bardos, "Spatial and Temporal Spectra of Periodic Functions for Spectrum Analysis," Appl. Opt., Vol. 23, p. 4269 (1984)
3. A. VanderLugt, "Use of Decimated Photodetector Arrays for Spectrum Analysis, Appl. Opt., Vol. 27, p. xxxx, May (1988)



---

**APPENDIX C**

**TEMPORAL FREQUENCIES OF SHORT AND EVOLVING PULSES IN  
INTERFEROMETRIC SPECTRUM ANALYZERS**

**Reprinted from Applied Optics**

**Volume 28, Pages 3800-3809, 15 September 1989**

---

# Temporal frequencies of short and evolving pulses in interferometric spectrum analyzers

P. H. Wisseman and A. VanderLugt

The output of an acoustooptic spectrum analyzer contains both spatial and temporal frequencies. In a previous paper, VanderLugt predicted a difference in the temporal frequency as a function of spatial frequency for a cw signal, a pulse evolving into the acoustooptic cell, and an isolated pulse moving through the cell. We provide a more complete analysis for each of the three input signals. An analysis is made of the effects of sidelobe movement in the evolving pulse signal. This analysis, combined with a mixed transform analysis, leads to an accurate prediction of the temporal frequency behavior and to new interpretations of the results presented in the original paper. In particular, our analysis accurately predicts the temporal frequencies associated with the evolving pulse. Experimental results validate the mathematical analysis for all three signal types.

## I. Introduction

For spectrum analysis of real-time signals we often use an acoustooptic cell to convert an electrical input signal into an optical signal that is a function of both space and time. When illuminated by coherent light, the optical system produces the Fourier transform of the input signal as a spatial distribution of light; each point at the output plane of the system corresponds to a specific spatial frequency. Furthermore, the temporal frequencies that characterize the input signal are retained at the output of the system. In a previous paper, VanderLugt analyzed the spectra resulting from three signals: a cw input signal; a pulse evolving into the acoustooptic cell; and an isolated pulse moving through the cell.<sup>1</sup> His analysis shows that these spectra differ in their temporal frequency as a function of which spatial frequency is detected, but those results were not confirmed experimentally. In this paper we present a more complete analysis of the relationship between the spatial and temporal frequencies for each of the three input signals by using a mixed transform. In addition, we analyze the effects of sidelobe move-

ment for the evolving pulse. We then experimentally confirm the results of the analysis for each signal.

In Sec. II we review the analysis of the spatial spectrum resulting from each of the three input signals. In Sec. III we discuss the effects of sidelobe movement for the evolving pulse and use a mixed transform to obtain a concise relationship between the spatial and temporal frequencies for each of the input signals. In Sec. IV we give an analysis of the baseband terms of the photodetector output which determine the requirements of the postdetection electronics used in the experiments. In Sec. V we describe the design of the optical and electronic systems used in the experiment. The experimental procedure and results are given in Sec. VI. A comparison of the present analyses with those originally put forth is given in Sec. VII, we illustrate this comparison with experimental results.

## II. Background

The relationship between spatial and temporal frequencies in an interferometric spectrum analyzer, as shown in Fig. 1, is analyzed in Ref. 1. Here we summarize the main results to serve as the basis for new analyses. A signal  $f(t) = s(t) \cos(2\pi f_c t)$  is applied to the acoustooptic cell in plane  $P_2$  so that the light amplitude leaving the cell in the direction of the first positive order is

$$A(x,t) = s(t - x/v) \exp[j2\pi f_c(t - x/v)], \quad (1)$$

where  $f_c$  is the center frequency of the acoustooptic cell and  $v$  is the velocity of the acoustic wave. We neglect unimportant amplitude and phase terms throughout

Both authors were with North Carolina State University, Electrical & Computer Engineering Department, Box 7911, Raleigh, North Carolina 27696; P. H. Wisseman is now at TRW, Advanced Technology Department, Redondo Beach, California 90278.

Received 19 December 1988.

0003-6935/89/133800-10\$02.00/0.

© 1989 Optical Society of America.

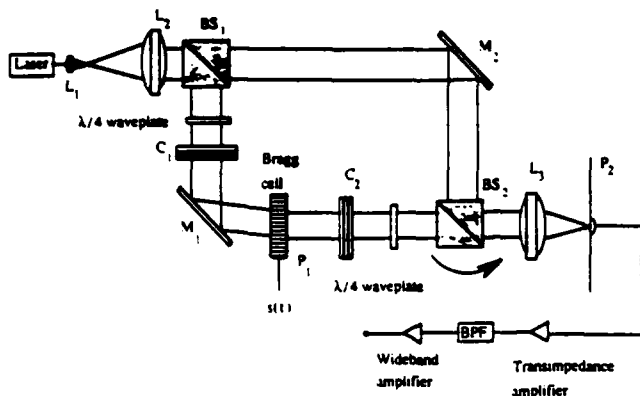


Fig. 1. Interferometric spectrum analyzer.

this analysis. Lens  $L_2$  performs a Fourier transform on the light leaving the cell, and the resulting amplitude distribution in plane  $P_2$  is given by the spatial Fourier transform

$$S(\alpha, t) = \int_{-L/2}^{L/2} s(t - x/v) \exp[j2\pi f_c(t - x/v)] \exp[j2\pi\alpha x] dx, \quad (2)$$

where  $L$  is the length of the acoustooptic cell and  $\alpha$  is the spatial frequency. We make the change of variables  $u = (t - x/v)$ , so that Eq. (2) becomes

$$S(\alpha, t) = v \exp(j2\pi\alpha vt) \int_{t-T}^t s(u) \exp[j2\pi(\alpha - \alpha_c)vu] du, \quad (3)$$

where  $T = L/v$  is the equivalent time duration of the acoustooptic cell, and we use the relationship  $f_c = \alpha_c v$  to produce Eq. (3). The amplitude distribution for a given input signal is obtained using either Eq. (2) or (3) depending on which is more convenient.

We consider three signals that demonstrate different relationships between the spatial and temporal frequencies. First, consider the signal for which  $s(t) = 1$ . This input is a cw signal of frequency  $f_c$  and unit amplitude. We use Eq. (3) to get

$$\begin{aligned} S(\alpha, t) &= v \exp(j2\pi\alpha vt) \int_{t-T}^t \exp[j2\pi(\alpha - \alpha_c)vu] du \\ &= L \exp(j2\pi f_c t) \text{sinc}[(\alpha - \alpha_c)L]. \end{aligned} \quad (4)$$

This result indicates that the spatial spectrum is a sinc function centered at  $\alpha_c$  with the amplitude proportional to the length of the cell. In addition, the entire spectrum has temporal frequency  $f_c$ , independently of the spatial frequency position in plane  $P_2$ .

The second signal is a pulse moving into the cell as shown in Fig. 2. This pulse has duration  $T_0 \gg T$  and carrier frequency  $f_c$ . If the time of arrival of the leading edge of the pulse is at  $t = 0$ , the leading edge moves to  $x = (-L/2 + vt)$  at time  $0 < t \leq T$ . If the pulse has a unit amplitude, Eq. (2) becomes

$$\begin{aligned} S(\alpha, t) &= \int_{-L/2}^{-L/2+vt} \exp[j2\pi f_c(t - x/v)] \exp[j2\pi\alpha x] dx \\ &= \exp[j2\pi(\alpha + \alpha_c)vt/2] \\ &\quad \times \text{sinc}[(\alpha - \alpha_c)vt]; \quad 0 \leq t \leq T. \end{aligned} \quad (5)$$

As for the cw signal, we find that the spatial spectrum

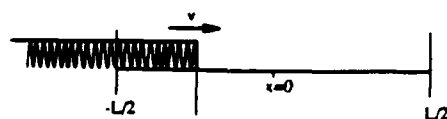


Fig. 2. Evolving pulse input signal.

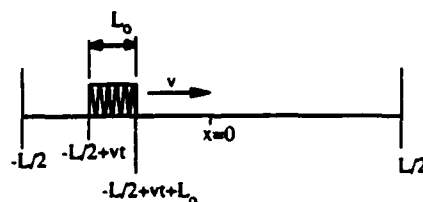


Fig. 3. Short pulse input signal.

is a sinc function centered at  $\alpha_c$ . However, the amplitude of the sinc function increases linearly as  $t$  increases, while its width decreases. The most interesting result is the relationship between the temporal and spatial frequencies. Whereas the temporal frequency is  $f_c$  over the entire spectrum for the cw signal, the temporal frequency is  $f_c$  only at the centroid of the sinc function (at  $\alpha = \alpha_c$ ) for the evolving pulse. In the sidelobes of the sinc function, the temporal frequency is equal to  $f = (\alpha + \alpha_c)v/2$ , which varies continuously with  $\alpha$ . According to this result, at a spatial frequency  $\alpha_c + \Delta\alpha$  the corresponding temporal frequency is  $f_c + \Delta f/2$ .

The third signal is a short pulse of duration  $T_0 \ll T$  and carrier frequency  $f_c$  that is completely within the acoustooptic cell. This signal, illustrated in Fig. 3, exists for the time period  $0 < t < (T - T_0)$ . Using Eq. (2), we find that the amplitude distribution in plane  $P_2$  is

$$\begin{aligned} S(\alpha, t) &= \int_{-L/2+vt}^{-L/2+vt+L_0} \exp[j2\pi f_c(t - x/v)] \exp[j2\pi\alpha x] dx \\ &= L_0 \exp(j2\pi\alpha vt) \text{sinc}[(\alpha - \alpha_c)L_0]; \quad T_0 \leq t \leq D(T - T_0). \end{aligned} \quad (6)$$

The spectrum of a short pulse is a sinc function, again centered at  $\alpha_c$ , whose amplitude and width are proportional to the pulse length  $L_0$  but independent of time. The temporal frequency for this signal is simply  $f = \alpha v$ . Therefore, at  $\alpha_c$  the temporal frequency is  $f_c$ , and at  $\alpha_c + \Delta\alpha$  the corresponding temporal frequency is  $f_c + \Delta f$ .

The results for these signals show interesting relationships between the spatial and temporal frequencies in the sidelobes of the spectra. For the cw signal, the temporal frequency of the sidelobes does not differ from that at the centroid of the spectrum. However, for the evolving pulse the change in the temporal frequency is  $\Delta f/2$  for a spatial frequency located  $\Delta\alpha$  from the centroid, and for a short pulse the change in temporal frequency is  $\Delta f$ . These results are explained in part by considering the motion of the midpoint of the pulse. For the cw signal, the velocity of the midpoint is zero because the signal always fills the cell. For the evolving pulse, the velocity of the midpoint is equal to  $v/2$ ,

while the velocity of the midpoint is equal to  $v$  for the short pulse.

The temporal frequencies of these spatial spectra are measured by heterodyne detection. This is achieved through an interferometric process between a reference probe and the spectrum of the signal.<sup>1</sup> In heterodyne detection, we retain the cross-product term from the intensity produced by the sum of the reference and signal beams:

$$I(\alpha, t) = |R(\alpha) + S(\alpha, t)|^2 \\ = |R(\alpha)|^2 + |S(\alpha, t)|^2 + 2 \operatorname{Re}[R(\alpha)S^*(\alpha, t)], \quad (7)$$

where  $R(\alpha)$  is a sinc function representing the reference probe and  $S(\alpha, t)$  is the spectrum due to the signal beam. The first two terms of Eq. (7) are at baseband. These terms are removed by bandpass filtering at the output of the photodetector. The third term is the cross-product term of interest and contains the temporal frequency information associated with the input signal.

### III. Sidelobe Movement and the Mixed Transform

The relationships between the temporal and spatial frequencies are based on deducing the temporal frequencies from the exponentials associated with the spatial transforms given the implicit requirement to take the real values of Eqs. (4), (5), and (6) in the detection process. These deductions raise some interesting issues. We expect that the temporal frequency content of the evolving pulse, after it transitions into a cw signal at  $t = T$ , would be similar to that of a cw signal. How, then, does the temporal frequency for a probe position at a particular spatial frequency shift from  $f_c = \Delta f/2$  to  $f_c$  during the transition? Furthermore, a short pulse behaves as an evolving pulse until its trailing edge enters the cell. Again, how does the frequency shift from  $f_c + \Delta f/2$  to  $f_c + \Delta f$  during the transition? To gain insight into these issues, we begin by considering the amplitude modulation induced by the sidelobe structure of the evolving pulse. Lee and Wight<sup>2</sup> provide an alternative analysis of the amplitude modulation for pulsed signals in a power spectrum analyzer.

The results in Eq. (5) indicate that, for a pulse moving into the acoustooptic cell, the sinc function at the output plane begins with a broad central lobe spread over the entire spatial frequency range and becomes narrower as the pulse fills the cell. If the probe is placed a considerable distance from  $\alpha_c$  outside the central lobe of the sinc function when the pulse has filled the cell, the amplitude modulation due to the movement of the sidelobes past the probe introduces frequency components at the output in the form of amplitude modulation.

Figure 4 shows the sinc function due to the evolving pulse as the pulse is moving into the cell and when the pulse has filled the cell. The sinc function is centered at  $\xi_c$ , corresponding to the spatial frequency  $\alpha_c$ , where  $\xi = \alpha \lambda F$ . From Eq. (5) we find that the distance between the nulls of the sinc function at some time  $t$  is

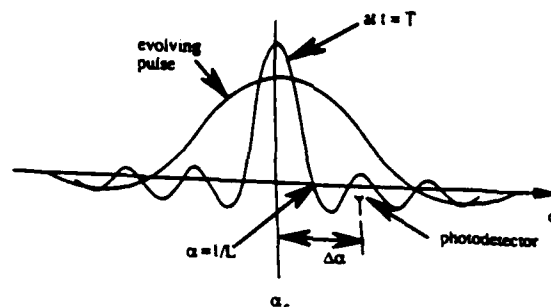


Fig. 4. Amplitude of the spectrum of the evolving pulse at two time instances.

given by  $\xi_0 = \lambda F/ut$ , where  $0 \leq t \leq T$ . We associate  $\xi_0$  with the wavelength  $\lambda_s$  of the sidelobe structure. We express the location of a probe located a distance  $\xi_p = \Delta \alpha \lambda F$  away from  $\xi_c$  in terms of the number of sidelobe wavelengths:

$$n = \frac{\xi_p}{\lambda_s} = \frac{(\Delta \alpha) \lambda F}{2 \lambda F/ut} = (\Delta \alpha) ut/2. \quad (8)$$

The instantaneous velocity of the sinc function at the probe is

$$v_s = \frac{d(2n\xi_0)}{dt} = \frac{-2n\lambda F}{ut^2}. \quad (9)$$

We substitute Eq. (8) into Eq. (9) to find that  $v_s = \Delta \alpha v \xi_0$ . The instantaneous frequency due to the amplitude modulation is the ratio of the velocity of the sinc function at the probe divided by the wavelength of the sinc function:

$$f = \frac{v_s}{\lambda_s} = \frac{(\Delta \alpha v) \xi_0}{2 \xi_0} = \frac{\Delta \alpha v}{2}. \quad (10)$$

This result shows that the frequency of the amplitude modulation is independent of time. It is not a function of the length of the evolving pulse. The spectrum given by Eq. (5) changes so that as the velocity of a null decreases the distance between nulls decreases by exactly the same factor to produce a fixed frequency. As shown by Eq. (10), the frequency introduced by the amplitude modulation is a function of the position of the probe.

This analysis has been largely intuitive, but the results have important implications. If the amplitude modulation interacts multiplicatively with the  $f_c + \Delta f/2$  temporal frequency term found from Eq. (5), the result is a carrier frequency at  $f_c + \Delta f/2$ , amplitude modulated by a frequency at  $\Delta f/2$ . We have made the association from Eq. (10) that  $\Delta f = \Delta \alpha v$ . An equivalent interpretation is that the signal results from the sum and difference frequencies at  $f = f_c$  and at  $f = f_c + \Delta f$ . This interpretation suggests that there are two frequency components at the output; one at the carrier frequency and one dependent on the position of the probe. These frequency components are also found in the other two spectra discussed. The cw signal produces a frequency  $f = f_c$  for all positions of the probe in the output plane, and the short pulse signal produces a

frequency  $f = f_c + \Delta f$  that is dependent on the probe position. Our second interpretation suggests that both frequency components are present while the evolving pulse is moving into the cell. After the pulse completely fills the cell to become a cw signal, only the  $f = f_c$  frequency component remains. On the other hand, when the evolving pulse terminates in a trailing edge to create a short pulse, only the  $f = f_c + \Delta f$  component is present.

This analysis provides insight into the transition between the three signals. However, to find a more precise solution that gives both the spatial and temporal frequency content of a signal, we use the mixed transform. The mixed transform is the Fourier transform of the input function in terms of both spatial and temporal frequencies<sup>3</sup>:

$$S(\alpha, f) = \iint_{-\infty}^{\infty} s(x, t) \exp[j2\pi(\alpha x - ft)] dx dt. \quad (11)$$

The mixed transform reveals the temporal frequency content at any position of a probe in the spatial frequency plane. We use the mixed transform to support the observations made in the preceding sections about the temporal frequency content for each of the three input signals.

Since the spatial frequency transform for each input signal was found in Sec. II, we find the mixed transform by completing the temporal part of the transform:

$$S(\alpha, f) = \int_{-\infty}^{\infty} S(\alpha, t) \exp[-j2\pi ft] dt. \quad (12)$$

For the cw input signal, the spatial frequency transform as given by Eq. (4) is used in Eq. (12) so that the mixed transform of the cw signal is

$$\begin{aligned} S(\alpha, f) &= L \operatorname{sinc}[(\alpha - \alpha_c)L] \int_{-\infty}^{\infty} \exp[-j2\pi(f - \alpha_c\nu)t] dt \\ &= LT \operatorname{sinc}[(\alpha - \alpha_c)L] \delta(f - \alpha_c\nu). \end{aligned} \quad (13)$$

This result shows that the temporal frequency consists of a  $\delta$ -function centered at  $f = \alpha_c\nu = f_c$  independently of the probe position and agrees with our observations in Sec. II. The  $\operatorname{sinc}[(\alpha - \alpha_c)L]$  function shows that the finite extent  $L$  of the spatial aperture causes spectral spreading in the spatial frequency domain, whereas  $\delta(f - \alpha_c\nu)$  shows that the temporal frequency content of the signal is pure.

For a short pulse signal, the spatial frequency transform is given by Eq. (6), and the mixed transform becomes

$$\begin{aligned} S(\alpha, f) &= L_0 \operatorname{sinc}[(\alpha - \alpha_c)L_0] \int_0^T \exp[-j2\pi(f - \alpha\nu)t] dt \\ &= L_0 T \operatorname{sinc}[(\alpha - \alpha_c)L_0] \operatorname{sinc}[(f - \alpha\nu)T]. \end{aligned} \quad (14)$$

As with the cw signal, a sinc function characterizes the spatial frequency distribution due to the short pulse length. In this case, however, the temporal frequencies are also distributed according to a sinc function instead of being concentrated entirely at  $f = 0$ , because

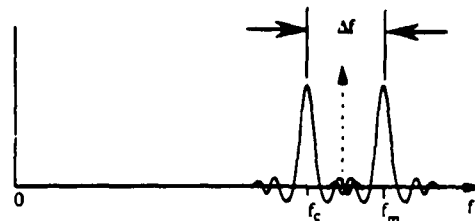


Fig. 5. Mixed transform of the evolving pulse.

the pulse also has a finite extent in time. For a probe at a spatial frequency corresponding to  $\alpha = \alpha_c + \Delta\alpha$ , the mixed transform shows a sinc function centered at the temporal frequency  $f = f_c + \Delta f$ , which agrees with observations made in Sec. II.

The spatial frequency transform for the evolving pulse is given by Eq. (5). The mixed transform becomes

$$\begin{aligned} S(\alpha, f) &= \int_0^T \nu t \exp[j\pi(\alpha + \alpha_c)\nu t] \\ &\quad \times \operatorname{sinc}[(\alpha - \alpha_c)\nu t] \exp[-j2\pi ft] dt \\ &= \frac{T}{\alpha - \alpha_c} [\operatorname{sinc}[(f - \alpha\nu)T] \\ &\quad - \exp[j\phi(\alpha, f)] \operatorname{sinc}[(f - \alpha_c\nu)T]], \end{aligned} \quad (15)$$

where  $\phi(\alpha, f)$  is a phase term. By a series expansion, it is easy to show that the result is finite for  $\alpha = \alpha_c$ . For a probe at  $\alpha = \alpha_c + \Delta\alpha$ , the result shows two temporal frequency components, one at  $f = f_c$  and one at  $f = f_c + \Delta f$  as shown in Fig. 5. This result supports the observations in which we accounted for the amplitude modulation due to the sidelobe motion as well as the frequency induced by the motion of the midpoint of the pulse. In addition to the strong frequency concentration at  $f_c$  and  $f_c + \Delta f$ , we see that weaker frequencies are generated by the sidelobe structure. We expect to observe these frequencies between the two dominant frequencies where the sidelobes of the two frequencies overlap, with their strength dependent on the phase relationship between the two terms of Eq. (15).<sup>4</sup>

#### IV. Baseband Terms

The intensity at the output, given by Eq. (7), contains three terms. In the preceding section we limited our discussion to the cross-product term, which contains the information of interest. In this section, we determine the temporal frequency content of the other two terms to validate our assertion that they are at baseband and can be removed by bandpass filtering. For this analysis we apply the mixed transform to the  $|R(\alpha)|^2$  and  $|S(\alpha, t)|^2$  terms. Again, we examine the three cases of a cw signal, a short pulse, and a pulse evolving into the cell. For all three signals the term  $|R(\alpha)|^2$  due to the reference beam has zero temporal frequency content since the reference beam is a plane wave that is not time dependent. Therefore, we focus on the temporal frequency content of  $|S(\alpha, t)|^2$ .

For the cw input signal, the spatial frequency dependence at the output is given by Eq. (4). The mixed transform of the  $|S(\alpha, t)|^2$  term becomes

$$\begin{aligned} |S(\alpha, f)|^2 &= \int_{-\infty}^{\infty} |S(\alpha, t)|^2 \exp(-j2\pi ft) dt \\ &= L^2 \operatorname{sinc}^2[(\alpha - \alpha_c)L] \int_{-\infty}^{\infty} \exp(-j2\pi ft) dt \\ &= L^2 \operatorname{sinc}^2[(\alpha - \alpha_c)L] \delta(f). \end{aligned} \quad (16)$$

The temporal frequency content of this term is a  $\delta$ -function at  $f = 0$ , which confirms that this term is at baseband. The spatial spectrum for a short pulse is given by Eq. (6), and the mixed transform is

$$\begin{aligned} |S(\alpha, f)|^2 &= \int_{-\infty}^{\infty} |S(\alpha, t)|^2 \exp(-j2\pi ft) dt \\ &= L_0^2 \operatorname{sinc}^2[(\alpha - \alpha_c)L_0] \int_0^T \exp(-j2\pi ft) dt \\ &= L_0^2 \operatorname{sinc}^2[(\alpha - \alpha_c)L_0] \operatorname{sinc}(fT), \end{aligned} \quad (17)$$

where we recognize that  $S(\alpha, t)$  exists only for a time interval of approximately  $T$  seconds. This result shows that most temporal frequencies are concentrated in a region  $f \leq 1/T$ , confirming that this term is also at baseband.

For the evolving pulse the spatial frequency dependence at the output is given by Eq. (5). The resulting mixed transform of the  $|S(\alpha, t)|^2$  term is

$$\begin{aligned} |S(\alpha, f)|^2 &= \int_{-\infty}^{\infty} |S(\alpha, t)|^2 \exp(-j2\pi ft) dt \\ &= \int_0^T (\nu t)^2 \operatorname{sinc}^2[(\alpha - \alpha_c)\nu t] \exp(-j2\pi ft) dt \\ &= \frac{T}{(\alpha - \alpha_c)^2} \left\{ \begin{array}{l} 2 \operatorname{sinc}(fT) - \exp(-j\phi) \operatorname{sinc}(f - \alpha\nu + \alpha_c)T \\ - \exp(-j\theta) \operatorname{sinc}(f + \alpha\nu - \alpha_c)T \end{array} \right\}. \end{aligned} \quad (18)$$

For a probe located at a spatial frequency  $\alpha = \alpha_c + \Delta\alpha$ , we have

$$|S(\alpha_c + \Delta\alpha, f)|^2 = \frac{T}{(\Delta\alpha)^2} \left\{ \begin{array}{l} 2 \operatorname{sinc}(fT) - \exp[-j\pi(f - \Delta\alpha)T] \operatorname{sinc}(f - \Delta\alpha\nu)T \\ - \exp[-j\pi(f + \Delta\alpha\nu)T] \operatorname{sinc}(f + \Delta\alpha\nu)T \end{array} \right\}, \quad (19)$$

which reveals that the temporal spectrum consists of sinc functions centered at  $f = 0$  and  $f = \pm\Delta f$ , where  $\Delta f = \Delta\alpha\nu$ . Again, by using a series expansion, we can show that the result is finite at  $\Delta\alpha = 0$ . These frequency components are at baseband, but the frequency spread is greater than in the other cases and depends on the position of the probe. The bandpass filter is designed for the worst case, which occurs when the probe is located at a position corresponding to the cutoff frequency of the acoustooptic cell. For an acoustooptic cell of bandwidth  $W$  centered at  $f_c$ , the maximum value of  $\Delta f$  is  $W/2$ . Therefore, the bandpass filter is designed for a passband of bandwidth  $W$ , centered at  $f_c$ , and a stopband for  $f < W/2$ . The bandpass filter must reject these frequency components so that the cross-product term is detected without crosstalk from the baseband terms.

## V. Optical System

The optical system used for this experiment is a modified Mach-Zehnder interferometer shown in Fig. 1. Light from a 6-mW He-Ne laser is expanded by a telecentric lens pair,  $L_1$  and  $L_2$ . Beam splitter  $BS_1$  divides the light into a signal beam, which contains the acoustooptic cell, and a reference beam. Cylindrical lens  $C_1$  focuses light into a horizontal line incident on the acoustooptic cell, while lens  $C_2$  collimates the light leaving the cell. Mirror  $M_1$  is adjusted so that light incident on the cell is at the Bragg angle and that diffracted light is parallel to the reference beam.

Quarterwave plates before and after the cell provide the required circularly polarized light incident on the cell and linearly polarize the light leaving the cell so that it interferes with the reference probe. The reference and signal beams are combined by beam splitter  $BS_2$ . This beam splitter is mounted on a rotating stage so that the position of the reference probe at plane  $P_2$  is adjustable. Finally, lens  $L_3$  provides the Fourier transform of the input signal and focuses the reference beam at the output of the interferometer onto a photodiode.

The acoustooptic cell is a model N45070 from Newport Electro-Optics Systems. This cell has a center frequency of 70 MHz and a 40-MHz bandwidth. The propagation time is  $T = 40 \mu\text{s}$ , with a propagation velocity of  $\nu = 617 \text{ m/s}$ . The input signal to the cell is provided by an IntraAction model DE-70M driver. This driver contains a voltage-controlled oscillator with an adjustable center frequency and a mixer for amplitude modulating the carrier by an external signal source. An HP model 8018A data generator is used as the signal source.

The output of the optical system is detected by a RCA C30831 PIN photodiode. After detection, the unwanted baseband terms are removed by a bandpass

filter, and the desired signal is amplified for measurement with an oscilloscope and spectrum analyzer. The transimpedance amplifier is a Signetics SE5212 whose nominal bandwidth is 100 MHz. However, the bandwidth depends on the capacitance of the photodetector and the values of the input resistance and capacitance of the device. The wideband amplifier is a Signetics NE5539 ultrahigh frequency operational amplifier. The maximum gain for the 90-MHz bandwidth required in our experiments is  $\sim 13 \text{ dB}$ , which is sufficient to provide a signal to the oscilloscope or spectrum analyzer for measurement.

The dynamic range is  $\sim 70 \text{ dB}$  in a 10-kHz bandwidth, which is equivalent to 34 dB at the full system bandwidth of 40 MHz. The measured dynamic range is less than the theoretical value due to aberration in

the system, especially in the cylindrical lenses, and rf noise in the circuit that could not be eliminated.

## VI. Experimental Results

To confirm our theoretical analysis, we want to observe the movement of the sidelobes of the spectrum for the evolving pulse. In addition to providing a known pulse duration within the cell, the aperture before the acoustooptic cell sharply truncates the incident beam and thereby produces large detectable sidelobes. We drive the cell with a cw signal and center the aperture on the Gaussian beam by checking that the nulls and peaks of the sidelobes are symmetrical. With an input signal consisting of 1- $\mu$ s pulses on a 33- $\mu$ s repetition period, the aperture is adjusted until the output of the photodetector has a 10- $\mu$ s duration, which corresponds to an aperture width of 6.17 mm.

We rotate the final beam splitter to move the reference probe to different positions in the spectrum so that we can measure the temporal frequency as a function of spatial frequency. Recall that the temporal frequency of a cw input signal is related to the spatial frequency in the Fourier plane by  $f = \alpha v$ . We begin by applying a cw input signal to the acoustooptic cell so that light is diffracted to the desired probe position. For example, to move the probe to a position corresponding to  $\alpha = \alpha_c + \Delta\alpha$ , as shown in Fig. 6, we set the temporal frequency of the input signal at  $f = f_c + \Delta f$ . We then adjust the probe to maximize the output of the photodetector. We estimate that we position the probe to within  $\pm 2 \mu\text{m}$  of the desired position, which corresponds to a  $\pm 17$ -kHz frequency uncertainty.

We calibrated the system at positions corresponding to the centers of the first three sidelobes and a position midway between the centroid and the first null for each of the pulse widths analyzed. The results of these calculations are shown in Table I. In this table,  $T_0$  is the duration of a pulse in the input signal,  $L_0 = vT_0$  is the corresponding length of the pulse in the acoustooptic cell, and  $\Delta\xi$  is the distance of the desired position of the probe from the centroid of the spectrum. The value  $\Delta f$  is the change in the frequency of the cw signal required to overlap the probe at  $\Delta\xi$ .

Since the maximum pulse duration is limited by the aperture, the value obtained for the 10- $\mu$ s pulse also apply for any pulses longer than 10  $\mu$ s, including the cw signal. For example, the third sidelobe for a 1- $\mu$ s pulse input signal is a distance  $\Delta\xi = 538.4 \mu\text{m}$  from the centroid of the spectrum. This corresponds to a temporal frequency of  $\Delta f = 3.5$  MHz. The frequency of a cw calibration signal required to position the probe at this position if  $f = f_c - \Delta f = 66.5$  MHz, where we use a 70-MHz center frequency and chose to position the probe on the low frequency side of the midband frequency.

The first spectrum of interest is that of a cw signal. The input signal is tuned to 70-MHz, and the position of the reference probe is adjusted to maximize the output of the photodetector. The output is observed on both an oscilloscope and an electronic spectrum analyzer, and the frequency and amplitude are mea-

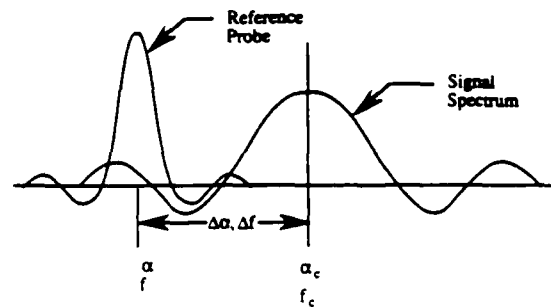


Fig. 6. Position of the reference probe in the signal spectrum.

sured for each of the positions in the spectrum given in Table I.

For the pulse input signals, a periodic modulation signal is necessary to provide enough time history of the output signal to ensure accurate frequency measurements by the test spectrum analyzer. We set the repetition period of the modulation signal at 33  $\mu\text{s}$ , which ensures that only one pulse at a time is in the aperture. The use of a periodic modulation signal, however, produces some artifacts in the spectrum analyzer measurements because only the fundamental frequency associated with the repetition rate and its harmonics are observed on the test spectrum analyzer; the continuous temporal spectrum of the pulses is, therefore, sampled at 30-kHz intervals.

The principal factors that affect the accuracy and resolution of the temporal frequencies measured by the test spectrum analyzer are the size of the reference probe, resolution bandwidth of the test spectrum analyzer, fundamental frequency of the periodic input signal, and calibration accuracy of the probe position. Since interference occurs only where the signal spectrum and reference beam probe overlap and are collinear, the effective size of the photodetector is determined by the width of the central lobe of the probe. For illustration, consider a short pulse input signal. Our analysis shows that the output contains all temporal frequencies in the band where the reference probe and signal spectrum overlap. The -3-dB bandwidth corresponding to the half-power width of the probe is, for a 15-mm aperture in the reference beam, equivalent to a 41-MHz band of temporal frequencies. The periodic nature of the input signals produces harmonics of the repetition frequency, which is  $\sim 30$  kHz. The

Table I. Frequency Shift Corresponding to Position in Spatial Frequency Spectrum

Pulse width $T_0$	Corresponding position $L_0$ (mm)	in sinc function	$\Delta\xi$ ( $\mu\text{m}$ )	$\Delta f$ (kHz)
10.0	6.17	Within main lobe	7.7	50.0
		First sidelobe	23.1	150.0
		Second sidelobe	38.5	250.0
		Third sidelobe	53.8	350.0
1.0	0.617	Within main lobe	76.9	500.0
		First sidelobe	230.8	1500
		Second sidelobe	384.6	2500
		Third sidelobe	538.4	3500

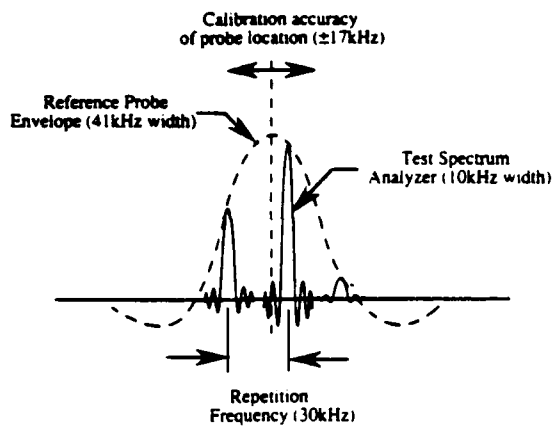


Fig. 7. Frequency resolution components of the measurement system.

temporal frequency spectrum observed by the test spectrum analyzer is, therefore, made up of samples at each harmonic frequency. The width of these samples is equal to the resolution bandwidth of the test spectrum analyzer, which is set at 10 kHz. The combination of all of these factors is shown in Fig. 7. The overall frequency measurement accuracy due to the  $\pm 17$ -kHz uncertainty in the probe position and the  $\pm 15$ -kHz uncertainty due to the sampling spacing is approximately  $\pm 23$  kHz.

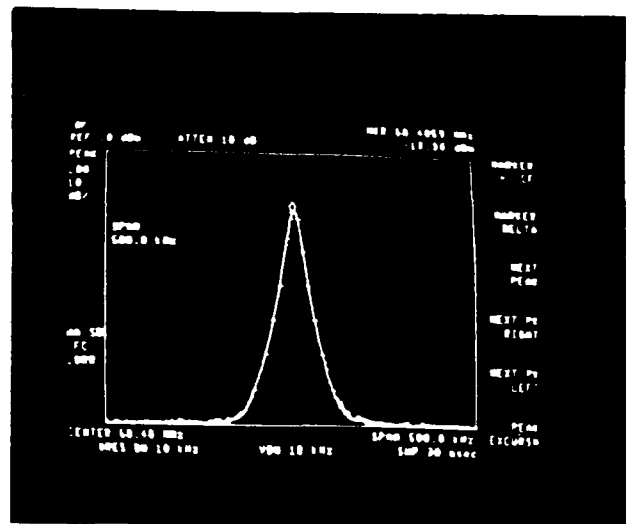
We begin with the experimental results for the cw signal. The spectrum of a purely evolving pulse is measured using a 10- $\mu$ s pulse. A 1- $\mu$ s pulse is used for the short pulse signal. For each pulse width, the frequency and amplitude at each position of the probe, as shown in Table I, are measured using the same method used for the cw signal. As new phenomena are observed, we introduce the experimental evidence. Where a particular phenomenon is common to all signals, we present it only in the first instance and discuss its significance.

#### A. Continuous Wave

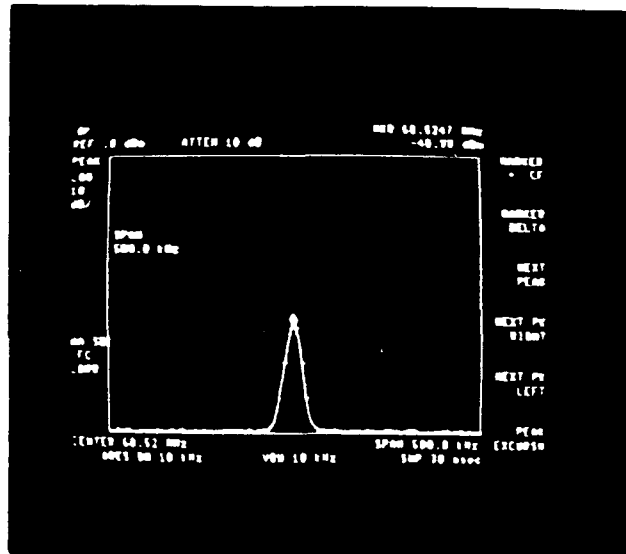
The results of the temporal frequency measurements for the cw signal are given in Table II. In this table, the quantity  $\Delta f = f_c = \Delta \alpha v$  is the difference between the carrier frequency and the frequency corresponding to the position of the reference probe. This calculated value is related to the spatial frequency of the spectrum and is expressed in terms of the corresponding temporal frequency so that it can be compared with the measured temporal frequency of the spectrum. The quantity  $\delta f$  is defined as  $\delta f = f_c - f_m$ , where  $f_m$  is the temporal frequency measured by the

Table II. Measured Values for cw Input Signal

$\Delta \alpha$ (mm)	$\Delta f$ (kHz)	$\delta f$ (kHz)	Power (dB)
0.0	0.0	0.0	46.1
0.078	48.0	(-2.0)	41.1
0.176	109.0	(9.0)	38.0
0.254	157.0	(7.0)	29.0
0.321	198.0	(-2.0)	29.7
0.402	248.0	(2.0)	21.5
0.572	353.0	(7.0)	23.4



(a)



(b)

Fig. 8. Spectrum of the cw signal: reference probe at (a)  $\Delta f = 0.0$  Hz and (b)  $\Delta f = 353$  kHz.

test spectrum analyzer. The values of  $\delta f$  in parentheses are equal to zero to within the accuracy of the test instrumentation, as discussed above. Also included in the table is the power of the output signal above the noise floor, measured using a 10-kHz resolution bandwidth. The small difference between the measured frequency and carrier frequency is within the 10-kHz resolution bandwidth of the spectrum analyzer. These results agree with the results predicted by Eq. (4).

Figures 8(a) and (b) show the resulting spectra for probe positions corresponding to  $\Delta f = 0$  and  $\Delta f = 353$  kHz, respectively, to illustrate the test data. For both probe positions we observe a peak, at the carrier frequency, whose width is determined by the resolution bandwidth of the test spectrum analyzer. The results from Table II for a cw input signal show that the



temporal frequency is equal to the carrier frequency for all positions of the reference probe.

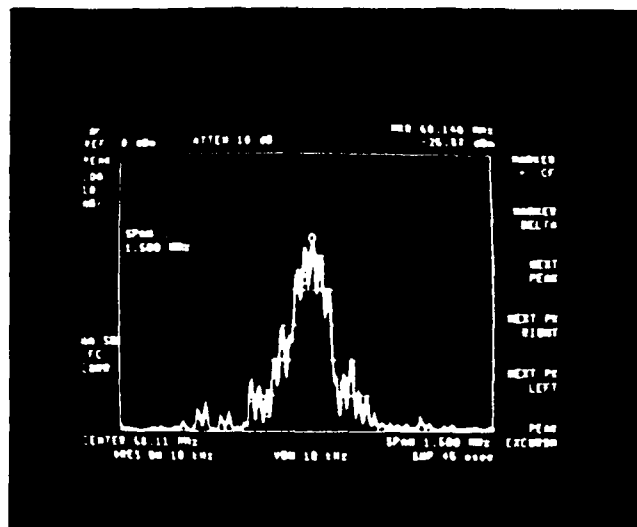
### B. The 10- $\mu$ s Pulse

The 10- $\mu$ s pulse represents a pure evolving and de-evolving pulse because the acoustooptic cell aperture is exactly 10- $\mu$ s long. This signal is of particular importance in confirming the predictions in Sec. III for the evolving pulse signal. The predicted results are that the temporal frequency consists of a component  $f_c$  at the carrier frequency and a component  $f_c + \Delta f$  at the frequency corresponding to the position of the probe. The expected values for the measured temporal frequencies are  $\delta f = 0$  and  $\delta f = \Delta f$  for each position of the probe. The results of the measurements for this signal are shown in Table III. Note that there are five values for the measured temporal frequency when the reference probe is positioned at  $\Delta f = 154$  kHz. The first two values, at -7 and 23 kHz, and the last two values, at 151 and 181 kHz, clearly illustrate frequency uncertainty resulting from the sample spacing of  $\pm 23$  kHz due to the periodic nature of input signal. Each of these pairs of values is, therefore, two samples of a single peak, indicating that there are peaks at approximately  $\delta f = 0$  and  $\delta f = \Delta f$ , as predicted by our analysis for an evolving pulse. The remaining value is at  $\delta f = 83$  kHz, which is within the 30-kHz system accuracy. This frequency is due to the sidelobes of the sinc functions in the mixed transform of an evolving pulse as given in Eq. (15).

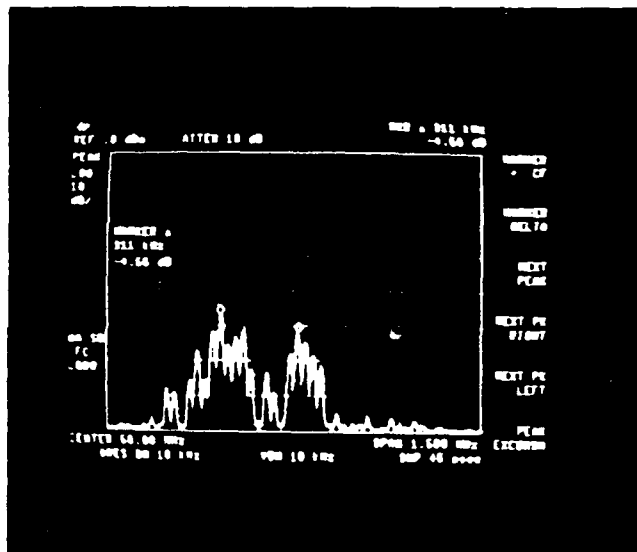
The values obtained at the various probe positions agree with the predicted results within the  $\pm 23$ -kHz accuracy of the test equipment. With the reference probe at the centroid of the spectrum, Fig. 9(a) shows that only the carrier frequency is detected, as predicted by the theory. The two strong frequency components as well as the weaker frequency components are clearly seen in the spectra for a probe position corresponding to  $\Delta f = 345$  kHz, as shown in Fig. 9(b).

### C. The 1- $\mu$ s Pulse

The 1- $\mu$ s pulse input signal is used to measure the spectrum of a short isolated pulse moving through the cell. The predicted results for the spectrum of this



(a)



(b)

Fig. 9. Spectrum of the evolving pulse: reference probe at (a)  $\Delta f = 0.0$  Hz and (b)  $\Delta f = 345$  kHz.

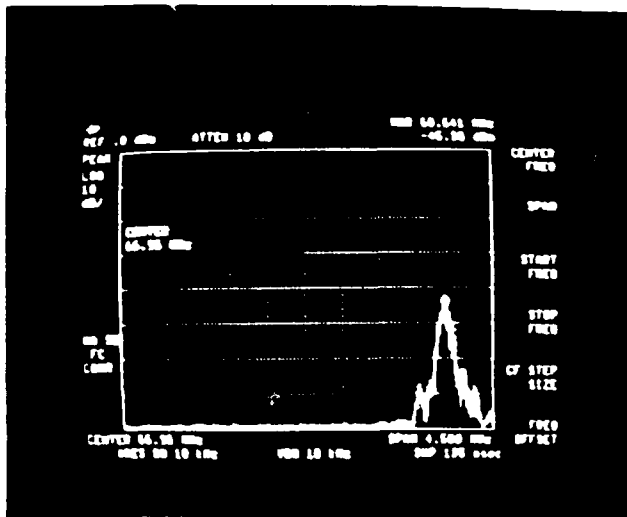
pulse, given by Eq. (6), are that there is only a single temporal frequency at  $f = f_c + \Delta f$ , which corresponds to the position of the reference probe. The measured values for this signal are shown in Table IV. The results in Table IV show that, at each position within the spectrum, the measured temporal frequency corresponds to the spatial frequency of the reference probe position to within the  $\pm 23$ -kHz accuracy of the system.

Table III. Measured Values for 10- $\mu$ s Pulse Input Signal

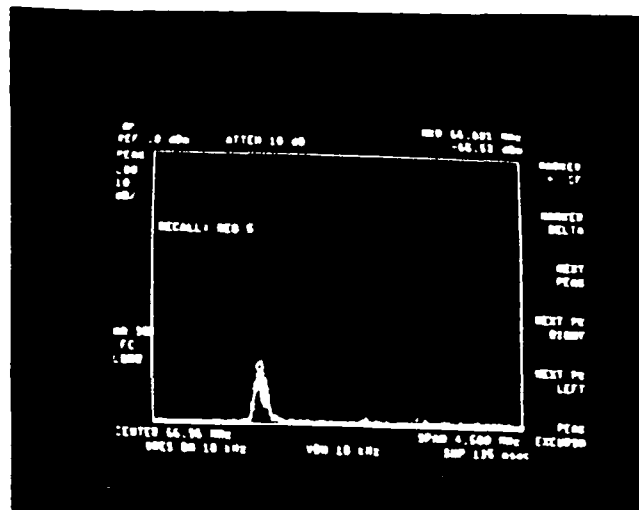
$\Delta\alpha$ (mm)	$\Delta f$ (kHz)	$\delta f$ (kHz)	Power (dB)
0.0	0.0	0.0	35.5
0.081	50.0	(23.0) 53.0	35.7 36.7
0.250	154.0	(-7.0) (23.0) 83.0 151.0 181.0	20.4 19.5 26.5 35.7 38.3
0.408	252.0	(23.0) 285.0	13.5 33.8
0.599	345.0	(17.0) 137.0 369.0	12.3 16.5 32.3

Table IV. Measured Values for 1- $\mu$ s Pulse Input Signal

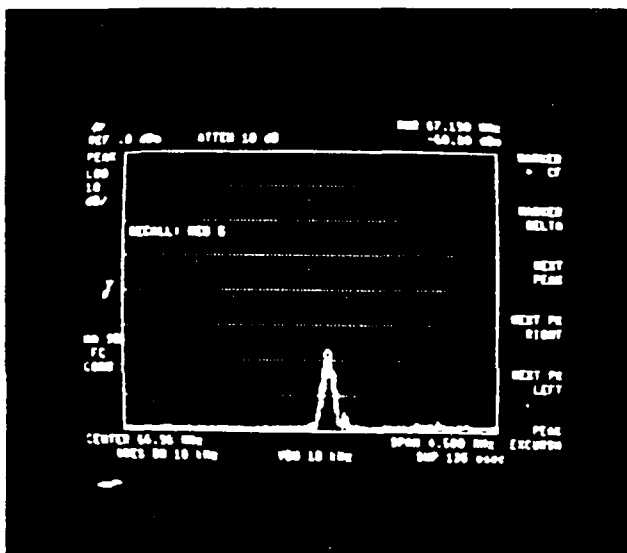
$\Delta\alpha$ (mm)	$\Delta f$ (kHz)	$\delta f$ (kHz)	Power (dB)
0.0	0.0	0.0	35.0
0.812	0.501	0.508	30.5
2.428	1.498	1.488	19.4
4.053	2.501	2.531	14.9
5.666	3.496	3.496	12.1



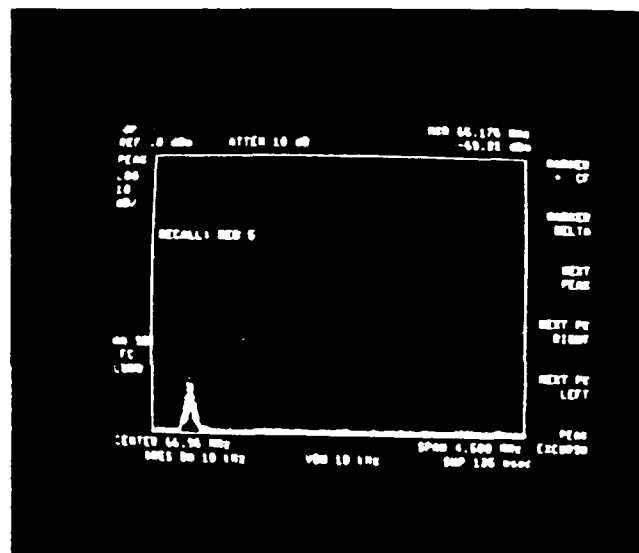
(a)



(c)



(b)



(d)

Fig. 10 Spectrum of the short pulse: reference probe at (a)  $\Delta f = 0.0$  Hz, (b)  $\Delta f = 1.5$  KHz, (c)  $\Delta f = 2.5$  MHz, (d)  $\Delta f = 3.5$  MHz.

The experimental results support the predictions made in our analysis. Figure 10 illustrates the change in the temporal frequency as the reference probe is moved away from the carrier frequency. Figure 10(a) is the spectrum at the centroid of the sinc function in the spatial frequency plane ( $\alpha = 0$ ). Figures 10(b), (c), and (d) show the spectrum when the probe is moved by  $\Delta\alpha = 2.428, 4.053,$  and  $5.666$  cycles/mm. The results show that the temporal frequencies are at  $\Delta f = 1.488, 2.531,$  and  $3.496$  MHz, which confirms that  $\delta f \approx \Delta f = \Delta\alpha v$ . We see that the temporal frequency observed at the output for a short pulse input signal is dependent only on the position of the probe within the spatial frequency spectrum. These results are extremely interesting because the temporal frequency is independent of the carrier frequency of the input signal as long as we satisfy the definition of a short pulse contained within the cell.

## VII. Time Waveform Representation

The analysis and experimental evidence support the fact that the spectrum of an evolving pulse consists of two frequencies spaced  $\Delta f$  apart. We return to the issue of why this temporal frequency behavior was not recognized in Ref. 1. This behavior could have been predicted if the original analysis had been carried forward one more step. For example, if we take the real part of Eq. (5), as is required to complete the analysis of the cross-product term, we have

$$S(\alpha, t) = vt \cos[2\pi(\alpha + \alpha_0)ct/2] \text{sinc}(\alpha + \alpha_0) t, \quad -\infty < t < \infty \quad (21)$$

We expand the sinc function to obtain

$$S(\alpha, t) = \frac{1}{2} [\cos[2\pi(\alpha + \alpha_0)ct/2] \text{sinc}[2\pi(\alpha + \alpha_0)ct/2] + \cos[2\pi(\alpha - \alpha_0)ct/2] \text{sinc}[2\pi(\alpha - \alpha_0)ct/2]] \quad (22)$$

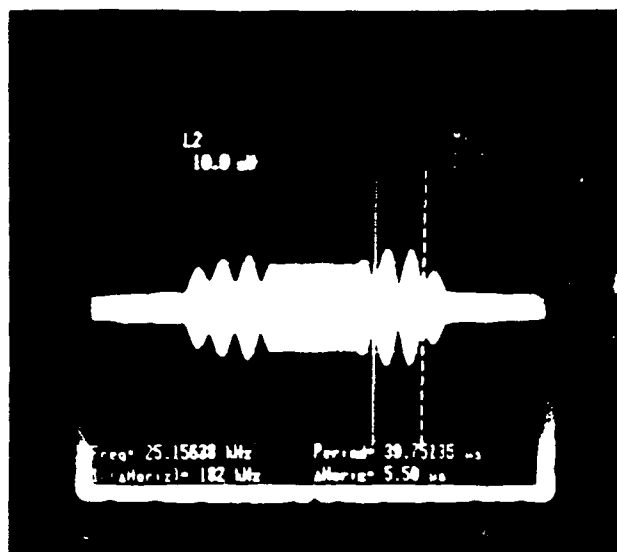


Fig. 11. Oscilloscope trace with probe at  $\Delta f = 345$  kHz for a  $20\text{-}\mu\text{s}$  pulse.

where we again ignore unimportant constants. Consider the situation where the reference probe is located at  $\alpha = \alpha_c + \Delta\alpha$ . We then find that

$$S(\alpha, t) = \frac{1}{\Delta\alpha} \cos[2\pi(\alpha_c + \Delta\alpha/2)vt] \sin[2\pi(\Delta\alpha/2)vt]. \quad (22)$$

This result clearly shows that the output has temporal frequency  $f_c + \Delta f/2$ , as noted before, modulated by a signal whose frequency is  $\Delta f/2$ . Its temporal decomposition is, of course, two frequencies spaced by  $\Delta f$  for the time interval  $0 \leq t \leq T$ .

To illustrate the time waveform represented by Eq. (22), we place the probe outside the central lobe so that we observe the amplitude modulation due to the side-lobe movement on an oscilloscope. We use a  $20\text{-}\mu\text{s}$  pulse signal to show the evolving pulse response for  $10\text{ }\mu\text{s}$ , the transition from an evolving pulse to a cw signal that exists for  $10\text{ }\mu\text{s}$ , and the response of an evolving pulse leaving the cell. The amplitude modulation for each portion of the output signal is seen in Fig. 11 for  $\Delta f = 345$  kHz. The frequency of the amplitude modulation at the beginning and end of the output is measured from the oscilloscope as  $182$  kHz. The expected value of the modulation frequency is  $\Delta f/2 = 172.5$  kHz. The error of the measured value is, therefore,  $5.2\%$ , which is reasonable for such measurements. The oscilloscope traces also show a sharp transition to the cw signal at  $10\text{ }\mu\text{s}$ , when the pulse fills the cell, and then back to the evolving pulse signal  $10\text{ }\mu\text{s}$  later as the trailing edge of the pulse passes through the cell. This result, more graphically than the others, answers the questions

raised earlier about how the temporal frequencies change as the input signal changes its form.

### VIII. Summary and Conclusions

Continuous wave signals are the most common type encountered in spectrum analysis applications. For this signal, the temporal frequency is equal to the frequency of the input signal for all spatial frequencies in the Fourier plane. Therefore, the temporal frequency is independent of the position of the probe. The short pulse and evolving pulse signals are less frequently encountered in spectrum analysis but are becoming more important because they form the basis of exotic signals used in modern communication systems. The temporal frequency content of these signals forms a significant signature which may help in separating them from cw signals.

Of the three signals that we studied, the most interesting is that of an evolving pulse because it forms a bridge between cw and short pulse signals. The analysis in Sec. III, which we confirmed by experiment, shows that the temporal frequency at the output when a pulse is moving into the cell consists of sinc functions at  $f = f_c$  and  $f = f_c + \Delta f$ . If the pulse is longer than the effective period of the acoustooptic cell, it becomes a cw signal when the pulse fills the cell, and all temporal frequencies except that at  $f = f_c$  disappear. If the pulse is not longer than the effective period of the cell, all the temporal frequencies except that at  $f = f_c + \Delta f$  disappear when the trailing edge of the pulse enters the cell. How quickly the frequency components disappear depends on the rise and fall times of the pulse edges.

A short pulse input signal moving through the cell results in an entirely different relationship between the spatial and temporal frequencies at the output. The temporal frequency is exactly equal to the frequency corresponding to the spatial frequency at the position of the probe. Therefore, the temporal frequency is independent of the carrier frequency of the input signal and depends only on the position of the probe in the spatial Fourier transform plane.

This work was supported by the U.S. Army Research Office.

### References

1. A. VanderLugt, "Interferometric Spectrum Analyzer," *Appl. Opt.* **20**, 2770-2779 (1981).
2. J. P. Y. Lee and J. S. Wight, "Acoustooptic Spectrum Analyzer: Detection of Pulsed Signals," *Appl. Opt.* **25**, 193-198 (1986).
3. A. VanderLugt and A. M. Bardos, "Spatial and Temporal Spectra of Periodic Functions for Spectrum Analysis," *Appl. Opt.* **23**, 4269-4279 (1984).
4. A. M. Bardos, Harris Corp., Melbourne, FL, private communication.

---

**APPENDIX D**

**ROLE OF PHOTODETECTORS IN OPTICAL SIGNAL PROCESSING**

**Reprinted from Applied Optics**

**Volume 27, Pages 2871-2886, 15 July 1988**

---

## Role of photodetectors in optical signal processing

G. W. Anderson, B. D. Guenther, J. A. Hyncek, R. J. Keyes, and A. VanderLugt

Optical signal processing applications place demands on photodetector arrays beyond those encountered in image sensing applications. We review the basic requirements and show that increased dynamic range and nonlinear decision operations that lead to reduced output data rates are the key improvements needed for both 1- and 2-D arrays. Arrays of high-speed photodetector elements with integrated postdetection circuitry are also needed. Although we suggest some possible methods for achieving these goals, our main objective is to stimulate the photodetector community to design and fabricate more useful devices.

### I. Introduction

A key attribute of optical processing is that computationally intensive operations can be performed using the highly parallel structure of processing architectures. With few exceptions, however, the input and output signals are functions of time; various input devices have been developed for converting time signals into space/time signals. In this paper, we focus on the output devices needed to convert the optically processed signals into time signals that can be further processed.

Most parallel optical processing architectures require an array of photodetector elements at the output of the system. Some performance parameters of currently available imaging devices are entirely acceptable. For example, linear arrays having up to 4096 elements integrated on a single chip are now available; this number is adequate for many 1-D processing applications. In contrast, we need  $2000 \times 2000$ -element area arrays now, but current 2-D arrays are limited to the order of  $1000 \times 1000$  elements. The evolution toward high-definition TV may provide the desired number of elements in the future. Throughout this

paper we distinguish between performance characteristics obtained through evolutionary developments and those achieved only through revolutionary developments; the latter are of greatest interest and importance for optical signal processing.

In Sec. II we describe the nature of some basic optical processing systems to establish required performance parameters of photodetector arrays. In Sec. III we review the basic device physics and current performance levels. We compare the noise mechanism for single-element devices with those of arrays and show that processing systems are not always quantum noise limited. In Sec. IV we describe some characteristics desired of future devices with particular attention to increasing the dynamic range and reducing the read-out rates.

### II. Background on Optical Data Processing

We provide a brief introduction to optical processing with particular emphasis on the photodetector component of the system. There are review papers that cover the subject in greater detail,<sup>1,2</sup> and special issues have been devoted to optical processing.<sup>3-6</sup> Optical signal processing has its roots in the work of Abbe who showed, at the turn of the century, how an image formed by a coherently illuminated microscope could be modified by altering the diffraction pattern produced by an object. This diffraction pattern, more generally called a spatial frequency distribution, is the Fourier transform of the object; this transform plays a central role in optical processing.

Until 1960, the outputs of optical processing systems were generally detected, recorded, and displayed using photographic film; we might say that the first 2-D photodetector array was photographic film. Film requires a certain number of photons per unit time, is sensitive to light in certain spectral ranges, and has a

G. W. Anderson is with U.S. Naval Research Laboratory, Washington DC 20375-5000; B. D. Guenther is with U.S. Army Research Office, Research Triangle Park, North Carolina 21211; J. A. Hyncek is with Texas Instruments, Inc., P.O. Box 225012, Dallas, Texas 75265; R. J. Keyes is with MIT Lincoln Laboratory, P.O. Box 73, Lexington, Massachusetts 02173; and A. VanderLugt is with North Carolina State University, Department of Electrical & Computer Engineering, Box 7911, Raleigh, North Carolina 27695.

Received 30 November 1987.

0003-6935/88/142871-16\$02.00/0.

© 1988 Optical Society of America.

dynamic transfer function (the H-D curve), a modulation transfer function, and a noise floor (often expressed in terms of granularity, grain size, or Selwyn's number). Although the terminology is different, the concepts are similar to those associated with photodetector arrays.

A key advantage of photographic film is that it has high spatial resolution which can be chosen to match that of the optical system. The major disadvantage of film is that it must be developed; as a result, it cannot support real-time operations. A vidicon or an image orthicon can be used as the output detector, but these devices have limitations such as inadequate dynamic range and geometric fidelity. The development of 1- and 2-D photodetector arrays based on photosensitive CCD structures provide a new flexibility of operation and have many desirable features. These devices were developed primarily for imaging applications such as electronic newsgathering, surveillance, scanning, or inspection. As a result, these devices are not tailored for use in optical processing applications.

We now describe some basic forms of optical processing systems; the algorithms we implement are generally related to either spectrum analysis or correlation. These examples are selected to identify those features of optical processing that place unusual demands on the photodetectors.

#### A. Spectrum Analyzers

Spectrum analysis is probably the most widely used algorithm in the physical sciences for gaining information about unknown signals. In the optical system shown in Fig. 1(a), an image is placed in plane  $P_1$  and illuminated by coherent light derived from a point source. A key feature of this system is that the complex valued light distribution in plane  $P_2$ , the back focal plane of lens  $L_2$ , is the Fourier transform of the light distribution in plane  $P_1$  (Ref. 7):

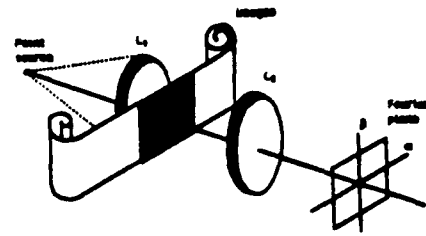
$$S(\alpha, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(x, y) s(x, y) \exp[j2\pi(\alpha x + \beta y)] dx dy, \quad (1)$$

where  $s(x, y)$  is the amplitude of the signal,  $a(x, y)$  is the aperture function which provides weighting to control sidelobe levels and establishes the boundaries of the signal to be processed,  $x$  and  $y$  are the spatial coordinates of plane  $P_1$ , and  $\alpha$  and  $\beta$  are the spatial frequency variables associated with the coordinates of plane  $P_2$ . Since all planes in an optical system have spatial coordinates, we relate the spatial frequency variable to the coordinates  $\xi$  and  $\eta$  in plane  $P_2$  by

$$\alpha = \frac{\xi}{\lambda F_2} \quad \beta = \frac{\eta}{\lambda F_2}, \quad (2)$$

where  $F_2$  is the focal length of lens  $L_2$ , and  $\lambda$  is the wavelength of light.

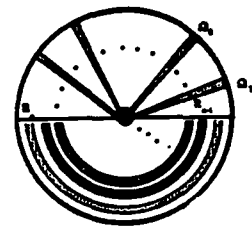
If  $s(x, y)$  contains some regular features such as the street pattern of a city, the spectrum  $S(\alpha, \beta)$  will show strong spectral content in orthogonal directions. The spacing and width of the sidelobe structure in this spectrum indicate the period of the street spacings. In contrast, the spectrum associated with natural terrain



(a)



(b)



(c)

Fig. 1. Spectrum analyses: (a) optical spectrum analyzer; (b) spectra of wave patterns; (c) ring/wedge detector geometry.

is generally more uniformly distributed at all spatial frequencies with no predominant peaks. As an illustration of 2-D spectrum analysis, we show in Fig. 1(b), a photograph of the ocean surrounding one of the Caribbean islands. Images of the surface of the ocean produce spectra that give information about wind direction and ocean depth. From the spectra shown in the insets, we note that strong diffraction occurs in directions normal to the wave patterns. Where there are two interfering wave patterns, two relatively strong diffraction patterns are present. Where the surface waves are propagating in a direction different from that of the major swells, due to a shift in the wind direction, the higher spatial frequencies have a direction differing from the lower frequencies. We arrange for small portions of the input to be illuminated sequentially; the spectra of these subregions are detected to give an indication of the sea state and its variation from one region to another.

##### (1) Special Photodetector Array

A photodetector array for detecting these spectral features consists of a set of wedge and annular photoconductive areas,<sup>8-10</sup> as shown in Fig. 1(c). In this device, one half of the area contains  $N$  photoconductive surfaces in the shape of wedges. The output signals  $\Omega_1, \Omega_2, \dots, \Omega_N$  indicate the degree to which  $s(x, y)$  has spectral content at specific angles. The other half of the detector consists of  $N + 1$  photodetector areas in the shape of rings. The output signals  $R_0, R_1, R_2, \dots, R_N$  indicate the relative energy present in  $s(x, y)$  at various spatial frequency bins. Since  $S(\alpha, \beta)$  is symmetric about the origin when  $s(x, y)$  is real valued, no information is lost in the detection process other than that which falls between the active areas. The object is classified by postprocessing the  $\Omega$ , and  $R$ ,

values according to algorithms developed for specific applications. The details of the postprocessing are not important here; the requirements on the photodetector are. We now consider some of these requirements.

### (2) Spectral Sensitivity

The operating wavelength is dictated by the diffraction efficiency of the input modulator, scattering in the optical system, and the availability of compact efficient sources. The normal source of coherent light is a laser; the photodetector should have a high sensitivity at the wavelength generated by the laser but need not have a broad spectral response. Powerful lasers such as water-cooled argon-ion lasers have strong spectral lines at 488.0 and 514.5 nm; power outputs are in the 0.2–20 W range. He-Ne lasers emitting at 632.8 nm are more often used; these lasers are air-cooled, compact, and reliable. They are generally less powerful, however, with typical outputs in the 10–50 mW range.

In recent years we have seen increased use of semiconductor lasers, such as injection laser diodes (ILDs), which emit light around 830 nm and produce 10–40 mW of power. More powerful ILDs having 200 mW of output power have recently been announced, and even more powerful ones are being developed. For the sake of illustration throughout this paper, we take  $\lambda = 632.8$  nm and let the laser power be 10 mW; these figures will not lead us too far astray in our calculations.

### (3) Number of Elements

An important class of spectrum analyzer uses acousto-optic devices to convert time signals into space/time signals suitable for real-time processing. Acousto-optic devices, often called Bragg cells, consist of an interaction media to which a transducer is bonded. The time signal  $s(t)$  drives the transducer so that pressure waves are launched into the interaction medium. The pressure wave induces a change in refractive index so that the light is modulated in both space and time by the traveling wave. Further details regarding the design parameters for Bragg cells can be found in the literature.<sup>11,12</sup>

The cell supports a range of frequencies from  $f_1$  to  $f_2$  to provide an rf bandwidth of  $W = f_2 - f_1$ . The transit time of a signal in the cell is  $T = L/v$ , where  $L$  is the length of the cell and  $v$  is the velocity of the acoustic wave; the time-bandwidth product of the cell is then simply  $TW$ . The time-bandwidth product establishes the number of elements required in the photodetector array; by the sampling theorem, we must have  $N \geq 2TW$  elements to avoid loss of information.

Common values for the  $TW$  of Bragg cells are of the order of 1000, which is almost independent of the interaction material or the bandwidth. For example, we may construct a Bragg cell having  $W = 100$  MHz and  $T = 10 \mu\text{s}$  or one with  $W = 1000$  MHz and  $T = 1 \mu\text{s}$ . A somewhat higher  $TW$  can be obtained when tellurium dioxide ( $\text{TeO}_2$ ) is used in a slow shear wave mode; in this case we might have  $W = 50$  MHz and  $T = 40 \mu\text{s}$  so that  $TW = 2000$ .

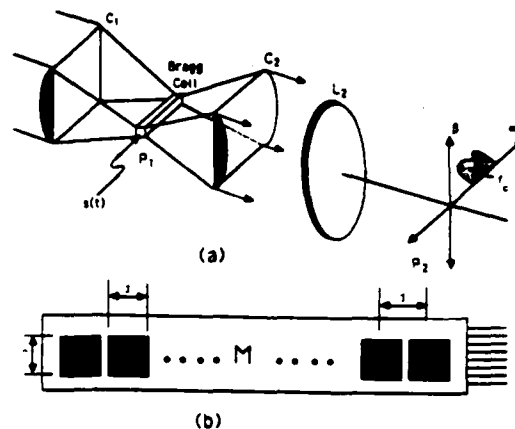


Fig. 2. Real-time spectrum analyzer: (a) the Bragg cell spectrum analyzer; (b) the geometry of the desired array.

### (4) Array Geometry

Having established the required spectral sensitivity and number of elements, we now consider the question of the array geometry. A Bragg cell spectrum analyzer is shown in Fig. 2(a). Cylindrical lenses  $C_1$  and  $C_2$  are used to illuminate efficiently the Bragg cell; in effect, they cause the cell to have an apparent height equal to that of the cylindrical lenses. Lens  $L_2$  with focal length  $F_2$  then produces, at plane  $P_2$ , the Fourier transform of the signal contained in the Bragg cell. Since there is no information in the vertical direction, all the light is focused onto the  $\beta = 0$  axis. The spectral content of the drive signal  $s(t)$  is displayed in the horizontal direction centered at a spatial frequency  $\alpha_c$  corresponding to the temporal frequency  $f_c$ . By ignoring nonessential terms, we have

$$S(\alpha, t) = \int_{-\infty}^{\infty} a(x)s(t - x/v) \exp(j2\pi\alpha x) dx, \quad (3)$$

where  $a(x)$  is the aperture function in plane  $P_1$ . From Eq. (3) we see that the spectrum is a function of both the spatial frequency  $\alpha$  and the current time  $t$ . In essence,  $S(\alpha, t)$  represents a sliding window Fourier transform of  $s(t)$  with  $a(x)$  representing the window function.

The appropriate photodetector array for such a 1-D spectrum analyzer is a linear array as shown in Fig. 2(b). The array consists of  $M$  active elements (shown shaded), each of width  $d'$  and height  $h$ . The detector pitch is  $d$  so that  $c = d'/d$  is the spatial duty cycle of the array elements. The angular separation between adjacent frequencies and the focal length of the transform lens determine the dimensions of the array. For a Bragg cell of length  $L$ , the angular separation between minimum resolvable frequencies is simply  $\Delta\theta = \lambda/L$  so that the frequencies are spaced at plane  $P_2$  a distance  $\Delta\theta F_2$ . We need at least two detector elements per frequency to satisfy the sampling theorem in the Fourier domain; we therefore conclude that

$$d = \frac{\lambda F_2}{2L}. \quad (4)$$

In a well-designed optical system the aberrations can be kept under control if the relative aperture, defined as  $L/F_2$ , is less than  $1/10$ . Thus an ideal pitch for the photodetector elements would be  $d \approx 10\lambda$ , which is of the order of  $3 \mu\text{m}$ . The pitch on currently available linear arrays is typically  $12 \mu\text{m}$  or more; as a result, the length of the optical system must be increased beyond the ideal. Although this may not be a problem for ground-based systems, because the volume of the optical system is generally a small fraction of that of the electronics, the situation is reversed in airborne systems where the volume and weight of the spectrum analyzer must be minimized.

In the vertical direction we have a different criterion for setting the value of  $h$ . The Bragg cell performance is optimized when the illumination height equals that of the transducer, which is generally of the order of  $100 \mu\text{m}$ . Since there is no sampling requirement in the vertical direction, the height of the photodetectors is set by the requirement that we collect the light efficiently, consistent with minimizing the length of the optical system. For typical geometries, this requirement leads to a range of  $h = 60 \mu\text{m}$  for the ideal case to  $h = 240 \mu\text{m}$  for the more typical case when  $d = 12 \mu\text{m}$ . The photodetector elements should, therefore, have aspect ratios of 10–20; alternatively, we can use cylindrical lenses to match the vertical height of the spectrum to that of the array. The required geometric accuracy for most spectrum analysis and correlation applications is that center spacings be within  $\pm 1\%$  of nominal with a cumulative error of  $\pm d/10$  over the length of the array.

#### (5) Readout Rate

We now consider the electronic format and readout rates for linear arrays. In applications such as spectrum analysis  $S(\alpha, t)$  may be a slowly varying function of time so that the sampling rate can be set fairly low without loss of information. Integration on the array, then, may range from a few milliseconds to a few seconds, provided that the detector is not saturated. If the integration time is very long, the contents of the array must be readout, digitized, stored, and accumulated in a postdetector memory whenever saturation is approached.

In other applications the spectrum may change rapidly so that we want to read the array once every  $T$  seconds to avoid loss of information. Unfortunately, the required temporal sampling rates often cannot be sustained by the CCD transfer rates. As an example, suppose that we implement a spectrum analyzer having a 400-MHz bandwidth and an effective processing time of  $T = 1.25 \mu\text{s}$  so that the time-bandwidth product is 500. We then need a 1000-element array to accurately sample the spectrum spatially. If we read the entire array in  $T$  seconds, the CCD output rate is  $800(10^6)$  samples/s; this rate is too high for both the CCD and the subsequent digital electronic postprocessing system to handle directly. A partial solution to this problem is to use multiple video lines to reduce

the rate per line; the aggregate rate, however, remains unaffected.

We need smart arrays in which the postprocessing functions are included on the photodetector chip whenever possible. If we could implement some logic functions at the element level, the transfer data rates and the complexity of the subsequent electronics would be drastically reduced. For example, suppose that we transfer information only if a photodetector element exceeds some preselected threshold. Or, suppose that we transfer information only if the instantaneous intensity exceeds some preset value. These operations require builtin circuitry for each element or set of elements in the array. The implications of developing such a capability are enormous because the transfer rates associated with processing a very wide-band received signal can then be reduced by factors of  $10^2$ – $10^3$  or even more.

#### (6) Dynamic Range

A key requirement of arrays is a large dynamic range; we often need a dynamic range of 60–70 dB. The output light intensity is proportional to the square of the amplitude  $A$ , which in turn is proportional to the square of the applied voltage  $V$ . Thus, in a power spectrum analyzer, we have a direct correspondence between the optical output power and the electrical input power. The photodetector current  $i_d$  is proportional to the detected optical power. The dynamic range of the photodetector is proportional to  $10 \log(i_d^2)$  so that if the detector and its circuitry can sustain a dynamic range of 60 dB, the rf power of the input signal can vary by only 30 dB; this result arises because  $i_d \propto V^2$  so that  $i_d^2 \propto V^4$  in a direct detection system.

A significantly different situation arises when we use a heterodyne technique in which the detected optical power is  $P_d \propto 2\sqrt{P_h P_s}$ , where  $P_h$  is the local oscillator power and  $P_s$  is the signal beam power. The current  $i_d$  from the detector is now proportional to  $\sqrt{P}$ , so that  $i_d \propto V$ ; in turn, the output SNR and dynamic range calculations are based on  $i_d^2 \propto V^2$ . As a result, a detector dynamic range of 60 dB is fully available to accommodate a 60-dB variation in the input rf signal. This heterodyne advantage can be obtained either spatially or temporally through optical interferometric schemes. The price for spatial heterodyning is the need for more photodetector elements; the price for temporal heterodyning is the need for an array of discrete elements that can be read out in parallel.

#### (7) Sensitivity and Power Levels

Photometry is a science whose terminology is strange. In the photodetector literature, we have quantities such as lumens, lux, phots, stilbs, apostilbs, and candelas, and we ratio these by feet, square feet, steradians, centimeters, etc. to get even stranger quantities such as a nit (a candela per square meter). We shall use watts as the measure of power and millimeters as the unit of distance.

In a spectrum analyzer we generally operate the Bragg cell at a diffraction efficiency of no more than



1%/frequency to reduce intermodulation products to an acceptable level. The rest of the system (lenses, mirrors, beam shaping, etc.) may be  $\sim 30\%$  efficient. Given a laser power in the range of 10–30 mW, we find that the maximum power that a photodetector will intercept is of the order of 30–90  $\mu\text{W}$ . The weakest signal we want to detect may be 60–70 dB below this level.

### (8) Blooming and Electrical Crosstalk

We usually read the array before the elements saturate. However, if we detect weak signals by means of longer integration times, we must ensure that the excess charge is properly drained away so that spillover into adjacent elements does not mask the weak signals. Typically, we want the crosstalk level to decrease at a rate of at least 10 dB/element away from a saturated element to a level of at least  $-70$  dB for all elements further away than the seventh element. Blooming is a more severe problem at longer wavelengths.

### (9) Linearity and Uniformity of Response

Single-element photodiodes have large linear dynamic ranges. At some intensity, however, saturation sets in, and the slope of the dynamic transfer curve decreases. Since we often introduce a compression scheme to facilitate the readout and display of the information, a high degree of linearity is not required. A more important characteristic is that the response is monotonic so that we can establish an inverse mapping that allows us to measure the input intensity to the required accuracy.

The saturation phenomena are slightly different for CCD arrays. The charge accumulates until the well is full; an additional charge is not readout so that there is a discontinuity in the derivative of the transfer curve. Since a unique inverse mapping cannot be obtained for this case, saturation in CCD arrays must be avoided.

A uniformity of response across the elements of an array of  $\pm 10\%$  is generally adequate; this degree of uniformity is easily met with current technology.

### B. Two-dimensional Spectrum Analyzers

We can use a 2-D spatial light modulator to format wideband electronic signals; this modulator replaces the photographic film used in the system shown in Fig. 1.<sup>13,14</sup> More recently, the trend is to use orthogonal Bragg cells to provide the desired format.<sup>15,16</sup> The output of the system is a 2-D spectrum wherein coarse frequencies are displayed on the horizontal axis; each coarse frequency in turn is further decomposed by the action of the second Bragg cell to provide a fine frequency resolution in the vertical direction that is of the order of  $1/T_0$ , where  $T_0$  is the total integration time. We generally need at least a  $1000 \times 1000$ -element photodetector array for this application; if we need to resolve a spatial carrier frequency to achieve a heterodyne detection, we need  $\sim 2$ –4 times as many elements in one of the directions.

This array must satisfy most of the performance measures of the linear array described in Sec. II.A.

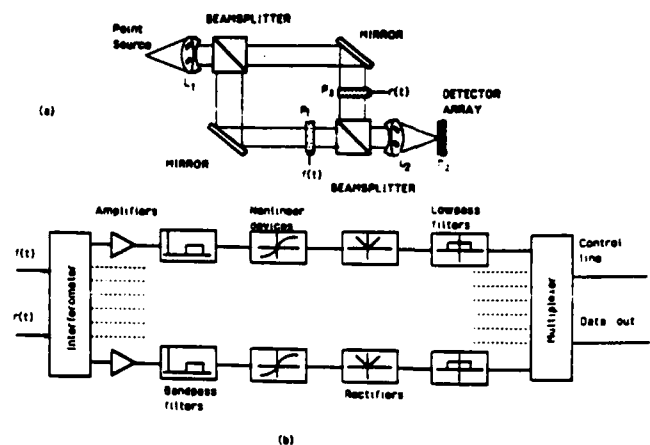


Fig. 3. Heterodyne spectrum analysis: (a) interferometric optical system; (b) detector circuitry (rf receiver).

The main difference is that the light intensity at the detector plane is significantly reduced. For example, we suffer a 1000-fold reduction in light due to decomposing a single coarse frequency into 1000 fine frequencies in the vertical direction. The lower light level is compensated by using longer integration times, which in turn leads to lower readout rates. For example, if we use  $1/30$  s as our readout period, the sample rate is  $60(10^6)$  samples/s. There are, however, applications where we would like to read the array at a rate of 1000 frames/s or more, so that rapidly changing signals such as frequency hoppers can be detected.

### C. Heterodyne Spectrum Analyzer

One way to overcome the dynamic range limitation is to heterodyne detect the complex spectrum.<sup>17,18</sup> Figure 3(a) shows an interferometric system in which the lower branch is equivalent to a power spectrum analyzer. In the Fourier domain the reference beam from the upper branch provides a distributed local oscillator that shifts all spectral components of the received signal to a fixed temporal frequency. The output of each photodetector is of the form  $B + A_k \cos(2\pi f_0 t + \phi_k)$ , where  $B$  is a bias term,  $A_k$  and  $\phi_k$  are the Fourier coefficient and phase of the  $k$ th frequency component, and  $f_0$  is a conveniently chosen IF. Each element in the array must, therefore, respond to a narrowband signal centered at  $f_0$ . The low-frequency terms, lumped into a bias term  $B$ , must be eliminated by filtering. As a result, we cannot use a conventional CCD photodetector structure; instead, we need a totally new kind of array in which we have parallel readout so that the signal from each element can be processed independently, as shown in Fig. 3(b). The basic advantage of the heterodyne approach is that the dynamic range is nearly doubled in decibels, so that a 30–35-dB dynamic range can be extended to 60–70 dB, exclusive of implementation losses.

### D. Correlation Applications

The physical demands on photodetector arrays, such as the number of elements, spacings, and sensitiv-

ities, are much the same for correlation applications as they are for spectrum analysis. The dynamic range requirement can be relaxed because the peak signal intensity is generally set by factors such as the target size, laser power, and optical efficiencies. The detected signal strength, therefore, does not vary as much as in spectrum analysis. There are, however, some new features of the detector array that are highly desirable.

In Fig. 4 we show a classical 2-D system used for pattern recognition. It is an extension of that shown in Fig. 1 with a spatial filter being placed in the Fourier domain  $P_2$  and a photodetector array placed in the image plane  $P_3$  of the input object; this output plane is sometimes called the correlation plane. The filter is constructed to match in both amplitude and phase the Fourier transform of a known signal. If this signal occurs in the data placed in the input plane, a strong correlation peak results at the image position of the signal in  $P_3$ . The photodetector must then detect the output light intensity and decide whether the correlation peak exceeds some preset threshold.

The photodetector array must contain as many elements as there are samples in the image. It is not uncommon to work with 125-mm wide film having 100-cycles/mm resolution; to match fully this capacity we would need a square array having 25,000 elements on a side to satisfy the sampling theorem. There are no arrays that meet the desired number of elements. But correlation systems for pattern recognition have some features that may help to solve this problem if the proper nonlinearity can be incorporated into the array. Consider a frame of imagery that contains several targets as shown in Fig. 5(a). A matched filter will detect signals when its orientation and scale are correct and collapse most of the signal energy into a  $2 \times 2$ -pixel space as shown in Fig. 5(b). The ratio of the number of pixels covered by the target to the number covered by the correlation peak is of the order of  $10^4$  for typical targets. Since the targets generally do not overlap, we could partition the correlation plane into  $4TW/10^4$  possible correlation regions and simply inquire as to which regions are active.

For the example given, we now require a photodetector array of only  $250 \times 250$  resolution elements. But these elements must be of a very special type that do not integrate the light over their surfaces. Instead, they must respond in a highly nonlinear fashion only when the intensity over a small part of their surface exceeds threshold. This requirement is more easily understood by considering, as in Fig. 5(c), an intensity profile across one scan line of the correlation plane. Suppose that the ratio of peak correlation signal to mean square noise is of the order of  $10^3$  and that the number of pixels contained in the target is  $10^4$ . If we were to use a photodetector surface whose response is linear, the integrated background intensity would be  $\sim 10$  times that of the contribution by the correlation peak, because, although the background noise intensity is small, there are  $10^4$  contributing elements to the result. We would then have converted a +30 dB into a -10-dB SNR condition; this is clearly unacceptable.

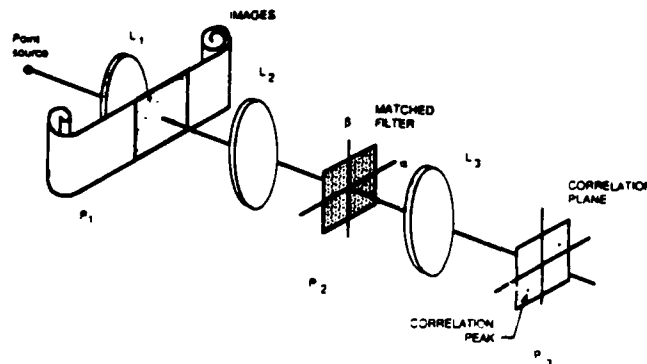


Fig. 4. Correlation system for pattern recognition.

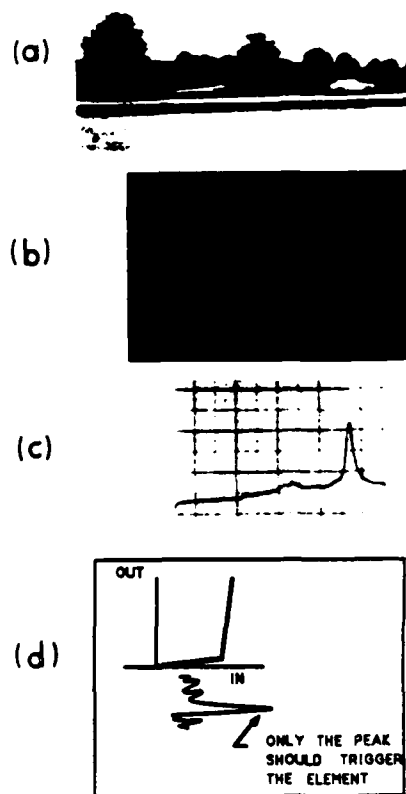


Fig. 5. Pattern recognition: (a) input scene; (b) correlation plane; (c) scan through correlation peak; (d) desired nonlinearity.

The desired transfer characteristic for each photodetector is shown in Fig. 5(d); when the light intensity is below a given value at a particular position, that element does not contribute to the output. When the intensity threshold is exceeded due to the presence of a correlation peak, the output goes to its maximum value as desired. Only a few of the  $100 \times 100$  possible positions in any one of the  $250 \times 250$  set of elements is, therefore, punctured, the remaining ones being below the breakdown voltage. If we had such a device for pattern recognition, we could very easily conceive of processing high-quality imagery—an application that we have basically ignored to date.

## E. Single-Element Devices

Finally, there are optical processing applications wherein a wideband single-element photodetector is needed. The format is preferably rectangular with a high aspect ratio (e.g., a detector 4 mm long by 0.40 mm high). Bandwidths up to 1 GHz are required, and the key problem is keeping the capacitance of the device under control while maintaining the bandwidth.

## F. Summary of Needs

We have briefly touched on various optical processing architectures that illustrate some of the requirements on photodetectors. There are, of course, many other applications such as adaptive filtering or ambiguity surface and Wigner distribution generation; these applications have requirements similar to those given above. Photodetector arrays are also needed in applications such as synthetic aperture radar processing or tomographic processing. We have defined five generic types of detector: (1) a wideband single-element detector; (2) a 1-D array of discrete (parallel readout) detectors; (3) a 1-D array of integrating (serial readout) detectors; (4) a 2-D array of discrete detectors; and (5) a 2-D array of integrating detectors. We can summarize these needs as follows:

### One-dimensional Arrays

At least 2096 elements in (CCD format).

At least 256 elements (discrete, wideband).

Multiple video lines (total rate of >100 MHz).

Wide dynamic range, low crosstalk, high linearity, low fixed pattern noise, low reflections, not polarization-dependent.

Desire either serial or random access readout of elements.

Temporal change detection from frame to frame.

Spatial rate of change detection over subregion of the array.

On-chip processing to calculate centroids, neighborhood operations, etc.

Bandwidths per element range from <1 MHz (serial readout) to >400 MHz (discrete fully parallel readout).

High sensitivity, particularly at light wavelengths of  $\approx 600$ –850 nm.

### Two-dimensional Arrays

Need 2000  $\times$  2000 element arrays.

Other specialized array formats:

3  $\times$  2096 for direction-of-arrival spectrum analysis;

3  $\times$  3 discrete for wideband signal acquisition with Doppler;

25  $\times$  25 for terminal guidance;

250  $\times$  250 for pattern recognition (thresholding and latching).

AC coupled lock-in amplifier effect.

Other requirements are as for 1-D arrays.

## III. Device Characteristics

A photodetector converts optical energy to electrical energy. Four basic types of detector can be used in an optical processor: 2-D vacuum tube sensors; single-

element solid-state sensors; 1-D solid-state detector arrays; and 2-D solid-state detector arrays. Vacuum tube detectors have advantages such as operation in a photon-noise-limited regime; however, we shall focus on the use of solid-state devices because of their small size, high reliability, low-voltage requirements, and the possibility of integrating the detector with signal processing electronics. There are three basic classes of solid-state photodetector: photoconductors; depletion layer detectors; and avalanche photodetectors.

### A. Elementary Device Physics

The device physics parameters which are critical to acousto-optic signal processing photodetectors are quantum efficiency, responsivity, pixel-to-pixel crosstalk, charge capacity, intrinsic detector transfer characteristics, detector dynamic range, and noise. These device parameters are important because of the high dynamic range ( $\approx 70$  dB or  $10^7$  in terms of photon flux) and high speeds required at the output of an optical processor. They have been treated in detail elsewhere<sup>19-21</sup> and consequently will be covered briefly here.

#### (1) Photoconductors

A photoconductor is a junctionless semiconductor device whose conductivity increases when exposed to light. Incident photons with energy greater than the semiconductor bandgap produce electron-hole pairs which contribute to the conductivity of the device until recombination takes place. The change in current density due to illumination is

$$\Delta J = e\mu\eta\phi\tau, \quad (5)$$

where  $e$  is the charge,  $\mu$  is the mobility of the charge carrier,  $\eta$  is the quantum efficiency,  $\phi$  is the light flux, and  $\tau$  is the lifetime of the charge carrier. The current gain in a dc biased photoconductor is given by

$$G = \tau/t_0, \quad (6)$$

where  $t_0$  is the transit time of the faster charge carrier (usually the electron) across the active detector region; gains of  $10^3$  or more can be obtained in silicon. For  $\tau \gg t_0$ , the bandwidth of the device is established by the carrier lifetime:

$$B = 1/\tau. \quad (7)$$

The gain-bandwidth product is given by Eqs. (6) and (7):

$$GB = 1/t_0, \quad (8)$$

and it is a constant for a given material and detector design.

The primary noise source in the photoconductor is Johnson noise due to the bias current, which restricts applications to the detection of light signals with limited dynamic range. New photoconductor designs such as photo-FETS and devices with very low doped materials show promise of overcoming these disadvantages. At present, however, photodiodes are the most widely used devices.

## (2) Photodiodes

The depletion layer photodiode is a reverse biased  $pn$  junction or Schottky barrier diode whose reverse current is modulated by charge carriers produced by photons in or near the depletion layer of the device. A typical device is shown in Fig. 6. A  $pn$  junction is reverse biased, creating an electric field of the order of  $10^4$  V/cm at the junction. This field depletes a region of width  $W$  of all carriers; this region is called the depletion region. When a photon with an energy greater than the bandgap energy is absorbed in the depletion region, an electron and hole are created. The electric field causes the carriers to drift rapidly to the electrodes, producing a photocurrent. Some photons may not be absorbed in the depletion region but penetrate into the  $n$ -doped material of Fig. 6. The holes (minority carriers), produced in a region of width  $L$ , are able to diffuse to the electrode before they recombine. These holes also contribute to the photocurrent. If  $\gamma$  is the absorption coefficient, the photon flux at a distance  $x$  into the  $n$ -doped material is

$$\phi = \phi_0 \exp(-\gamma x), \quad (9)$$

where

$$\phi_0 = P_0/h\nu \quad (10)$$

is the number of incident photons in a beam with an optical power of  $P_0$  and a frequency of  $\nu$ . Any photons existing beyond a distance  $x = W + L$  do not contribute to the photocurrent and are lost. We have ignored absorption in the electrical contact layer and reflections at the various interfaces; these effects may be included by modifying  $\phi_0$ . We discuss the impact of these factors in the following paragraphs.

The quantum efficiency is defined as the number of charge carriers produced by the detector divided by the number of photons incident on the detector:

$$\begin{aligned} \eta &= \left[ \frac{\phi_0 - \phi(W+L)}{\phi_0} \right] = (1-r) \{ 1 - \exp[-\gamma(W+L)] \} \\ &= (1-r) \left[ 1 - \frac{\exp(-\gamma W)}{1 + \gamma L} \right]. \end{aligned} \quad (11)$$

Here we explicitly display the modification needed to include the effects of reflections at the air-detector interface. The constant  $r$  is the Fresnel reflection coefficient at the surface of the detector. We continue to ignore reflections at the other interfaces. Reduction of reflection losses at the external surface of the detector through the use of an antireflection coating will increase the quantum efficiency by increasing the number of photons reaching the depletion layer. The photocurrent density generated by a detector is

$$J = e\phi_0\eta \approx e\phi_0 \left[ 1 - \frac{\exp(-\gamma W)}{1 + \gamma L} \right] (1-r). \quad (12)$$

To maximize the photocurrent it is obvious that  $\gamma W$  and  $\gamma L$  should be as large as possible. The depletion layer width is inversely proportional to the doping concentration:

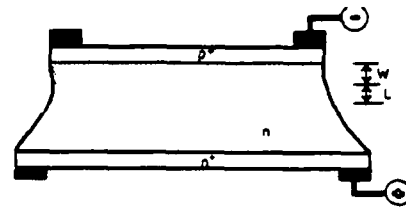


Fig. 6. The  $pn$  junction used as a photodiode.

$$W = \sqrt{\frac{2\epsilon}{e} (V_a + V_c) \left( \frac{1}{N_a} + \frac{1}{N_d} \right)}, \quad (13)$$

where  $V_a$  is the applied voltage,  $V_c$  is the contact voltage,  $N_a$  is the acceptor concentration, and  $N_d$  is the donor concentration. The drift component of the current is a fast transport mechanism, as short as 0.1 ns for  $W = 10 \mu\text{m}$  in silicon, while the diffusion component is relatively slow, 40 ns in silicon for  $L = 10 \mu\text{m}$ . Thus for high speeds  $L$  should be kept small, while  $W$  should be large. The junction capacitance, which affects the speed of the diode through the RC time constant of the output circuit, is given by

$$C = \frac{eA}{W} = \frac{A}{\sqrt{\frac{2}{e\epsilon} (V_a + V_c) \left( \frac{1}{N_a} + \frac{1}{N_d} \right)}}, \quad (14)$$

where  $A$  is the diode junction area. As can be seen, low doping and large applied fields increase  $W$  and thereby increase the frequency response of the detector. For silicon we have  $V_c = 0.6$  V, and for GaAs we have  $V_c = 1.1$  V. An intrinsic material would be used to maximize the value of  $W$  because there are few donors or acceptors and  $N_a$  and  $N_d$  are nearly zero. The width  $L$  is a function of the minority carrier lifetime and the diffusion rate of minority carriers.

The absorption coefficient consists of three coefficients. Only the interband absorption coefficient  $\gamma_B$  contributes to the quantum efficiency. The other two coefficients are the free carrier absorption  $\gamma_C$  and the scattering loss  $\gamma_S$ . The processes that give rise to these coefficients remove photons from the incident radiation without generating current. The quantum efficiency is thus reduced by the factor

$$\frac{\gamma_B}{\gamma_B + \gamma_C + \gamma_S}. \quad (15)$$

The photodetector's performance is also a function of the wavelength of the illumination. The absorption coefficient is a function of wavelength so that wavelengths much longer than the bandgap  $E_g$  will be associated with a small  $\gamma_B$ . The quantum efficiency will drop as the illuminating wavelength shifts to the long wavelength side of the bandgap due to deep penetration of photons into the substrate. If the wavelength is fixed and we decrease the bandgap, it seems that there is no limit to the improvement in detector performance. However, if the bandgap of the semiconductor is much smaller than the energy of the illuminating wavelength, an excessive number of thermally generat-

ed charge carriers are produced, resulting in a large dark current  $I_d$ .

As mentioned earlier, absorption in the  $p^+$  layer reduces the number of photons reaching the detector region, requiring  $\phi_0$  to be modified:

$$\phi_1 = \phi_0 \exp(-\gamma_p \cdot x_p).$$

The  $p^+$  layer, shown in Fig. 6, introduces a short-wavelength cutoff by absorbing photons near the surface. This limitation can be overcome by illuminating the device in Fig. 6 from below. A second technique for overcoming the short-wavelength limit is to replace the  $p^+$  layer by a metal film, producing a Schottky barrier photodiode, whose operation is essentially the same as the  $p^+n$  junction.

To reduce the series resistance, and thus some of the noise, the  $p$  region is heavily doped. A heavily doped  $n$  region is added to the back of the intrinsic layer to serve as a second contact. When intrinsic material is used the result, shown in Fig. 6, is a PIN diode. The high purity of the intrinsic layer allows a large depletion width to be created at low voltages, which minimizes the noise current.

### (3) Avalanche Photodiodes

If the reverse bias voltage is set near the avalanche breakdown point ( $\approx 100$ – $400$  V), the gain of the detector can be raised above one. A region of high field of the order of  $10^5$  V/cm accelerates the charges to sufficient energy to create new carriers through impact ionization. The gain of this device is defined as the ratio of the multiplied current to the photon-induced current:

$$G = \exp[(\beta_n + \beta_p)W], \quad (16)$$

where  $\beta_n$  and  $\beta_p$  are the ionization rates for the negative and positive charge carriers, respectively. The gain can also be written as a function of device parameters; the maximum gain is given by

$$G_{\max} = \sqrt{\frac{V_b}{nIR}}, \quad (17)$$

where  $V_b$  is the breakdown voltage,  $n$  is an experimentally determined parameter,  $I$  is the saturation current, and  $R$  is the series resistance of the diode (including space charge effects). The avalanche gain is a function of the bias voltage, which is set at the breakdown voltage for this discussion.

$$G = \frac{1}{1 - \left[ \frac{V_b - IR}{V_b} \right]^n}. \quad (18)$$

The dependence of the gain on the bias voltage can lead to nonlinearities when a large dynamic range is required. The breakdown voltage is a function of the temperature, and thus the avalanche gain varies with temperature. These effects require careful control of high bias voltages. Breakdown can occur at lower voltages due to edge breakdown or microplasma generation at localized defects so that the device gain is limited.

When  $\beta_p \ll \beta_n$ , so that only the electrons contribute to the ionization process, the bandwidth is maximized, and excess noise produced by the multiplication process is minimized. The relative size of the two ionization rates is characterized by the ratio

$$k = \beta_p / \beta_n. \quad (19)$$

As  $k$  approaches one, the gain becomes a strong function of the applied electric field, which makes it difficult to manufacture devices with uniform characteristics. Nonuniform materials increase the randomness of the collision ionization, leading to an increase in the noise in the detector. The excess noise term is of the form

$$G^2 = kG + (1 - k) \left[ 2 - \frac{1}{G} \right]. \quad (20)$$

If  $k = 1$ , the excess noise term is equal to  $G$  and  $x = 1$ . If  $k = 0$ , the excess noise term is a factor somewhere between 1 and 2. Typical  $k$  values of 0.02 in silicon yield an excess noise of 4 at a gain of 100, although values 2.5 times smaller have been reported.<sup>20</sup> Other noise processes and bandwidth limitations are the same as the limits of a PIN diode. Bandwidth decreases as gain increases because the avalanche process takes time to develop.

A number of problems are associated with avalanche photodiodes. The major problem is the production of a device with a dislocation-free substrate to raise the effective breakdown voltage level and a doping concentration controlled to 0.1% to reduce the effects of random fluctuations. A second problem is obtaining enough surface passivation to reduce edge breakdown. The passivation also reduces surface leakage change with age, which is a major cause of reliability problems. Arrays of these devices are more difficult to produce because of the need for guard ring structures to prevent edge breakdown.

### B. Arrays

The detectors discussed above must be assembled into either 1- or 2-D arrays before they can be used in an optical processing application. The detector arrays may be instantaneous or integrating devices whose outputs are addressed in parallel or in serial output lines, respectively. The detector array devices may be photoconductor arrays, charge coupled device (CCD) arrays, avalanche photodiodes (APDs), PIN diodes, or Schottky barrier diodes. If the arrays are nonintegrating, the signal from each detector is amplified by a single amplifier circuit which may be a logarithmic amplifier to provide output signal compression to handle high dynamic ranges. If the arrays are integrating devices, they may be addressed by MIS shift registers or by CCD parallel-in serial-out shift registers.

Two special problems arise when detectors are assembled into arrays. One is optical in origin and is called pixel-to-pixel crosstalk. The second is electronic in origin and is called fixed pattern noise. Photon produced electrons generated in one detector element and collected in another cause pixel-to-pixel

**crosstalk.** Unfortunately, techniques used to isolate individual pixels reduce the quantum efficiency. Fixed pattern noise arises from variations in the performance parameters of the array elements.

In Fig. 7 the photodiodes are isolated by a region of opposite carrier concentration, here denoted as a *n* region, often called the tub. The thickness of the tub region below the depleted layer is important in limiting pixel-to-pixel crosstalk due to both thermally and photon generated carriers. Carriers formed due to absorption of photons below the tub layer do not contribute to crosstalk because they are not collected by any photodiode.

The quantum efficiency is affected by the techniques used to create the individual pixels. All photons absorbed between the pixels can contribute to the quantum efficiency but would also contribute to optical crosstalk. If grooves are etched between the pixels or if channel stops formed by diffusion or ion implantation exist, as is commonly the case in CCD imagers, the pixel-to-pixel isolation is improved, but the quantum efficiency is decreased. A deeper depletion region will result in the collection of more charge, overcoming the loss associated with isolation, but will increase the crosstalk arising from carriers produced deep in the device. A tradeoff must be made between crosstalk and quantum efficiency.

Other problems associated with the readout process are holdover crosstalk, blooming, and fixed pattern noise. If the integration period is of the order of a microsecond, some of the charge from the previous signal may remain; this is called holdover or temporal crosstalk. When the incident light is very high, blooming may be caused by charge spilling from one pixel to another. A number of methods of handling blooming have been developed such as adding a drain to the sensor structure to remove excess charge.<sup>22</sup> Fixed pattern noise occurs because of nonuniformities in gain or the dc offset in different detector elements. Leakage of signals from switching circuits also contribute to the fixed pattern noise. The dynamic range and measurement accuracy is reduced by these pixel-to-pixel nonuniformities.

#### IV. Discussion of Representative Architectures

To achieve the high performance required of photodetectors, we examine the limits encountered in constructing such detectors and the discovery of new approaches that could improve their performance. The device structures are usually divided into two categories. First, are nonintegrating detectors, where the photon flux is continuously converted into an electrical output signal. The second are integrating detectors, where the charge produced by the incident photons is integrated and then converted into the output signal. The later category can be further subdivided into two subgroups. One subgroup uses a feedback amplifier connected to the detector. The second subgroup uses an array of detectors such as in a CCD or CID device; the optical signal is converted into a charge packet, stored in a potential well, and is later trans-

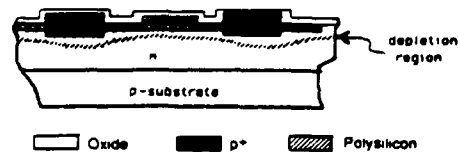


Fig. 7. Crosstalk/depletion region.

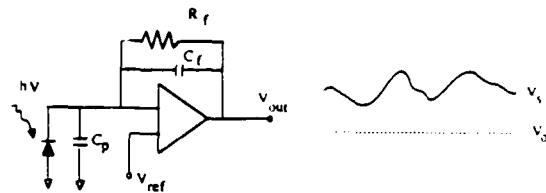


Fig. 8. Feedback amplifier.

ferred into a charge-to-voltage converting structure common to a group of detectors or to the whole array. To understand the fundamental limits of operation, functional differences, and common features of each of these detectors, we briefly describe their principles of operation.

The nonintegrating detector usually consists of a photodiode connected to a feedback amplifier as shown in Fig. 8. This device has a wide dynamic range due in part to the feedback loop which maintains a constant bias across the photodiode, irrespective of the magnitude of the photocurrent that is generated in the detector. The dynamic range of this arrangement is the ratio of the maximum voltage which can be permitted across the feedback resistor before reaching saturation to the minimum noise floor.

#### A. Signal-to-Noise Ratios

The minimum noise floor consists of thermal noise and shot noise caused by the average current in the circuit. The shot noise term is

$$SN = 2e(I_s + I_d + I_h)B_n R_L G^m, \quad (21)$$

where *e* is the charge of an electron, *B<sub>n</sub>* is the noise bandwidth, *R<sub>L</sub>* is the load resistance, *G<sup>m</sup>* = *G<sup>2+x</sup>* is a gain factor characteristic of avalanche photodetectors, and *G<sup>x</sup>* is the excess noise introduced earlier. Typical values for *m* are 2.3–2.5 for silicon devices and 2.7–3.0 for III–V alloy devices. The average currents are *I<sub>s</sub>* due to the signal and *I<sub>d</sub>* due to the dark current. We also include a current *I<sub>h</sub>*, generated when we add a local oscillator to the signal to operate in a heterodyne detection mode. When we operate in a direct detection mode, the current *I<sub>h</sub>* is equal to zero. The thermal noise is

$$TN = 4kTB_n, \quad (22)$$

where *k* is Boltzmann's constant, *T* is the temperature in degrees kelvin, and *B<sub>n</sub>* is the postdetection bandwidth. The signal-to-noise ratio is then

$$SNR = \frac{(i_s^2)G^2 R_L}{2e(I_s + I_d + I_h)B_n R_L G^m + 4kTB_n}, \quad (23)$$

where  $\langle i_s^2 \rangle$  is the time average of the current produced by the signal, and  $G$  is the gain of an avalanche photodetector.

The value of the load resistance is determined by the upper cutoff frequency  $f_c$  required to pass the signal:

$$R_L = \frac{1}{2\pi c_d f_c} \quad (24)$$

where  $c_d$  is the capacitance of the photodetector. We substitute Eq. (24) into Eq. (23) and rearrange terms to obtain

$$\text{SNR} = \frac{\langle i_s^2 \rangle G^2}{2e(I_s + I_d + I_h)B_n G^m + 8\pi k T c_d B_n f_c} \quad (25)$$

A few comments regarding Eq. (25) are in order so that we understand how the SNR is to be calculated under the wide range of photodetector applications described in Sec. II. First, consider the class of nonintegrating devices. In some cases the signal may be at baseband so that  $B_n \approx f_c$ ; in other cases the signal may be narrowband so that  $B_n \ll f_c$ . In either event, we want to select a photodetector gain that maximizes the dynamic range using the minimum laser power. This occurs when the shot noise dominates the thermal noise:

$$G^m \approx \frac{4\pi k T c_d f_c}{e(I_s + I_d + I_h)} \quad (26)$$

which reveals that avalanche diodes are most useful when the signal is wideband (large  $f_c$ ) or when we use direct detection ( $I_h = 0$ ). From Eq. (25) we see that, since  $m > 2$ , any gain larger than that indicated by Eq. (26) will result in a deterioration of the dynamic range. In heterodyne applications the local oscillator is usually sufficiently strong so that shot noise dominates the thermal noise (large  $I_h$ ); avalanche detectors then give way to PIN detectors, for which  $G = 1$ .

Second, consider integrating photodetector array circuits which usually have low bandwidths so that the shot noise dominates the thermal noise and avalanche operation is of no advantage. We can then simplify the SNR expression to

$$\text{SNR} = \frac{\langle i_s^2 \rangle}{2e(I_s + I_d + I_h)B_n} \quad (27)$$

The minimum noise occurs when  $I_s \approx I_d + I_h$ . For direct detection applications, we see that the system performance is dark current noise limited; for heterodyne detection the performance is local oscillator current noise limited.

The dynamic range is the ratio of the maximum signal power to the minimum noise power. In a nonintegrating application, the maximum signal power is determined by the available laser power or by an allowable departure from linearity as saturation is approached. In an integrating application, the maximum signal power is set by the saturation level as determined by the charge capacity of the diode. Since the dynamic range is an important parameter in optical processing, we give the results for the four principal modes of operation:

Direct detection: *integrating*

$$DR = \frac{\langle i_s^2 \rangle}{2eI_d c_d B_n} \quad (28a)$$

Heterodyne detection: *integrating*

$$DR = \frac{\langle i_s^2 \rangle}{4e(I_d + I_h)c_d B_n} \quad (28b)$$

Direct or heterodyne

Detection: *nonintegrating*

$$DR = \frac{\langle i_s^2 \rangle G^2}{16\pi k T c_d B_n f_c} \quad (28c)$$

From these results, we might conclude that the dynamic range does not increase when heterodyne detection is used, contrary to our earlier claims. The advantage of heterodyne detection becomes apparent when we realize that the photodetector current in heterodyne detectors is linearly related to the optical signal amplitude instead of to the optical signal intensity, as is the case for direct detection.

Suppose, for example, that  $P_h$  is the optical power at the detector from the local oscillator and that  $P_s$  is the optical power from the signal. The induced photocurrent is then

$$i_s = S[P_h + P_s + \sqrt{2P_h P_s} \cos(2\pi f_0 t)], \quad (29)$$

where  $S$  is the sensitivity of the photodetector in A/W and where we have introduced a temporal IF  $f_0$  to help separate the desired third term from the first two-bias terms. The average currents due to the local oscillator and to the signal are  $I_h = SP_h$  and  $I_s = SP_s$ . These terms contribute to the shot noise; they do not appear in the numerator of our SNR expressions because they are typically eliminated by a bandpass filter. After filtering we see that the signal current is proportional to  $\sqrt{P_s}$ , so that the system is linear in optical signal amplitudes. Without the heterodyne action provided by the local oscillator, we see that the system is linear in optical signal intensities.

From these results, we see that the SNR is always smaller than the dynamic range, a direct result of the Poisson's statistics of the noise in these photon detecting systems. In devices such as CCDs, for example, where the charge is generated electrically by filling a potential well, the noise in each signal charge packet is approximately constant and independent of the number of input electrons. It is, therefore, useful to extend the dynamic range by using a nonlinear element in the feedback of the amplifier without sacrificing the SNR. Equation (23) remains approximately valid in this case since it does not directly contain the feedback element parameters.

The possibility of extending the dynamic range has been recognized by researchers who successfully designed and tested systems containing logarithmic amplifiers or piecewise linear amplifiers.<sup>20-22</sup> The dynamic range obtained approaches 80 dB in power units, which is satisfactory for optical signal processing applications. One problem encountered is a variation of the bandwidth with the signal level, as suggested by

Eq. (24). This problem could in principle be eliminated if the feedback capacitance  $c_f$  shown in Fig. 8 is made variable so that the  $c_f/R_f$  product is kept constant. It is not difficult to incorporate a voltage variable capacitor (varicap) into the feedback loop, which changes its value by an order of magnitude. It may not be practical, however, to compensate exactly the variations in  $R_f$  by variations in  $c_f$ , and some excess bandwidth must be designed into the system. This will, unfortunately, increase the noise floor of the amplifier somewhat.

Integrating detectors which use an amplifier with feedback will have a comparable performance. The operation of such a detector-amplifier system can be understood from the circuit given in Fig. 9. The signal is integrated on the feedback capacitor  $c_f$  which is reset after the readout is completed. The SNR and dynamic range are expressed as before, except that the noise bandwidth is now defined as  $B_n = 1/(2T_r)$ , where  $T_r$  is the time interval between successive readouts of the diodes (the integration time). By analogy with the technique just described, we can also extend the dynamic range by using a nonlinear element in the feedback loop. The varicap could be easily incorporated into an integrated detector-amplifier array. An advantage of this approach is the flexibility in selecting the integration time to always use the full dynamic range of the amplifier. An additional advantage available in integrating systems is the possibility of reducing the effects of the thermal noise. This can be achieved by a well-known correlated double sampling signal processing method.<sup>23</sup> In this approach a difference of two signal readings is formed between the outputs just after the reset and at the end of the integration period  $T_r$ . The difference signal theoretically does not contain the thermal noise component; in practice, however, some amount of the noise is still present.

The most critical problem of insufficient dynamic range, however, is encountered in the detector arrays which operate on the CCD principle. These devices convert the input photon flux into charge, which is stored in potential wells and later, after completion of integration, transferred into the charge-to-voltage conversion device. The chief advantage of using CCD arrays in optical signal processing stems from the high packing density of sensing elements and, therefore, the capability of building large arrays.

The difficulty of extending the dynamic range in these detectors results from the linearity of the photon-to-charge conversion. Amplifiers with nonlinear feedback elements are not easily incorporated into each photosite, and the dynamic range extension must be therefore performed directly in the charge domain before the array is read out. We now illustrate the general principle of such a charge domain dynamic range extension. A cross section of the device is shown in Fig. 10; it contains a long narrow photosite with a drain gate on one side. The drain gate interfaces the photosite with the  $n^+$  drain and allows the charge to be drained out from the photosite. Another gate, not shown in the drawing, runs along the length of the

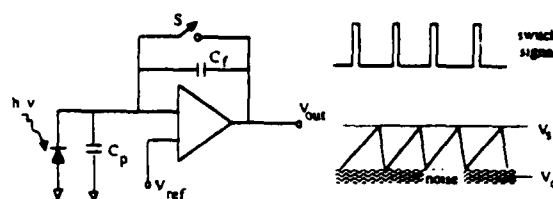


Fig. 9. Integrating detector with feedback.

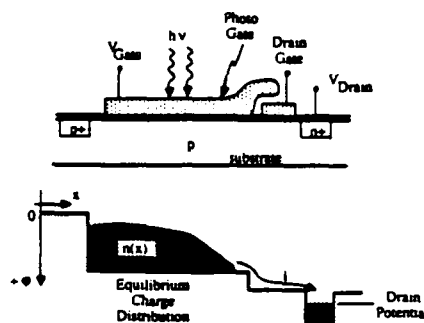


Fig. 10. Photodetector with drain gate.

photosite and interfaces it with an element of the readout CCD register. This gate transfers the charge from the photosite to the readout register at the end of the integration period. During the integration period the drain gate is open and the charge is allowed to be continuously drained out. However, due to the relatively long length of the photosite there will be an equilibrium charge distribution along the length, since the new carriers, which are constantly generated at a given rate, cannot be instantly drained out. The equilibrium formation time, which determines the minimum integration time, can be found from the carrier diffusion time along the length of the photosite. The equation describing the charge distribution is

$$\frac{d}{dx} \left\{ eD \frac{dn}{dx} + \mu \frac{e^2}{c_0} n \frac{dn}{dx} \right\} = -I_p \quad (30)$$

where  $n$  represents the charge concentration in electrons/mm<sup>2</sup>,  $I_p$  is the photocurrent density in A/mm<sup>2</sup>, and  $c_0$  is the equivalent charge storage capacitance in F/mm<sup>2</sup>. The rest of the symbols have their usual meaning. At the end of the integration period the drain gate is closed and the charge is quickly transferred into the CCD readout register via the transfer gate. Since one pixel of the CCD register spans the whole photosite length, the charge transfer time can be much shorter than the integration time, which minimizes errors due to the redistribution and integration of unwanted signal. The total amount of the transferred charge can be obtained by integrating Eq. (30). The boundary conditions are

$$\begin{aligned} \text{At } x = 0, \quad \frac{dn}{dx} &= 0; \\ \text{At } x = L, \quad n &= 0. \end{aligned} \quad (31)$$

The solution for  $n(x)$  will be



$$n(x) = -\frac{kTc_0}{e^2} + \sqrt{\frac{kTc_0}{e^2} + (L^2 - x^2)I_s \frac{c_0}{\mu e^2}} \quad (32)$$

The total number of electrons transferred will then be

$$N = W \int_0^L n(x) dx, \quad (33)$$

which, after evaluation of the integral, becomes

$$N = \frac{WL}{2} \left[ \frac{kTc_0}{e^2} \right] \left\{ \left( \frac{a^2 + 1}{a} \right) \arcsin \left[ \frac{a}{\sqrt{a^2 + 1}} \right] - 1 \right\}, \quad (34)$$

where  $W$  is the width of the photosite. In Eq. (34) we introduce a parameter  $a^2 = I_s L^2 \mu / (D^2 c_0)$ , where  $D$  is the carrier diffusion constant, to simplify the notation. This result can be further simplified for  $a \gg 1$  and the final expression for  $N$  rewritten as

$$N = \frac{\pi}{4} \left( \frac{WL^2}{e} \right) \sqrt{I_s c_0 \mu}. \quad (35)$$

From this result we observe that the signal detected at the array is now proportional to the square root of the input photocurrent. This feature doubles the dynamic range. The available CCD well capacity of the transport registers and the charge conversion amplifier capacity will be efficiently utilized. There are other possibilities of charge domain signal processing which can lead to logarithmic or piecewise linear transfer characteristics.<sup>20,21</sup> The detailed description of such devices, however, is beyond the scope of this paper.

The rate at which ADs can convert analog signals into a digital format is a strong function of the number of binary bits per sample. Currently, ADs can digitize a 4-bit sample at 1-GHz rate, 8 bits at 200 MHz, 12 bits at 20 MHz, and 16 bits at  $\sim 100$  kHz. Precise measurement accuracy (large number of bits) leads to low sample rates, but the machine intelligence community is pushing for higher sample rates. Detection devices which have a logarithmic response to radiation greatly compress the voltage scale presented to the AD. The first binary bits provide a reasonable measurement accuracy for small signals; the last bits represents large signal changes and hence poor absolute accuracy.

If measurement accuracy at high levels is not a critical issue, a single 6-bit ADC, sampling at a 200-MHz rate, should be capable of digitizing the outputs of many optical processors. Using a parallel second AD, in which the first bit is set at slightly less than the square root of the maximum expected dynamic range, would recover measurement accuracy at high intensities if so desired.

To conclude this section, we stress that there is little difference between the integrating and nonintegrating detector schemes in performance as well as in the possibility of dynamic range extension by utilization of nonlinear elements. The possibility of dynamic range extension in CCD arrays is particularly attractive, since the high pixel packing density and the low noise of charge detection amplifiers can be fully utilized for the large data rates present in optical signal processing applications.

## B. Data Rate Reduction

The computational power of optical processing results in high output data rates unless steps is taken to introduce on-chip processing that includes nonlinear or decision operations. We discuss three broad possible methods to achieve the desired data rate reduction: (1) on-chip processing for arrays that output electronic signals; (2) methods for segmenting arrays, along with the possibility of on-chip processing; and (3) new array types in which the output of the detector remains in an optical format for further processing.

### (1) Optical-In Electrical-Out Arrays

Signal processing within the detector chips may significantly reduce the output data rates to the post-processor. Some of the desired functions are:

Video amplitude compression (perhaps programmable).

Temporal change detection from frame to frame.

Spatial change detection along the elements of the array.

Dynamically programmable spatial convolutions for such tasks as centroiding or corner detection, using cellular blocks of up to  $7 \times 7$  elements.

Random access to any subregion of the array to isolate and dynamically track selected activity.

Threshold levels that may be either globally or locally set and adjusted adaptively to achieve constant false alarm rates.

Methods for synchronous detection to remove the strong bias terms or background that arise in some processing operations.

Analog first-in first-out (FIFO) or last-in first-out (LIFO) memory.

While some of these functions have already been demonstrated or designed at low kilohertz rates, the tremendous data rates produced in some real-time processing operations require frame rates that are 2 orders of magnitude faster (e.g., 100 MHz to 1 GHz), which implies that the various analog operations cited must be performed in 10–1 ns at each element in the array.

Such a tremendous leap in capabilities is likely to require new materials or new techniques such as quantum well structures. Some of the more critical difficulties involve readout schemes with contradictory demands for large bandwidth, low noise, and low power, commensurate with the anticipated load. Important characteristics such as quantum efficiency, gain, offset, and threshold must be uniform both within each chip and from chip to chip. To improve performance, it may be useful to provide a 3-D photodetector structure so that the entire pixel area can be used for imaging the light signal; the complex electronic processing and multiplexing circuitry needed to perform the on-chip operations would then be placed on the opposite side of the chip. A first implementation may be the use of backside illumination with the photosignal coupled to signal processing circuitry on the front side. A more general and longer range implementation would be the use of multiple silicon-on-insulator layers with

the photodetector being formed in the top layer and the circuitry being implemented in the intermediate layer and substrate. It is important that the photodetector imaging layer be compatible with the laser source. Thus, if the laser wavelength is  $\approx 850$  nm, either a very thick silicon imaging layer must be used to obtain a high quantum efficiency, or some other semiconductor, such as germanium, with a high absorption coefficient must be used. Silicon/germanium heterostructure materials or superlattices may be useful for this purpose.

An important family of optical processors operate in a heterodyne fashion to increase the useful dynamic range of the system or to measure both the amplitude and phase of the light distribution. In either case the IF may range from 1 MHz typically to as high as 200 MHz. For these applications, the detector array must be nonintegrating, and the readout structure is generally a set of parallel outputs. An acousto-optic channelizer typically produces 100–1000 narrowband channels, each centered at an IF. By synchronously demodulating each IF channel on the photodetector chip, we can obtain the rf envelope of each channel at baseband, which is more suitable for other channel processing/detection functions on the photodetector chip.

In some applications, we would prefer to operate on signals that have not been downconverted from their original rf band. A programmable rf photodetector could sum the rf photodetector currents from a linear array to a common rf output, possibly with controlled summing gain at each photodetector element. By placing this rf photodetector in the frequency plane of a coherent optical processor, we can implement programmable filter functions such as frequency excision (notch rejection filters), tunable high-Q bandpass filters, and filter equalization functions. A high-contrast (40–80-dB) on/off ratio for each photodetector pixel is desired to perform various matched filter functions.

## (2) Segmented Arrays

Nature has designed a number of imaging systems with specially designed detector planes that reduce the data rates that must be processed. An example is the eye, where only the central viewing zone, called the fovea, has high resolution. The cone detectors, located in the fovea, provide color vision while the rod detectors, located almost exclusively outside the fovea, have no color sensitivity. The image plane is nonuniformly sampled, and many detectors (e.g., the rods) are interconnected at a number of levels. Thus the detector plane of a vertebrate is arranged to reduce the data rate transmitted to the brain.

We wish to explore the possibility of interconnecting the detector plane of an optical processor so that we may reduce the data rate of the signals provided by the optical processor. In general, we expect to find that the requirements placed on the detector array for the optical processor will differ from those of an imaging system. One attempt to design a special detector for an optical processor resulted in the ring-wedge detec-

tor discussed in Sec. II and illustrated in Fig. 1(c). This detector, made on a single silicon wafer, has a sensitivity of 0.25 A/W at 633 nm and has found application in the on-line inspection of products such as hypodermic needles.

Some optical processors use conventional detectors but modify the illumination at the detector plane by using a mask. Size distribution measurements can be made by placing a mask in the back focal plane of a lens and by illuminating a medium containing scattering objects whose size distribution is of interest. At least two commercial devices make use of this approach.<sup>24</sup> In the following discussion we wish to generalize the concept of specially designed detector systems for optical processing.

Some systems, such as those using a quadrant detector for tracking, use a feedback system to bring the optical signal to the optimum position on the detector. This is similar to the operation of the eye where the image of interest is brought to the part of the detector plane where the high-resolution elements are located. More generally, an array with a random access capability allows the user to define a small window of active detector elements and to move that window around on the focal plane. If an important event is known to exist in a given region, that region can be sampled at a faster rate while letting regions of lower interest integrate for longer periods.

One of the easiest modifications of a detector array to accomplish is segmenting the array of  $N$  elements into subarrays. Each subarray is given the capability of processing the signal at a local level to reduce the amount of data it will report to the optical processor output. If we have  $K$  subarrays, each of size  $M$ , the total number of detectors we must scan if there had been no subdivisions is  $N = KM$ , and the data rate would be

$$\frac{NQ}{C} = \frac{KQM}{C}, \quad (36)$$

where  $Q$  is the number of quantization levels of the signal, the  $C$  is the system clock speed. The assumption is that the system is limited to a data rate of

$$R = \frac{MQ}{C}. \quad (37)$$

By dividing the detector array into  $K$  subarrays we reduce the data rate to the desired level. In each subarray, some processing is done to simplify the amount of data by  $S$ , where  $S < 1$ , so that each subarray provides  $B = SM$  data points. A time  $T$  is spent performing this processing so that we cannot scan the subarrays at a rate faster than  $K/T$ , where  $T = IC$  and  $I > 1$  can be thought of as an effective integration factor. The result of the processing at the subarray level is to reduce the total data rate to

$$\frac{KBQ}{T} = \frac{KSMQ}{IC} = \frac{KS}{I} R. \quad (38)$$

Thus processing at the subarray level, characterized by the ratio  $S/I$ , can have a major impact on the process-

ing rate needed at the postdetection stage of the system.

An example may help support this simplified analysis of the problem. First, let us assume that we use a 1-D array of detectors in an optical spectrum analyzer. Suppose that we have a linear array of 4000 detectors. If sixty signals are encountered during a frame time of 1 ms, we need a data rate of 81.92 Mbits/s, using a 20-bit word to achieve the desired accuracy. Suppose now that we divide the array into thirty-two subarrays of 125 detectors each. We also assume that we will not encounter more than two signals in each subarray so the chance of saturating our subarray is small. Each subarray will report out the position and size of each signal it encounters. We digitize the signal according to the following rule:

5 bits → label for subarray;

7 bits → label for position in subarray;

20 bits → magnitude of signal (60-dB dynamic range).

Following these rules, the total bit rate for the same sixty signals is only 2.56 Mbits/s, a reduction in the data rate of 42 times. The identification of the pixel within the subarray can be done with a simple counter and would not measurably increase the data rate of the subarray processor. The label of each subarray does not change and involves no processing.

Another optical processing problem is the identification and location of an object in the field of view of a 2-D optical processor. This problem is usually solved by performing a correlation operation, as we discussed earlier. For most complicated signals, the correlation function is nearly a delta function, and we need only locate the position of maximum light intensity. To measure the position of a correlation peak, we determine the intensity and position of the correlation signal. For output signals of this type, a detector based on the lateral photoeffect can be used.<sup>25</sup> The device has a floating *pn* junction and the photovoltage is measured by contacts placed on either side of the junction. A spot of light illuminating the center of the junction produces no voltage, but a spot of light close to one of the contacts produces a voltage proportional to the distance from the center of the junction. The polarity of the photovoltage depends on which side of the center the light spot is found.

The lateral photoeffect detectors are usually designed as a linear or circular detector. Other configurations could be envisioned such as a spiral or a zigzag configuration. More important than the configuration of the detector is the method used where multiple signals may produce multiple correlation peaks. Each signal will produce a correlation peak located at a position corresponding to the position of the signal in the input field of view. Segmentation of the output plane may also be a solution to this problem. Each segment would contain a lateral photoeffect detector. The size of the detector is dictated by the expected size of the pattern and the probability of finding two patterns within a given region.

### (3) Optical In/Optical Out Focal Plane Detectors

Some optical processors are capable of producing outputs with two spatial dimensions and one temporal dimension, which may require spatial sampling with  $1000 \times 1000$  samples and temporal sampling at rates approaching 1 GHz. No practical sensor is foreseen which would be capable of bringing out all the data electronically. A possible solution is to develop a detector in which the output is an optical form to provide the required output bandwidth and parallelism. The optical output detector is envisioned as a 2-D array of elements which respond to light, provide gain, perform additional operations (preferably nonlinear) in response to some external control signal, and retain the signal in an optical form. The optical output provides the opportunity for additional optical processor stages to follow. The optical output may be coherent or incoherent; it could even be of a different wavelength from the input. We see a broad applicability for such devices because they represent a large variety of three-terminal nonlinear optical devices. The avenues of approach to such devices may include research for GaAs/GaAlAs superlattice materials.

## V. Summary and Conclusions

Photodetector arrays have been, and continue to be, designed primarily for imaging applications. Optical signal processing applications place new and more stringent requirements on these devices. We need less blooming, faster recovery, less electrical readout noise, more video output lines, segmented arrays, and so forth. The major improvements, however, must be in dynamic range and in techniques for reducing the output data rates.

We have briefly reviewed typical optical processing applications, described the basic physics of the devices, and have offered some suggestions for improving the dynamic range and for reducing the output data rates. It is our hope that the ideas presented here will result in the development of devices that are better tailored to the needs of optical computing systems.

This paper was the result of an Army Research Office Palantir study. These studies address the physical foundations of approaches to solutions of important technological problems with the aim of stimulating new avenues for progress toward their solution. The participants in the study consisted of the authors of this paper. One of the authors (G.W.A.) wishes to acknowledge partial support by the Department of the Navy and (R.J.K.) the partial support by the Department of the Air Force.

## References

1. A. VanderLugt, "Coherent Optical Processing," Proc. IEEE 62, 1300 (1974).
2. J. W. Goodman, "Operations Achievable with Coherent Optical Information Processing Systems," Proc. IEEE 65, 39 (1977).
3. Special Issue on Optical Computing, Proc. IEEE 65 (Jan. 1977).
4. Special Section on Acousto-optic Signal Processing, Proc. IEEE 69 (Jan. 1981).
5. Special Issue on Optical Computing, Proc. IEEE 72 (July 1984).

6. Special Section on Optical Pattern Recognition, *Opt. Eng.* **23**, (Nov./Dec. 1984).
7. A. VanderLugt, "Operational Notation for the Analysis and Synthesis of Optical Data Processing Systems," *Proc. IEEE* **54**, 1055 (1966).
8. N. George, J. Thomasson, and A. Spindel, "Photodetector for Real Time Pattern Recognition," U.S. Patent 3,689,772 (1970).
9. J. T. Thomasson, T. J. Middleton, and N. Jensen, *Proc. Soc. Photo-Opt. Instrum. Eng.* **45**, 257 (1974).
10. G. Lukes, *Proc. Soc. Photo-Opt. Instrum. Eng.* **45**, 265 (1974).
11. A. Korpel, "Acousto-optics—A Review of Fundamentals," *Proc. IEEE* **69**, 48 (1981).
12. E. H. Young, Jr., and S-K Yao, "Design Considerations for Acousto-optics Devices," *Proc. IEEE* **69**, 54 (1981).
13. R. Doyle and W. Glenn, "Remote Real-time Reconstruction of Holograms Using the Lumatron," *Appl. Opt.* **11**, 1261 (1972).
14. "Real-time Modulator for Coherent Optical Processing," Final Report AFAL-TR-73-88, Environmental Institute of Michigan (May 1973).
15. See, for example, W. T. Rhodes, "Acousto Optic Signal Processing: Convolution and Correlation," *Proc. IEEE* **69**, 65 (1981).
16. T. M. Turpin, "Spectrum Analysis Using Optical Processing," *Proc. IEEE* **69**, 79 (1981).
17. A. VanderLugt, "Interferometric Spectrum Analyzer," *Appl. Opt.* **20**, 2770 (1981).
18. A. VanderLugt and A. M. Bardos, "Spatial and Temporal Spectra of Periodic Functions for Spectrum Analysis," *Appl. Opt.* **23**, 4269 (1984).
19. D. F. Barbe, "Imaging Devices Using the Charge-Coupled Concept," *Proc. IEEE* **63**, 38 (1975).
20. G. M. Borsuk, "Photodetectors for Acousto-optic Signal Processors," *Proc. IEEE* **69**, 100 (1981).
21. G. M. Borsuk, G. W. Anderson, and F. J. Kub, "Photodetectors for Acousto-optic Signal Processing," *Proc. Soc. Photo-Opt. Instrum. Eng.* **639**, 2 (1986).
22. S. G. Chamberlain and J. P. Y. Lee, "A Novel Wide Dynamic Range Silicon Photodetector and Linear Imaging Array," *IEEE Trans. Electron. Devices* **ED-31**, 175 (1984).
23. M. H. White, D. R. Lampe, F. C. Blaha, and I. A. Mack, "Characterization of Surface Channel CCD Image Arrays at Low Light Levels," *IEEE J. Solid-State Circuits* **SC-9**, 1 (1974).
24. MICROTRAC, Leeds and Northrup, North Wales, Penn. and Particle and Droplet-Size Distribution Analyzer, Type ST 1800, Malvern Instruments, Malvern, England.
25. J. T. Wallmark, *Proc. IRE* **45**, 474 (1957); J. I. Alferov, V. M. Andreev, E. L. Portnoi, and I. I. Protasov, *Sov. Phys. Semicond.* **3**, 1103 (1970).

---

**APPENDIX E**

**ACOUSTO-OPTIC PHOTONIC SWITCH**

**Reprinted from Optics Letters**

**Volume 14, Pages 1177-1179, 1 November 1989**

---

# Acousto-optic photonic switch

Dan Owen Harris and A. VanderLugt

Department of Electrical and Computer Engineering, North Carolina State University, Daniels Hall, Box 7911, Raleigh, North Carolina 27695

Received June 19, 1989; accepted August 18, 1989

An acousto-optic architecture is presented to implement a nonblocking space-division switch with  $O(N)$  complexity. Signal degradation is minimal so that the switch is suitable for nonregenerative application within optical networks; it is also capable of rapid reconfiguration. Experiments for a  $1 \times 4$  switch show an insertion loss ranging from 4.6 to 5.6 dB, a worst-case signal-to-cross-talk ratio of better than 30 dB, and a reconfiguration time of 1.46  $\mu\text{sec}$ .

Photonic space-division switches can be classified into three distinct categories: planar, masking, and deflecting. Planar switches are based on the multistage networks used in electronic switching, substituting single-element spatial light modulators as optical crosspoints.<sup>1,2</sup> These switches are blocking and have  $O(N \log N)$  complexity. In addition, insertion loss and optical cross talk become prohibitively large for large  $N$ , and broadcasting is difficult. In a masking architecture each beam in an input-port array is spread to span an entire output-port array. A mask, typically some two-dimensional spatial-light-modulator array, is dynamically configured to regulate transmission of the source beams to specified output ports.<sup>3,4</sup> This architecture is nonblocking and provides broadcast capability. Its complexity, however, is  $O(N^2)$ , and insertion loss and optical cross talk become prohibitively large for large  $N$ . In deflecting architectures, light is steered to the proper output ports by an array of deflectors; the deflectors can either be reflective surfaces or diffraction gratings. Since deflected beams may pass through each other without interacting, and only one deflector is required per input port, a nonblocking switch can be realized with  $O(N)$  complexity. With these architectures, insertion loss and optical cross talk can be small even for large  $N$ . Also, we can broadcast information with diffracting switches by superimposing or space multiplexing gratings that provide the proper point-to-point connections.

We report here on the implementation of a deflecting photonic switch based on acousto-optic diffraction. In our configuration, shown in Fig. 1, light from an input fiber collimated along the  $z$  axis interacts with one channel of a multichannel cell. An acoustic wave in the cell, created by the application of a rf signal to the piezoelectric transducer, deflects the light at an angle in the  $x$ - $z$  plane that is proportional to the rf frequency. To access a given port, we set the rf signal to the appropriate frequency. The deflected beams reach the output fibers by means of a Fourier-transform lens.

This general architecture was proposed in 1986 by VanderLugt, who developed architectures based on the Fresnel transform, the Fourier transform, and a

hybrid Fresnel/Fourier transform.<sup>5</sup> Huang *et al.* recently reported experimental performance of the Fresnel implementation.<sup>6</sup> The Huang implementation is nonblocking, provides multicasting capability, and has a fast reconfiguration time; however, two significant problems exist. First, the addressing is accomplished through the use of an internal,  $N^2$ -complexity rf switch; this negates the desired  $O(N)$  complexity throughout the system. Second, the insertion loss in the switch is high, ranging from 12.6 to 17.6 dB, depending on the operating conditions. This loss is predominantly due to 9 dB of mode mismatch loss between the input and output fibers.

In our implementation the rf signals used for addressing are created by dedicated digital-frequency synthesizers. Since no internal switch is needed to regulate the rf inputs and only one synthesizer is needed per input port, we preserve the  $O(N)$  complexity throughout the deflecting architecture. Our analyses show that a Fourier-transform system reduces the mode mismatch loss significantly compared with the Fresnel-transform implementation. Therefore the fiber-to-fiber coupling is much more efficient, and this

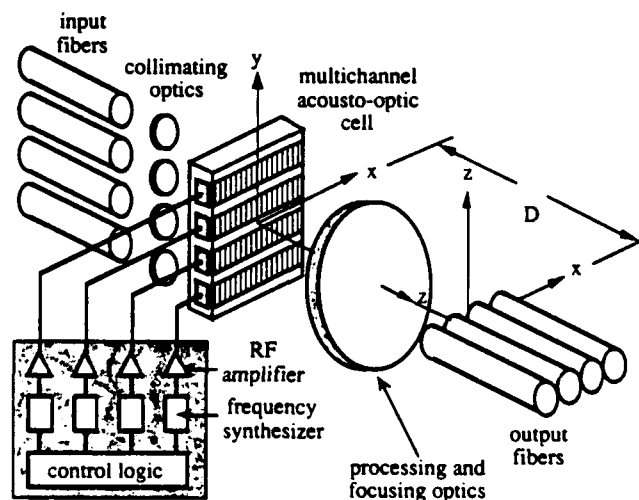


Fig. 1. Proposed acousto-optic photonic switch architecture for a  $4 \times 4$  implementation.

configuration is appropriate for use in nonregenerating applications.

Diffraction is caused by the finite extent of the deflected beam, which manifests itself as a spreading of the profile. As the extent of the profiles enlarge, the beams overlap, causing optical cross talk that fundamentally limits the number of distinguishable output ports. For most applications, a signal-to-cross-talk ratio of 30 dB should be sufficient. A straightforward derivation shows that the worst-case signal-to-cross-talk ratio SCR can be approximated by

$$\text{SCR} \approx \frac{1}{2} \frac{P_{ii}}{\sum_{k \neq i} P_{ik}}, \quad (1)$$

where  $P_{ii}$  is the signal power at the  $i$ th output port and  $P_{ik}$  is the cross talk between the  $i$ th and  $k$ th ports.

In our cross-talk calculations we represent the collimated beam profiles by the truncated Gaussian amplitude function  $a(x) = \exp[-2A(x/L)^2] \text{rect}(x/L)$ , where  $L$  is the length of the acousto-optic cell. We calculated cross-talk levels at various distances  $D$  from the acousto-optic cell for apertures characterized by  $A = 2, 4, 8$ , and  $16$ . For a 30-dB signal-to-cross-talk ratio in a diffraction-limited system, we found that far-field diffraction with  $A = 8$  maximized the number of distinguishable ports.

In our configuration we collimate light from the input fibers with an array of lenses, each having a focal length  $F_1$ . A single lens of focal length  $F_2$  creates the Fourier transform of the deflected light, which, in effect, results in an inverted image of the appropriate input fiber at the desired output port. The transverse magnification of this system is  $M_T = F_2/F_1$ ; efficient coupling can be made between similar fibers by setting  $F_1 = F_2$ . For applications where the input fiber core is smaller than that of the output fiber, we may increase the magnification of the system proportionally without incurring significant spot-size or numerical-aperture-related loss.

The causes of insertion loss in our switch include fiber-to-fiber coupling loss and inefficiency in acousto-optic diffraction. The mode mismatch loss can be estimated using well-known optical fiber splice loss theory; reflection and restricted-aperture related losses must be calculated independently. Acousto-optic cell diffraction efficiencies for cells with a moderate bandwidth  $W$  range as high as 50–80%, while efficiencies for high-bandwidth cells are approximately 10%.<sup>7</sup> We also need to consider losses imparted by our power-combining scheme at the fiber outputs; depending on the application, this loss could range from negligible to a worst case of  $1/N$ .

For closely packed fiber outputs, the focal length of the transform lens is governed by

$$F_2 = \frac{v(N-1)s}{\lambda W}, \quad (2)$$

where  $v$  is the acoustic velocity in the cell,  $s$  is the fiber cladding diameter, and  $\lambda$  is the wavelength of light. We divide Eq. (2) by  $F_1 = L/2NA$ , where  $NA$  is the numerical aperture contained within the  $e^{-8}$  point on

the far-field pattern of the input fiber, and rearrange to get

$$N - 1 = \frac{\lambda F_2}{2NA s F_1} \frac{LW}{v} = \frac{TW}{\rho_s}. \quad (3)$$

Here  $T = L/v$  is the transit time of the cell and  $\rho_s = s/d_0$ , where  $d_0 = \lambda M_T/2NA$  is the width of a Rayleigh resolution element in the output plane. For given input and output fiber geometries, the number of ports may be increased by increasing the cell bandwidth or the transit time. Note that for an output-port duty cycle of one half in a diffraction-limited system,  $\rho_s$  must be greater than 4 to yield a worst-case signal-to-cross-talk ratio greater than 30 dB.

Finally, we consider the reconfiguration time, which is a combination of frequency synthesizer setup time and acousto-optic cell transit time. The synthesizers require 145 nsec to change the rf frequency. The transit time is the time interval required for the cross talk to settle to its designated value and is calculated using

$$T = \frac{L}{L_r} T_r, \quad (4)$$

where  $T_r$  is the reconfiguration rise time, i.e., the time required for the output power to go from 10% to 90% of its final value, and  $L_r$  is the cross-sectional width of the collimated beam that contains 80% of the optical intensity. Note that since the number of output ports is proportional to  $TW$ , the transit time may be reduced by simply increasing the cell bandwidth.

The experimental system, shown in Fig. 2, emulates a  $1 \times 4$  building block with output ports consisting of a closely packed linear array of 62.5/125 graded-index fibers. The input light has a wavelength of 633 nm and is delivered to the acousto-optic cell by a fiber designed to be single mode at 1300 nm. In terms of fiber types and operating wavelength, our system closely resembles that used by Huang, so direct comparisons of loss and cross talk are appropriate. The light exiting this fiber is collimated by a 16-mm focal-length lens into a beam with  $A = 8$  and  $L = 5.1$  mm. The acousto-optic cell is made from flint glass and has

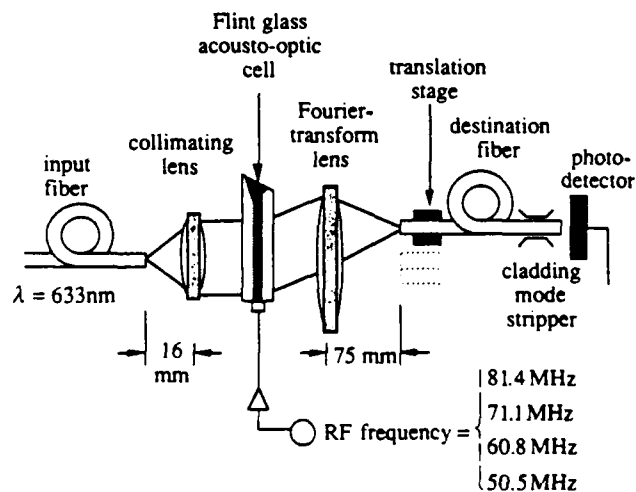


Fig. 2. Top view of the apparatus used in the  $1 \times 4$  acousto-optic switch experiment.

**Table 1. Experimental Insertion Loss and Worst-Case Signal-to-Cross-Talk Ratio (SCR) for the  $1 \times 4$  Acousto-Optic Photonic Switch**

Output Port	$\eta$ (dB)	Insertion Loss (dB)	SCR (dB)
1	1.6	4.8	33
2	2.5	5.6	30
3	1.2	4.6	31
4	1.2	4.6	32

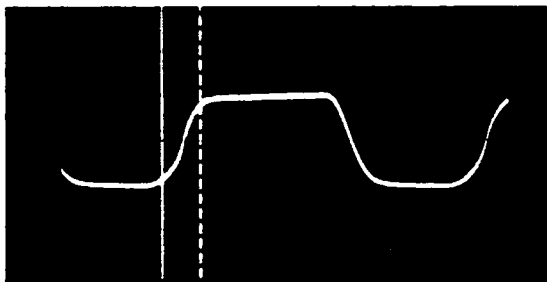


Fig. 3. Rise time of the  $1 \times 4$  acousto-optic switch.

a center frequency of 70 MHz, a bandwidth of 40 MHz, a typical diffraction efficiency of 0.8 at 633 nm, and an acoustic velocity of 3.9 km/sec. Driver frequencies for the four outputs are 50.5, 60.8, 71.1, and 81.4 MHz, utilizing 30.9 MHz of the cell bandwidth. A transform lens with a focal length of 75 mm, combined with the rf frequency resolution, results in an output-port separation of  $125.3 \mu\text{m}$ . A single fiber is located in the Fourier-transform plane; it is moved to various locations along the  $\xi$  axis, modeling the placement of the four output ports. The length of the output fiber is approximately 2 m, so a glycerine-based cladding mode stripper is used at the exit end.

The results of loss and cross-talk measurements are summarized in Table 1. They indicate losses ranging from 4.6 to 5.6 dB, values that are relatively low. Mode mismatch loss in this experiment was less than 1.2 dB for all output ports, in accordance with our theory. In this particular configuration, we incurred unnecessary losses because the focal length of the collimating lens was not optimized, causing light loss by the acousto-optic cell aperture. Previous experimen-

tal research in which we used laser sources directly showed that total insertion losses below 3 dB can be achieved if these aperture losses are avoided. Worst-case signal-to-cross-talk ratios were calculated using relation (1). They exceeded 30 dB for every port, which is consistent with the theory.

To determine the reconfiguration rise time, we square-wave modulated the rf input, then measured the rise time of the photodetector output as shown in Fig. 3. We found this rise time to be approximately 425 nsec, while  $L_r$  was approximately 1.65 mm. Substituting these values into Eq. (4), we find that the transit time is  $1.31 \mu\text{sec}$ , which is exactly the value of  $L/v$ . Additionally, the ratio of  $TW$  to  $N - 1$  is 13.5, while the total reconfiguration time for the switch is  $1.46 \mu\text{sec}$ . Use of an optimum focal-length collimator should reduce the reconfiguration time to approximately  $1 \mu\text{sec}$ .

This acousto-optic architecture provides a means of implementing a nonblocking switch with  $O(N)$  complexity throughout and extremely low insertion loss. It imposes minimal degradation to the optical signals and is suitable for nonregenerative switching within optical networks; it is also capable of rapid reconfiguration. Experimental results for a  $1 \times 4$  switch indicate low insertion loss, signal-to-cross-talk ratios exceeding 30 dB, and a reconfiguration time of the order of  $1 \mu\text{sec}$ .

This research was supported by the U.S. Army Research Office.

## References

1. M. Kondo, Y. Ohta, M. Fujiwara, and M. Sakaguchi, *IEEE J. Quantum Electron.* QE-18, 1759 (1982).
2. K. M. Johnson, M. R. Surette, and J. Shamir, *Appl. Opt.* 27, 1727 (1988).
3. A. A. Sawchuk, B. K. Jenkins, C. S. Raghavendra, and A. Varma, *Computer* 20, 50 (1987).
4. A. R. Dias, R. F. Kalman, J. W. Goodman, and A. A. Sawchuk, *Opt. Eng.* 27, 955 (1988).
5. A. VanderLugt, *Proc. Soc. Photo-Opt. Instrum. Eng.* 634, 51 (1986).
6. P. C. Huang, W. E. Stephens, T. C. Banwell, and L. A. Reith, *Electron. Lett.* 25, 252 (1989).
7. A. Korpel, *Proc. IEEE* 69, 48 (1981).



---

APPENDIX F

ACOUSTO-OPTIC PHOTONIC SWITCH: AN OPTICAL CROSSBAR ARCHITECTURE

Reprinted from Proceedings of the SPIE

Volume 1178, Pages 221-230, 1989

---

## Acousto-optic photonic switch: an optical crossbar architecture

Dan Owen Harris and A. VanderLugt

North Carolina State University, Department of Electrical and Computer Engineering  
Daniels Hall, Box 7911, Raleigh, NC 27695-7911

### ABSTRACT

We describe a crossbar switch architecture based on acousto-optic beam deflection. The architecture has  $O(N)$  hardware complexity throughout, while exhibiting minimal insertion loss, low crosstalk, and fast reconfiguration. Because of the small amount of signal degradation imposed by this switch, it is suitable for nonregenerative applications within fiber-optic networks. By increasing hardware complexity, broadcasting can also be achieved within the framework of this architecture. We report experimental performance of a  $1 \times 4$  switching element. Insertion loss ranges from 2 - 6 dB, worst-case signal-to-crosstalk ratios are in excess of 30 dB, and reconfiguration times are on the order of one microsecond.

### 1. INTRODUCTION

A crossbar switch<sup>1</sup> is a space-division switch characterized by two distinct properties. First, the crossbar provides *full connectivity* within an interconnection network, that is, a potential path exists between each input port and every output port. A more stringent requirement for the crossbar states that no path between an input and any unconnected output port may be blocked by any other input-output connection. A switch of this type is called *nonblocking*. Finally, in a generalized crossbar, it is possible to connect one input to all the output ports simultaneously, effectively *broadcasting* information. A more limited version of this capability is called *multicasting*, in which a simultaneous connection between one input and more than one, but not all, outputs may exist. The switching power of the crossbar does not come without a price, however. With currently used electrical technologies<sup>2</sup>,  $O(N^2)$  crosspoints are required to implement a generalized crossbar, while well-known nonblocking networks require at least  $O(N^{3/2})$  crosspoints. As  $N$  becomes large, these networks become impractical.

Since beams of light pass through each other without interacting, optical technologies are inherently nonblocking; this property of optics may be used to construct crossbar switches with an  $O(N)$  complexity. Photonic switches can also alleviate information flow bottlenecks imposed by electronic switches within optical communication networks. Furthermore, if the loss and crosstalk imposed by the switch are low enough, then potential exists for nonregenerative switching in fiber-optic networks.

### 2. ACOUSTO-OPTIC SWITCHING ARCHITECTURE

Acousto-optic cells have lower losses and faster reconfiguration times than most other spatial light modulator technologies. The deflecting nature of acousto-optic diffraction can be used to produce a nonblocking,  $N \times N$  space-division switch with  $O(N)$  complexity. Also, loss and crosstalk levels in the acousto-optic switch can be small enough for use in nonregenerative applications, even when  $N$  is large.

## 2.1. General architecture

We propose to implement an acousto-optic photonic switch using the architecture shown in Figure 1. In this configuration, light is delivered to the switch by input ports arranged in a linear

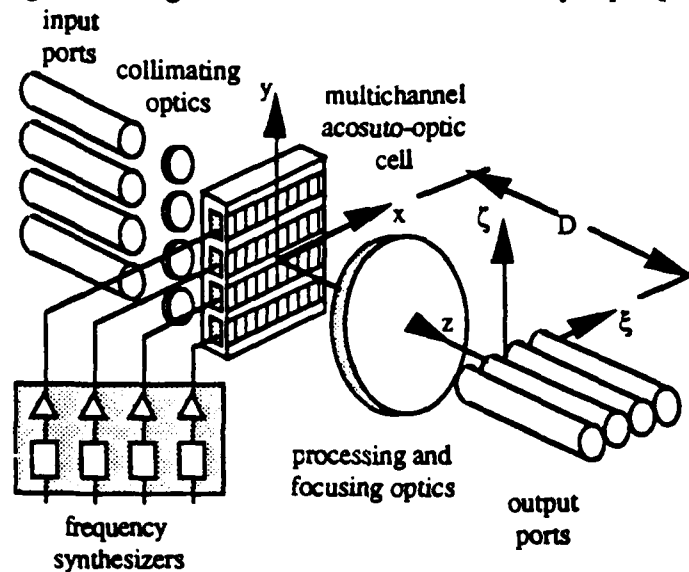


Figure 1. Acousto-optic switching architecture for a 4x4 implementation.

array. Light from each input port is collimated along the z-axis, and each collimated beam interacts with one channel of a multichannel acousto-optic cell. An acoustic wave, created by the application of an RF signal to a piezoelectric transducer, propagates along the x-axis in each channel. The interaction between the light and the acoustic wave results in a portion of the light being deflected in the x-z plane; this deflection angle is proportional to the RF frequency. Therefore, to access a given output port, we simply adjust the RF frequency so that light is deflected to that port.

Since only one acoustic channel is required per input port, and there is never any path contention, the architecture is nonblocking with  $O(N)$  hardware complexity. As  $N$  becomes large, this results in greatly reduced complexity compared to electrical technologies. The total number of input ports that can be served is unlimited in theory; however, commercially available multichannel cells are presently made with up to 128 channels. The number of output ports is limited by the time-bandwidth product of the acousto-optic cell; current cell technologies can support well in excess of 100 output ports.

The RF signals used for addressing are created by dedicated digital frequency synthesizers. Since only one synthesizer is needed per input port for applications where multicasting capability is not present, we preserve the  $O(N)$  complexity of the deflecting architecture. With the synthesizers we have specified, there is a 140 nsec delay between the time control logic requests a change in the RF frequency and the time the change begins to occur; after the transition begins, only 5 nsec of settling time is required. Multicasting can be accomplished by either superimposing continuous waves or space multiplexing pulses of the appropriate carrier frequencies within the acoustic channel. The former method has been demonstrated experimentally<sup>3</sup> and incurs a splitting loss. It also increases hardware complexity, because multicasting to  $M$  outputs requires  $M$  monotone frequency synthesizers per acoustic channel. In addition to the splitting loss, space multiplexing of pulses brings about a  $1/M$  reduction in the number of output ports; although only one frequency synthesizer per acoustic channel is required, hardware and software overhead increase because

some mechanism for periodically changing the RF input to the cell must be present.

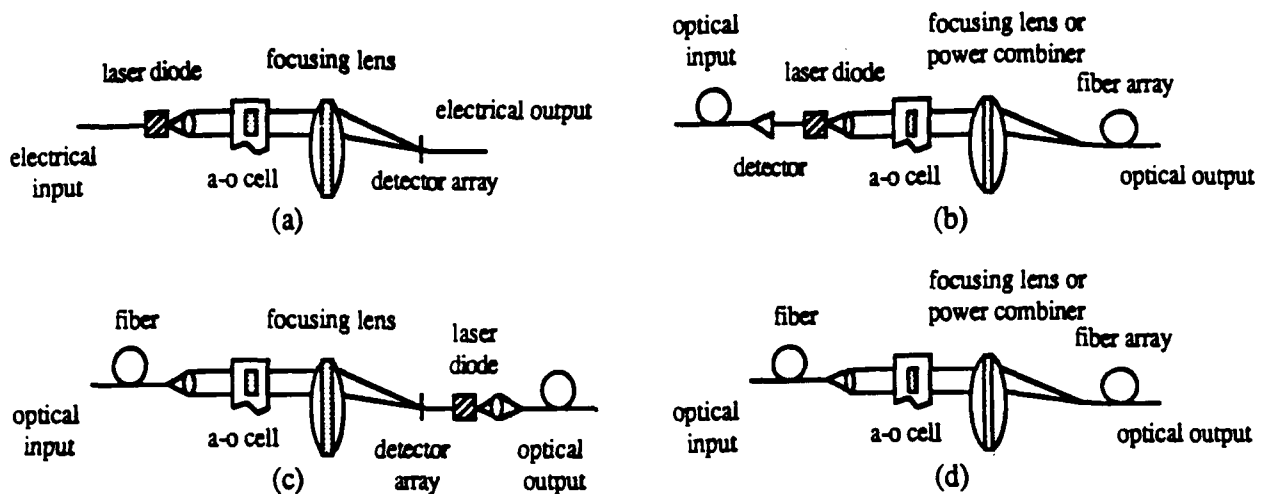


Figure 2. Input and output port configurations.

## 2.2. Input and output port configurations

The acousto-optic photonic switch can be used in a broad range of applications which span many types of physical input and output port configurations. Figure 2 provides a graphical summary of the most practical of these combinations.

Our crossbar switch may also be desirable in some purely electrical communications network applications. Figure 2(a) illustrates this case, where the electrical signal is used to drive a laser diode. In general, packaging of the input optics will be simplified if light from the laser is delivered by an optical fiber; since many commercial laser diodes are equipped with fiber-optic pigtailed, this is easily accomplished. The light from the fiber pigtail is collimated and fed to the acousto-optic cell, then horizontally deflected toward the proper output port. The output light is also vertically focused onto the detector array axis by a lens.

Several input/output port configurations may be used in optical networks. The first, which is diagrammed in Figure 2(b), is appropriate for routing in packet switched networks. Here, the optical input is converted to an electrical signal, the addressing information is read, and a new packet header is generated. The packet is then converted back to an optical signal using a laser diode source. Again, light from the laser is collimated and delivered to the cell, but in this case, light from the cell must be directed onto an output axis composed of optical fibers. In some applications, the limited numerical aperture of the fibers may cause inefficient power coupling if vertical focusing is used, so some type fiber-optic power combiner is in order. Unfortunately, power combiners typically have efficiencies of  $1/N$ ; ultimately, a better means of vertically collapsing beams onto the output fiber axis is needed.

The second scheme is given in Figure 2(c). In this implementation, light is collimated directly from the input fiber and applied to the cell. Addressing information must either be supplied through a parallel electrical network, or header information can be excised by tapping a portion of the optical signal, then delaying it long enough for the necessary processing to take place. A detector array is used at the output, so vertical focusing is again an efficient means of guiding the beams to the output axis. After the light has been converted to electricity by the detector, an electrical-to-optical conversion is necessary before the signal can be fed to the output fiber.

A third possible combination which utilizes photon-to-electron and electron-to-photon conversions is not shown; it is a combination of the schemes shown in Figures 2(b) and 2(c), whereby a set of conversions is used on the input side to extract addressing information, and another set is employed to avoid the numerical aperture coupling problem on the output axis. We should note that all three of the optical network switching schemes we have mentioned thus far are considered to be nonregenerating even though photon/electron conversions are present. This is because the converted signals are sent through the switch in their native form, and hardware to decide individual bit values and generate new, distortionless pulses is not required.

The final optical network configuration is the case where the signal remains in photonic form throughout the switch as shown in Figure 2(d). Here light is supplied by a fiber and deflected to a fiber-optic output array. This configuration can reduce hardware costs, especially in high bit rate applications, because no photon/electron conversions are required. However, both the addressing and output coupling constraints that were described above apply to this configuration.

### 2.3. The 1xN switching element

The NxN switch we have described will be implemented with a multichannel acousto-optic cell. The basic characteristics are similar to those of N identical 1xN switching elements. Each element must be capable of efficiently deflecting light from an input port to an one of an array of detectors or fibers. The basic optical system we use to perform this function is shown in Figure 3. In this system, light from the far-field distribution of the input fiber is collimated by the first lens; here, the distribution of the collimated light is actually the Fourier transform of the light which exits the fiber core. After the collimated light is deflected by the acousto-optic cell, a second lens creates the Fourier transform of the beam, which results in an inverted image of the input fiber at the desired output port. The transverse magnification of this imaging system is given by  $M_T = F_2/F_1$ , where  $F_1$  and  $F_2$  are the focal lengths of the collimating and transform lenses, respectively. The ratio of focal lengths of the two lenses, and consequently, the magnification, is chosen such that the output spot is approximately the same size as the fiber or detector aperture; in this manner, most of the power in the deflected beam will be collected at the output port. We note that combinations of cylindrical lenses may be used if magnifications in the vertical and horizontal directions need to be different, or if elliptical illumination of the acousto-optic cell is required.

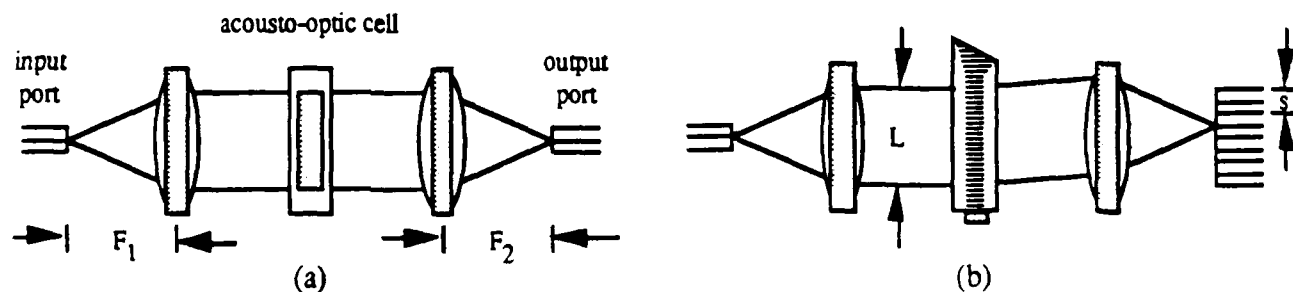


Figure 3. Optical schematic of (a) side and (b) top views of a 1x4 switching element.

## 3. THEORY

VanderLugt has described the basic operation of multichannel acousto-optic switches for both Fresnel and Fourier diffraction<sup>4</sup>. We now investigate the character of crosstalk, insertion loss, and reconfiguration time in the switch, and also develop design procedures based on the 1xN switching element.

### 3.1. Crosstalk

In this switch, spillage of optical power into inappropriate output ports is due to three different effects: optical crosstalk occurs from the diffraction and eventual overlapping of the deflected optical beams, acoustic crosstalk arises from a similar spreading of acoustic waves among channels in the cell, while electrical crosstalk originates from electromagnetic coupling of RF signals in the transducers and output signals in detector arrays. We expect to minimize the effect of acoustic wave spreading by using small input beams interacting with sound near the transducers, while electrical crosstalk can be controlled through proper shielding and isolation. Future research will include a thorough investigation of these effects. At this point, a close examination of the optical crosstalk due to diffraction is in order to help determine the optimum physical configuration of the architecture.

Diffraction occurs in the acousto-optic switch due to the finite extent of the intensity profile of the deflected beam, manifesting itself as a spreading of the profile. As the profiles enlarge, the beams overlap, producing optical crosstalk. Crosstalk may limit the packing density of ports along the output axis and, since the acousto-optic cell has a fixed range of deflection angles, can reduce the maximum number of distinguishable output ports. Optical crosstalk is controlled by altering the diffracted intensity profile; this is done by either modifying the input beam profile or moving the output ports further from or closer to the acousto-optic cell. To find the optimum combination of aperture weighting, output fiber placement, and port separation, we must determine the maximum number of output ports that can be distinguished for a given crosstalk level based on a required signal-to-crosstalk ratio; for most applications, a signal-to-crosstalk ratio of 30 dB should be sufficient. Crosstalk can be modeled as an incoherent summation, and a straightforward derivation shows that the worst-case signal-to-crosstalk ratio can be approximated by

$$SCR \approx \frac{1}{2} \frac{P_{ii}}{\sum_{k \neq i} P_{ik}}, \quad (1)$$

where  $P_{ii}$  represents the signal power at the  $i^{\text{th}}$  output port, and  $P_{ik}$  the crosstalk between the  $i^{\text{th}}$  and  $k^{\text{th}}$  ports. This represents the entirety of additive noise for systems with fiber inputs and outputs. For systems using detectors, the reciprocal of the overall signal-to-noise ratio can be approximated by the sum of the reciprocals of the signal-to-crosstalk ratio and the individual receiver signal-to-noise ratios.

In our crosstalk calculations, input beam profiles were represented by the truncated Gaussian amplitude function  $a(x) = \exp[-2A*(x/L)^2]*\text{rect}(x/L)$ , where  $L$  is the length of the acousto-optic cell; crosstalk levels were calculated for illumination apertures characterized by  $A=2, 4, 8, \text{ and } 16$ . We determined the  $P_{ik}$ 's for various values of  $L$  by calculating the Fresnel intensity pattern for a given weighting and distance  $D$  from the acousto-optic cell, and integrating the pattern over the appropriate region for the output port width and separation. For a 30 dB signal-to-crosstalk ratio in a diffraction limited system, we found that the number of distinguishable output ports increases as we increase  $D$ ; the output port count is maximized when the fibers are placed in the diffraction far-field, which reinforces the use of our Fourier plane imaging system. Also, the required port separation is optimized at  $A=8$ ; for an output duty cycle of one half, the required Fourier transform plane separation is approximately six Rayleigh resolution elements for  $A=4$  and  $A=16$ , and dipping to four when  $A=8$ . Since TW elements can be resolved in the Fourier plane, where TW is the time-bandwidth product of the acousto-optic cell, the maximum number of output ports is given by

$$N_{\max} - 1 = \frac{TW}{4}. \quad (2)$$

From this relation, we see that we can serve more output ports by increasing either the acousto-optic cell bandwidth  $W$  or the transit time  $T$ . Note that  $T=L/v$ , where  $v$  is the velocity of sound in the cell.

### 3.2. Insertion Loss

Losses in the switch are principally caused by reflections, acousto-optic cell diffraction efficiency, and, in the case of fiber-optic output ports, coupling loss. Acousto-optic cell diffraction efficiencies for moderate bandwidth cells ( $W \approx 100$  MHz) can range from 50 - 80%, while diffraction efficiencies for large bandwidth cells ( $W \approx 1000$  MHz) are about 10%.<sup>5</sup> Fiber-to-fiber coupling loss from spotsize and numerical aperture mismatch can be calculated using techniques developed within splice loss theory. We must also consider power combining loss at the output ports; this is application dependent, and can range from being negligible to a worst-case efficiency of  $1/N$ .

### 3.3 Reconfiguration time

Finally, we consider switching speed. We discussed the set-up time of our frequency synthesizers, but a more fundamental limitation of switching speed is the acousto-optic cell reconfiguration time. The time required to transfer the optical power to the output fiber is defined as the interval needed for the power to increase from one-tenth to nine-tenths its final value; this interval is known as the rise time. In terms of the acousto-optic system, this is the time required for one point on the acoustic wave to traverse the cross sectional width of the collimated beam which contains 80 percent of the optical intensity. Recall from previous discussion, however, that we want to truncate the aperture at the  $e^{-8}$  points in order to maximize the number of output ports. If we follow this criterion, rise time and reconfiguration time are not equivalent because, even though the power is effectively transferred within the rise time, the crosstalk will not have settled to its specified levels. Therefore, the true acousto-optic cell reconfiguration time is the transit time  $T$ . Since the rise time is easily measured, it is convenient to calculate the transit time using

$$T = \frac{L}{L_r} T_r, \quad (3)$$

where  $T_r$  represents the rise time and  $L_r$  the corresponding portion of the input aperture.

### 3.4. Switching element design

The focal length of the transform lens is governed by

$$F_2 = \frac{v(N-1)s}{\lambda W}, \quad (4)$$

where  $\lambda$  is the wavelength of light and  $s$  is the spacing between output ports. The focal length of the collimating lens can also be written in terms of optical system parameters as

$$F_1 = \frac{L}{2NA}, \quad (5)$$

where  $NA$  is the numerical aperture contained within the  $e^{-8}$  point on the far-field pattern of the input fiber. From this expression, we see that  $F_1$  should be chosen to be as small as possible in order to minimize  $L$ , and ultimately, reconfiguration time. Also, for multimode fiber inputs,  $F_1$  must be large enough to provide adequate collimation of the input light; this is necessary because poor collimation reduces acousto-optic cell diffraction efficiency. If we divide equation (4) by equation (5) and rearrange, we find that

$$N - 1 = \frac{\lambda F_2}{2NA s F_1} \frac{LW}{v} = \frac{TW}{\rho_s}. \quad (6)$$

Here,  $\rho_s = s/d_0$ , where  $d_0 = \lambda M_T / 2NA$  is the width of a Rayleigh resolution element in the output plane. Therefore, all we need to know to calculate  $\rho_s$  is the output port separation, the operating wavelength, the numerical aperture of the input fiber, and the maximum transverse magnification that will allow efficient coupling from input to output port. Once this parameter is calculated, we can specify the acousto-optic cell bandwidth that will provide the desired combination of output port count and reconfiguration time. We must take care to ensure that  $\rho_s$  is large enough to guarantee the

specified signal-to-crosstalk ratio. After the appropriate cell bandwidth has been determined, the focal lengths of the two lenses are calculated using equation (4) and the expression for transverse magnification.

#### 4. EXPERIMENTAL PERFORMANCE

We conducted experiments to verify the predicted performance of our acousto-optic photonic switch architecture. In these experiments, we make measurements to determine the insertion loss, worst-case signal-to-crosstalk ratio, and acousto-optic cell reconfiguration time for a switch with a small core input fiber and larger core output fibers.

The experimental system, shown in Figure 4, emulates a single input switch with four output ports consisting of a close packed linear array of 62.5/125 graded index fibers. The acousto-optic cell is made from flint glass and has a center frequency of 70 MHz, bandwidth of 40 MHz, typical diffraction efficiency of 80% at 633 nm, and acoustic velocity of 3.9 km/sec. It is driven with RF signals of 50.5, 60.8, 71.1, and 81.4 MHz (which deflect light to output ports 1, 2, 3, and 4, respectively), making use of 30.9 MHz of the available cell bandwidth. A transform lens with a focal length of 75 mm, when combined with the RF frequency resolution, results in an output port separation of  $125.3 \mu$ . A single fiber is located in the Fourier transform plane so that it can be moved to various locations along the  $\xi$ -axis, accurately modeling the placement of each of the four

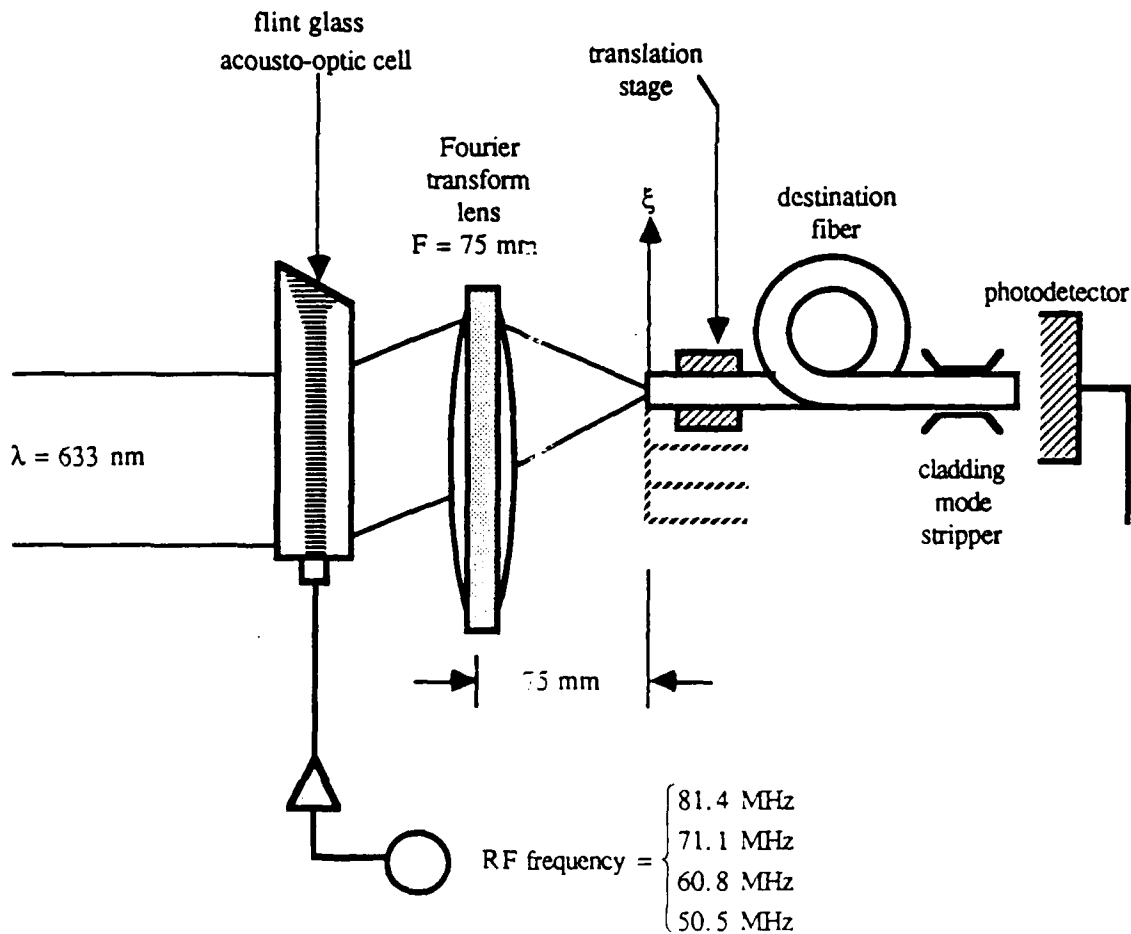


Figure 4. Apparatus used in 1x4 experiments.



output ports. The length of the fiber used in this experiment is approximately two meters; a glycerine-based cladding mode stripper is used at the exit end of the fiber to ensure that all optical power falling on the photodetector originates from the fiber core.

The input light has a wavelength of 633 nm; we use visible light in our experiments so that we may more easily identify any corruption of the output beams. We perform this experiment for two different input scenarios. In Experiment 1, we deliver light to the acousto-optic cell by a fiber designed to be singlemode at 1300 nm. The light leaving this fiber is collimated by a 16 mm focal length lens into a beam characterized by  $A=8$  and  $L=5.1$  mm. For Experiment 2, we took the beam directly from our laser and rendered it into a collimated beam characterized by  $A=8$  and  $L=3.2$  mm. This was intended to simulate light which is collimated from a 633 nm singlemode fiber by a 10 mm focal length lens.

Loss and crosstalk are derived from power measurements from all of the output ports for the case when each port is addressed. These results are summarized in Table I. For the first experiment, insertion losses ranged from 4 - 6 dB, while in the second, they ranged from 2 - 4 dB; in both experiments, the losses are exceptionally low. Losses in the first experiment were higher than in the second for several reasons. First, about 1 dB was lost in 16 mm collimation lens in the first experiment, while in the second experiment, loss between the laser and acousto-optic cell was negligible. Second, the diffraction efficiency was slightly better in the second experiment; since the transducer height is of the acousto-optic cell is only 2 mm, a higher percentage of the incident beam passes through the sound column in the second experiment. Finally, the coupling efficiency of the output beam to the fiber is about 0.5 dB better in the second experiment because the f-number of the output beam is less in the second experiment, and the loss from numerical aperture mismatch is reduced. The the combination of spotsize and numerical aperture mismatch loss experienced by both our systems was minimal, ranging from 0.5 - 1.5 dB.

Worst-case signal-to crosstalk ratios were calculated using equation (1). The crosstalk for Experiment 1 is nearly what we expect from a Gaussian beam, while those seen in the second experiment were slightly higher than predicted. We measured the profile of the laser used in the experiment and found that the shape was not as smooth as a Gaussian; this roughness causes the sidelobes in the diffraction pattern to increase, resulting in higher crosstalk. We feel that the signal-to-crosstalk ratios obtained in Experiment 1 are most representative of our switching

output port	Experiment 1			Experiment 2		
	$\eta$ (dB)	Insertion Loss (dB)	SCR (dB)	$\eta$ (dB)	Insertion Loss (dB)	SCR (dB)
1	1.6	4.8	33	1.1	2.6	29
2	2.5	5.6	30	2.2	3.6	26
3	1.2	4.6	31	0.9	2.4	27
4	1.2	4.6	32	0.8	2.4	30

Table I. Measured insertion loss and worst-case signal-to-crosstalk ratios for the 1x4 experiments.



Figure 5. Acousto-optic cell rise times for (a) Experiment 1 and (b) Experiment 2.

architecture, since they are derived from a Gaussian beam. These values are less than 30 dB for all ports.

The reconfiguration time was calculated using the rise time of the acousto-optic cell system. We determined the rise time of the system by modulating the RF driver signal with a square wave, displaying the time-varying output of the photodetector on an oscilloscope, and measuring the rise time from the display. From Figure 5, we see that this rise time is approximately 425 nsec for Experiment 1 and 275 nsec for Experiment 2; we also found  $L_T$  to be 1.65 mm and 1.00 mm for Experiments 1 and 2, respectively. Substituting these values into equation (3), we may deduce that the transit time is 1.31  $\mu$ sec for Experiment 1 and 880 nsec for Experiment 2, which compares well with their respective 1.31  $\mu$ sec and 820 nsec values of  $L/v$ . Total reconfiguration times for the first and second experiments would be 1.46 and 1.03  $\mu$ sec, respectively.

## 5. CONCLUSION

The acousto-optic photonic switch we propose is the only architecture we know that can be used to implement a nonblocking switch with just  $O(N)$  complexity throughout. Assuming efficient output power combination can be achieved, it imposes minimal degradation to the optical signals which pass through it, and is suitable for nonregenerative switching within optical networks; it also capable of rapid reconfiguration. We conducted an experiment to verify insertion loss, worst-case optical crosstalk, and acousto-optic cell reconfiguration time for a 1x4 switching element with singlemode fiber input and multimode fiber output. Insertion loss for the switch was approximately 2-6 dB, worst-case signal-to-crosstalk ratio was found to exceed 30 dB, and acousto-optic cell reconfiguration time was on the order of a microsecond.

## 6. ACKNOWLEDGEMENTS

We acknowledge helpful discussions with C. S. Anderson, K. A. Emig, J. R. Jones, and A. A. Nilsson, and thank S. L. Hassett, S. R. Stumpf, and T. W. Whitehead for supplying optical fibers and ancillary equipment used in the experiments. This work was supported by the U. S. Army Research Office.

## 7. REFERENCES

1. A. A. Sawchuk, B. K. Jenkins, C. S. Raghavendra, and A. Varma, "Optical crossbar networks", *IEEE Computer Mag.*, Vol. 20, p. 50 (1987).
2. B. E. Briley, *Introduction to Telephone Switching*. Reading, MA: Addison-Wesley Publishing Co., Inc. (1983).
3. P. C. Huang, W. E. Stephens, T. C. Banwell, and L. A. Reith, "Performance of a 4x4 optical crossbar utilising acousto-optic deflector", *Electron. Lett.*, Vol. 25, p. 252 (1989).
4. A. VanderLugt, "Fresnel transforms and optical computing", SPIE, *Optical and Hybrid Computing*, Vol. 634, p. 51 (1986).
5. A. Korpel, "Acousto-optics — a review of the fundamentals" *Proc. IEEE*, Vol. 69, p. 48 (1981).

---

APPENDIX G

ACOUSTO-OPTIC PHOTONIC SWITCH

Reprinted from Optics News

Volume 27, Pages 49-50, 1989

---

It was shown that the lift-off technique does not degrade the bandwidth and the quantum efficiency of In-GaAs/InP p-i-n photodiodes.<sup>5</sup> The lifted-off devices, with a 2  $\mu\text{m}$  absorption layer thickness and a  $(24 \times 24) \mu\text{m}^2$  mesa area, showed a bandwidth of 13.5 GHz and an internal quantum efficiency of 90% at 1.3  $\mu\text{m}$  wavelength. These results are similar to those obtained on devices of the same dimensions that were not lifted-off.

The transfer of semiconductor devices without degradation in their performance leads to the exciting possibility of integrating high performance photodetectors with optical waveguides in materials other than semiconductors. We fabricated GaAs MSM photodetectors on both glass and LiNbO<sub>3</sub> waveguides (see figure), and detected a photocurrent in response to light in the waveguide, demonstrating the optical coupling between the two.<sup>6-7</sup> The coupling, which was not optimized by waveguide and detector design, gave rise to an absorption coefficient of 40  $\text{cm}^{-1}$  for a guided mode in a proton-exchanged LiNbO<sub>3</sub> waveguide; similar results were obtained on ion-exchanged glass waveguides. The ability to obtain good optical coupling between a transferred semiconductor device and an underlying waveguide lays the foundation for a new generation of integrated devices. Potential applications of this new technique have only begun to be explored.

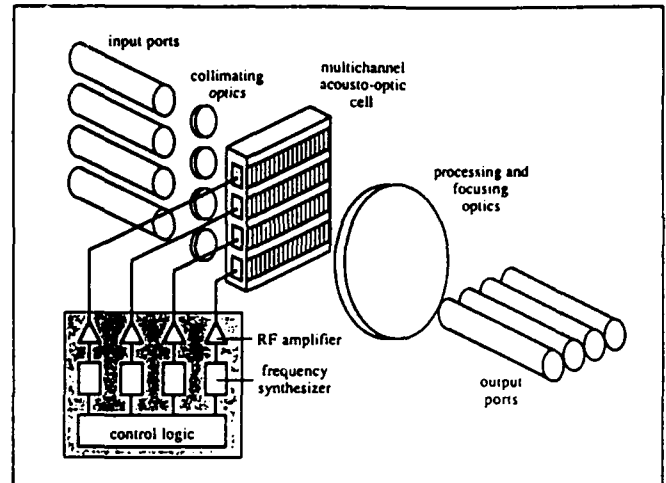
#### REFERENCES

1. W.K. Chan, J.H. Abeles, K.C. Nguyen, R. Bhat, and M.A. Koza, *IEEE Photonics Technol. Letts.*, 1, 65, 1989.
2. H. Schumacher, H.P. LeBlanc, J. Soole, and R. Bhat, *IEEE Electron Dev. Lett.*, 9, 607, 1988.
3. J.B.D. Soole, H. Schumacher, R. Esagui, H.P. LeBlanc, R. Bhat, and M.A. Koza, to be published, *Appl. Phys. Letts.*
4. E. Yablonovitch, T.J. Gmitter, J.P. Harbison, and R. Bhat, *Appl. Phys. Letts.* 51, 2222, 1987.
5. H. Schumacher, T.J. Gmitter, H.P. LeBlanc, R. Bhat, E. Yablonovitch, and M.A. Koza, *Device Research Conference*, Boston, June 1989.
6. W.K. Chan, A. Yi-Yan, T.J. Gmitter, L.T. Florez, J.L. Jackel, E. Yablonovitch, R. Bhat, and J.P. Harbison, *Device Research Conference*, Boston, June 1989.
7. A. Yi-Yan, W.K. Chan, T.J. Gmitter, L.T. Florez, J.L. Jackel, E. Yablonovitch, R. Bhat, and J.P. Harbison, *IEEE Photonic Technol. Lett.* 1, 379, 1989.

## Acousto-optic photonic switch

*D.O. Harris and A. VanderLugt,  
North Carolina State University*

In recent years, there has been interest in developing photonic switches.<sup>1-2</sup> A promising class of space-division architectures is based on dynamic beam steering, where input light is deflected to one of several output



*A 4 × 4 acousto-optic photonic switch.*

ports. Such an architecture uses the three-dimensional processing capability of optics, allowing nonblocking  $N \times N$  switches to be constructed with only  $N$  deflectors. In addition to the low hardware complexity, low losses are possible even for large switches because light must encounter only one deflector, regardless of the number of output ports. Optical crosstalk is due to diffractive spreading of output beams in this type of switch and, typically, only nearest neighbor contributions are significant; therefore, crosstalk can also be limited to low levels when  $N$  is large.

We have proposed and demonstrated the performance of a deflecting photonic switch based on acousto-optic technology.<sup>3</sup> The switch has good loss and crosstalk characteristics and is capable of rapid reconfiguration. In this architecture (see figure), light from a vertical array of input fibers is collimated, with each of the collimated beams interacting with a single channel in a multichannel acousto-optic (AO) cell. An acoustic wave, created by the application of an RF signal to the AO channel, induces a horizontal linear shift in the spatial phase of the light beam. This phase shifted beam then passes through a Fourier transform lens, which creates a horizontally offset image of the input fiber in the output plane. The amount of horizontal offset in the image of the input fiber is equal to the slope of the linear phase shift, which, in turn, is proportional to the frequency of the applied RF signal. Since the output ports are arranged in a horizontal line, any output port can be accessed through proper selection of the RF frequency.

In this new approach, a dedicated digital frequency synthesizer is used to drive each of the AO channels. Since the switch is nonblocking and only one hardware element is required per input port, the device implements a normal crossbar with an unprecedented hardware complexity of

N. Also, since an imaging system is used to couple light from input fiber to output, efficient power transfer can be facilitated through proper selection of the transverse magnification. The transform lens will automatically focus the light vertically and can be used to direct power onto the output axis.

This configuration is convenient, since only one lens is required in the output optics of the entire  $N \times N$  switch. For applications where output fibers have insufficient numerical aperture to collect the light, we can use a conventional photodetector regeneration technique or use a separate transform lens for each input port and an  $N \times 1$  fiber-optic power combiner.

We have verified the performance of a  $1 \times 4$  switching element for an  $8 \mu\text{m}$  core input fiber and  $62.5 \mu\text{m}$  core output fibers. For an optical wavelength of  $633 \text{ nm}$ , insertion loss ranges from 2–6 dB and worst-case signal-to-crosstalk ratios are about 30 dB; for reasons outlined above, the loss and crosstalk do not increase significantly when the number of ports is increased. Experimental switching times are about  $1 \mu\text{sec}$  when 30.9 MHz of RF bandwidth is used; this  $1 \mu\text{sec}$  reconfiguration time can be maintained for more output ports by increasing the RF bandwidth of the AO cell by about 10 MHz for each additional port.

#### REFERENCES

1. A.A. Sawchuck, B.K. Jenkins, C.S. Raghavendra, and A. Varma, Optical crossbar networks, *Computer*, Vol. 20, No. 6, pp. 50-60, 1987.
2. S.D. Personick, Photonic switching: technology and applications, *IEEE Comm. Mag.*, Vol. 25, No. 5, pp. 5-9, 1987.
3. D.O. Harris and A. VanderLugt, Acousto-optic photonic switch, *Opt. Lett.*, Vol. 14, pp. 1177-1179, 1989.

## 16-channel heterodyne broadcast network at 155 Mb/sec

*R.E. Wagner, R. Welter, W.B. Sessa,  
and M.W. Maeda, Bellcore*

There has been considerable interest this year in exploring heterodyne reception techniques for use in fiber distribution networks, especially broadband-ISDN networks. These will likely be implemented using conventional fiber technology, but could later be upgraded using heterodyne receiver technology. Heterodyne systems provide several key advantages: the ability to transmit many channels simultaneously, to tune the receiver to select a single channel, and to detect smaller signals. The current thinking is that B-ISDN networks will use a double-star topology, employ a SONET digital transmission standard,

and also include broadcast video.

Three years ago, preliminary experiments with a 10-channel heterodyne system were conducted,<sup>1</sup> using a vibration-isolated laboratory bench and manipulators for aligning each laser. A recent demonstration<sup>2-3</sup> performed in early 1989 concurrently with other similar demonstrations<sup>4-5</sup> takes the technology one step further. It shows the practical feasibility of heterodyne broadcast networks using commercially available components with rack-mounted equipment.

We have demonstrated a heterodyne lightwave broadcast network that transmits 16 digital video channels each modulated at the SONET rate of 155.52 Mb/sec.<sup>2</sup> The 16 signals are transmitted individually to a  $16 \times 16$  fiber star coupler, where they are combined and split into 16 output fibers. A heterodyne receiver at one of the output fibers can be tuned to select a desired video signal.

Each of the transmitters incorporates a commercially available DFB laser package, with the laser devices selected to have linewidths of less than 30 MHz and wavelengths that fall within a 1 nm band centered at 1540 nm. The laser package also includes an optical isolator with 30 dB of isolation, plus a thermoelectric cooler element. The temperatures of the laser packages are adjusted to set the optical frequencies of the individual lasers for a 10 GHz channel spacing. To prevent the optical frequency of one laser from drifting into the assigned frequency band of another, the laser frequencies were locked relative to the 1509.554 nm absorption line of ammonia gas. Frequency control was achieved using a personal computer to monitor a scanning Fabry-Perot interferometer and to set the temperature and drive current of each laser, stabilizing it to within 200 MHz of its assigned optical frequency.

The lasers are modulated in a Frequency-Shifted-Keyed (FSK) format with a special Alternate Mark Inversion (AMI) code. This code was selected to avoid the degradations in system performance that typically occur with DFB lasers due to a non-uniform FM response.

The heterodyne receiver used a balanced polarization-diversity configuration to prevent deep signal fades that would normally be caused by polarization changes at the receiver input. It was designed to accommodate an IF linewidth of 100 MHz, which required each laser to have a linewidth of less than 50 MHz—easily achieved with typical DFB lasers. The receiver was tuned by changing the temperature of the local oscillator laser to select the desired signal.

With 6 dB of margin in the network and 10 dB for the fiber (40 km at 0.25 dB/Km), there is 12 dB remaining for a second star coupler with a 16-way power split, thus potentially serving 256 subscribers simultaneously. Still further splitting was demonstrated by putting an Er-doped

---

APPENDIX H

OPTIMUM SAMPLING OF FRESNEL TRANSFORMS

Submitted to Applied Optics

---

# **Optimum Sampling of Fresnel Transforms**

by

**A. VanderLugt**

**North Carolina State University**

**Electrical and Computer Engineering Department**

**Daniels Hall, Box 7911**

**Raleigh, NC 27695**

## **Abstract**

The opportunity to process signals in domains other than the time or frequency domains arises naturally in coherently illuminated optical systems that produce Fourier transforms. It is well known that  $N$  samples are sufficient to represent the information content in the object, image, and Fourier planes. We extend these results to show that we can accurately represent the signal in any Fresnel plane of a coherently illuminated optical system with exactly  $N$  samples provided that we use a specified sampling technique.



## 1 Introduction

Signal processing operations are generally performed in either the time or the temporal frequency domains. For example, spectrum analysis is a process in which signal features are easily detected using Fourier transforms to display the frequency content of the signal.

Correlation is extensively used in signal detection and can be implemented in either the time or the frequency domain, the choice resting on performance/cost considerations. Sometimes processing is done simultaneously in both domains such as in radar processing to display range and doppler information, or in displaying Wigner-Ville distributions. In optical processing, the equivalents of the time and temporal frequency domains are the spatial and spatial frequency domains of a coherently illuminated optical system.

Processing signals in domains other than the time or frequency domains has not been studied extensively. In an optical Fourier transform systems there is a continuum of domains available for implementing processing operations between the object plane, that contains a time or a spatial signal, and the Fourier transform plane, that contains the corresponding frequency information. Each plane between the object and Fourier planes is a Fresnel transform of the object. The characteristics of the Fresnel transform is a unique mixture of those of the space and frequency distributions; it resembles one or the other depending on the location of the Fresnel plane.

The sampling theorem is an important signal processing tool for characterizing the complexity of signals or systems. For example, Toraldo used the sampling theorem to determine the degrees of freedom in an image and the capacity of an optical channel, as did Linfoot in his study of optical images.<sup>1,2</sup> O'Neill and Walther used sampling theory to study the importance of phase in optical systems,<sup>3,4</sup> and Barakat, at about the same time, applied sampling to optical diffraction theory as a computational tool.<sup>5</sup> These authors applied the sampling theorem to either the image plane or the aperture plane of incoherently illuminated systems. Little attention was given to sampling in the Fourier plane of coherently illuminated systems until the 1970's.<sup>6,7</sup>

---

Marks, Walkup and Hagler subsequently applied similar analyses to space-variant systems.<sup>8</sup>

Since there is no equivalent of a fixed time base in optical systems, the image size and the required sample spacing are functions of magnification. The optical invariant, as derived in geometrical optics, accounts for system magnification and guarantees that the information in the image is the same as that for the object, if the system is free of aberrations. Hence, if  $N$  samples (or pixels) are sufficient to accurately sample the object, we know that  $N$  samples, whose size and spacing are scaled by the magnification, are also sufficient to sample the image.

The question arises as to whether  $N$  samples are sufficient to accurately measure the intensity of a Fresnel transform. In this paper we show that a fixed number of samples is sufficient to measure the Fresnel transform at any plane, but only under the condition that the sampling is nonuniformly distributed in space. The detailed sampling distribution is dependent on the bandwidth of the object and the particular Fresnel plane being sampled.

In Section 2 we use a simple optical system to illustrate the basic principles for sampling Fresnel transforms. In Section 3 we develop the optimum sampling distribution for the restricted case of when the information capacity of the system is maximized; such a system images an object at an infinite conjugate plane. In Section 4 we remove this restriction and derive the optimum sampling procedure for the completely general case of a finite conjugate imaging system. In Section 5 we find the maximum and minimum spatial frequencies in any Fresnel transform plane in a generalized imaging system.

## 2 A Fourier Transform System

From the sampling theorem, we know that a signal whose time bandwidth product is  $TW$  can be accurately characterized by  $N=2TW$  samples, where  $T$  is the time duration of the signal and  $W$  is its bandwidth. The sample spacing is therefore  $T_0=T/N=1/(2W)$ . The corresponding notion in an optical system is that the sampling distance  $d_0$  for a bandlimited object having a cutoff spatial frequency  $\alpha_{CO}$  is  $d_0=1/(2\alpha_{CO})$ . If the object has length  $L$ , the number of samples required is  $N=L/d_0$ . For a two dimensional object, the required number of samples is  $N=N_x N_y$ , where  $N_x$  and  $N_y$  are number of samples required in the  $x$  and  $y$  directions.

---

The sampling spacings and spatial frequencies must be associated with the intensities of optical signals, since it is the intensity that is observed and measured. In a coherently illuminated optical system, we deal with amplitude distributions to which we can also attribute spatial frequencies and sample spacings. These quantities are not, however, observable or directly measurable. We assume throughout these analyses that the object plane is coherently illuminated with a collimated beam of monochromatic light. The sampling strategies developed here, however, always refer to intensity functions.

## 2.1 The Fourier Domain

Consider first the Fourier transform system of Figure 1, which we analyze in one dimension. An object  $f(x)$  in the front focal plane  $P_1$  of the lens is illuminated by a plane wave of coherent light. At plane  $P_2$  the information from the object is dispersed by a diffraction process to form the Fresnel transform:<sup>9</sup>

$$g(u) = \int_{-\infty}^{\infty} \text{rect}(x/L) f(x) e^{-j\frac{\pi}{\lambda F}(u-x)^2} dx, \quad (1)$$

where  $F$  is the distance from plane  $P_1$  to plane  $P_2$  and  $\text{rect}(x/L)$  shows that the object has length  $L$ . The lens multiplies the Fresnel transform by a quadratic phase function  $\exp[j(\pi/\lambda F)u^2]$ , where  $F$  is the focal length of the lens, to produce  $h(u)$  at plane  $P_3$ :

$$h(u) = g(u) e^{j\frac{\pi}{\lambda F}u^2}. \quad (2)$$

A second Fresnel transform similar to (1) is applied to  $h(u)$  and produces the Fourier transform at plane  $P_4$ :<sup>10</sup>

$$\begin{aligned} F(\xi) &= \int_{-\infty}^{\infty} h(u) e^{-j\frac{\pi}{\lambda F}(\xi-u)^2} dx \\ &= \int_{-\infty}^{\infty} \text{rect}(x/L) f(x) e^{j\frac{2\pi}{\lambda F}\xi x} dx. \end{aligned} \quad (3)$$

Although the configuration shown in Figure 1 is the simplest to analyze, the Fourier transform exists under a much wider range of conditions, such as with a different position of plane  $P_1$  or using divergent or convergent illumination.<sup>10</sup>

In the object plane  $P_1$  all information is confined to the region defined by the aperture  $L$  and the sample spacing is  $d_0=L/N$ . In the Fourier plane  $P_4$  all information passes through an aperture

$2\xi_{CO}$ . A previous study shows that the information capacity of the system is optimized when  $2\xi_{CO}=L$ , so that the physical size of the object and Fourier planes are equal.<sup>6</sup> A related study shows that the packing density for a generalized imaging system is highest in either plane  $P_1$  or  $P_4$ , depending on the cutoff frequency  $\alpha_{CO}$ .<sup>7</sup> When the capacity of the system is maximized, the packing density is the same in both planes and the required sample spacing is also equal at both planes. Under these conditions, the lens aperture  $A$ , as shown in Figure 1, transmits both the undiffracted and diffracted light so that  $A=2L$ . Hence, the spatial extent of the Fresnel transform at planes  $P_2$  or  $P_3$  is twice that of either the object or the Fourier transform.

## 2.2 The Fresnel Domains

The increased spatial extent of the Fresnel transform would be of little concern if the sample spacing could be increased correspondingly, so that the number of samples required remains the same; unfortunately this is not the case because the spatial frequency  $\alpha_f$  in some Fresnel planes is at least as large as  $\alpha_{CO}$  in the object plane. To illustrate this point, consider the impulse response  $r(u)$  of the system in the region between planes  $P_1$  and  $P_2$ , as shown in Figure 2:

$$r(u) = \int_{-\infty}^{\infty} d(x) e^{-j\frac{\pi}{\lambda D}(u-x)^2} dx, \quad (4)$$

where  $d(x)$  is the impulse response function for a bandlimited function and  $D$  is the distance from plane  $P_1$  to the plane of observation. This integral cannot be evaluated in closed form but a working approximation for a sampling function  $d(x)=\text{sinc}(x/d_0)$  is that<sup>9</sup>

$$r(u) = \begin{cases} e^{-j\frac{\pi}{\lambda D}u^2} & ; \quad |u| \leq \theta_{CO}D, \\ 0 & ; \quad \text{else.} \end{cases} \quad (5)$$

Thus, the energy due to the impulse response, which is equivalent to a sampling function, is contained within a cone whose apex angle is  $2\theta_{CO}$  as shown in Figure 2. The connection between the physical cutoff angle  $\theta_{CO}$  and the cutoff spatial frequency  $\alpha_{CO}$  is such that the boundaries of the cone represent the rays that pass through  $\xi=\pm\xi_{CO}$  at the Fourier plane of Figure 1.

The proper sampling distribution for an arbitrary Fresnel plane is somewhat complicated to solve in detail because the integrals cannot be solved in closed form, but the solution is simplified by using the approximation given in (5). Consider the spatial frequency at any plane

intermediate to planes  $P_1$  and  $P_2$  generated by a pair of samples separated by  $nd_0$ , where  $1 \leq n \leq (N-1)$ , and centered on the optical axis at  $x=0$ . The object distribution is therefore

$$f(x) = d(x - nd_0 / 2) + d(x + nd_0 / 2), \quad (6)$$

so that the intensity at a plane an arbitrary distance  $D$  from the object is

$$\begin{aligned} I(u) &= \left| \int_{-\infty}^{\infty} f(x) e^{-j \frac{\pi}{\lambda D} (u-x)^2} dx \right|^2 \\ &= \left| \int_{-\infty}^{\infty} d(x - nd_0 / 2) e^{-j \frac{\pi}{\lambda D} (u-x)^2} dx + \int_{-\infty}^{\infty} d(x + nd_0 / 2) e^{-j \frac{\pi}{\lambda D} (u-x)^2} dx \right|^2 \\ &= \left| e^{-j \frac{\pi}{\lambda D} (u - nd_0 / 2)^2} + e^{-j \frac{\pi}{\lambda D} (u + nd_0 / 2)^2} \right|^2 \\ &= 2[1 + \cos(2\pi nd_0 u / \lambda D)]. \end{aligned} \quad (7)$$

From (7) we find that the spatial frequency  $\alpha_f$  of the intensity, in a general Fresnel plane, due to two samples in the object plane is

$$\alpha_f = \frac{nd_0}{\lambda D}. \quad (8)$$

This result is subject to the constraint that  $nd_0 \leq L$ , so that the points lie within the region that defines the object, and the very important constraint that light from the two samples overlap at the Fresnel plane so that interference takes place.

When  $D$  is small, so that the Fresnel plane is near the object plane, the spatial frequency  $\alpha_f$  is produced only by two closely spaced samples because light rays from widely spaced samples do not overlap since  $\theta_{CO}$  has finite values. As the Fresnel plane moves away from the object plane, the separation between the samples that produce the cutoff frequency increases correspondingly. When  $n=N$ , the separation between samples is  $(N-1)d_0=L$  and, by constructing cones for both of the edge sample points of the object in Figure 3, we note that light from these two points overlap only when  $D=F$  so that the maximum spatial frequency, in any Fresnel plane between  $P_1$  and  $P_2$ , is  $\alpha_f=L/\lambda F=\alpha_{CO}$ .

Since the Fresnel diffraction pattern at the lens plane has at least twice the spatial extent as the object, but the same required sample spacing, we need at least  $2N$  samples to accurately measure  $I(u)$ , if we assume that the samples are uniformly spaced. For a two-dimensional object, we require four times the number of samples required to represent the Fresnel transform and this factor of four increases the computational burden on digital postprocessing operations.

### 3 The Optimum Sampling Distribution for a Maximum Capacity System

The key to exploring other sampling strategies for Fresnel transforms lies in finding the distribution of spatial frequencies in any plane of the system. The spatial frequency at any position in the Fresnel plane can be obtained analytically, but it is most easily explained with the aid of a graphical construction. Suppose that the object  $f(x)$  consists of only two samples, one fixed at  $x=+L/2$  while the other is free to assume any other position in  $P_1$  of Figure 1. From (8) we note that the spatial frequency at any plane is proportional to the angle subtended by the two samples as measured from the observation plane.

From Figure 3a we find that the maximum sample spacing of  $(N-1)d_0$  in plane  $P_1$  creates the maximum frequency at the Fresnel plane  $P_2$  but the minimum amount of overlap. The maximum spatial frequency is produced where the *interior* rays cross so that  $\alpha_f=L/\lambda F=\alpha_{CO}$ , but the region of overlap at the Fresnel plane is found from straight forward geometrical calculation as only  $R=L-(N-1)d_0=d_0$ . From Figure 3b, we find that the minimum sample spacing of  $d_0$  in plane  $P_1$  creates the minimum frequency at the Fresnel plane  $P_2$ , in accordance with (8). The amount of overlap at the Fresnel plane is, however, at a maximum value. The spatial frequency is  $\alpha_f=d_0/\lambda F$  and the region of overlap is  $R=L-d_0$ .

The distribution of spatial frequencies throughout out the Fresnel plane is therefore a function of the relative separation between samples of the object and the absolute position of the samples in the object. Since a typical object contains samples uniformly distributed throughout plane  $P_1$ , the Fresnel transform is the sum of the contributions from all samples, taken in a pairwise fashion, in the object plane. For example, consider the response in the Fresnel plane to any closely spaced pair of sample points from the object. As we see from Figure 3b, the region

of overlap in the Fresnel plane  $P_2$  is always centrally located opposite the position of the pair of samples. Samples having large separations produce higher frequencies at the Fresnel plane, but the samples are constrained as to a small range of positions near the optical axis in plane  $P_1$ ; hence, the region of overlap in plane  $P_2$  decreases, vanishing when the separation approaches  $L$  as shown in Figure 3a. On the other hand, closely spaced samples can occur nearly anywhere in plane  $P_1$  so that the region of overlap in plane  $P_2$  is large. As a result, low spatial frequencies are present throughout the region  $|u| \leq L$  in plane  $P_2$  but high spatial frequencies exist only near the optical axis in a region for which  $|u| \approx d_0$ .

We now observe an important difference regarding the distribution of spatial frequencies in various planes in the system. In the object plane the highest spatial frequency can occur anywhere; since we have no *a priori* knowledge of its position we must sample the object uniformly. A similar argument holds for the Fourier plane so that the information in the Fourier plane must also be uniformly sampled. However, from the ray diagram associated with Figure 3 for the spatial frequencies in the lens plane, we see that all spatial frequencies are present in the Fresnel plane only at  $u=0$  and that the highest spatial frequency at any position in the Fresnel plane decreases as  $|u| \rightarrow L$ .

The distribution of spatial frequencies as a function of the variable  $u$  in the Fresnel plane is found by a simple extension of the graphical solution given above. Since the range of the positions of a pair of samples in the object plane decreases linearly as the spacing between samples increases, the range of positions in the Fresnel plane containing the corresponding spatial frequency also decreases linearly. Hence the relationship between the maximum spatial frequency found at any spatial position at the lens plane is simply the triangular function

$$\alpha_f(u) = \frac{L}{\lambda F} \left[ 1 - \frac{|u|}{L} \right]. \quad (9)$$

The maximum spatial frequency at the lens plane is therefore  $\alpha_f(0) = L/\lambda F$  when  $u=0$  and is  $\alpha_f(L) = 0$  when  $u = \pm L$ . The fact that spatial frequencies in Fresnel planes are not uniformly distributed as a function of spatial position suggest that we should use a nonuniform sampling distribution at the Fresnel plane.

The optimum sampling distribution at plane  $P_2$  is a dense sample spacing near the optical axis where the spatial frequencies are highest with an increase in the sample spacing as we move away from the axis where the spatial frequencies are lower. From a sampling viewpoint, we can represent the frequency distribution in an arbitrary Fresnel transform by a chirp function

$$c(u) = 1 + \cos\left[\frac{\pi}{\lambda F}(L - u)^2\right], \quad (10)$$

shown in Figure 4a. The frequency distribution of the chirp is identical to that of the Fresnel transform for an arbitrary object. The chirp function has its maximum frequency at  $u=0$  and its minimum frequency at  $u=L$ ; the rate of change is linear as required by (9).

The optimum sampling of the chirp function establishes the optimum sampling distribution for the Fresnel transform at plane  $P_2$  for any object. We apply the Nyquist sampling criterion that requires two samples per highest frequency. Since the highest frequency in the Fresnel plane, as a function of  $u$ , is given by (9) or equivalently by (10), the required sample spacings are a function of the position  $u$ . For the chirp function (10), the optimum sample spacings are shown in Figure 4b, where alternate samples are staggered in the vertical direction for clarity. As desired, the sample spacing is small near the optical axis where the frequencies are high in the Fresnel transform, and the sample spacing increases linearly as we move away from the optical axis, as required by (9).

The minimum sample spacing is found by noting that the phase of the chirp at  $u=0$  is  $\pi L^2/\lambda F$ . The width of the first half cycle of the chirp is obtained by decreasing the phase of the cosine by  $\pi$  and solving the relationship

$$\frac{\pi}{\lambda F}(L - u)^2 = \frac{\pi L^2}{\lambda F} - \pi \quad (11)$$

for  $u$ :

$$u = L - \sqrt{L^2 - \lambda F}, \quad (12)$$

which is the value of the minimum sample spacing  $d_{f \min}$ . We can rewrite (12) as



$$\begin{aligned}
u &= L - L\sqrt{1 - \frac{\lambda F}{L^2}} \\
&= L\left[1 - \sqrt{1 - \frac{1}{L\alpha_{co}}}\right] \\
&= L\left[1 - \sqrt{1 - \frac{2}{N}}\right] \cong \frac{L}{N} \cong d_0.
\end{aligned} \tag{13}$$

Thus, the minimum sampling spacing in any Fresnel plane between  $P_1$  and  $P$  is the same as that for the object.

The maximum sample spacing is found by noting that the phase of the chirp at  $u=L$  is zero. The width of the next to the last half cycle of the chirp is found by incrementing the phase of the cosine by  $\pi$  and solving the relationship

$$\frac{\pi}{\lambda F}(L - u)^2 = \pi \tag{14}$$

for  $u$ :

$$u = L - \sqrt{\lambda F}. \tag{15}$$

The maximum sample spacing  $d_{f \max}$  is found by subtracting the value of  $u$  given by (15) from  $L$  to find that  $d_{f \max} = \sqrt{\lambda F}$ . As a note in passing,  $d_{f \max}$  is also equal to the radius of the first dark ring of a Fresnel zone pattern.

Suppose that we have an object whose length is  $L=100\text{mm}$  and whose cutoff frequency is  $\alpha_{co}=0.25\text{cycles/mm}$  so that the line bandwidth product is  $LBP=L\alpha_{co}=25$  and the number of samples required is  $N=50$ . If we were to uniformly sample the Fresnel transform, we would require  $N=100$  samples because the highest frequency is  $\alpha_f=\alpha_{co}$  and the length of the Fresnel plane is  $2L$  as compared to  $L$  for the object plane. To illustrate the optimum sampling distribution for the Fresnel transform, we set the parameter  $\lambda F=400$  and find that  $d_0=d_{f \min}=2\text{mm}$  and that  $d_{f \max}=20\text{mm}$  in agreement with the results shown in Figure 4b. In the sampling scheme developed here, the total number of samples needed in the Fresnel plane is determined by multiplying the average frequency of the chirp by the chirp length. Since the average spatial frequency at the Fresnel plane of length  $2L$  is  $0.125\text{cycles/mm}$ , we find that  $N=50$ . The number of samples at the Fresnel planes  $P_2$  and  $P_3$  in Figure 1 is therefore the same as the number for the object plane  $P_1$  or the number for the Fourier plane  $P_4$ . This nonuniform

sampling pattern is completely independent of the exact structure of the object and therefore can be used in any application.

We now prove that the required number of samples is fixed at all Fresnel planes in the system and develop the optimum sampling strategy in an arbitrary plane. We initially confine our attention to the region between planes  $P_1$  and  $P_2$  in Figure 1. First, from the ray diagram of Figure 3, we see that the maximum frequency is the same at all planes because at least one pair of points in the object plane produce a spatial frequency  $\alpha_f = \alpha_{co}$  in all Fresnel planes. The region of overlap decreases linearly as  $D$  increases so that a smaller portion of the Fresnel domain must be sampled with sample spacing  $d_0$  as we move from  $P_1$  towards  $P_2$ . On the other hand, the total extent of the Fresnel transform also increases linearly as  $D$  increases. These features are illustrated diagrammatically in Figure 5.

The diagram in Figure 5 is called a *space/frequency diagram* because it plots the maximum spatial frequency as a function of position in each Fresnel plane between the object and lens planes. At the object plane the object is regularly sampled, with sample spacing  $d_0$  because the maximum spatial frequency  $\alpha_{co}$  is, in general, uniformly distributed throughout the object. As we progress toward the plane of the lens, the space/frequency diagram becomes trapezoidal; the central region must be uniformly sampled with sample spacing  $d_0$  while regions where the maximum spatial frequency gradually goes to zero is nonuniformly sampled. The optimum sampling within the two end regions is found in the same way as before by appending chirp functions with appropriate chirp rates to the central region.

The number of samples  $N_f$  in any Fresnel plane between the object and the lens plane is found with the aid of Figure 5 which shows the trapezoidal shaped section of the space/frequency diagram between planes  $P_1$  and  $P_2$ . The number of samples is the sum of the uniformly spaced samples in the region for which  $|u| \leq (L/2 - \theta_{co}D)$  and the nonuniformly spaced samples in the outer regions of the trapezoid:

$$N_f = \frac{L - 2\theta_{co}D}{d_0} + \frac{4\theta_{co}D}{2d_0} = \frac{L}{d_0} = N, \quad (16)$$

where  $2d_0$  is the sample spacing for the average spatial frequency in the nonuniformly sampled region. From (16) we find that the number of samples, when the optimum sampling technique is used, is the same in all planes between the object and the lens, as was to be shown. We recognize that the areas of all the space/frequency sections are equal, consistent with this conclusion.

The space/frequency diagram for the Fresnel planes between the lens and the Fourier plane is the mirror image of that given in Figure 4, with the lens plane being the plane of axial symmetry. This completes the proof that the required number of samples in all Fresnel planes in a Fourier transform system is constant, provided that the optimum sampling procedure is used and provided that the Fourier transform system is structured for maximum information capacity.

#### 4 The Sampling Strategy for an Unconstrained Imaging System

We now extend these sampling results to all Fresnel plane in an imaging system having finite magnification. We also remove the restriction that the information capacity of the system is maximized so that the object and Fourier plane do not necessarily have the same size. A general imaging system, without the maximum capacity constraint, evolves from the Fourier transform system of Figure 1 when the object plane is moved away from the lens so that the image is formed at a finite distance on the opposite side of the lens. Such a system is shown in Figure 6a for the situation where the magnification  $M = -2$ . We begin the analysis by tracing rays for an object that has a high cutoff frequency. These rays are the same ones that formed the cone of light produced by the sampling function  $d(x)$  in Figure 2. The ray trace for the two edge samples as shown in Figure 6a is called a *scissors diagram*; it is helpful for quickly determining the boundaries of the light as it passes through the system and for easily locating the Fourier plane. The space/frequency diagram shown in Figure 6b shows the distribution of frequencies at various sections throughout the system. We first discuss the optimum sampling distribution on the object side of the lens, followed by a similar analysis for those planes on the image side of the lens.

---

#### 4.1 The Object Side of the Lens

The triangular space/frequency section located between the object and lens planes is the same as that at the plane where the lens was located in the Fourier transform system of Figure 1. For finite imaging conditions, the lens is located further away from the object plane so that the light continues to disperse between this section and the lens plane. To maintain an optimum sampling distribution in this region, we must modify the sample spacing obtained so far. We begin by noting that the highest spatial frequency  $\alpha_f$ , as given by the hyperbolic function (8), is now less than  $\alpha_{co}$  because these Fresnel planes are beyond the distance where the marginal rays first cross. The angle subtended by the two edge samples therefore decreases below that needed to produce the cutoff spatial frequency. The section of the space/frequency distribution is still trapezoidal by virtue of the arguments given above, but the maximum frequency in the central part of the trapezoid is smaller.

A typical section of the space/frequency distribution in this region is shown in Figure 7. The number of samples  $N_f$  in plane between the triangular section and the plane of the lens is given by the sum of the uniformly spaced samples in the region for which  $|u| \leq (-L/2 + \theta_{co}D)$  and the nonuniformly spaced samples in the outer regions of the trapezoid:

$$N_f = \frac{2(-L/2 + \theta_{co}D)}{d_f} + \frac{2L}{2d_f} = \frac{2\theta_{co}D}{d_f}, \quad (17)$$

where  $d_f$  is the sample spacing in the uniformly sampled region and  $2d_f$  is the average sample spacing in the nonuniformly sampled region. We use the fact that  $\theta_{co} = \lambda\alpha_{co}$  and that the sample spacing for the spatial frequency  $\alpha_f$  is  $d_f = \lambda D/L$  in (17) to find that  $N_f = N$ , which shows that  $N$  samples are also sufficient to represent the signal at any Fresnel plane on the object side of the lens, even under the most general imaging condition.

#### 4.2 The Image Side of the Lens

From the scissors diagram in Figure 6a, we observe that, at the plane of the lens, the angles subtended by the edge samples of both the object and image planes are the same, as required by the imaging condition and by the optical invariant. Hence, there is continuity in the space/frequency diagram across the lens plane. As we move away from the lens toward the

Fourier plane, the sections of the space/frequency diagram remain trapezoidal, and the required sample spacing  $d_f$  increases because the angle subtended by the extreme samples in the image increases. In any of these planes, we apply the optimum mixture of uniform/nonuniform sampling as described above.

At the Fourier plane, the space/frequency section becomes rectangular and the sample spacing is

$$d_F = \frac{2F_{co}}{N} = \frac{2\lambda F \alpha_{co}}{N} = \frac{\lambda F}{L}, \quad (18)$$

which shows that the sample spacing in the Fourier plane is *completely independent* of the spatial frequency content of the object; it is dependent only on the length of the object. Note that when we use a single lens to image the object plane with finite magnification, we cannot necessarily achieve the conditions for optimum information capacity so that, in general,  $d_F \neq d_0$ . The sample spacing in the Fourier plane may be greater or less than  $d_0$ , depending on the geometry of the system as we explore further in Section 5.

As we progress from the Fourier plane to the image plane, the space/frequency sections remain trapezoidal, while the maximum spatial frequency continues to increase because the edge samples of the image subtend larger angles as we proceed toward the image plane, leading to the higher spatial frequencies. The maximum spatial frequency continues to increase until we reach the triangular space/frequency section. From this plane to the image plane, the sections are trapezoidal but the maximum frequency decreases linearly to its final value of  $\alpha_{co}/|M|$ , where  $M$  is the magnification of the system. In turn, the sample spacing at the image plane is also  $|M|d_0$  as required by geometrical optics and the optical invariant.

### 5 Maximum and Minimum Spatial Frequencies

We have seen that the optimum sampling distribution in any Fresnel plane is to use a nonuniform sample spacing, generally combined with some regions of uniform sampling. The question arises as to which planes require the smallest and largest sample spacings. To develop the answer, consider the scissors and space/frequency diagrams given in Figure 8, which is the same system configuration as given in Figure 6, except that the object bandwidth is much lower.

The sample function size and spacing at the object plane is larger than that for the high bandwidth system which leads to a smaller diffraction angle  $\theta_{CO}$ , as represented by the cones in the scissors diagram. Since the interior rays from the edge samples do not cross on the object side of the lens, the space/frequency sections are all trapezoids from the object to the lens plane. Furthermore, the maximum frequency at all planes in this region is  $\alpha_{CO}$ . As in the high bandwidth system, there is continuity of the space/frequency diagram across the lens plane.

On the image side of the lens, the exterior rays from the edge samples cross, leading to the first triangular space/frequency section shown in Figure 8b. Between the lens and this section, the maximum frequency increases linearly. This relationship is visualized by noting that, in contrast to the argument given for the high bandwidth system, light propagating to the edge sample points at the image do not overlap at the plane of the lens. If we consider the Fourier plane the fulcrum of the scissors and the edge samples at the image as the handles, we need to close the scissors (i.e., move the samples closer together) until overlap occurs at the lens plane. Once overlap is achieved at the lens plane, we more easily see that the continuity of spatial frequencies across the lens plane holds. We can then open the scissors as we progress toward the first triangular plane; at this plane the scissors are fully open. Since the angle between the samples increases linearly as we progress along the optical axis from the lens to the first triangular section, so too does the highest spatial frequency.

As we progress between the triangular sections, we find that the spatial frequency continues to increase in a hyperbolic fashion because the distance between the samples is fixed at  $lMl$  while the distance to the image plane decreases. From the second triangular plane to the image, we need to again close the scissors to achieve overlap; the maximum spatial frequency in these planes therefore decreases linearly to its final value of  $\alpha_{CO}/lMl$ . Note that the space/frequency diagram for the low bandwidth object as shown in Figure 8b is completely nested within that of the wideband object as shown in Figure 6b.

We now sharpen the qualitative results obtained so far to find the planes for which the highest spatial frequency is at a maximum or minimum and to quantify the magnitude of these

---

frequencies. First, it is clear that if the magnification is  $|M| < 1$ , the frequency at the image plane must be higher than that at the object plane. Based on the general description given in Section 4, the maximum spatial frequency must be on the image side of the lens if  $|M| < 1$ . The remaining question concerns the situations when  $|M| \geq 1$ .

Since the maximum frequency always occurs at the optical axis by virtue of (9), it is useful to plot the maximum frequency that occurs at the axis. Consider the scissors diagram in the upper part of Figure 9 for a wide bandwidth object (solid rays) and a low bandwidth object (dotted lines). As an example to quantify the spatial frequencies, consider a system which has an object length  $L=1\text{mm}$  and is configured to provide a magnification  $M=-2$ . If the focal length of the lens is  $8\text{mm}$ , the distance from the object plane to the lens plane is  $(1-1/M)F=12\text{mm}$  and the distance from the lens to the image plane is  $(1-M)F=24\text{mm}$ . To find the path of the maximum frequency at the optical axis for each plane in the system, we begin with the plot labeled B in the lower part of Figure 9. In this case the cutoff frequency at the object plane is  $\alpha_{co}=400\text{cycles/mm}$  and the maximum frequency remains at this value until we reach the plane at which the marginal rays first cross. From this plane to the lens plane the maximum frequency decreases, following the hyperbolic curve

$$\alpha_f = \frac{L}{\lambda D}, \quad (19)$$

where  $D$  is the distance from the object plane to the plane in question. From the lens plane towards the image plane, the maximum frequency path follows a second hyperbolic curve

$$\alpha_f = \frac{-ML}{\lambda D}, \quad (20)$$

where  $D$  is now the distance from the plane in question to the image plane. We follow this curve until we reach the plane where the marginal rays cross for the second time in the scissors diagram. From that plane to the image plane the maximum frequency path follows the straight line to the final value of  $\alpha_{co}/|M|$ .

For an object with a lower bandwidth, such as that shown in path C, the general behavior is the same as that just described. From the scissors diagram, we see that the first and second crossover planes for the marginal rays occur closer to the lens so that the maximum frequency

path spends less time on the hyperbolas. At a somewhat lower object bandwidth, such as that shown as path D, the crossover planes are both on the image side of the lens. In this case the path of the maximum frequency does not intersect the first hyperbola so that the maximum frequency is the same at every plane between the object and lens planes as argued in Section 4.3. The maximum frequency path then linearly increases until it intersects the hyperbola on the image side of the lens and remains on this hyperbola until the plane of the second crossover is reached. The maximum frequency then decreases linearly until the image plane is reached.

As the object bandwidth decreases, the maximum frequency path spends less and less time on the hyperbola. This phenomena reaches its extreme condition for path E, the condition for which the object bandwidth is zero; the object is therefore a simple truncated plane wave. The width of this wave remains constant, to within the approximation given by ( 5), between the object and the lens. The light then begins to focus, reaching its smallest spatial extent in the Fourier plane; it then expands to fill the image plane. This result nicely illustrates that the spatial frequency at the Fourier plane is completely independent of the object bandwidth; it is purely a function of the object length. For example, the sample size at the object and image planes is large for path E, whereas the sample spacing at the Fourier plane is always given by ( 18).

For all the object bandwidths considered so far, the maximum spatial frequency occurs at the second crossover plane on the image side of the lens. For an object with a much higher bandwidth, such as that shown in path A, we find that the maximum frequency shifts to the object side of the lens. We now consider the general condition for which this shift occurs. The maximum frequency on the image side of the lens must satisfy the hyperbolic relationship given by ( 20). We need to find a general formula to find the distance D from the image plane at which the crossover occurs. By tracing one of the exterior rays, we find, after considerable but straightforward algebraic manipulations, that the crossover distance is

$$D = \frac{-ML}{\frac{L}{F} - \frac{2\theta_{co}}{M}}, \quad (21)$$

which, when substituted into ( 20), yields the result that the shift occurs when



$$\alpha_{co} \geq \frac{M}{(1+M)} \frac{L}{\lambda F}. \quad (22)$$

For the parameters given above, this shift occurs when  $\alpha_{co}=500$ cycles/mm, which is consistent with the results shown in Figure 9.

The maximum spatial frequency must occur in one of four possible planes: the object plane, the Fourier plane, the image plane, or the second crossover plane where the space/frequency diagram has a triangular section. As a note in passing, it is easy to prove that *there are exactly two crossover planes in any system (except for the degenerate case when the object bandwidth is at its lowest value so that the two crossover planes coalesce at the Fourier plane) and that the second crossover plane must be on the image side of the lens*. The first crossover plane may be on either side of the lens, depending on the object bandwidth, and is due to the interior marginal rays produced by two samples. The second crossover is due to the exterior marginal rays produced by the samples.

It is quite simple to find the plane containing the minimum spatial frequency at the optical axis. For low bandwidth objects, the image plane contains the minimum frequency. As the object bandwidth increases, we find that the minimum frequency shifts to the lens plane; tracing the interior rays on the image side of the lens, we find that this shift occurs when

$$\alpha_{co} \geq \frac{-M^2}{(M-1)} \frac{L}{\lambda F}. \quad (23)$$

The minimum frequency occurs at the lens plane for all object bandwidths that satisfy (23).

An analysis similar to that used in connection with Figure 9 applies for the case when the magnification is less than one. If  $M=-1/2$ , both the scissors diagrams and the space/frequency diagrams are quite different, however, from those given in Figure 9. We cannot therefore simply read the space/frequency diagrams from right-to-left. The reason is that we have illuminated the object with collimated light so that the central rays associated with the cones that describe the jaws of the scissors do not open and close in the same way on the object and image sides of the lens. The central rays are always normal to the object plane whereas they pivot about the axis at the Fourier plane on the image side of the lens. The even more general condition in which the

illumination is divergent or convergent produces results similar to those already derived. The primary difference is that the position of the Fourier plane will shift to a new position, causing an adjustment in all the other planes for which space/frequency sections were shown; the object and image plane positions are, of course, not affected by the type of illumination. The formulae and calculations needed to describe this even more general case are straightforward and are not given here.

## 6. Summary and Conclusions

The optical invariant ensures that the information content in the object and image planes of an optical system are equal;  $N$  samples are sufficient to accurately measure the signals in these planes. It is easy to show that the Fourier plane can be added to the list, since all the information must pass through a restricted aperture if the object is bandlimited. In this paper, we show that  $N$  samples are sufficient to sample a signal in any Fresnel plane as well, provided that a specified nonuniform sampling distribution is followed. We show that the highest spatial frequencies in any Fresnel transform is concentrated near the optical axis so that the samples must be most closely spaced in this region. We have shown that the highest possible spatial frequency may occur in one or more of four planes in a generalized imaging system: the object plane, the Fourier plane, the second crossover plane, or the image plane.

This nonuniform sampling is similar to the the visual system in which the region of greatest acuity for the eye is at the optical axis. It may have application to showing how to properly dilute arrays in other wavelength region such as discrete element, phased arrays in the microwave region. The optimum sampling procedure reduces the required number of samples by up to a factor of four, for the two-dimensional case, leading to less computations in applications such as image restoration.

This work was supported by the U.S. Army Research Office.

Last revision Thursday, February 15, 1990

---

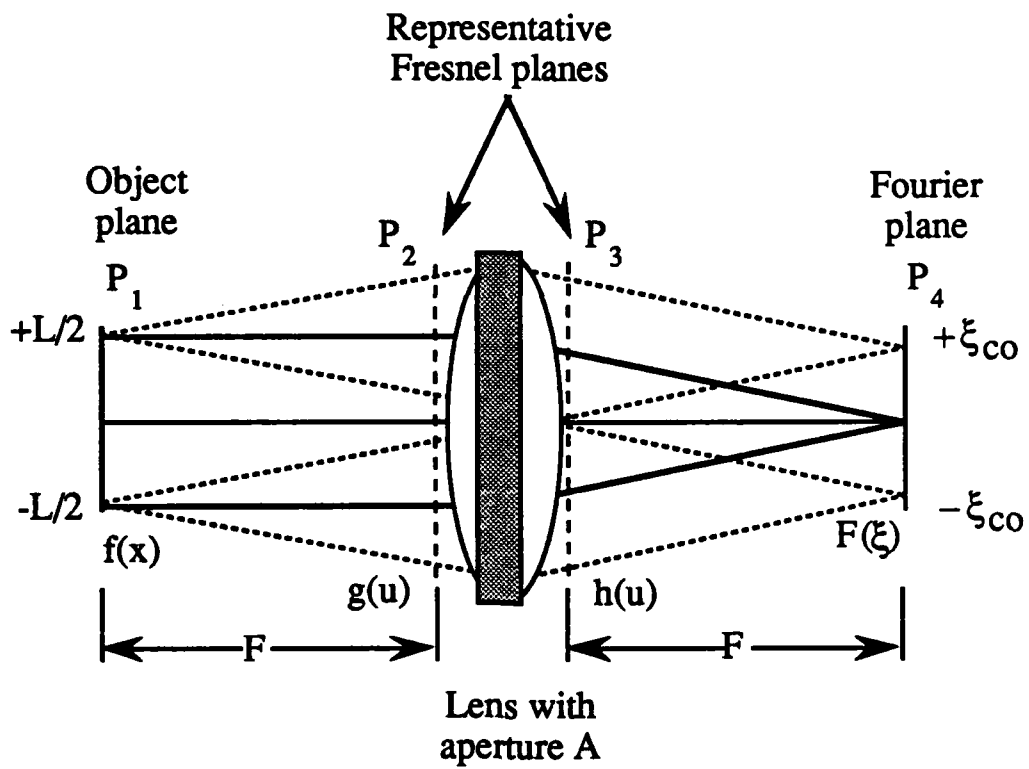


Figure 1: A Fourier Transform System

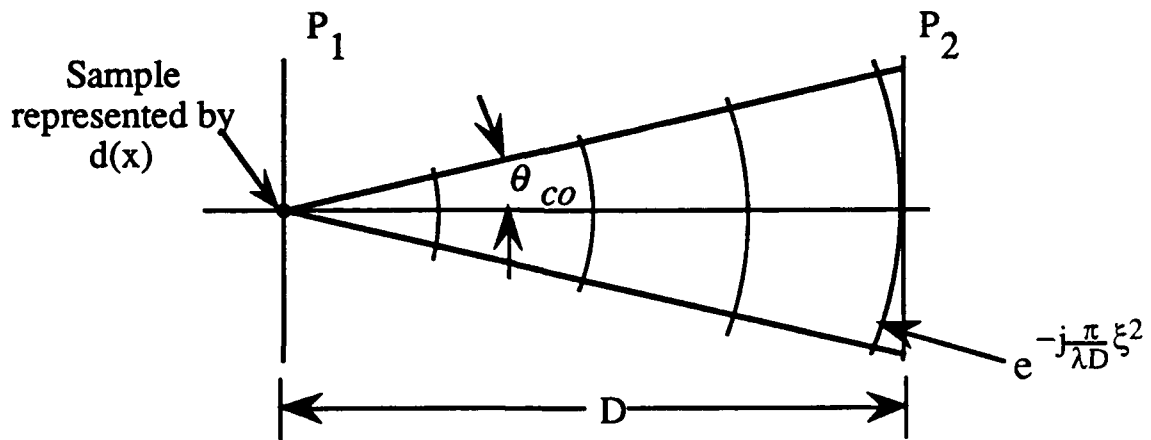


Figure 2: The Impulse of a Sample of the Object

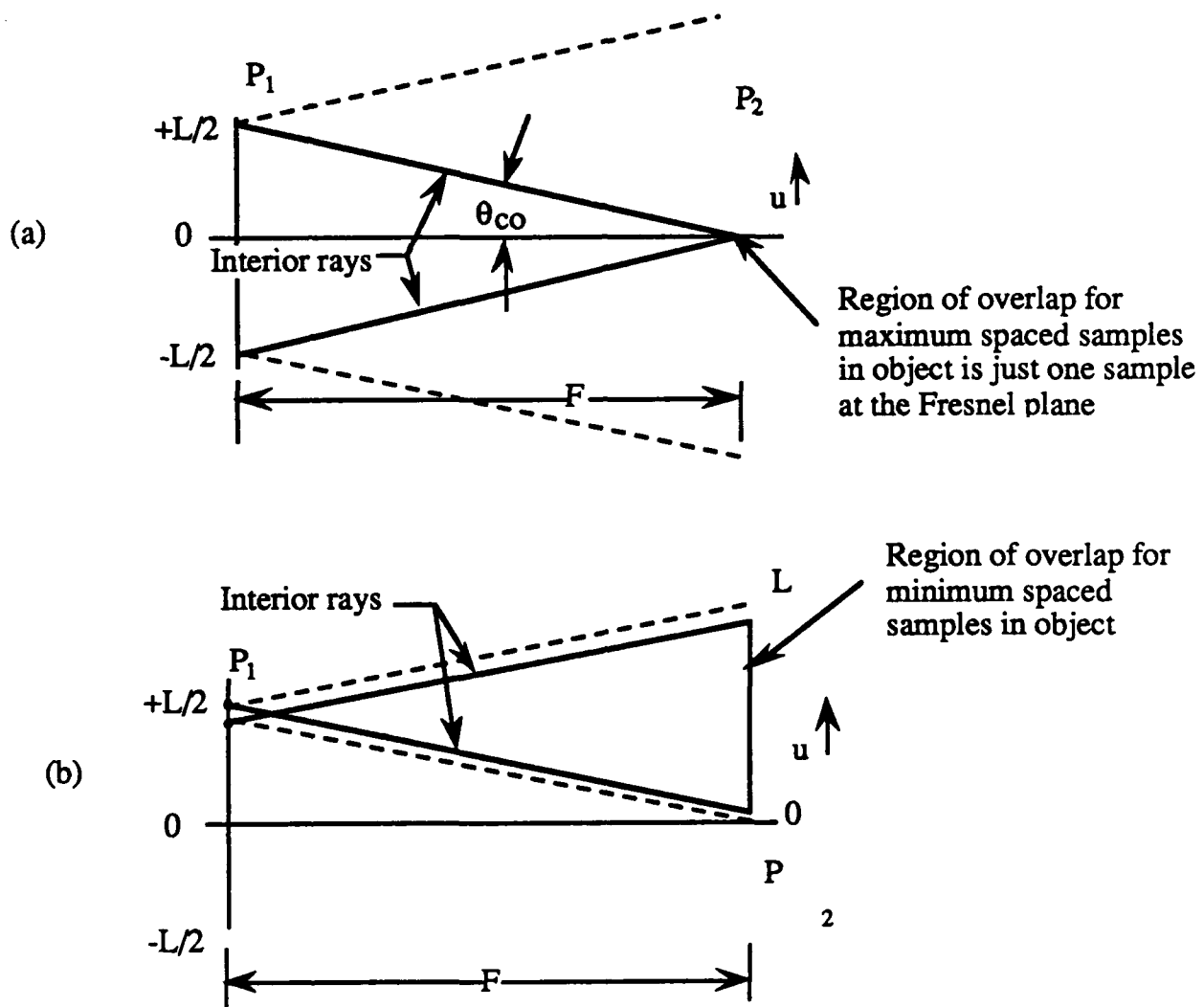


Figure 3: Sample Spacing to Produce (a) Maximum Frequency but Minimum Overlap and (b) Minimum Frequency but Maximum Overlap

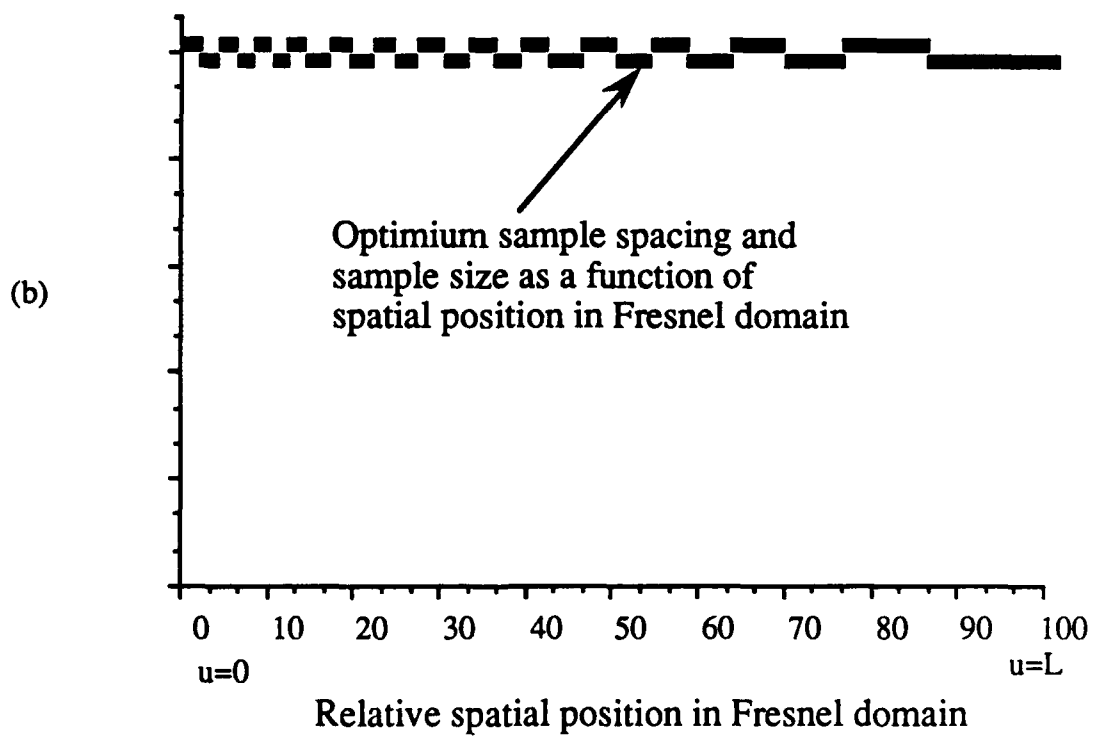
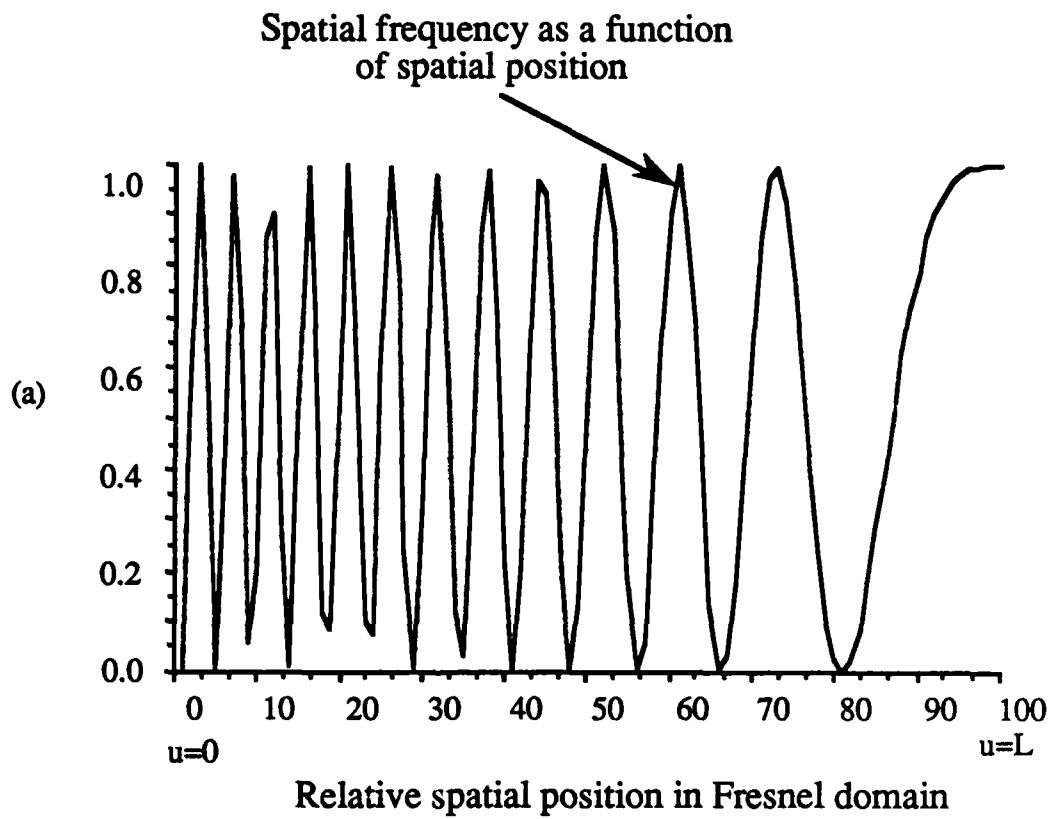


Figure 4: Optimum Sampling Spacing for the Fresnel Transform

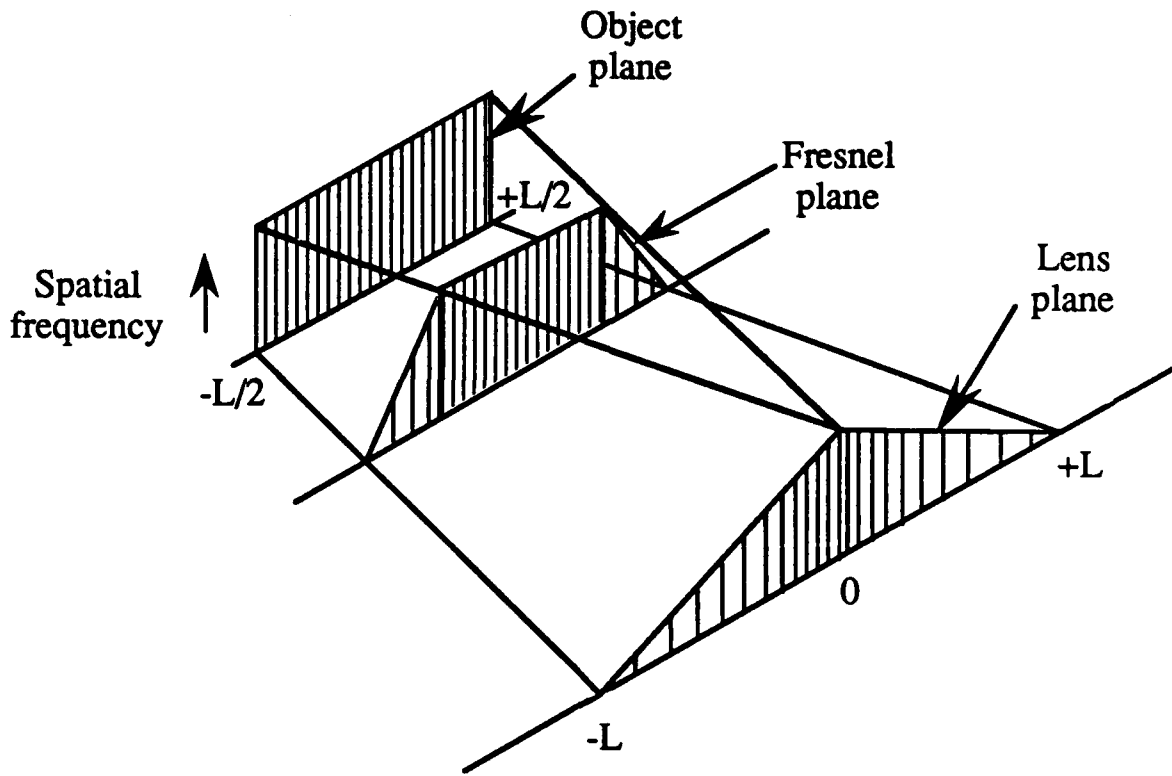


Figure 5: Optimum Sampling for Fresnel Transforms

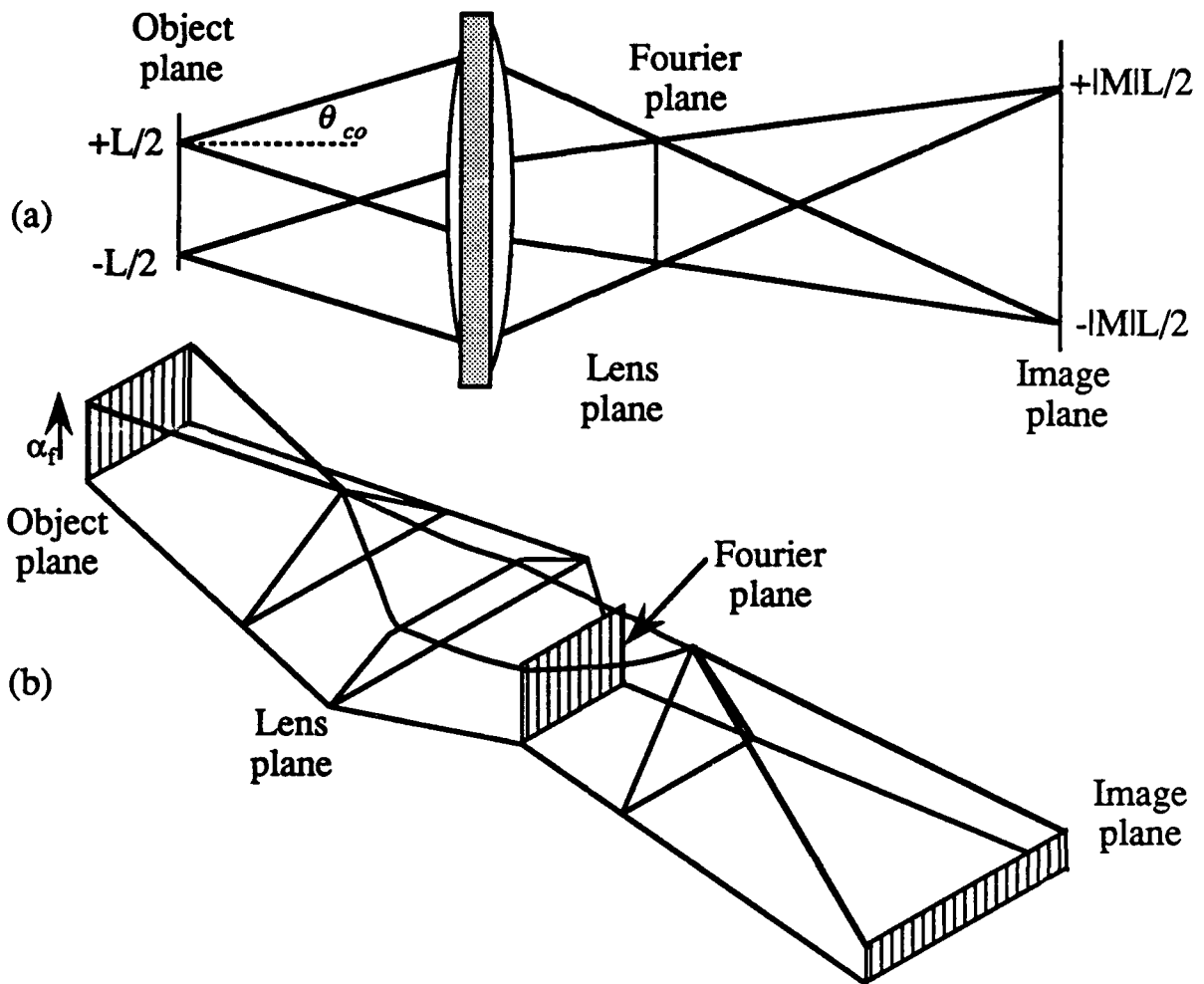


Figure 6: High Bandwidth Imaging System

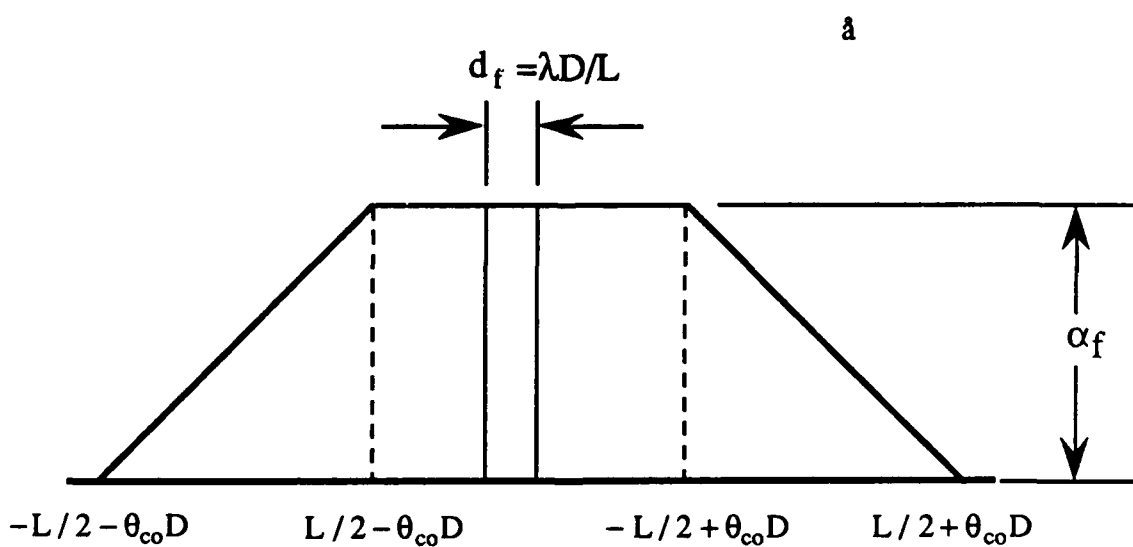


Figure 7: General Trapezoidal Region

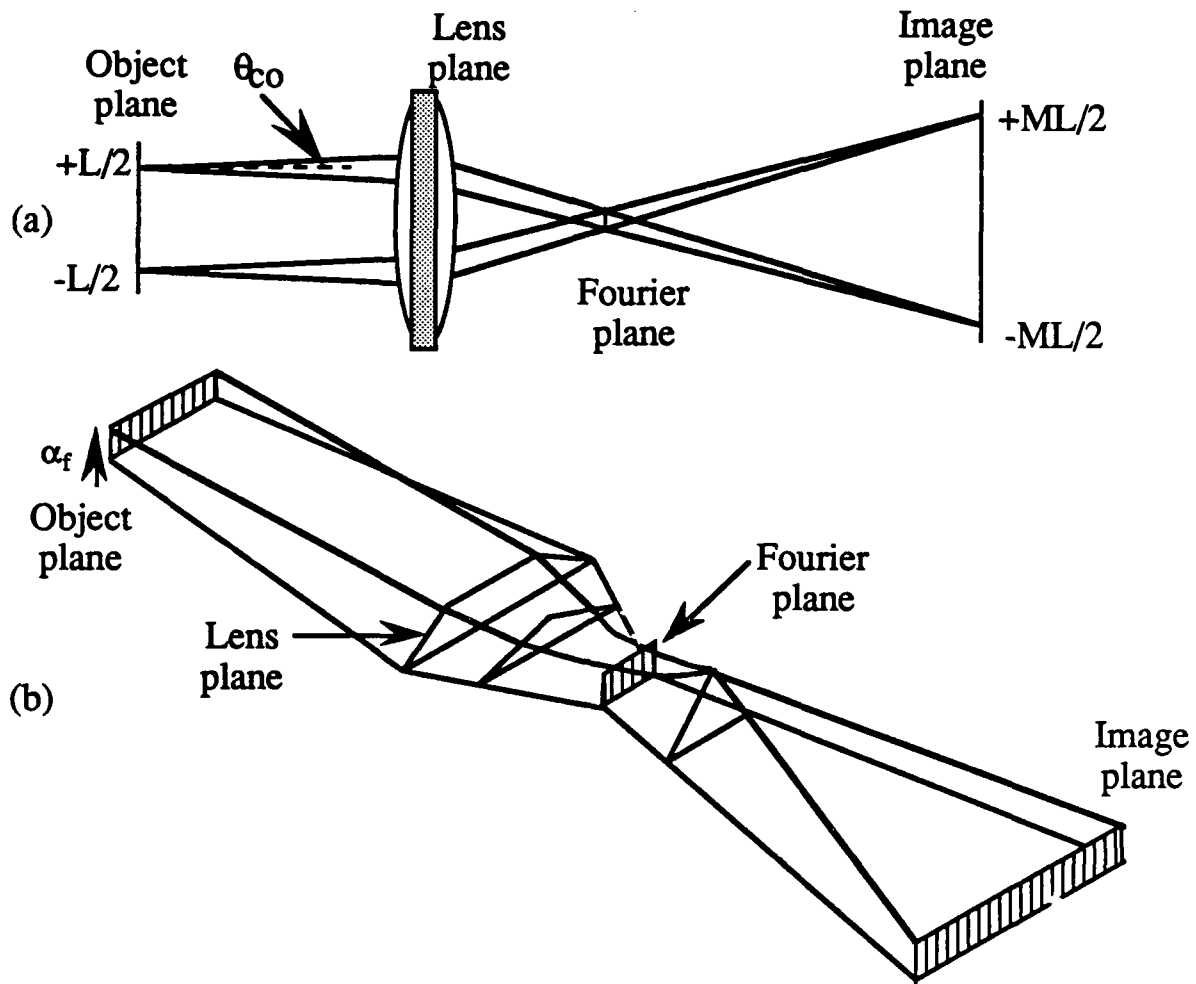


Figure 8: Low Bandwidth Imaging System



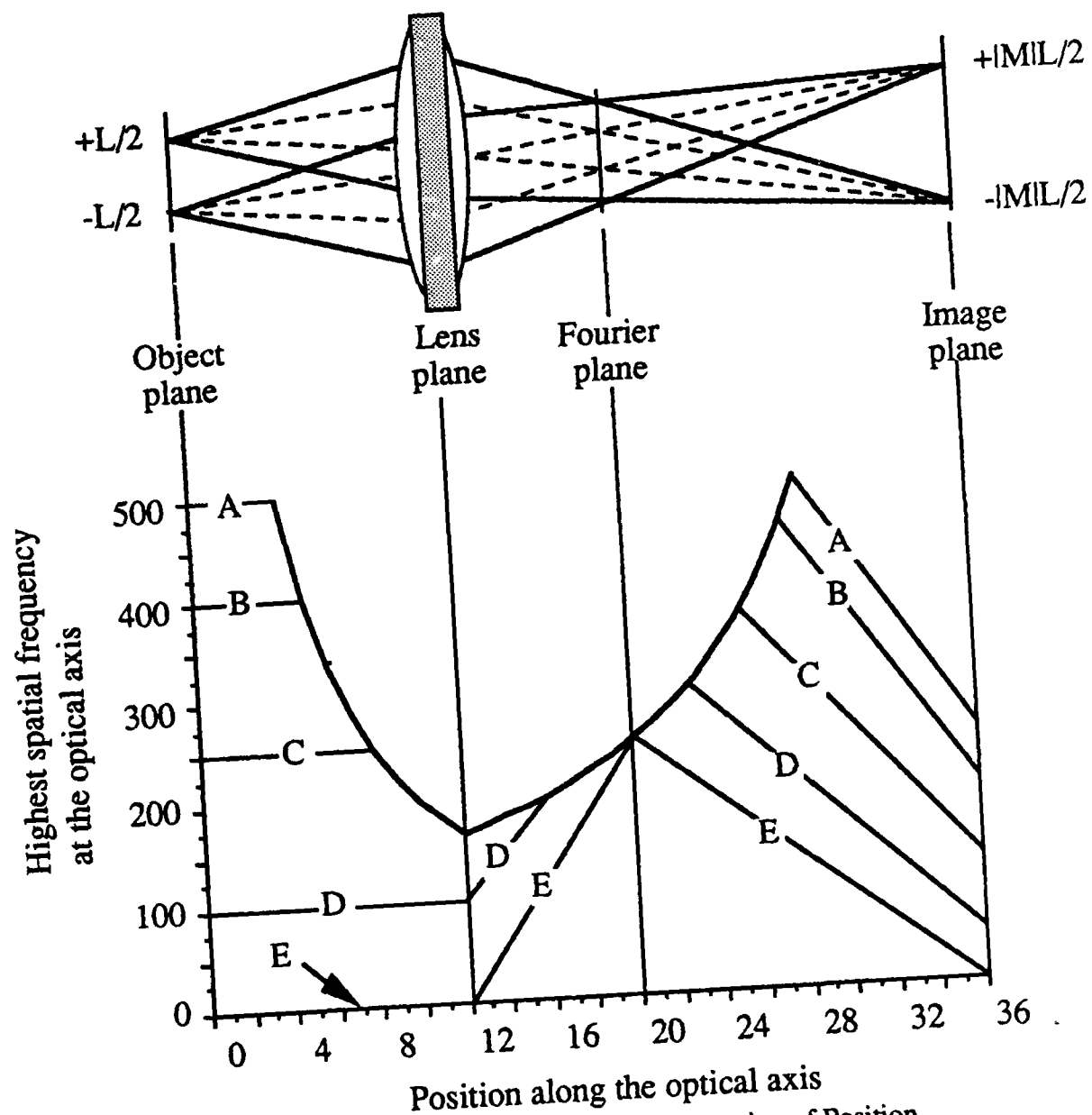


Figure 9: Maximum Frequency as a Function of Position

1. G. Toraldo di Francia, "Resolving Power and Information," J. Opt. Soc. Am., Vol. 45, p. 497, (1955); "Capacity of an optical channel in the presence of noise," Optica Acta, Vol. 2, p. 5, (1955)
2. E. H. Linfoot, "Information Theory and Optical Images," J. Opt. Soc. Am., Vol. 45, p. 808, (1955)
3. E. L. O'Neill and A. Walther, "The question of phase in image formation," Optica Acta, Vol. 10, p. 33, (1963)
4. A. Walther, "The question of phase retrieval in optics," Optica Acta, Vol. 10, p. 41, (1963)
5. R. Barakat, "Application of the Sampling Theorem to Optical Diffraction Theory," J. Opt. Soc. Am., Vol. 54, p. 920 (1964)
6. A. VanderLugt, "Design relationships for holographic memories," Appl. Opt., Vol. 12, p. 1675, (1973)
7. A. VanderLugt, "Packing density in holographic systems," Appl. Opt., Vol. 14, p. 1081 (1975)
8. R.J. Marks, J.F. Walkup, and M.O. Hagler, "A sampling theorem for space-variant systems," J. Opt. Soc. Am., Vol. 66, p. 918 (1976)
9. A. VanderLugt, "Fresnel Transforms and Bragg Cell Processors," Appl. Opt., Vol. 24, pp. 3846 (1985)
10. A. VanderLugt, "Operational Notation for the Analysis and Synthesis of Optical Data Processing Systems," Proc. IEEE, Vol. 54, p. 1055 (1966)