OPTICAL SIGNAL PROCESSING

U.S. Army Research Office
DAAL03-90-G-0039

FINAL REPORT
27586-F

A. VanderLugt
Principal Investigator
30 July 1993

North Carolina State University
Raleigh, North Carolina 27695

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	31 July 1993	Final Report 5 Feb 1990 to 30 June 1993

4. TITLE AND SUBTITLE
Optical Signal Processing

5. FUNDING NUMBERS
DAAL03-90-G-0039

6. AUTHOR(S)
A. VanderLugt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Electrical and Computer Engineering
North Carolina State University
PO Box 7911
Raleigh, NC 27695-7911

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 27586.21-PH

11. SUPPLEMENTARY NOTES
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.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A quasi-realtime adaptive processing system was used to correct the multipath distortion found in wideband digital radios. The measured power spectral density of the input signal was used to adaptively select one of eight equalization filters which reduce the residual distortion to less than 3.6 dB even for the most severe channel distortion. A related adaptive system was used for signal excision in which we removed narrowband interference from wideband signals with minimum signal distortion. We built an 8x8 acousto-optic switch in a multimode fiber-optic system. Insertion loss is approximately 2-4 dB, signal-to-crosstalk ratio is better than 25 dB, and the reconfiguration time is 880 nsec. We detected short pulses by using the Fresnel transform. We detected pulses as short as the theoretical limit of 20 nanoseconds for this system, and separated by as little as 60 nanoseconds or by as much as 17 nanoseconds. We considered all possible acousto-optic scanning configurations and classified them into four basic types. A consistent set of design relationships for each of the scanning configurations has been developed and presented in both tabular and graphic forms from which a preliminary design is obtained.

14. SUBJECT TERMS
Optical signal processing, channel equalization, adaptive optical processing, Fresnel transforms, optical switches, acousto-optics, short pulse detection, spatial filtering.

15. NUMBER OF PAGES
192

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

TABLE OF CONTENTS

LIST OF APPENDICES

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 investigate methods for performing quasi-realtime adaptive processing with primary emphasis on the equalization of high speed communication channels. These techniques are also applicable for adaptively setting notches in a signal excisor to remove the effects of signal distorting jammers. Such operations are useful when the adaptation rate is fairly slow so that continuous closed loop operation is not needed. A second objective is to investigate the time-variant and frequency-variant properties of the Fresnel transform to make better use of optical processing capabilities, and to develop generalized methods for analyzing these systems so that new architectures and processing algorithms can be developed. In particular, we investigate the application of these techniques to detecting signals such as short pulses, spread spectrum, and other low-probability-of-intercept signals. The third objective is to develop models to analyze and experimentally verify the performance of acousto-optic switches. Diffraction pattern modeling and determination of crosstalk levels are needed for various combinations of switch size, longitudinal and lateral output channel placement, reconfiguration time, and detector size. Signal losses, distortion, and noise introduced by the switching element determine the bit error ratio. A prototype to demonstrate performance levels has been built and tested.

The research work descried in this report covers the period from 5 February 1990 to 30 June 1993. The research effort has resulted in several innovative optical processing techniques for improved performance. The major accomplishment can be divided into three areas: (1) studies relating to adaptive optical signal processing, (2) studies relating to the time-variant and frequency-variant properties of the Fresnel transform, and (3) the analysis and experimental work relating to an acousto-optic switch. 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 are included in the Appendices.

2.0 Adaptive Optical Signal Processing

One part of this study was to develop architectures for quasi-realtime adaptive processing. Closed loop adaptive processing was studied as a part of a previous contact (under DAAG-29-80-C-0149) and demonstrated the implementation of 30dB notches in a 50MHz bandwidth.[1,2] Such systems are very sensitive to slight phase changes caused by vibrations or thermals drifts, particularly when the loop bandwidth is fairly wide. There are many applications, however, in which systems can be updated incrementally in what we call a quasi-realtime adaptive fashion. These systems behave more like those in which the filters are updated slowly to maintain the desired steady state condition of the system.

T.P. Karnowski developed methods for constructing a general spatial filtering operation. Since the required impulse response is one-dimensional when we processes time signals, we obtain a continuous amplitude response from a binary two-dimensional light modulator by using an area modulation technique. Karnowski used area modulation to produce the required positive amplitudes, and a half-wave plate to produce the required negative amplitudes in a conventional Mach-Zehnder interferometeric system. Although area modulation eliminates the nonlinearities normally associated with amplitude recordings, certain imperfections in the system such as aberrations and a Gaussian illumination profile cause what are, in effect, amplitude nonlinearities. Fortunately, these nonlinearities are not signal dependent and we found ways that they can be compensated by using a look-up table that is generated when the system is calibrated. Experiments using photographic film produced excellent results, while those using a computer controlled liquid crystal display were less satisfactory due to the low contrast of the particular display used. Nevertheless, as better spatial light modulators become available, this techniques should prove useful for generating adaptive impulse responses. He completed work on this topic for his Master's Thesis "Area Modulation Filter in an Interferometric Acousto-Optic Signal Processor" and published it in Applied Optics[3]; it is included in Appendix A.

C. S. Anderson developed optical methods for digital radio channel equalization. An adaptive acousto-optic equalizer has been constructed which corrects the multipath distortion found in digital radios for telecommunications. A multipath channel emulator was constructed with RF electronics to distort test signals that are then corrected by the optical processor. The optical system uses an acousto-optic power spectrum analyzer to measure the power spectral density of the input, distorted signal. A gradient search algorithm determines both the magnitude and phase characteristics of the distortion. The channel estimator then adaptively selects one of eight equalization filters recorded onto an area modulated mask. These filters are designed to equalize any possible channel with less than 3 dB of residual magnitude distortion remaining. Since the magnitude distortion of the channel can be as high as 25 dB, this represents at least a

22 dB improvement in performance. Both the channel estimation and filtering sections of the processor were tested and showed that the channel estimator selected the correct equalization filter 95% of the time; the other 5% occurred for distortions that were in the transition regions between two filters. Selection of the incorrect filter in these regions increases the residual distortion by only 0.75 dB. The frequency response of the eight equalization filters was also tested and found to be in good agreement with the theoretical response. The channel estimation and the channel filtering portions of the processor were then combined and allowed to adapt automatically. Tests of the adaptive processor showed that the equalizer could maintain the residual distortion to less than 3.6 dB even when the most severe channel distortion was present. This work has received supplemental funding from BellSouth Enterprises and his thesis "Hybrid Acousto-Optic and Digital Channel Equalization for Microwave Digital Radios" has been accepted in partial fulfillment of the PhD degree. One paper has been published on this research effort[4] and a second paper has been accepted for publication; these papers are included in Appendix B and Appendix C.

R. N. Ward investigated a related form of a quasi-realtime adaptive system for signal excision. Frequency plane excision is one method for removing narrowband interference from wideband signals. He modified an acousto-optic system, which uses frequency plane excision, to provide control signals to the spatial light modulators which perform the notching of the jammer. The major system modification was to add a CCD photodetector array to determine the magnitude and frequency of the jammers. The experiments involve modulating a baseband pulse to an intermediate frequency, excising various frequency components, demodulating the pulse, and then characterizing the residual distortion on the pulse. He studied the distortion of short pulses produced by the removal of various frequency components of pulses having various widths and analyzed the resulting distortion at baseband. These experiments showed that the distortion is most severe when the excision is performed in the mainlobe of the sinc function spectrum of the pulse, while the distortion decreases rapidly when the excision is in the sidelobes of the pulse. The results of his work was the topic for his Master's Thesis "The Excision of Narrowband Interference from Wideband Signals using an Open-Loop Adaptive Technique in an Acousto-Optical Signal Processor," and published the results in Applied Optics[5] as shown 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. D.O.Harris performed an extensive analysis on a free space propagating optical switch, based on acousto-optic cell technology. He carried out 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 ±1Hz 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 ranges from 2 to 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. He studied the question of output power combining in single mode fibers and proved that no passive power combining scheme can be better than 1/N efficient, where N is the number of output ports. Active combiners, such as ones based on acousto-optic deflection, can be completely efficient, but only with a significant increase in hardware complexity. He also performed analyses and conducted experiments to verify insertion loss, worst-case optical crosstalk, and acousto-optic cell reconfiguration time predictions for an 8x8 switch. The measured values are in good agreement with the theory. He used Close network theory to show how multiple switches can be concatenated to produce larger switches. Finally, he developed several useful methods to provide multicasting and broadcasting capabilities for the switch. Supplemental funds has been obtained from BellSouth Enterprises in support of this work. Harris included the results of this research in his PhD Thesis "Acousto-Optic Photonic Switch"; this effort has been published[6,7] and is included in Appendices E and F.

4.0 Fresnel Transform Studies

A. VanderLugt 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, he 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. He 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. 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. He 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. 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 published[8,9] and is included as Appendices G and H.

A. VanderLugt also 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. 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. This work was published[10] and included in Appendix I. An expanded continuation of this theoretical work has been published in Applied Optics[11] and is included in Appendix J.

M. J. Wardlaw performed experimental work on the detection of short pulses by Fresnel preprocessing. In the experimental system, the data acquisition and control is centralized within a Macintosh FX computer. Two specialized I/O cards are used; one card provides an IEEE-488 interface capability while the other provides both digital and analog I/O capability. The IEEE-488 interface is connected to an arbitrary function generator, a spectrum analyzer, and a programmable oscilloscope. Using this equipment and the Fresnel preprocessor and optical spectrum analyzer he constructed, Wardlaw could detect pulses as short as the theoretical limit of 20 nanoseconds for this system and separated by as little as 60 nanoseconds or by as much as 17 microseconds. He distinguish short pulse signals from cw signals by taking measurements at two different Fresnel planes and subtracting the spectrum of one measurement from that of the other.

Since spectral contributions from the cw signals are essentially the same for the two cases, they are removed by the subtraction process. The spectral contribution from the pulses is retained, however, because there is a known shift in the frequency of the sinusoidal interference pattern. To remove the cw signals, we implemented a frequency-variant operation on the output of the spectrum analyzer in which the frequency difference between the elements of a positive/negative doublet frequency response is a linear function of the frequency. Multiple sets of pulses may be present within the acousto-optic cell at the same time. To test the performance of the system under this condition, we generated five pulses within a time interval equal to about 40% of the cell transit time. By recognizing that fringe patterns will be generated from each pulse pair, we can predict that the number of combinations, or pulse pairs, available to generate frequencies at the output of the Fresnel preprocessor and, therefore, at the output of the spectrum analyzer. For our scenario, five pulses, taken two at a time, yield 10 possible frequencies. The experimental results nicely confirmed the theoretical calculations and formed the basis for his MS thesis entitled "Detection of Short Duration Pulses by Fresnel and Fourier Transformations." A paper detailing the results of this work has been accepted by Applied Optics for publication[12] and is included in Appendix K.

5.0 Other Studies

Acousto-optic devices use wideband chirp signals to scan light beams at very high scan rates. The rich variety of conditions under which an acousto-optic scanner can be implemented and the interlocking nature of the performance parameters often leads to ad hoc solutions in which conflicts arise among the required performance specifications. A. VanderLugt has considered all possible scanning configurations and classified them into four basic types. A consistent set of design relationships for each of the scanning configurations has been developed and presented in both tabular and graphic forms from which a preliminary design is obtained. Examples are given for how these design relationships are used. A paper covering this work has been published in Applied Optics[13] and is included in Appendix L.

A surprising number of traditional topics, such as the Rayleigh resolution criterion, spatial filtering, bandlimited signals, the sampling theorem, phase contrast microscopy, and white light holography can be explained using simple interference theory. These basic results are then easily extended to the N-source case to introduce diffraction theory, and facilitates the teaching of the elements of modern optics to junior and senior level students. This work has been published[14] and is included in Appendix M. An algorithm for efficiently and accurately calculating Fresnel transforms has been accepted by Optical Engineering and is included in Appendix N.

The detection of femtosecond pulses is generally accomplished indirectly through an indirect method using the autocorrelation function; that is, the pulse is used to measure its own

duration. A. VanderLugt and T.W. Powell began to look at the possibility of measuring the duration of femtosecond pulses by creating a simple spatial interference pattern from the Fresnel transforms of two sources; the pulse length is found by counting the number of fringes in the resulting spatial interference pattern. Several optical implementations for generating the spatial interference patterns were evaluated and several signal representations were considered to help in the evaluation, but some issues arose relating to the validity of the usual notions of geometrical and physical optics when applied to short pulses. These issues are currently under evaluation. An optical breadboard system capable of detecting the spatial/temporal characteristics of pulses has been built and is being tested using the femtosecond lasers being built by Dr. B. D. Guenther and his students at Duke University.

6.0 Miscellany

A. VanderLugt presented a paper on "Sampling of Fresnel transforms" at the Technical Symposium on Advances in Optical Information Processing IV, held in Orlando, FL, 18-20 April, 1990 .

A. VanderLugt presented a paper on "Short pulse detection by acousto-optic processing" at the Technical Symposium on Advances in Optical Information Processing V, held in Orlando, FL, 21-24 April, 1992 .

A. VanderLugt served as Chairman of a Foreign Applied Sciences Assessment Center panel on Soviet Research Related to Optical Processing. The major activity of the panel, all of whom are experts in this field, is their review of published Soviet literature, to assess their capabilities, and to recommend possible changes in US policy.

At the request of the Director of Physics, U.S. Army Research Office, A. VanderLugt helped review the progress made at the Photonics Center at West Point. This visit was on April 30-May 1, 1992 to evaluate their progress against their objectives.

A. VanderLugt finished a 604 page text *Optical Signal Processing* which is the first text that covers the full range of topics from pattern recognition and target identification to wideband signal processing. This text, published by Wiley Interscience, is used both as a reference book for researchers in optical signal processing and as a textbook for senior level undergraduate and graduate level courses.

The participating personnel on this research study were Dr. A. VanderLugt, Principal Investigator, D.O. Harris, T.P. Karnowski, R.S. Ward, and C.S. Anderson, M.J. Wardlaw, S.M. Devonshire, T. W. Powell, Jr, and H-S. Wu:

D.O. Harris received his Doctor of Philosophy degree in December, 1990, with a thesis entitled "Acousto-Optic Photonic Switch," partially fulfilling the requirements for the PhD degree.

T.P. Karnowski received his Masters of Science degree in December, 1990, with a thesis entitled "Area Modulation Filter in an Interferometric Acousto-Optic Signal Processor," partially fulfilling the requirements for the MSc degree.

R.S. Ward received his Masters of Science degree in May, 1991, with a thesis entitled "The Excision of Narrowband Interference from Wideband Signals using an Open-Loop Adaptive Technique in an Acousto-Optical Signal Processor," partially fulfilling the requirements for the MSc degree.

C.S. Anderson received his Doctor of Philosophy degree in December, 1991, with a thesis entitled "Hybrid Acousto-Optic and Digital Channel Equalization for Microwave Digital Radios," partially fulfilling the requirements for the PhD degree.

M.J. Wardlaw received his Masters of Science degree in December, 1992, with a thesis entitled ""Detection of Short Duration Pulses by Fresnel and Fourier Transformations," partially fulfilling the requirements for the MSc degree.

7.0 References

1. A. VanderLugt, "Adaptive Optical Processor," Applied Optics, Vol. 21, p. 4005 (1982)
2. W. R. Beaudet, A. M. Bardos and A. VanderLugt, "Stability Considerations for Adaptive Optical Filtering, Applied Optics, Vol. 25, p. 2314 (1986)
3. T.P. Karnowski and A. VanderLugt, "Generalized Filtering in Acousto-Optic Systems using Area Modulation," Applied Optics, Vol. 30, pp. 2344-2353 (1991)
4. C.S. Anderson and A. VanderLugt, "Hybrid Acousto-Optic and Digital Equalization for Microwave Digital Radio Channels," Optics Letters,Vol. 15, pp.1182-1184 (1990)
5. R.N. Ward and A. VanderLugt, "Signal Distortion in an Adaptive Excision System," Optical Engineering, Vol. 31, pp.606-613 (1992)
6. D.O. Harris, "Multichannel acousto-optic crossbar switch," Applied Optics, Vol. 30, pp. 4245-4256 (1991)
7. D.O. Harris and A. VanderLugt, "Multichannel acousto-optic crossbar switch with arbitrary signal fan-out," Applied Optics, Vol. 31, pp. 1684-1686 (1992)
8. A. VanderLugt, "Sampling of Fresnel transforms," Proc. SPIE, Vol. 1296, Advances in Optical Information Processing IV, pp. 246-252 (1990)
9. A. VanderLugt, "Optimum Sampling of Fresnel Transforms," Applied Optics, Vol. 29, pp. 3352-3361 (1990).
10. A. VanderLugt, "Short pulse detection by acousto-optic processing," Proc. SPIE, Vol. 1704, Advances in Optical Information Processing V, pp. 146-152 (1992)

11. A. VanderLugt, C. S. Anderson and P. W. J. Melsa, "Time-delay detection of short pulses by Fresnel and Fourier transformations," Appl. Opt., Vol 32, p. 3761-3771 (1993)

12. M. J. Wardlaw, "Detection of short pulses by Fresnel preprocessing," Accepted for publication by Applied Optics.

13. A. VanderLugt and A.M. Bardos, "Design relationships for acousto-optic scanning systems," Applied Optics, Vol. 31, pp. 4058-4068 (1992)

14. "Interference and Diffraction," B.D. Guenther and A. VanderLugt, IEEE Trans. Education, Vol. 35, pp.126-132 (1992)

APPENDIX A

GENERALIZED FILTERING IN ACOUSTO-OPTIC SYSTEMS USING AREA MODULATION

Reprinted from Applied Optics

Volume 30, Pages 2344-2353, 10 June 1991

Generalized filtering in acoustooptic systems using area modulation

T. P. Karnowski and Anthony VanderLugt

We report on a technique for implementing general filtering operations in acoustooptic signal processing systems. We use a binary recording method called area modulation to reduce linearity problems associated with spatial light modulators that operate by controlling the amplitude of light. We present an analysis of this method, we report on experiments with an acoustooptic system to verify the analysis, using photographic film and a liquid crystal display to implement the area modulation masks, and we discuss the limitations of the technique.

I. Introduction

Acoustooptic signal processors perform operations such as spatial filtering, spectrum analysis, correlation, and adaptive filtering. Some processing applications need spatial filters whose phase and amplitude responses are changed frequently or controlled adaptively in real time. These filter functions are often difficult to implement with accurate response characteristics. We describe an acoustooptic processing architecture which uses the second spatial dimension of the optical system to accommodate an auxiliary 2-D spatial light modulator to implement a generalized filtering operation.

Area modulation has been applied to spectral analysis, optical computing, point spread function synthesis, and correlator applications[1-4] and reduces the effects of amplitude nonlinearities associated with conventional amplitude spatial modulation techniques. In an acoustooptic processor, the input signal is moving through space so that implementing the desired filter as a stationary impulse response is particularly attractive. We generate the impulse response with a real time spatial light modulator, such as a liquid crystal display, whose response can be changed adaptively through computer control.

In Sec. II we introduce the area modulation technique and give a heuristic description of the operation of the processor. In Sec. III we develop the processor transfer function, and in Sec. IV we discuss the effects of quantization, aberrations, beam apodization, photodetector size, and phase errors on the operation. We report the results of filtering experiments in Sec. V.

II. Background

Figure 1 shows a modified Mach-Zehnder interferometer in which the signal branch contains a slow shear TeO_2 acoustooptic cell of length L, driven by $f(t) = s(t)\cos(2\pi f_c t)$, where $s(t)$ is a baseband signal and f_c is a carrier frequency. The optical signal corresponding to the positive diffracted order is represented by

$$f(x,t) = a(x)s\left(t - \frac{T}{2} - \frac{x}{v}\right)\exp\left[j2\pi f_c\left(t - \frac{T}{2} - \frac{x}{v}\right)\right], \quad (1)$$

where $a(x)$ is the aperture weighting function, $T = L/v$ is the length of the time signal within the acoustooptic cell, and v is the acoustic velocity. Lens L_2 produces the Fourier transform of $f(x,t)$ at plane P_2:

$$F(\alpha,t) = \exp\left[j2\pi f_c\left(t - \frac{T}{2}\right)\right]\int_{-\infty}^{\infty} a(x)s\left(t - \frac{T}{2} - \frac{x}{v}\right) \times \exp[j2\pi(\alpha - \alpha_c)x]dx, \quad (2)$$

where the spatial frequency variable in Fourier space is identified by α. The transform $F(\alpha,t)$ is centered at a spatial carrier frequency $\alpha_c = f_c/v$, in plane P_2, because of the temporal carrier frequency f_c.

A generalized filtering operation can be implemented as an impulse response $h(x,t)$ in plane P_0, where the time dependence of the impulse response is assumed to be slowly varying relative to the signal $s(t)$. Lens L_2 produces the optical Fourier transform of $h(x,t)$ at plane P_2:

The authors are with North Carolina State University, Electrical & Computer Engineering Department, Raleigh, North Carolina 27695-7911.

Received 8 August 1990.
0003-6935/91/172344-10$05.00/0.

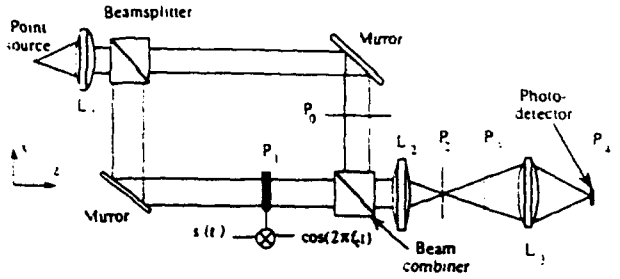


Fig. 1. Optical system for implementing generalized filters.

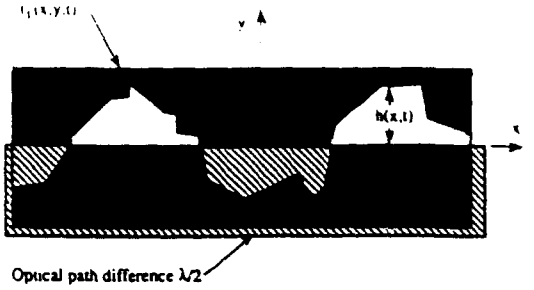


Fig. 2. Unbiased phase plate area modulation function.

$$H(\alpha,t) = \int_{-\infty}^{\infty} h(x,t)\exp(j2\pi\alpha x)dx. \tag{3}$$

The intensity at plane P_2 is the squared magnitude of the sum of $F(\alpha,t)$ and $H(\alpha,t)$:

$$I(\alpha,t) = |F(\alpha,t) + H(\alpha,t)|^2. \tag{4}$$

We use a large photodetector to collect all the light at plane P_2, expand $I(\alpha,t)$ into its component parts, and find that the photocurrent is proportional to

$$g(t) = \int_{-\infty}^{\infty} [|H(\alpha,t)|^2 + |F(\alpha,t)|^2 + 2\,\mathrm{Re}[H^*(\alpha,t)F(\alpha,t)]]d\alpha. \tag{5}$$

We are generally interested only in the cross-product term in the integrand. The first two terms of Eq. (5) are at baseband, where they are rejected by a bandpass filter that accepts the cross-product term. The bandpass filter has a center frequency f_c and a bandwidth equal to twice that of the baseband signal $s(t)$. By proper choice of the function $h(x,t)$, we implement an arbitrary filter as shown by the third term of Eq. (5).

Although the amplitude of the impulse response could be implemented by a transmittance function, it is difficult to obtain the desired degree of linearity or the required negative values. Because the signals in this acoustooptic system are basically 1-D, we can use the second dimension to implement area modulation techniques for generating the impulse response. In the area modulation technique, we create a 2-D function $t_f(x,y,t)$, using a spatial light modulator such as a liquid crystal display, whose transmittance is a function of both space and time. The transmittance of the spatial light modulator is either zero or one, so that when the transmittance is integrated in the vertical direction, we create an analog signal $h(x,t)$ that has, in principle, no amplitude distortion. The area of the spatial light modulator is partitioned into an upper half, producing the positive values, and a lower half, producing the negative values through a π phase shift. Since the liquid crystal display is controlled electronically, we can adaptively change the impulse response as needed.

An arbitrary filtering operation, such as one in which the filter is asymmetric about its central frequency, requires that the impulse response be complex valued. Such filters can be realized, however, with real valued impulse response functions if we first induce symmetry in the frequency domain by reflecting the desired filter function about its highest frequency. To produce the desired filter, we center the spectrum of the signal to be processed on that half of the filter function produced by $h(x,t)$ that has the required asymmetric response; the other half of the filter function is not used.

Felstead described four types of area modulation.[1] For our purposes we consider only the unbiased phase plate technique shown in Fig. 2. The height of the area modulation window above the x-axis is proportional to $h(x,t)$. The transparent plate has an optical path difference of $\lambda/2$ which imparts a π phase shift in the light transmitted below the x-axis, shown in Fig. 2 as the crosshatched region. The transmittance for this modulation technique is represented by

$$t_f(x,y,t) = \mathrm{rect}\left(\frac{x}{L}\right)\mathrm{rect}\left\{\frac{y - \frac{h(x,t)}{2}}{h(x,t)}\right\}\mathrm{sgn}(y), \tag{6}$$

where

$$\mathrm{sgn}(y) = \begin{cases} +1 & y \geq 0, \\ -1 & y < 0. \end{cases} \tag{7}$$

The first rect function in Eq. (6) bounds the function in the horizontal direction. The second rect function specifies the vertical aperture height by the term in the denominator, while the numerator specifies the vertical displacement or center of the aperture. In the remainder of the paper we use the term sgn plate to identify the optical element that creates the π phase shift needed to produce the negative values for the impulse response.

The normal mode of operation is to place the area modulator function $t_f(x,y,t)$, as given by Eq. (6), in plane P_0. The phase response of typical area modulators, however, is not well controlled and produces errors in the realized filter function. We make an important modification to our system by moving the magnitude portion of the area modulator $t_f(x,y,t)$ to plane P_3, while keeping only the sgn plate inside the interferometer. This method has the advantage that, since both the reference and signal beams pass through the spatial light modulator, path differences that cause unwanted phase modulation are eliminated. Since the transmittance of the area modulator has only binary values, the desired amplitude modulation is preserved as shown in Sec. III. The only optical element required in the reference beam is therefore the sgn plate which can be made by coating a high quality

optical flat with a $\lambda/2$ coating over half of its surface to produce the required negative values of $h(x,t)$.

III. Processor Analysis

In Fig. 1 we show that the acoustooptic cell is imaged by lens L_2 at plane P_3, where it passes through the area modulator mask $t_i(x,y,t)$. The sgn plate in plane P_0 is also imaged at plane P_3 by lens L_2 to form the complete area modulator mask. The signal beam at plane P_3, after we remove all but the positive diffracted order by a spatial filter in plane P_2, is

$$f(x,t) = a_s(x,y)s\left(t - \frac{T}{2} - \frac{x}{v}\right) \times \exp\left[j2\pi f_c\left(t - \frac{T}{2} - \frac{x}{v}\right)\right] t_i(x,y,t) \exp[j\phi(x,y)], \quad (8)$$

where $a_s(x,y)$ is the amplitude profile function of the illumination beam, $t_i(x,y,t)$ is the transmittance of the spatial light modulator as a function of time, $\phi(x,y)$ is a quadratic phase term caused by the imaging operation, and we ignore the coordinate reversal in passing from plane P_1 to plane P_3. Similarly, the reference beam at plane P_3 is

$$r(x,t) = a_r(x,y)t_i(x,y,t)\ \mathrm{sgn}(y)\psi(x,y) \times \exp(-j2\pi\alpha_c x) \exp[j\phi(x,y)], \quad (9)$$

where $a_r(x,y)$ is the amplitude profile of the illumination beam, $\exp(-j2\pi\alpha_c x)$ is a linear phase shift imparted to the reference beam by rotating the beam combiner, the $\mathrm{sgn}(y)$ function represents the phase shifting element, and $\psi(x,y)$ is a term included to represent the aberrations in the interferometer.

Although the detection could be performed by an area photodetector in plane P_3, the required detector must be as large as the spatial light modulator. Such large photodetectors typically have large capacitances, which limit the bandwidth of the system. We therefore prefer to detect in a Fourier plane where the required photodetector area is much less. Lens L_3 produces the Fourier transforms of Eqs. (8) and (9), which are summed at plane P_4, square-law detected, and integrated over α and β by the photodetector; β is the spatial frequency variable normal to α in plane P_4. The two terms resulting from the squared magnitude of the signal beam and the reference beam alone are baseband components that are removed by a bandpass filter as discussed before. The third term is the information bearing signal, given by the cross-product term

$$I_3(\alpha,\beta,t) = 2\,\mathrm{Re}\left[\mathcal{F}\left\{a_s(x,y)s\left(t - \frac{T}{2} - \frac{x}{v}\right) \times \exp\left[j2\pi f_c\left(t - \frac{T}{2} - \frac{x}{v}\right)\right] t_i(x,y,t)\right\} \times \mathcal{F}^*\{a_r(x,y)t_i(x,y,t)\ \mathrm{sgn}(y)\psi(x,y) \exp(-j2\pi\alpha_c x)\}\right], \quad (10)$$

where $\mathcal{F}$ denotes the 2-D spatial Fourier transform.

We represent the photodetector area as $\mathrm{rect}(\alpha/2\alpha_{co}) \times \mathrm{rect}(\beta/2\beta_{co})$, where α_{co} and β_{co} are the spatial frequencies at the edges of the photodetector. We expand the Fourier transform operation in integral form to show that the output current $i_3(t)$, produced by $I_3(\alpha,\beta,t)$, is

$$i_3(t) = \int\int \mathrm{rect}(\alpha/2\alpha_{co})\,\mathrm{rect}(\beta/2\beta_{co}) \times \mathrm{Re}\left[\int\int a_s(u,w)s\left(t - \frac{T}{2} - \frac{u}{v}\right) \times \exp\left[j2\pi f_c\left(t - \frac{T}{2} - \frac{u}{v}\right)\right] t_i(u,w,t) \exp[j2\pi(\alpha u + \beta w)]\,du\,dw \times \int\int a_r^*(x,y)t_i^*(x,y,t)\ \mathrm{sgn}(y)\psi^*(x,y) \times \exp(j2\pi\alpha_c x) \exp[-j2\pi(\alpha x + \beta y)]\,dx\,dy\,d\alpha\,d\beta, \quad (11)$$

where we ignore unessential scaling factors. Next, we combine the integrals to obtain

$$i_3(t) = \mathrm{Re}\int\int\int\int\int\int \mathrm{rect}(\alpha/2\alpha_{co}) \times \mathrm{rect}(\beta/2\beta_{co})a_s(u,w)s\left(t - \frac{T}{2} - \frac{u}{v}\right)\exp\left[j2\pi f_c\left(t - \frac{T}{2} - \frac{u}{v}\right)\right] \times t_i(u,w,t)a_r^*(x,y)t_i^*(x,y,t)\ \mathrm{sgn}(y)\psi^*(x,y) \times \exp(j2\pi\alpha_c x) \exp[j2\pi[\alpha(u-x) + \beta(w-y)]]\,dx\,dy\,du\,dw\,d\alpha\,d\beta. \quad (12)$$

We recognize that the integration over α and β produces a sinc function, so the output becomes

$$i_3(t) = \mathrm{Re}\int\int\int\int \mathrm{sinc}[2\alpha_{co}(u-x)]\mathrm{sinc}[2\beta_{co}(w-y)] \times a_s(u,w)s\left(t - \frac{T}{2} - \frac{u}{v}\right)\exp\left[j2\pi f_c\left(t - \frac{T}{2} - \frac{u}{v}\right)\right] \times t_i(u,w,t)a_r^*(x,y)t_i^*(x,y,t)\ \mathrm{sgn}(y)\psi^*(x,y) \times \exp(j2\pi\alpha_c x)\,dx\,dy\,du\,dw. \quad (13)$$

Since the photodetector collects all the light from plane P_2, no information at the Fourier plane is lost and we can therefore treat the sinc functions as delta functions. We use the sifting property of the delta function to find that

$$i_3(t) = \mathrm{Re}\left\{\int\int a_s(x,y)a_r^*(x,y)|t_i(x,y,t)|^2\ \mathrm{sgn}(y)\psi^*(x,y) \times s\left(t - \frac{T}{2} - \frac{x}{v}\right)\exp\left[j2\pi f_c\left(t - \frac{T}{2} - \frac{x}{v}\right)\right]\exp(j2\pi\alpha_c x)\,dx\,dy\right\}. \quad (14)$$

Finally, we assume that apodization $a(x,y)$ of the signal and reference beams is generally identical, because of the common light source, so that

$$i_3(t) = \mathrm{Re}\left\{\exp\left[j2\pi f_c\left(t - \frac{T}{2}\right)\right]\int\int |a(x,y)|^2|t_i(x,y,t)|^2\psi^*(x,y) \times \mathrm{sgn}(y)s\left(t - \frac{T}{2} - \frac{x}{v}\right)\exp[j2\pi(\alpha_r - \alpha_c)x]\,dy\,dx\right\}. \quad (15)$$

The central result [Eq. (15)] is a convolution integral, with x/v being the displacement variable, and with y available as an extra dimension used in realizing the magnitude of the desired filter function through area modulation. For a binary function, $|t_i(x,y,t)|^2 = t_i(x,y,t)$ and Eq. (15) confirms that any aberrations caused by the spatial light modulator are eliminated as claimed in Sec. II.

We illustrate that Eq. (15) represents an arbitrary filtering operation, since the value of $t_i(x,y,t)$ is under our control, with a simple filtering operation. Sup-

pose that the apodization function $a(x,y)$ is separable into the product $a(x)a(y)$. We incorporate the area modulation function in the limits of the integration over y, an operation justified by the nature of the area modulator:

$$i_3(t) = \mathrm{Re}\left\{\exp\left[j2\pi f_c\left(t-\frac{T}{2}\right)\right]\int |a(x)|^2 s\left(t-\frac{T}{2}-\frac{x}{v}\right) \times \exp[j2\pi(\alpha_r-\alpha_c)x]\int_0^{h(x,t)} |a(y)|^2\psi^*(x,y)\,dy\,dx\right\}, \quad (16)$$

where $h(x,t)$ is the height of the area modulation function. For an aberration-free system, $\psi(x,y)$ is unity. Furthermore, we use a uniform apodization function so that the output current becomes

$$i_3(t) = \mathrm{Re}\left\{\exp\left[j2\pi f_c\left(t-\frac{T}{2}\right)\right]\int h(x,t)s\left(t-\frac{T}{2}-\frac{x}{v}\right) \times \exp[j2\pi(\alpha_r-\alpha_c)x]\,dx\right\}, \quad (17)$$

which, for $\alpha_r = \alpha_c$, clearly shows the desired convolution. For illustrative purposes, we examine the integral when the applied signal $s(t)$ is a constant so that Eq. (17) becomes the Fourier transform

$$i_3(t) = \mathrm{Re}\left\{\exp\left[j2\pi f_c\left(t-\frac{T}{2}\right)\right]\int h(x,t)\exp[j2\pi(\alpha_r-\alpha_c)x]\,dx\right\}, \quad (18)$$

and the output becomes

$$i_3(t) = H(\alpha_r-\alpha_c,t)\cos\left[2\pi f_c\left(t-\frac{T}{2}\right)\right]. \quad (19)$$

If the carrier frequency α_c is set equal to α_r, we see that the envelope of the output of the system is equal to $H(0,t)$; that is, it is the filter response for an input signal at zero frequency, as expected, because $s(t)$ is a constant whose temporal frequency is zero. We probe the frequency response of the filter by scanning the carrier frequency over the full frequency range of the system.

IV. Nonideal Performance

Several factors limit the performance of the system. The issues we analyze are the effects of quantization, undesired phase modulation, beam apodization, photodetector size, and the effect of errors in the phase inverting element. As we shall see, although the area modulation technique intrinsically avoids the nonlinearities associated with amplitude spatial modulators, some nonlinearities are generated by nonideal properties of the process.

A. Quantization

A liquid crystal display has a pixel structure in the x- and y-directions that sets the resolution limit of the device. In the x-direction, a pixel is equivalent to a sample so that we satisfy the Nyquist sampling theorem if we assign two samples to a period of the highest spatial frequency in the signal. In the y-direction, the total number N of samples affects the accuracy to which we can approximate the desired value of $h(x,t)$. If, for example, the liquid crystal display has $N = 256$ samples in the vertical direction, the amplitude of $h(x,t)$ has a quantizing error of one part in 512 for the largest signal value. Smaller signal values have correspondingly larger percentage errors. The value of N also sets the limit on the dynamic range of the impulse function because it sets the ratio of the largest to the smallest value.

B. Aberrations

The aberration function $\psi(x,y)$, as incorporated in Eq. (9), is of the form

$$\psi(x,y) = \exp\left[j\frac{2\pi}{\lambda}\theta(x,y)\right], \quad (20)$$

where $\theta(x,y)$ is the function describing the aberrations caused by all the elements of the interferometer. Note that aberrations contributed by any element outside the interferometer are common to both signal and reference beams and do not contribute to the output, as shown by Eq. (15). To illustrate the effects of aberrations in the x- and y-directions, we separate $\theta(x,y)$ into $\theta(x) + \theta(y)$. We set $|a(x)|^2 = \mathrm{rect}(x/L)$, while the limits of integration on y range from 0 to $h_{\max}(x,t)$; we fix the impulse response $h(x,t)$ in time by ignoring its adaptive qualities for the moment. We bring the $\theta(x)$ term out of the integration over y in Eq. (16) to yield

$$i_3(t) = \mathrm{Re}\left\{\exp\left[j2\pi f_c\left(t-\frac{T}{2}\right)\right]\int_{-L/2}^{L/2} s\left(t-\frac{T}{2}-\frac{x}{v}\right) \times \exp\left[-j\frac{2\pi}{\lambda}\theta(x)\right]\int_0^{h(x)}\exp\left[-j\frac{2\pi}{\lambda}\theta(y)\right]dy\,dx\right\}, \quad (21)$$

where we set $\alpha_r = \alpha_c$ so that the reference and signal beams are centered at the same spatial frequency. We define the inphase and quadrature components of the impulse response as

$$h_c(x) = \int_0^{h(x)}\cos\left[\frac{2\pi}{\lambda}\theta(y)\right]dy, \quad (22)$$

$$h_s(x) = \int_0^{h(x)}\sin\left[\frac{2\pi}{\lambda}\theta(y)\right]dy. \quad (23)$$

With these definitions, we express Eq. (21) as

$$i_3(t) = \mathrm{Re}\left\{\int_{-L/2}^{L/2}[h_c(x)-jh_s(x)]s\left(t-\frac{T}{2}-\frac{x}{v}\right) \times \exp\left[j2\pi f_c\left(t-\frac{T}{2}\right)\right]\exp\left[-j\frac{2\pi}{\lambda}\theta(x)\right]dx\right\}. \quad (24)$$

We take the real part of Eq. (24) and find that the resulting signal becomes

$$i_3(t) = \int_{-L/2}^{L/2} s\left(t-\frac{T}{2}-\frac{x}{v}\right)h_c(x)\cos\left[2\pi f_c\left(t-\frac{T}{2}\right)-\frac{2\pi}{\lambda}\theta(x)\right]dx - \int_{-L/2}^{L/2} s\left(t-\frac{T}{2}-\frac{x}{v}\right)h_s(x)\sin\left[2\pi f_c\left(t-\frac{T}{2}\right)-\frac{2\pi}{\lambda}\theta(x)\right]dx. \quad (25)$$

The impulse response $h(x)$ includes a newly generated, undesired quadrature component $h_s(x)$, caused by phase modulation in the y-direction, in addition to the desired part $h_c(x)$ that would result if $\theta(y) = 0$. We also note that the aberrations in the x-direction cause a phase modulation of the carrier frequency.

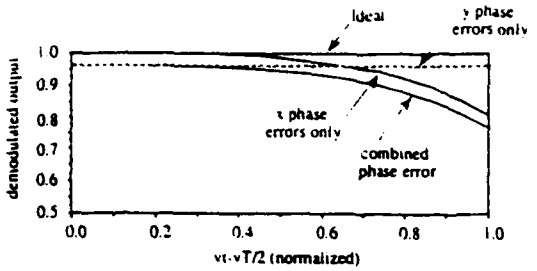


Fig. 3. System output after synchronous demodulation.

The result of these aberrations is a distortion of the desired system impulse response. To gain further insight into the nature of the distortion, we let $s(t)$ become a delta function, so that the essential form of $h(x)$ is revealed. Using the sifting theorem and a trigonometric identity, we find that

$$i_3(t) = \sqrt{h_c^2\left(vt - \frac{vT}{2}\right) + h_s^2\left(vt - \frac{vT}{2}\right)} \times \cos\left[2\pi f_c\left(t - \frac{T}{2}\right) - \frac{2\pi}{\lambda}\theta_x\left(vt - \frac{vT}{2}\right) - \tan^{-1}\frac{h_s\left(vt - \frac{vT}{2}\right)}{h_c\left(vt - \frac{vT}{2}\right)}\right], \quad (26)$$

where $\theta_x(vt - vT/2)$ denotes the x-component of the aberration and the argument $vt - vT/2$ shows that the delta function traces out the impulse response function and converts the spatial impulse response $h(x)$ into a temporal impulse response $h(vt - vT/2)$. The result of a synchronous demodulation of Eq. (26) gives the result

$$v_d(t) = \sqrt{h_c^2\left(vt - \frac{vT}{2}\right) + h_s^2\left(vt - \frac{vT}{2}\right)} \times \cos\left[\frac{2\pi}{\lambda}\theta_x\left(vt - \frac{vT}{2}\right) + \tan^{-1}\frac{h_s\left(vt - \frac{vT}{2}\right)}{h_c\left(vt - \frac{vT}{2}\right)}\right]. \quad (27)$$

Since the scanning of the impulse response function causes $h(x) \rightarrow h(vt - vT/2)$, this result indicates that the impulse response is distorted in amplitude.

To illustrate the worst-case signal degradation, we assign $h(x)$ its maximum value. For analytical purposes, we model the aberrations with a polynomial function whose first nonlinear term is given by

$$\theta(y) = \frac{\lambda}{M}\left(\frac{2y}{L}\right)^2. \quad (28)$$

This function shows that the maximum phase error at $y = L/2$ is λ/M, where an allowable value of M is to be found. In Fig. 3 we plot the ideal response, normalized to unity, as a constant for all values of time. The effect of phase errors in the y-direction for any value of M is found from Eq. (21) to simply reduce the magnitude of the output which does not affect the accuracy of the implemented filter function. If $M \geq 4$, the attenuation is not more than 1 dB. The more serious impact, as shown in Fig. 3, is caused by phase errors in the x-direction. In all cases we assume that the thickness

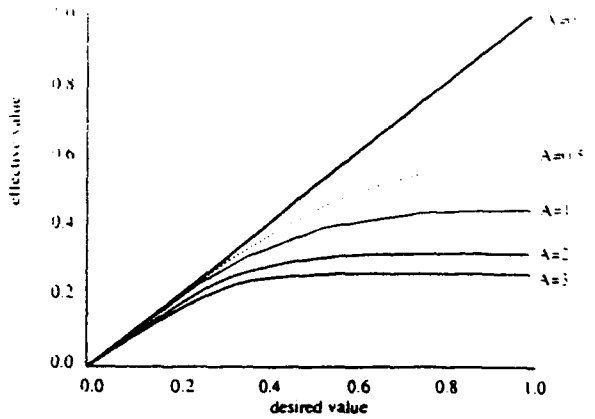


Fig. 4. Compression due to apodization in the y-direction.

function $\theta(vt - vT/2)$ in Eq. (27) is positive, to represent worst-case performance.

We therefore find that, while the linearity of the spatial light modulator should not be an issue in area modulation, aberrations introduce a constraint on the linearity of the impulse response functions. Fortunately, these nonlinearities are not signal dependent. They are functions purely of spatial positions in the optical system and may be partially corrected, when the system is calibrated, by the use of a lookup table to compensate the height of the area modulator. The number of resolution elements N of a spatial light modulator sets the limit on the degree to which distortions in the values of $h(x)$ are corrected, while still producing the required dynamic range for the impulse response function. An alternative method is to control the amplitude of each pixel position, through a lookup table, to compensate the nonlinearity. This fixed pattern, which can be used in conjunction with control over the height of the area modulator, is also not signal dependent so that no new nonlinearities are introduced.

C. Beam Apodization

Another possible detrimental effect on the performance of the system performance is the effect of the beam apodization $a(x,y)$. The effects of this term on the x-integration are an easily predicted multiplicative distortion, since Eq. (16) shows that the desired impulse response $h(x,t)$ is replaced by $|a(x)|^2 h(x,t)$. This weighting also can be corrected, if necessary, by using a lookup table because the distortion is a function only of the geometry of the system. If the illumination profile is Gaussian, the height of the area modulator near the edges must be proportionally increased relative to those at the center.

The apodization effect in the y-direction, however, is less clear because the signal is integrated in this direction. Since laser beams naturally have a Gaussian profile, a commonly used aperture function is

$$a(y) = \mathrm{rect}\left(\frac{y}{L}\right)\exp\left[-2A\left(\frac{y}{L}\right)^2\right]. \quad (29)$$

We integrated $|a(y)|^2$ from 0 to the desired value of $h(x)$

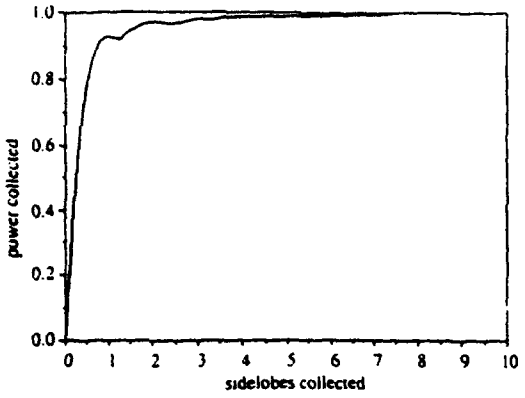


Fig. 5. Power collected from a sinc distribution.

to obtain the effective $h(x)$ value. Figure 4 shows the resulting curves for apodizations of $A = 0, 0.5, 1, 2$, and 3. The plots show a compression effect which also affects the linearity of the area modulation technique. Obviously, the best apodization in the y-direction is a small value for A; too small a value for A, however, leads to poor optical power usage.

D. Photodetector Size

The next nonideal situation involves the size of the photodetector. All area modulators have limited spatial resolution as set by the pixel size. The highest frequency in the α-direction is limited to $\alpha_{co} = 1/2d_0$, where d_0 is the minimum resolvable sample size of the spatial light modulator. If the photodetector has an insufficient width to collect all the diffracted light, the result is equivalent to low pass filtering the signal.

The effect of a limited photodetector size in the vertical direction is not as obvious. We first calculate the Fourier transform of a positive area modulation signal in the vertical direction:

$$P(\beta,x) = \int_{-\infty}^{\infty} t(x,y)\exp(j2\pi\beta y)dy = \int_{0}^{h(x)} \exp(j2\pi\beta y)dy, \quad (30)$$

where the apodization function $|a_r(y)|^2$ has been set equal to unity. We perform the integration with respect to y to obtain

$$P(\beta,x) = \frac{\exp[j2\pi\beta h(x)] - 1}{j2\pi\beta}, \quad (31)$$

which reduces to

$$P(\beta,x) = h(x)\exp[j2\pi\beta h(x)]\,\mathrm{sinc}[\beta h(x)]. \quad (32)$$

If the value of $h(x)$ is small, the sinc function is broad, with a wide main lobe. If the photodetector is too small to collect all the sidelobes, the output signal is reduced somewhat. In Fig. 5 we show the relative output as a function of the number of sidelobes collected. To avoid amplitude distortion of the signal, we ensure that the photodetector collects light to at least the fourth sidelobe of the sinc function produced by a single pixel. Such a detector therefore collects light to at least the eighth sidelobe for any other signal level, producing a negligible impact on signal quality. The net effect is that a signal level represented by a one pixel height is slightly distorted, but the distortion is well within the quantizing error limits.

E. Nonideal sgn Plate Responses

A nonideal sgn plate may have an error in the thickness of the coating used to produce the negative values for $h(x)$. Suppose that the plate has a phase of 0 rad above the x-axis and a phase of $\pi \pm \Delta e$ rad below the axis. Such an error causes a distortion in the impulse response because the plate does not produce the proper balance between the positive and negative values. This error is most easily detected by measuring the residual average value, $|H(0)|$, in the Fourier plane when $s(t) = 1$ and the area modulator is a clear aperture.

We determine the suppression of the average value of $h(x)$ as a function of Δe. By integrating over y, we find that the contribution to $H(0)$ from the region above the x-axis is

$$\int_0^{L/2} dy = \frac{L}{2}. \quad (33)$$

The contribution from the region below the axis is

$$\int_{-L/2}^{0} \exp[j(\pi + \Delta e)]dy = \frac{L}{2}[\cos(\pi + \Delta e) + j\sin(\pi + \Delta e)]. \quad (34)$$

The normalized magnitude of the sum of Eqs. (33) and (34) yields

$$|H(0)| = \tfrac{1}{2}\sqrt{[1 + \cos(\pi + \Delta e)]^2 + \sin^2(\pi + \Delta e)} = \sin(\Delta e/2). \quad (35)$$

If the plate were perfect, the average value $|H(0)|$ would be zero. From Eq. (35) we find that an error of only 10° results in $|H(0)| = 0.087$ instead of the desired value of 0. The thickness of the plate must therefore be controlled, to within a value established by the quantization limit, to produce accurate impulse response functions.

V. Experimental Results

We implemented impulse responses with two types of spatial light modulator: photographic film and a liquid crystal display. Film provides an ideal method to establish the ultimate performance of the system for comparison purposes. Although the liquid crystal display has not been used in this type of application before, several authors have explored the use of liquid crystal displays in other optical processing applications.[5-7]

Before the optical system was constructed, we modified an inexpensive liquid crystal TV display, following the procedure given in Ref. 6. We removed the hinge on the TV screen to allow the liquid crystal display to open to 90°. We removed the plastic diffuser plate and the internal polarizers and cleaned the remaining glue from the screen with acetone. We did not apply index matching windows to the liquid crystal display because we placed it outside the interferometer, where its phase errors are not important.

A. System Measurement Without the sgn Plate

We used a network analyzer to calibrate the acoustooptic system response without a spatial light modulator or the sgn plate. The network analyzer generates a single frequency reference signal that is swept over the frequency band to probe the response of the system. The output from the photodetector is compared to the reference signal to determine the amplitude and phase response of the system. To account for the time delay between the center of the illuminating beam and the edge of the transducer in the acoustooptic cell, we placed an equivalent delay in the reference signal from the network analyzer. The phase response is therefore more easily assessed because the steep linear phase shift across the frequency band, due to a long time delay, is eliminated.

For the system test, the aperture function $a(x)$ is a rectangular function whose width is 2 mm, corresponding to a time window of $T = 3.24\ \mu s$. The trace in Fig. 6(a) shows the main lobe of the system response and part of the first sidelobes. This result shows that the width of the main lobe is 600 kHz, consistent with the acoustooptic aperture of $T = 3.24\ \mu s$. The x-direction apodization for our experiments was approximately $A = 0.15$, which is nearly uniform illumination. Thus, we expect the system response to be a sinc function, with sidelobe peaks of -13 dB relative to the peak main lobe level, as confirmed by the experimental results given in Fig. 6(a).

The system phase response over a 1-MHz band is shown in Fig. 6(b). The network analyzer sweep time is ~40 ms. Since the exposure time of the network analyzer display was set to 100 ms, Fig. 6(b) reveals a second trace that shows the effects of vibrations, which caused the system response to change slightly from sweep to sweep. A linear phase component of ~225°/MHz is present due to a small mismatch in compensating the time delay between the center of the illuminating beam and the transducer of the acoustooptic cell. The trace shows a sharp transition at three divisions from the center of the display, which indicates the 180° phase shift at the sidelobe nulls located at 59.7 and 60.3 MHz. Some ripple is also noticeable in the trace in the linear region of the response. This feature appeared in all the phase traces and is probably due to feedthrough in the delay line used to cancel the long time delay caused by the acoustooptic cell. Figure 6(c) shows the magnitude response of the system over an expanded bandwidth of 10 MHz. The overall response is the expected sinc function; the system noise level is evident at the higher sidelobe positions where the signal spectrum is small.

B. System Measurement with the sgn Plate

We positioned the sgn plate in the reference beam and adjusted the vertical position of the plate until the network analyzer indicated the minimum value of the spectrum at the center frequency of 60 MHz, which is equivalent to the location of $H(0)$. Figure 7(a) shows the magnitude response with the sgn plate placed in the system, from which we deduce that $|H(0)| \approx 0.2$ so

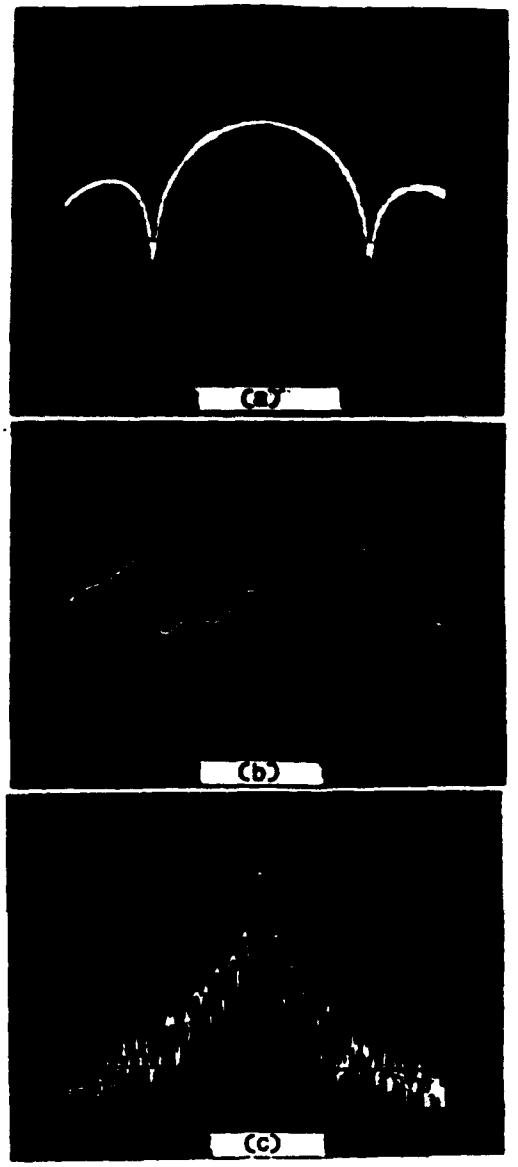


Fig. 6. System response without the sgn plate: (a) magnitude response over 1 MHz; vertical scale, 10 dB/division; horizontal scale, 100 kHz/division; (b) phase response over 1 MHz; vertical scale, 90°/division; horizontal scale, 10 dB/division; (c) magnitude response over 10 MHz; vertical scale, 10 dB/division; horizontal scale, 1 MHz/division.

that $\Delta e = 23°$, which is outside the manufacturers specification that the coating thickness was $180 \pm 10°$. A comparison of the system phase response with the sgn plate, shown in Fig. 7(b), with the phase response of Fig. 6(b) shows that the aberrations induced by the sgn plate are small.

C. Photographic Spatial Light Modulator Results

To test the system with an ideal spatial light modulator, we used photographic film to introduce some impulse responses whose frequency responses can be analytically predicted. These functions are a sinusoidal impulse response and a sinc function impulse response.

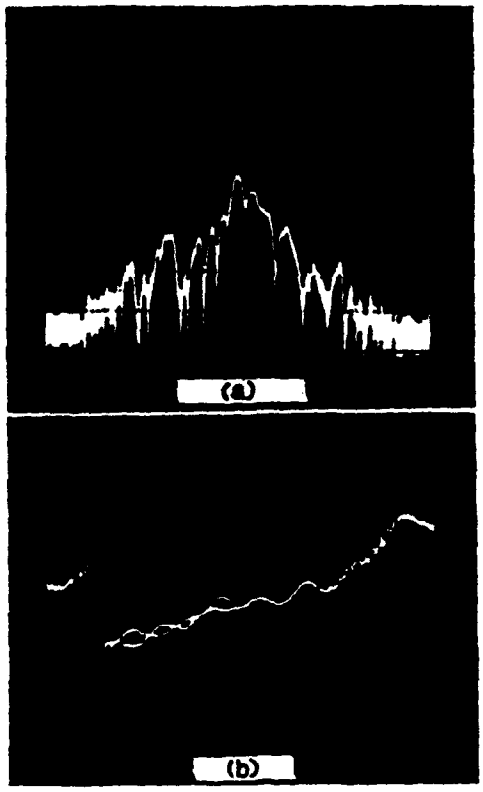


Fig. 7. System response with the sgn plate: (a) magnitude response of the whole plate over 10 MHz; vertical scale, 10 dB/division; horizontal scale, 1 MHz/division; (b) phase response of the whole plate over 1 MHz; vertical scale, 90°/division; horizontal scale, 100 kHz/division.

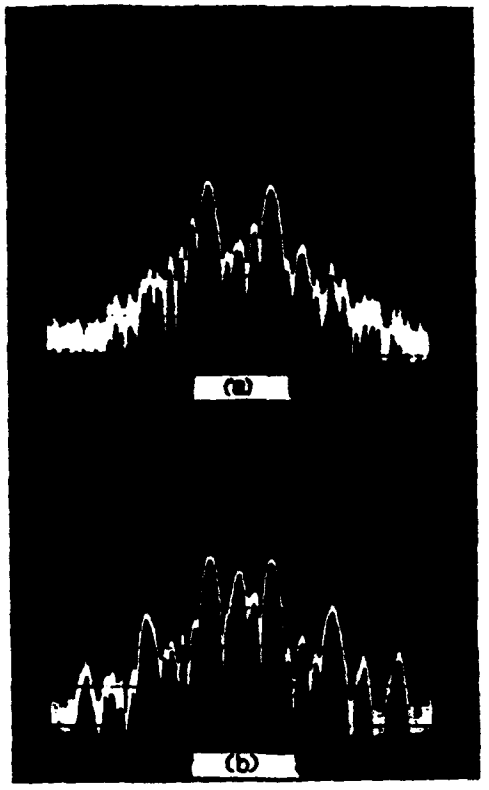


Fig. 8. Response of the sinusoidal filter of 0.36-cycles/mm spatial frequency: (a) magnitude response of the sinusoid with a peak-to-peak aperture of 8 mm over 10 MHz; vertical scale, 10 dB/division; horizontal scale, 1 MHz/division; (b) magnitude response of the sinusoid with a peak-to-peak aperture of 28 mm over 10 MHz; vertical scale, 10 dB/division; horizontal scale, 1 MHz/division.

1. *Sinusoidal Impulse Response*

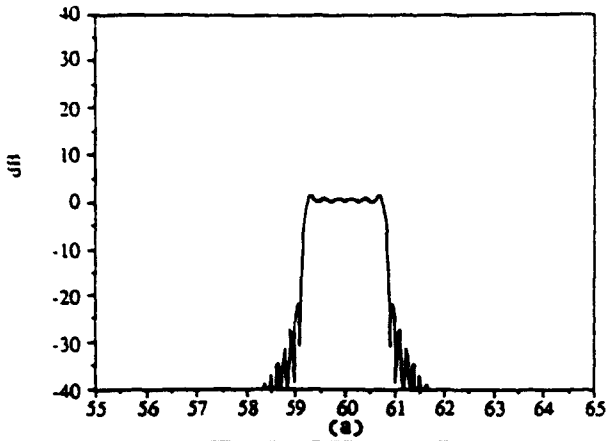
We implemented a sinusoidal impulse response of 0.36-cycles/mm spatial frequency, corresponding to a temporal frequency of 0.8 MHz. We used two versions of this signal. The first version has a peak-to-peak aperture height of 8 mm, for which the beam apodization at the maximum height is exp(−0.012). The resulting spectrum is shown in Fig. 8(a), and the expected frequency components at 59.2 and 60.8 MHz are evident. The center frequency, representing the average value of the sinusoid, is low as expected.

To illustrate the effects of nonlinearities introduced by aberrations and beam apodization, we implemented the same sinusoidal function, but increased the peak-to-peak amplitude to 28 mm, for which the beam apodization at the maximum height is exp(−0.15). The resulting spectrum, shown in Fig. 8(b), still shows the desired frequency content at 59.2 and 60.8 MHz, but the central component at 60 MHz is only 4 dB below the level without the impulse response mask. Furthermore, we see considerable frequency content at the harmonics of 0.80 MHz, particularly at the third and fifth harmonics of 60 ± 2.4 and 60 ± 4 MHz.

Harmonics are evidence of nonlinearities that may be due to one or more of the sources discussed in Sec. IV. For the larger amplitude impulse response, we observed wavefront distortion at the output plane of the order of $\lambda/2$, particularly at the edges of the lower section of the interference pattern. Using a computer program, we determined the predicted sizes of the third and fifth harmonics using the $\lambda/2$ observed wavefront distortion and the apodization of exp(−0.15) at the maximum height of $h(x)$. The predicted third and fifth harmonic sizes were −16 and −33 dB below the first harmonic, which agrees reasonably well with the observed values, given the roughness of the distortion measurement and the simplicity of the model. From the calculations, we find that most of the distortion is due to the phase errors.

2. *Sinc Impulse Responses*

The next impulse response was a sinc function representing an equivalent temporal impulse response sinc($t/0.57\ \mu s$). The frequency response is a rectangular function whose width is 1.76 MHz, as shown in Fig. 9(a); the sidelobes beyond the edges of the rectangular function are consistent with those caused by a truncation of the sinc function by the finite aperture of the acoustooptic cell. Figure 9(b) shows the experimental results produced by a sinc impulse response whose peak amplitude is 5 mm high. The results show a passband of 1.8-MHz width, which agrees well with the theoretical value of 1.76 MHz. The sidelobe nulls are



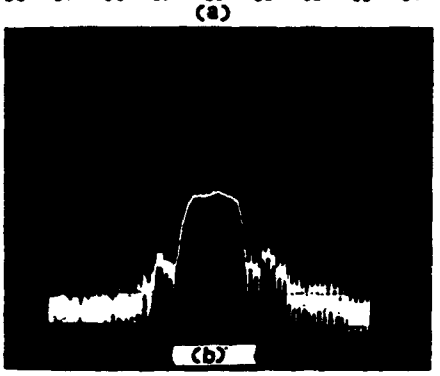


Fig. 9. Response to the sinc filter of 2.5-mm main lobe width: (a) expected magnitude output; (b) magnitude response of the filter 5 mm high over 10 MHz; vertical scaling, 10 dB/division; horizontal scaling, 1 MHz/division; (c) phase response of the filter 5 mm high over 5 MHz; vertical scaling, 90°/division; horizontal scaling, 500 kHz/division.

not as well defined in this result as those in Fig. 9(a) because of the system noise level, but the sidelobe envelope follows the expected form. The passband shows a ripple of ~2 dB caused by the truncation of the sinc function. The phase response, shown in Fig. 9(c) over a 5-MHz band, has a slowly varying ripple component at one division on either side of the central axis of the display.

D. Liquid Crystal Display Results

We used a computer to write a square-wave grating on the liquid crystal display, at a spatial frequency of 1.35 cycles/mm, to show the diffraction capability of the display. We expect the output to be a sampled sinc function with the first diffracted component at 3 MHz, as confirmed by the experimental result shown in Fig. 10. The diffracted orders 3 MHz from the central order are clearly shown.

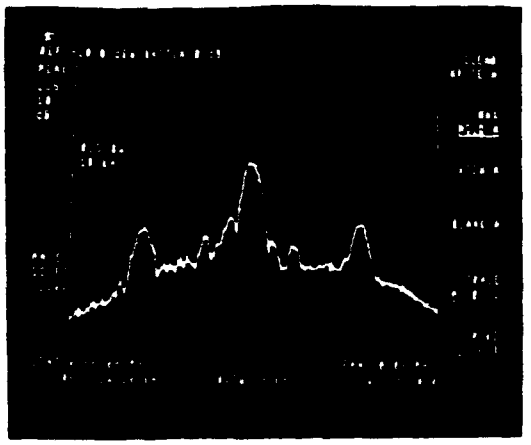


Fig. 10. Response with grating of 1.35-cycles/mm spatial frequency written to a LCTV over 10-MHz bandwidth; vertical scaling, 10 dB/division; horizontal scaling, 1 MHz/division.

If the ON transmittance is equal to one and the OFF transmittance is equal to zero, so that the contrast ratio of the display is equal to one, we expect the magnitude of the first diffracted order of the grating to be down by ~4 dB relative to the main lobe. In the case of the liquid crystal display, however, additional attenuation of the diffracted orders is encountered because the ON transmittance was measured as only 0.056 and the OFF transmittance was measured as 0.021; the contrast ratio of the display is therefore rather low. The results given in Fig. 10 show responses that are 16 and 18 dB below the central order, partly because of the low contrast levels and partly because of the higher than desired OFF transmittance.

We wrote sinc functions on the liquid crystal display to further shape the output spectrum. Unfortunately, these experiments were less successful. The resulting output traces show evidence of some modulation, but the spectra are difficult to measure accurately. The performance of the liquid crystal display was disappointing for two main reasons. The first problem was that the low contrast ratio of the unit, coupled with the available signal-to-noise ratio in our wide bandwidth experimental system, made the liquid crystal display results difficult to measure. An additional problem was the fact that the OFF regions of the liquid crystal light modulator allowed too much of the incident illumination to pass, producing errors in the output signal. If the area modulation mask is high contrast, the background illumination represents a much smaller deviation from the desired output.

VI. Conclusions

Acoustooptic processors are useful for applications where computationally intensive integral operations must be performed at rapid speeds. The extension of the acoustooptic processor to arbitrary filtering operations can expand the capability of these architectures.

Area modulation is a particularly useful technique for implementing arbitrary filter operations in acoustooptic systems. The feasibility of the technique was suggested by the analysis we performed and verified by the experimental results. The photographic spatial light modulator results were particularly encouraging. While the liquid crystal display results were not as successful, other spatial light modulators are available which could offer better performance in a more sophisticated system. A particularly important performance parameter is that the transmittance in the OFF state be low. These systems could operate as programmable filters at frequencies and bandwidths limited only by the detection circuitry and by acoustooptic cell technology.

We thank C. S. Anderson, D. O. Harris, and R. N. Ward for helpful discussions and advice. This work was supported by the U.S. Army Research Office.

References

1. E. B. Felstead, "Optical Fourier Transformation of Area-Modulated Spatial Functions," Appl. Opt. 10, 2468–2475 (1971).
2. E. B. Felstead, "Some Optical Correlations," Research Report 10-2, Electrical Engineering Department, Queens U., Kingston, Ontario (Apr. 1970); "A Simple Real-Time Incoherent Optical Correlator," IEEE Trans. Aerosp. Electron. Syst. AES-3, 907–991 (1967).
3. A. W. Lohmann and W. T. Rhodes, "Two-Pupil Synthesis of Optical Transfer Functions," Appl. Opt. 17, 1141–1151 (1978).
4. B. J. Pernick, "Area-Modulated Recordings for One-Dimensional Optical Signal Processors," Opt. Eng. 27, 452–455 (1988).
5. H. K. Liu and T. H. Chao, "Liquid Crystal Television Spatial Light Modulators," Appl. Opt. 28, 4772–4780 (1989).
6. A. M. Tai, "Low-Cost Spatial Light Modulator with High Optical Quality," Appl. Opt. 25, 1380–1382 (1986).
7. G. D. Boreman and E. R. Raudenbush, "Characterization of a Liquid Crystal Television Display as a Spatial Light Modulator for Optical Processing," Proc. Soc. Photo-Opt. Instrum. Eng. 639, 41–46 (1986).

APPENDIX B

HYBRID ACOUSTO-OPTIC AND DIGITAL EQUALIZATION FOR MICROWAVE DIGITAL RADIO CHANNELS

Reprinted from Optics Letters

Volume 21, Pages 1182-1184, 1 November 1991

Hybrid acousto-optic and digital equalization for microwave digital radio channels

C. S. Anderson and A. VanderLugt

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

Received June 22, 1990; accepted August 22, 1990

Digital radio transmission systems use complex modulation schemes that require powerful signal-processing techniques to correct channel distortions and to minimize bit error rates. We propose combining the computation power of acousto-optic processing and the accuracy of digital processing to produce a hybrid channel equalizer that exceeds the performance of digital equalization alone. Analysis shows that a hybrid equalizer for 256-level quadrature amplitude modulation (QAM) performs better than a digital equalizer for 64-level QAM.

Microwave digital radios currently use 64-level quadrature amplitude modulation (64-QAM) signals to achieve data rates of as much as 140 Mbits/sec. Improvements in the performance of the next generation of radios require excellent channel control to increase further the number of amplitude levels, as well as the data rates, while maintaining bit error rates. With digital equalizers the amount of improvement is limited by the adaptation algorithm and the number of equalizer taps.[1] To achieve a significant increase in data rates, other equalization techniques are needed that are not restricted by digital hardware speed. We propose combining the computation power of acousto-optic processing and the accuracy of available digital processing to produce a hybrid channel equalizer that satisfies these goals.

The concept of hybrid channel equalization is illustrated in Fig. 1, which is a block diagram of a digital radio. An optical equalizer operates at the intermediate frequency (IF) to correct channel distortion partially prior to a final digital equalization of residual errors. Since the digital postprocessor must correct only residual distortion, better overall equalization is achieved without increasing the number of taps.

Channel distortion is primarily caused by multiple transmission paths due either to atmospheric refraction or to ground reflections. The frequency response of the channel distortion can be written as

$$H(f) = \sum_{i=1}^{N} w_i \exp(-j2\pi f \tau_i), \tag{1}$$

where w_i and τ_i are the amplitude and delay of each path and N is the total number of paths. Rummler claimed that no more than three separate paths were present on a line-of-sight radio link[2,3]; experimental verification has shown this to be true for 99% of all measured distortions.[4] Rummler also showed that delay differences between two of the three paths could be adequately represented by a fixed delay of 6.3 nsec, even though the actual delays range from 2 to 15 nsec.[5]

With these simplifications, the modeled frequency response of the distorted channel becomes

$$H(f) = a\{1 - b \exp[-j2\pi(f - f_0)\tau]\}, \tag{2}$$

where a, b, and f_0 are all complex functions of the original values of w_i and τ_i. In this relation $\tau = 6.3$ nsec is the fixed delay difference chosen by Rummler. From Eq. (2), we find that the magnitude and phase characteristics are

$$|H(f)| = a\{1 + b^2 - 2b \cos[2\pi(f - f_0)\tau]\}^{1/2} \tag{3}$$

and

$$\theta(f) = \tan^{-1}\left\{\frac{b \sin[2\pi(f - f_0)\tau]}{1 - b \cos[2\pi(f - f_0)\tau]}\right\}, \tag{4}$$

where a is a gain factor, f_0 determines the notch location of $H(f)$, and b is the ratio of the power in the longer delay path relative to that in the direct path. Magnitude and phase plots shown in Fig. 2 illustrate the effect of varying b and f_0 for a 20-MHz channel. At the receiver, the optimum linear filter for correcting channel distortion in the presence of noise is[6]

$$H_{\text{opt}}(f) = \frac{A[H(f)]^{-1}}{1 + \dfrac{R_{nn}(f)}{R_{ss}(f)|H(f)|^2}}, \tag{5}$$

where $R_{nn}(f)$ is the spectral density of the noise, $R_{ss}(f)$ is the spectral density of the signal, and A is a complex-valued constant that accounts for both gain and

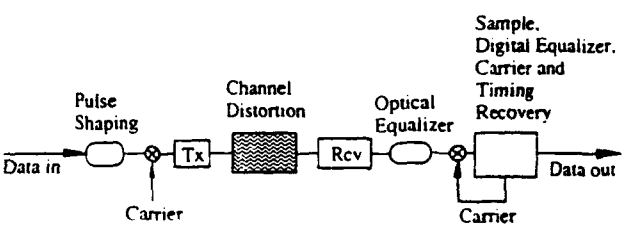


Fig. 1. Schematic of the digital radio, showing the transmitter (Tx) and the receiver (Rcv), with hybrid equalization.

0146-9592/90/211182-03$2.00/0

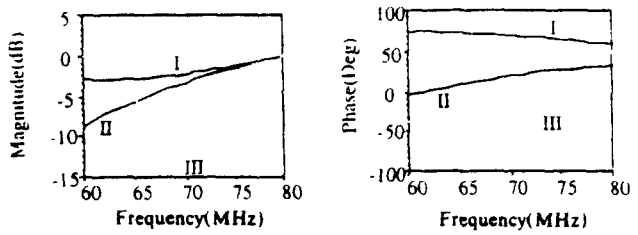


Fig. 2. Magnitude and phase plots for Rummler's channel model: case I, $b = 0.5$ and $f_0 = 62$ MHz; case II, $b = 0.9$ and $f_0 = 50$ MHz; case III, $b = 1.1$ and $f_0 = 70$ MHz.

time delay. This filter ensures that the noise is not too highly amplified when the value of $|H(f)|$ is small. On the other hand, if the spectral density of the distorted signal $R_{ss}(f)|H(f)|^2$ is much larger than $R_{nn}(f)$, the result in Eq. (5) is closely approximated by the inverse filter $[H(f)]^{-1}$. Since this condition is generally met in digital radios, even for notch depths of 20 dB, the acousto-optic equalizer is designed to implement $[H(f)]^{-1}$.

To adapt the equalizer to varying channel conditions, we first estimate the unknown channel parameters without interrupting the communication system. We use power spectral measurements of the received signal to solve Eq. (3) for the maximum-likelihood estimates of b and f_0. An ambiguity in the estimation of b arises, however, because the shapes of the power spectra for $b_1 = b$ and $b_2 = 1/b$ are identical, with the only difference occurring in the phase characteristics that are complex conjugates. For example, an estimation using the power spectrum for case I, as shown in Fig. 2, results in the two estimates of $b_1 = 0.5$ and $b_2 = 2$, and we must determine which of these two distortions is present. A particularly useful solution is to process the received signal with equalizers based on both b_1 and b_2 and choose the data that are best corrected. This dual processing approach is conveniently implemented in optical systems, since both equalizers are obtained from a single filter owing to the conjugate relationship between the diffracted orders of the filter.[7]

We want to decide which filtered signal is best without coherently demodulating both data streams. If the estimates of b_1, b_2, and f_0 are accurate, one filter output will be equalized, while the other will experience twice the channel phase distortion. To decide which is best, we square-law detect each IF signal and low-pass filter the result to produce a baseband signal described by

$$s(t) = i^2(t) + q^2(t), \tag{6}$$

where $i(t)$ and $q(t)$ are the baseband in-phase and quadrature data signals, respectively. If $i(t)$ and $q(t)$ are well-corrected QAM signals, $s(t)$ will contain discrete amplitude levels when measured at the peak sampling instants. For example, a 16-QAM signal has -3, -1, 1, and 3 as the possible amplitude levels of each channel, so that every output sample of Eq. (6) will have an amplitude of 2, 10, or 18. On the other hand, if the QAM signal is distorted, the output sample values will be more broadly distributed. To determine which equalizer output to select, we count a large number of samples around one of the expected amplitude levels. The equalizer with the largest count is deemed best, and its signal is converted to baseband and demodulated.

We simulated this hybrid approach by using a 16-QAM signal (20-dB carrier-to-noise ratio) that was corrupted with the distortion shown in Fig. 2, case II, and then filtered with both candidate equalizers. The output of each was processed according to Eq. (6) and sampled at the proper instants. Figure 3 shows a histogram of the samples after 1000 data points were collected. As expected, the histogram of the equalized signal is localized at levels 2, 10, and 18, but the histogram of the signal from the other filter shows that the amplitudes are more uniformly distributed. The correct equalizer was selected in 95% of all trials, and the incorrect selection of 5% occurs only when the difference between the two distortions is small.

The acousto-optic processor shown in Fig. 4 estimates the channel conditions and selects the best correction filter. Power spectral measurements are obtained by detecting the Fourier transform of the received signal in an auxiliary frequency plane P_2'. The output of the photodetector array in plane P_2' is digitized and processed to implement the estimation algorithms for b and f_0.

To realize complex filter functions with real-valued, nonnegative photographic masks, we encode the amplitude and phase of the equalizers onto a spatial carri-

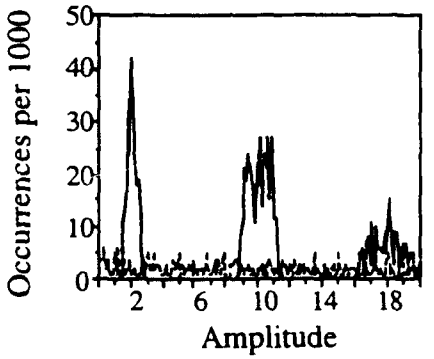


Fig. 3. Histogram of data samples for matched (solid curve) and mismatched (dotted curve) equalizers.

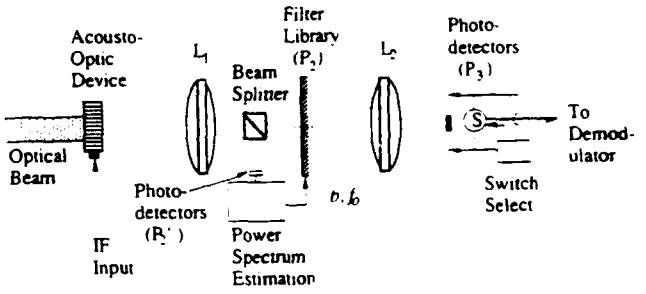


Fig. 4. Dual-frequency-domain equalizer with channel estimation.

Table 1. Comparison of Minimum Mean-Square Error for the Digital and Hybrid Equalizers

	Normalized Mean-Square Error	
QAM No.	Digital Equalization	Hybrid Equalization
64	0.0089	0.0011
256	0.0361	0.0045
1024	0.1445	0.0181

er frequency. The distorted signal is simultaneously processed with both the equalizer and its conjugate as seen by writing the amplitude transmittance of the mask as

$$t(x) = 1 + \frac{1}{|H(x)|}\cos[2\pi\alpha_c x + \theta(x)]$$
$$= 1 + \frac{1}{|H(x)|}\exp[j\theta(x)]\exp(j2\pi\alpha_c x) + \frac{1}{|H(x)|}\exp[-j\theta(x)]\exp(-j2\pi\alpha_c x), \quad (7)$$

where α_c is the spatial carrier frequency, $H(x)$ and $\theta(x)$ are given by Eqs. (3) and (4), and we have ignored scaling factors. The last two terms of Eq. (7) represent the equalization filters that have conjugate phase relationships. After being Fourier transformed by lens L_2, the two filtered signals are spatially separated because the spatial carrier frequency α_c produces different linear phase components for each term. By placing a photodetector at each output location, both equalizers operate in parallel. A separate digital processor compares the squared output of both equalizers to set the switch position that feeds the correct data to the demodulator.

The number of filters required to equalize the channel over the expected ranges of b and f_0 is estimated by specifying the amount of residual IF spectral distortion that is tolerable at the output of the acousto-optic equalizer. Since the system contains both optical and digital equalizers, the requirements on each can be slightly relaxed. Suppose that a 3-dB residual amplitude error can be readily corrected by a digital equalizer and that we store a finite number of filters in the library to cover the range of b values. After optical equalization, the residual amplitude error in decibels can be written as

$$e_{res}(f) = \left| 10 \log \left\{ \frac{1 + b^2 - 2b\cos[2\pi(f - f_0)\tau]}{1 + (b + \Delta b)^2 - 2(b + \Delta b)\cos[2\pi(f - f_0)\tau]} \right\} \right|, \quad (8)$$

where Δb is the difference between the value of b for a given filter and its true value for the channel. Given a value for b, the maximum allowable Δb is found by requiring that $e_{res}(f)$ is less than 3 dB over the filter bandwidth. We found that eight filters are sufficient to equalize the full range of channel distortions. No additional filters are needed to cover the range of values of f_0 since the center frequency of the IF data signal can be changed to compensate for the notch position.

The number of samples needed to represent accurately each filter function in frequency is determined by comparing the actual frequency response to a series of samples smoothed by an interpolation function. We found that the filter is accurately represented with 25 samples across the 20-MHz bandwidth. To equalize all possible notch positions, we must record the filter over a 40-MHz bandwidth so that the total number of samples is $N_s = 2 \times 25$. In addition, the complex filters require a spatial carrier frequency so that they can be recorded onto a material that supports only nonnegative values.[7] Suppose that the carrier frequency is six times the maximum frequency of the equalizer frequency response, so that the total number of samples needed to record a single filter is $N_s = 6 \times 2 \times 25 = 300$. Multiplying the number of samples by the number of filters to handle the range of b values leads one to the requirement of 2400 samples, which is low compared with that of most optical holograms.[8] Computer-generated holograms are practical for use in this application because the required filter functions are easily calculated using Eqs. (3) and (4), and the small number of samples can be produced with available computer printers.

Finally, we compare the minimum mean-square error of an 11-tap digital equalizer with and without optical preprocessing. The optical equalizer was selected from one of the eight library filters and included some additive amplitude and phase ripple. For both receivers we calculated the minimum mean-square error for 64-, 256-, and 1024-QAM signals and normalized the error relative to the power at the minimum modulation level. From the results listed in Table 1, we see that a hybrid equalizer for 256-QAM has a minimum mean-square error of 0.0045, which is approximately half the minimum mean-square error of 0.0089 for a digital equalizer, operating alone, for 64-QAM. This eightfold increase in performance can be used in one of several ways: to increase the symbol rate, to increase the number of QAM levels, or to decrease the bit error rate. Experimental research to validate the calculations presented here is in progress.

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

References

1. J. K. Chamberlain, F. M. Clayton, H. Sari, and P. Vandamme, IEEE Commun. Mag. 24, 43 (1986).
2. W. D. Rummler, in *Proceedings of the International Conference on Communication* (Alger, Toronto, 1978), Vol. 3, p. 47.5.1.
3. W. D. Rummler, Bell Syst. Tech. J. 58, 1037 (1979).
4. P. Balaban, AT&T Tech. J. 64, 2525 (1985).
5. J. Sandberg, IEEE Trans. Antennas Propag. AP-28, 743 (1980).
6. H. L. VanTrees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968), pp. 467–481.
7. A. VanderLugt, IRE Trans. Inf. Theory IT-10, 139 (1964).
8. W. H. Lee, Appl. Opt. 9, 639 (1970).

APPENDIX C

ACOUSTO-OPTIC CHANNEL EQUALIZATION FOR MICROWAVE DIGITAL RADIOS

Accepted for publication in Applied Optics

Acousto-Optic Channel Equalization for Microwave Digital Radios

by

Chris S. Anderson

North Carolina State University

Electrical and Computer Engineering

Please send all correspondence to :

Chris Anderson

Dynetics, Inc.

P.O. Drawer B

Huntsville, AL. 35816

(205) 922-9230

ABSTRACT

Digital radio transmission systems use complex modulation schemes that require powerful signal-processing techniques to correct channel distortions and to minimize bit error rates. In this paper, acousto-optic processors are used, in conjunction with adaptive digital equalizers, to reduce the bit error rate of the digital radio receiver. The acousto-optic processor implements an inverse channel filter that is rapidly adaptable to time varying distortions. A specific architecture is identified and a laboratory system is tested to verify the ability of the processor to track and correct time-varying channels. Computer simulations are used to show that the hybrid acousto-optic and digital equalizer allows a four-fold increase in the modulation capacity of radio, relative to all digital equalization, while improving the bit error rate performance.

Key Words:

Acousto-optics

Adaptive

Channel Equalization

Digital Radios

Inverse Filters

Spectrum Analysis

I. Introduction

Microwave digital radios are an important transmission medium for telecommunications, partly because of their compatibility with digital switching and routing technologies [1]. To support the increasing demand for high quality digital service, the data capacity of the radio links must be increased, which requires excellent control over channel characteristics. Digital equalizers are used in the radio receiver to combat channel distortion but they do not support the high level modulation formats anticipated for future systems.

Recently we proposed a receiver technique that uses a combination of analog acousto-optic processing in the intermediate frequency portion of the radio along with digital processing at baseband[2]. While combined analog and digital processing is commonly performed in digital radios, it has not been fully exploited as a channel equalization technique. Examples of hybrid receivers include the combination of a digital equalizer with a square-root of Nyquist filter[3], with an intermediate frequency slope equalizer[4], or with a resonant circuit that is the inverse channel filter for minimum phase distortion[5]. Each of these approaches performs well for some channel types but cannot correct for all expected distortions. Using the adaptive, acousto-optic processor, we are able to track and correct for all channel types after which a digital equalizer further reduces the effect of channel distortion. Using the synergetic effect of the analog and digital filter, the receiver bit error rate is lowered.

In this paper, we describe the acousto-optic processor and quantify its benefit to the digital radio. We begin in Section II by introducing the basic features of microwave digital radios and identifying a hybrid acousto-optic and digital processor that optimizes their performance. In Section III, we analyze an acousto-optic architecture that implements the analog portion of the hybrid equalizer and in Section IV, we develop a channel estimation technique that adapts the optical processor to time varying distortions. Experimental testing of the acousto-optic system is described in Section V and the results of a computer

simulation are used, in Section VI, to measure the performance of a hybrid equalized digital radio.

II. Microwave Digital Radios

The major components of a microwave digital radio link are a transmitter, an atmospheric transmission channel and a receiver as illustrated in Figure 1. Input into the transmitter are multiple channels of digital telecommunication data, which may contain information from voice, video or data signals. The incoming digital signal is converted to a train of narrow analog pulses, spaced T seconds apart. This analog signal has a high frequency content which must be reduced by pulse shaping filters to match the allotted bandwidth. The output of the pulse shaping filter is modulated in quadrature and radiated from an antenna.

At the receiver, the signal is mixed to an intermediate frequency of either 70 or 140 MHz, depending on the particular radio. The signal is gain controlled, to compensate for attenuation, and may be partially corrected for distortion by an analog channel equalizer such as the acousto-optic processor. After analog filtering, the signal is coherently demodulated to baseband and processed with a matched filter that maximizes the signal-to-noise ratio. The incoming waveform is sampled and digitized at the symbol rate T, equalized with an adaptive digital filter, and decision thresholded to recover the message signal.

A. Multipath Channel Distortion

The received signal can be affected by the atmospheric channel in three ways: it can be time delayed, attenuated, and distorted. Time delay is expected because the signal propagates over an average distance of 25 miles. Signal attenuation occurs because the entire radio wave is not collected by the receiving antenna and because atmospheric

scattering reduces the signal strength. The most important atmospheric effect is multi-path channel distortion, which corrupts the shape of the transmitted pulse and causes bit errors.

Multipath distortion occurs when multiple portions of the radio wave are refracted, by the atmosphere, into the collecting antenna[6]. The frequency response for N_p separate multipaths can be written as[7]

$$H(f) = \sum_{i=1}^{N_p} w_i e^{-2\pi f \tau_i}, \qquad (1)$$

where w_i and τ_i are the attenuation and time delay of each path respectively. Testing of line-of-sight radio links has shown that the number of paths is less than or equal to three for more than 99% of all measured distortions[7]. By reducing N_p to three and assuming that the time delays between two paths are nearly equal, Rummler[8] showed that the channel distortion is given by

$$H(f) = a(1 - be^{-j2\pi(f-f_0)\tau}), \qquad (2)$$

where a, b, and f_0 are all complex functions of the original w_i and t_i and $t = 6.3$ ns is the fixed delay difference chosen by Rummler. From Eq. (2), we find that the magnitude and phase characteristics are

$$|H(f)| = a\left(1 + b^2 - 2b\cos(2\pi(f - f_0)\tau)\right)^{\frac{1}{2}}, \qquad (3)$$

and

$$\theta(f) = \tan^{-1}\left[\frac{b\sin(2\pi(f - f_0)\tau)}{1 - b\cos(2\pi(f - f_0)\tau)}\right]. \quad (4)$$

Magnitude and phase plots are shown in Fig. 2 and illustrate the effect of varying b, which controls the notch depth, and f_0, which controls the notch location. Since τ is on the order of 6 ns[8], the multiple notches predicted by Eq. (3) are spaced by 166.67 MHz so that at most one notch corrupts the 20 MHz bandwidth signal. As b approaches one, the notch becomes infinitely deep at $f=f_0$.

An interesting complex conjugate relationship exists between channel distorstions having b values that are reciprocally related; i.e.,$b_1=1/b_2$. Physically, a b value larger than one corresponds to the longer delay multipath having a larger amplitude at the receiver than the shorter delay path. This type of distortion, known as non-minimum phase distortion, is known to occur in approximately 50% of all observed channels[7]. Since the minimum and non-minimum distortions are effectively a complex conjugate pair, their magnitude distortions are identical but their phase distortions have opposite sign. It is, therefore, counterproductive to equalize a channel distortion with the wrong equalizer type since a doubling of the phase distortion will occur. Most analog equalizers produced to date have either have been designed for minimum distortions or they do not produce phase correction at all. As we show in Section III, it is straightforward to produce an equalizer for both distortion types using the acousto-optic approach.

B. Hybrid Acousto-Optic and Digital Receivers

The acousto-optic processor uses a bank of eight filters that are designed to reduce the channel distortion prior to digital equalization; five filters are for minimum phase

distortion and three are for non-minimum distortion. For eight different b values, we implement a near optimum linear filter described by[9]

$$RCV(f) = \frac{H^*(f)}{|H(f)|^2} TR^*(f)$$
$$= \frac{F^*(f)}{|H(f)|^2}, \qquad (5)$$

where TR(f) is the frequency response of the transmit filter. Equation (5) represents the inverse channel filter multiplied by the transmit filter. Although inverse filters enhance the noise when used for deep notches, the subsequent digital equalizer compensates to minimize the expected error between the transmitted and received waveform[10].

Figure 3 shows the expected error for an 11-tap digital equalizer with and without the acousto-optic processor; 11-taps are used in current digital radios. The error is plotted for both receivers as a function of channel notch depth and for a f_0 value of 4 MHz. When the acousto-optic processor is not used, the short tap length digital equalizer cannot adequately correct the distortion and produces large amounts of error. The reason for this poor performance is that the 11-tap equalizer does not have the degrees of freedom (taps) to correct the deep notch in the magnitude response. As designed, the acousto-optic equalizer uses the equivalent of 70 taps (time-bandwidth product) and can correct the severe notch. When used after the acousto-optic equalizer, the digital processor must correct for only small amounts of residual distortion and noise.

Also shown in Fig. 3 is the mean square error produced by an optimum linear receiver[9]. Although the acousto-optic processor uses only eight filters, the hybrid equalized receiver performs very well relative to this bound. As we show in Section IV,

the error reduction relative to the digital equalizer significantly lowers the bit error rate of the hybrid equalized radio. As the transmitted waveform becomes more complex in order to increase data rates, this improved equalization will become critical to maintain acceptable bit error rates.

III. Channel Filtering

We now identify an acousto-optic processor that implements the required filters described by Eq. (5). To select the best architecture, we consider issues such as filtering in the time or frequency domain, heterodyne versus homodyne detection, and the practical effect of optical aberrations. Once the architecture is selected, we analyze the output signal as a function of the input and an area modulated mask containing the inverse filters; practical implementation of these masks is discussed in this section.

A. Architecture Selection

Several acousto-optic architectures are available for filtering electrical signals with a fixed photographic mask[11-13]. These include time domain architectures where the mask is located in the image plane of the acousto-optic cell, and Fourier domain architectures where the mask is located at the Fourier plane. Another important architecture choice is the generation of the reference beam which is necessary for recovering the filtered signal on the original intermediate carrier frequency f_c [15].

Figure 4 shows both the time and frequency domain architectures. In the time domain system, lenses L_1 and L_2 image the acousto-optic signal onto the impulse response mask. Lens L_3 integrates and places the mask output onto a high speed photodetector, which then produces the desired filtered signal. A more commonly used architecture is the Fourier domain system shown in Fig. 4b. This architecture filters the signal by altering its spectral characteristics with the mask, after which the optical signal is Fourier transformed and detected. Theoretically, the outputs of these two systems can be made identical with appropriately selected mask functions.

In practice, two primary differences exist between the architectures which makes the time domain system more attractive for our application. First, the time domain processor uses a spatial carrier frequency, recorded onto the mask, to make the diffracted and undiffracted acousto-optic signals collinear; this is necessary for homodyne detection[15]. In the Fourier domain processor, however, the diffracted and undiffracted beams are spatially separated at the mask and cannot be made collinear by the addition of a carrier frequency. As Fig. 4b shows, an additional grating is required for collinearity thereby increasing the number of components. The second advantage of the time domain system is that mask phase errors, caused by spatial nonuniformities in the mask substrate, do not effect the processor output. This beneficial result arises because both signal and undiffracted beams pass through the same optical non-uniformities and after square-law detection, the phase error cancels from the homodyne signal. In the Fourier system, the diffracted and undiffracted beams are separated at the mask and suffer different phase distortion. When square-law detected, the two unequal distortions do not cancel and directly effect the output.

The final architecture choice is between homodyne and heterodyne detection[15]. As Fig. 4 shows, the homodyne system uses the undiffracted acousto-optic beam as the photodetector reference. Since normal acousto-optic diffraction angles are small, the diffracted and the undiffracted beam travel nearly the same path, allowing both beams to be processed with the same optical components. Because this almost common path architecture minimizes the sensitivity to thermal and vibrational instabilities, the homodyne detection scheme was selected.

To address one of the eight impulse responses recorded onto the mask, we use a galvonometer mirror in the back focal plane of lens L_1 to fold the optical path and to change the beam angle in the vertical direction; this mirror is also located in the front focal plane of L_2. Once the channel conditions are determined and the best filter selected, the angle of the

mirror is set to address the appropriate mask element. This mirror angle is held constant until the channel conditions require a different mask element, after which the mirror angle is switched; a more complete architecture schematic is shown in Section V.

The eight inverse filters are chosen to implement Eq. (5) for different b values but for a fixed notch location f_0. To account for the range of expected notch locations, the incoming electrical signal is frequency shifted with a mixer to align the center of the notch and the peak of the filter response. After the acousto-optic filter, the electrical signal is mixed back to the intermediate frequency of the radio with the same local oscillator used in the first mixer.

B. Input/Output Relationship

The selected acousto-optic architecture is a space integrating correlator and has been extensively analyzed in the literature[15]. In this section, we modify existing analyses to account for binary, area modulated masks. The optical signal incident on the mask can be written as[15]

$$f_+(x,y) = \left[1 + jm_0 g(t - T_a/2 + x/v) e^{j2\pi f_c [t - T_a/2 + x/v]}\right]. \tag{6}$$

where m_0 is the modulation index, $g(t)$ is the baseband input signal, T_a is the acoustic time aperture, and v is the acoustic velocity. To produce a homodyne photodetector output, the diffracted and undiffracted beams are made collinear and overlapping by a grating in the mask. For area modulation, we write the binary mask transmittance as

$$m(q,p) = h(q,p) grat(q,p), \tag{7}$$

where q and p are the spatial coordinates, $h(\rho,q)$ encodes the filter impulse response, and

$$grat(q,p) = \mathrm{Rect}\left(\frac{p-H/2}{H}\right)\sum_i \mathrm{Rect}\left(\frac{2f_c}{v}\left[q - i\frac{v}{f_c}\right]\right) + \mathrm{Rect}\left(\frac{p+H/2}{H}\right)\sum_i \mathrm{Rect}\left(\frac{2f_c}{v}\left[q - i\frac{v}{f_c} + \frac{1}{2}\right]\right). \tag{8}$$

A few cycles of this grating are shown in Fig. 5. A shifted grating is used in Eq. (8) so that light incident on the bottom half of the mask is diffracted with a π phase shift relative to the light diffracted from the top half; this technique is similar to the *sgn* plate used by Karnowski[16]. By encoding the height of the grating in the upper and lower portions of the mask, we can produce bipolar filters.

Suppose that the mask filter function is split into two sections given by

(9)

where h_+ is for regions over the top half of the grating and h_- is for the bottom half. After multiplying the mask in Eq. (7) by the optical signal in Eq. (6) and collecting all collinear terms, we find that the detector output current is proportional to

(10)

where some unimportant factors have been neglected and the equivalent filter function is defined as

$$h_{eq}^{+}\left(\frac{\xi}{\nu}\right) = \int_{-\infty}^{\infty} |h_{+}(\xi,\gamma)| d\gamma; \tag{11}$$

the same definition is used for h_{eq}^{-} by replacing h_{+} in Eq. (11) with h_{-}. Even though the individual filter functions in Eq. (11) are positive, Eq. (10) shows that the total impulse response can be bipolar. If all functions are real-valued, Eq. (10) reduces to

$$i_{\text{det}} = \cos\left(2\pi f_c[t - T/2] + \frac{\pi}{2}\right) \int_{-\infty}^{\infty} \left[h_{eq}^{+}(\chi) - h_{eq}^{-}(\chi)\right] g(t - T/2 + \chi) d\chi, \tag{12}$$

which is the envelope of the input signal filtered with the impulse response given in Eq. (5).

C. Mask Production

The area modulated impulse responses were produced using photo-lithography masks for VLSI circuits. Several companies manufacture these components based on a computer mask description. Our mask functions were converted into forms accessible by AUTOCAD, which is a high precision graphics program. The completed AUTOCAD graphic was used to produce a chromium mask, where the dark grating bars are made opaque with a chrome coating and the clear areas are left uncoated. The measured attenuation of the opaque bars is greater than 50 dB relative to the clear apertures. Other

masks, such as those using film absorption, are available but the attenuation is not as large allowing some light leakage that alters the impulse response of the filter.

IV. Channel Estimation

To select the best filter from the eight element library, we must rapidly identify the channel distortion parameters a, b, and f_0 and then set the mirror angle in the optical processor to address the appropriate filter. If the value of these parameters are known, the complete magnitude and phase response of the channel is characterized and the best mask element can be determined. Fortunately, these channel parameters can be determined from the power spectral density of the communication signal which, in turn, can be measured by time averaging the output of an acousto-optic power spectrum analyzer. In this section, we show that this operation produces an unbiased an consistent estimate of the power spectral density and then determine the three channel parameters from this estimate. In addition, we introduce a technique for differentiating between minimum and non-minimum phase distortion.

A. Acousto-Optic Power Spectral Density Estimation

The acousto-optic power spectrum analyzer shown in Fig. 6 is a well developed system for measuring the frequency content of a signal[17]. Unfortunately, the *power spectrum* of a random signal is a poor estimate of the statistical *power spectral density* and leads to erroneous estimates of the channel parameters[18]. By temporally integrating the output of the spectrum analyzer, however, an estimate of the power spectral density is obtained, from which a, b, and f_0 are determined.

To show this result, we write the optical signal at the photodetector array as[15]

$$G(\alpha,t) = \int_{-L/2}^{L/2} g(t - T_a/2 + x/v)e^{j2\pi\alpha x}dx, \tag{13}$$

where L is the acousto-optic cell length and a is proportional to the spatial position along the detector array. The CCD detector array square-law detects the instantaneous power spectrum and, at any instant in time, the incremental charge added to the detector is proportional to

$$|G(\alpha,t)|^2 = \int_{-L/2}^{L/2}\int_{-L/2}^{L/2} g(t-T_a/2+x/v)g^*(t-T/2+y/v)e^{j2\pi\alpha x}e^{-j2\pi\alpha y}dxdy, \tag{14}$$

where some unimportant scaling factors are neglected. Since the input signal is random, this power spectrum is random and its instantaneous value does not adequately estimate the power spectral density.

The expected value of the instantaneous power spectrum or, equivalently, its ensemble average, is given by

$$E\left\{|G(\alpha,t)|^2\right\} = \int_{-L/2}^{L/2}\int_{-L/2}^{L/2} E\left\{g(t-T_a/2+x/v)g^*(t-T_a/2+y/v)\right\}e^{j2\pi\alpha x}e^{-j2\pi\alpha y}dxdy, \tag{15}$$

where $E\{\}$ is the expectation operator. If $g(t)$ is stationary, which is a valid assumption in digital radios, the expectation in Eq. (15) is the statistical autocorrelation function R_{gg} and allows Eq. (15) to be rewritten as[19]

$$E\left\{|G(\alpha)|^2\right\} = \int_{-L/2}^{L/2}\int_{-L/2}^{L/2} R_{gg}\left(\frac{y-x}{v}\right)e^{j2\pi\alpha x}e^{-j2\pi\alpha y}dxdy. \tag{16}$$

The time dependance is dropped in this equation because of the stationary assumption for the signal statistics. To further simplify Eq. (16), we make the variable transformation $\xi = (y - x)/v$ on the integral over x and reverse the order of integration. By integrating with respect to y first, Eq. (16) can be evaluated as two separate integrals, producing[20]

$$E\left\{|G(\alpha,t)|^2\right\} = v\int_{-T_a}^{0} R_{gg}(\xi)e^{j2\pi\alpha v\xi}\int_{-L/2}^{v\xi+L/2} dyd\xi + v\int_{0}^{T_a} R_{gg}(\xi)e^{j2\pi\alpha v\xi}\int_{v\xi-L/2}^{L/2} dyd\xi. \quad (17)$$

After performing the integrals over y and rearranging terms, we obtain

$$E\left\{|G(\alpha,t)|^2\right\} = \frac{1}{T}\int_{-T}^{T} R_{gg}(\xi)\left[1-\frac{|\xi|}{T}\right]e^{j2\pi\alpha v\xi}, \quad (18)$$

which shows that the expected value of the instantaneous *power spectrum* is the Fourier transform of the autocorrelation function multiplied by a triangular window. Since, by definition, the power spectral density is the Fourier transform of the autocorrelation function, Eq. (18) is equal to the convolution of the true *spectral density* with $sinc^2(T\alpha)$. By choosing the appropriate time aperture of the acousto-optic cell, the convolution with $sinc^2(T\alpha)$ has little effect on the shape of the power spectral density; we use a 1.25 μs time aperture.

Although a single power spectrum is an unbiased estimate, it may have little resemblance to the actual spectral density. Stated differently, the error between the instantaneous power spectrum and the power spectral density is, on average, zero but the

error variance may be quite large. To reduce the error variance, we introduce the time integrated power spectrum of

$$\left\langle |G(\alpha)|^2 \right\rangle_{T_i} = \frac{1}{T_i} \int_0^{T_i} |G(\alpha,t)|^2 \, dt, \tag{19}$$

where T_i is the detector integration time. In Reference 8, we show that this unbiased estimate of the power spectral density is also consistent.

B. Parameter Estimation

With the power spectral samples digitized from the acousto-optic system, the channel parameters a, b and f_0 are estimated and used to select the best inverse channel filter. First, the spectral samples are normalized by the known response of the pulse shaping filters so that the resulting data is samples of $|H(f)|^2$. Next, an error function, given by

$$\begin{aligned} \varepsilon &= \frac{10}{N} \sum_{i=1}^{N} \left[\log(M_i) - \log\left(|H[f_i]| \right) \right]^2 \\ &= \frac{10}{N} \sum_{i=1}^{N} \left[\log(M_i) - \log\left(a^2 \left[1 + b^2 - 2b\cos\left(2\pi [f - f_0]\tau \right) \right] \right) \right]^2 . \end{aligned} \tag{20}$$

is defined and minimized with respect to a, b and f_0. In Eq. (20), M_i and f_i are the spectral measurements and their frequency locations respectively. The parameter values that produce the smallest error in Eq. (20) are the desired channel estimates and are found by using a steepest descent algorithm[9]. After the channel parameter are found, a look-up table

is used to select the best filter from the eight element library and to set the mixer frequency to align the channel notch. Since digital radio channel conditions change on the order of 10-100 milliseconds, these operations need only be performed a few times a second.

C. Minimum/Non-Minimum Determination

Using only power spectral measurements, we cannot distinguish between minimum ($b<1$) and non-minimum ($b>1$) phase distortion, since these two channels differ only in their frequency domain phase. A blind digital equalization technique to identify the appropriate phase condition when the channel passes through a severe notch that causes the adaptive digital equalizer to diverge from the optimum solution. As an example, suppose that radio operation begins when little or no distortion is present. Initially, all distortions are assumed to have minimum phase. The acousto-optic processor applies only minimum phase filters and the digital equalizer acquires lock. During normal operation, the hybrid equalizer tracks the minimum phase channel until a catastrophic distortion event causes the digital processor to lose lock.

Once the digital tap weights diverge from the optimum solution, the filter must be retrained by again using blind equalization. Unfortunately, we are unsure of the distortion phase type so the acousto-optic processor must switch between two filters which are the complex conjugates of one another; i.e., one is for minimum phase fades and the other is for non-minimum phase. One thousand symbols are processed for the minimum phase filter and if the digital equalizer re-acquires lock, the acousto-optic processor uses only those filter types. If the blind equalization algorithm does not converge, however, the non-minimum phase filter is applied and another 1000 symbols are processed. Filter switching continues until channel conditions improve to the point that the hybrid equalizer converges to the minimum mean square error solution.

Even with this switching algorithm, the hybrid equalizer recovers more quickly from catastrophic distortion than the all digital equalizer. The digital equalizer, without acousto-optic filtering, must wait until conditions improve substantially before re-acquiring lock. The acousto-optic processor in the hybrid equalizer, however, reduces the channel distortion processed by the digital equalizer and allows much faster convergence.

V. Experimental Results

In this section, we experimentally verify that the acousto-optic processor estimates and corrects for channel distortion described by Eq. (2). Our experiments are subdivided into three areas and include 1) frequency response measurements of the inverse filters, 2) channel estimation measurements, and 3) channel tracking and equalization tests. When the results of these experiments are combined, they prove that the acousto-optic processor provides the necessary adaptive analog filtering needed for the hybrid equalizer.

Figure 7 shows the laboratory acousto-optic processor that combines both the filtering and channel estimation branches discussed in Sections III and IV. Our first tests measured the frequency response recorded onto the mask and verified that inverse channel filters could be produced. Frequency responses were measured with a HP 8654A network analyzer, digitized, and then transferred to a Macintosh computer. Figure 8 shows the measured and theoretical frequency responses for three of the eight inverse filters; the other five responses perform similarly. The error between the measured and the theoretical magnitude response, which is caused by slight imperfections in the mask and optical system, is small. Measured phase responses are in excellent agreement with theory, even as the phase transition becomes very sharp. Notice that for the b=b=1.047(1/0.955) equalizer, the phase slope is opposite from the minimum phase filters thus verifying the non-minimum phase equalizer. This frequency response data is also used in Section VI to simulate the acousto-optic processor in a hybrid equalized digital radio.

The second test verified that the channel estimator could select the appropriate filter from the library. A multipath channel distortion circuit was developed that emulates the frequency response given in Eq. (2). Figure 9 shows a plot of the estimated notch depth versus the actual notch depth for two notch locations. The horizontal axis is subdivided into filter regions within which the listed filter is best. For example, the third filter, which is a exact inverse filter for a b value of 0.92, is the best filter when the notch depth is between 18.7 dB (b=.88) and 25 dB (b=.944). The vertical axis is also subdivided into regions within which a notch depth estimate causes the listed filter to be selected. Shaded regions indicate areas where the best filter is selected while the unshaded regions are where a sub-optimum filter is chosen.

From Fig. 9, we see that estimator chose the best filter in 37 out of the 39 test cases, or approximately 95% of the time; we believe that this number of trials provides a good estimate of performance if the test cases are selected uniformly along the notch depth axis. The two incorrectly chosen filters occurred at transition regions between filters, which is not surprising since the margin of error is small at these points. Incorrectly choosing the filter at the transition region increases the distortion out of the acousto-optic filter by around 1 dB, which is small relative to the unfiltered channel.

The last experiment verified that the adaptive acousto-optic processor tracks time varying channels. Test signals were input into the acousto-optic processor after being distorted by the channel emulator. The output of the spectrum analyzer was digitized and the channel estimation algorithms were performed using a Macintosh computer. To begin the experiment, the channel emulator was set to provide no distortion, after which the notch depth was gradually increased. The acousto-optic processor estimates channel conditions and automatically applies the best filter. To quantify the performance, the root mean square distortion, remaining after acousto-optic filtering, was calculated and found to be below

1.5 dB. This result shows that the acousto-optic processor accurately corrects the channel leaving only small amounts of residual distortion.

VI. Simulation Results

We now test the performance of the hybrid and all digital equalizers in a digital radio by measuring their bit error rates via a simulation. To accurately model the hybrid equalizer in the simulation, the frequency response data for the acousto-optic processor is used as the analog filtering response. The digital radio can use either 64, 256, or 1024-level quadrature amplitude modulation; as the number of modulation levels doubles, so too does the amount of data transmitted per pulse. Figure 10 shows the bit error rate of the all digital radio and of the hybrid equalized radio as a function of the channel notch depth. As expected from the mean square errors presented earlier, the hybrid bit error rate is considerably lower than for the receiver using a pulse shaping filter. For example, a 31 dB notch, centered at f_0=0, produces a bit error rate 27.5 dB higher in the conventional 256-level system as compared to its hybrid counterpart. Other simulation results showed that the hybrid acousto-optic and digital equalizer allows a four-fold increase in the modulation capacity of radio, relative to all digital equalization, while improving the bit error rate performance.

VII Conclusions

We have identified a hybrid acousto-optic and digital equalizer and quantified its benefits as compared to conventional equalization. An adaptive acousto-optic processor was developed that tracks time varying channels and applies an inverse channel filter to remove the distortion. The processor corrects notches while producing less than 1.5 dB or root-mean-square residual distortion. Both conventional and hybrid equalized radios were simulated and the latter was found to support a four-fold increase in data capacity while

maintaining equivalent bit error rate performance. This large improvement can be used to increase the number of telecommunication channels or to reduce bit error rates.

VIII Acknowledgements

We thank A. VanderLugt, T. Karnowski, and R. Ward for their helpful suggestions and insights during this research effort. This work was funded by the Army Research Office and by Bell South Enterprises.

FIGURES

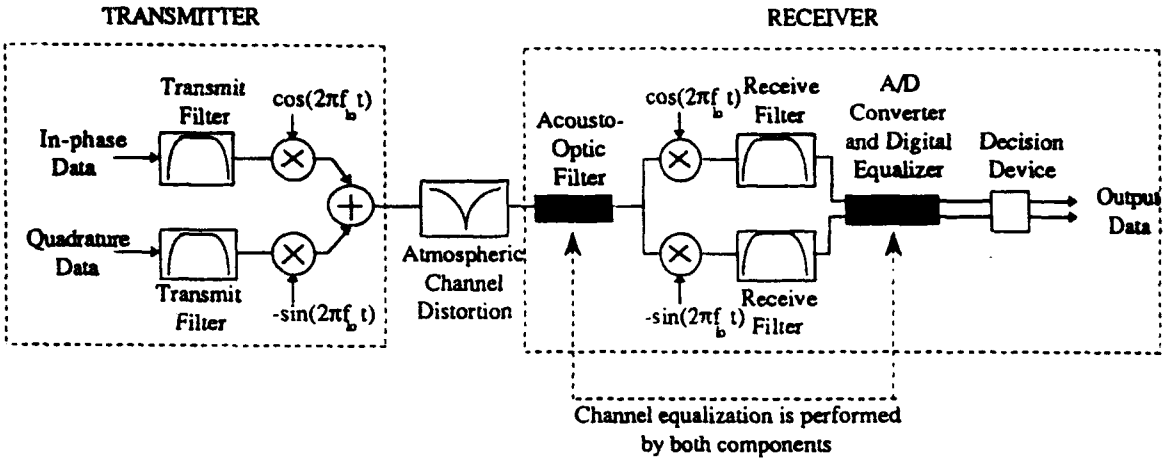


Figure 1. Schematic of a microwave digital radio with acousto-optic equalizer.

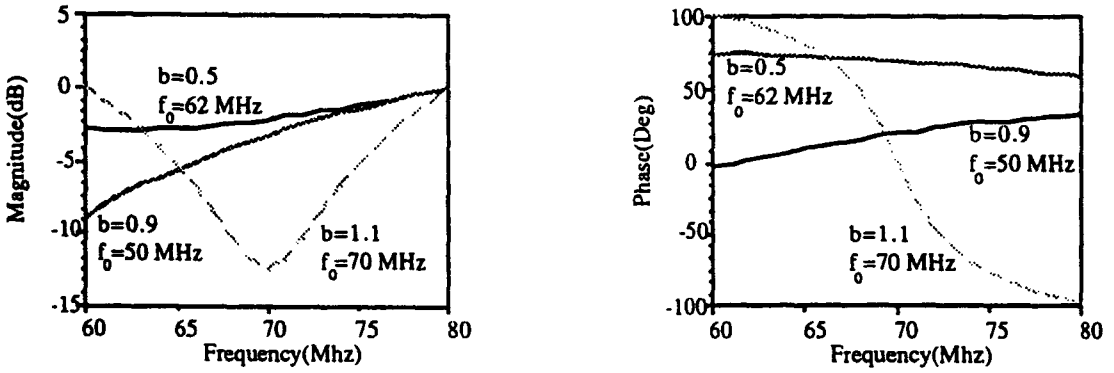


Figure 2. Magnitude and phase plots for Rummler's channel model.

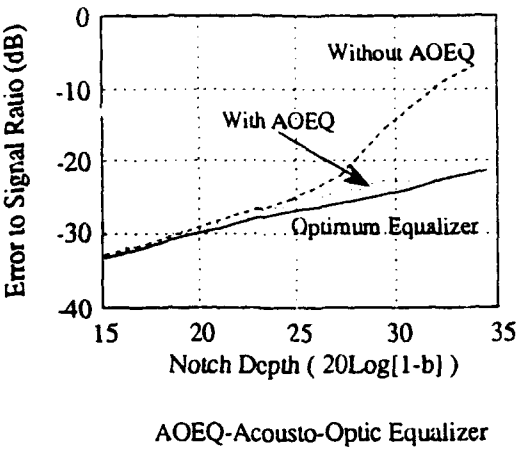
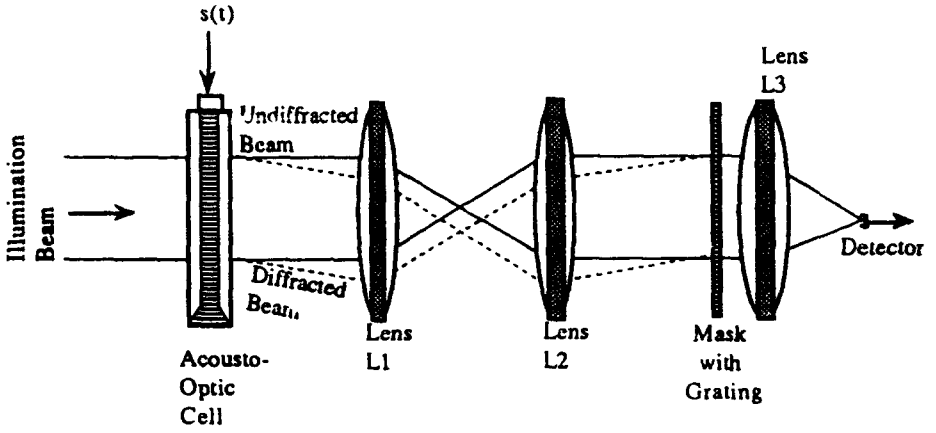
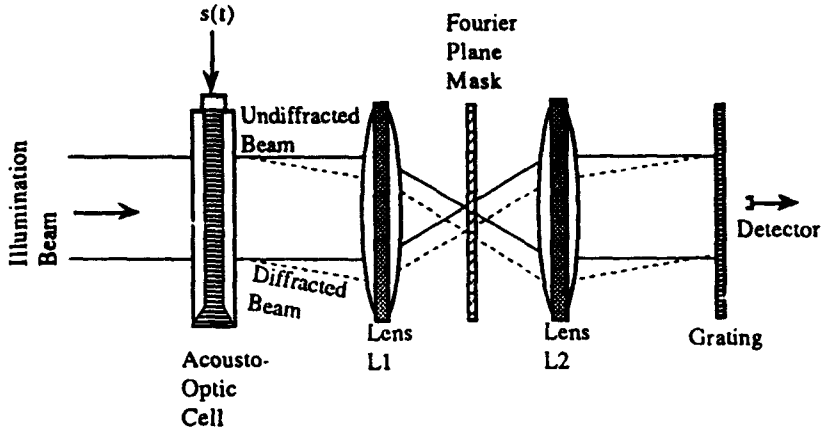


Figure 3. Mean square error output from an 11 tap equalizer. The notch location is 4 MHz. The lower curve is the smallest error than can be obtained from any linear processor.



(a)



(b)

Figure 4. Top view two acousto-optic architectures for filtering electrical input signal. Both architectures use homodyne detection.

a) Time domain filtering

b) Fourier domain filtering

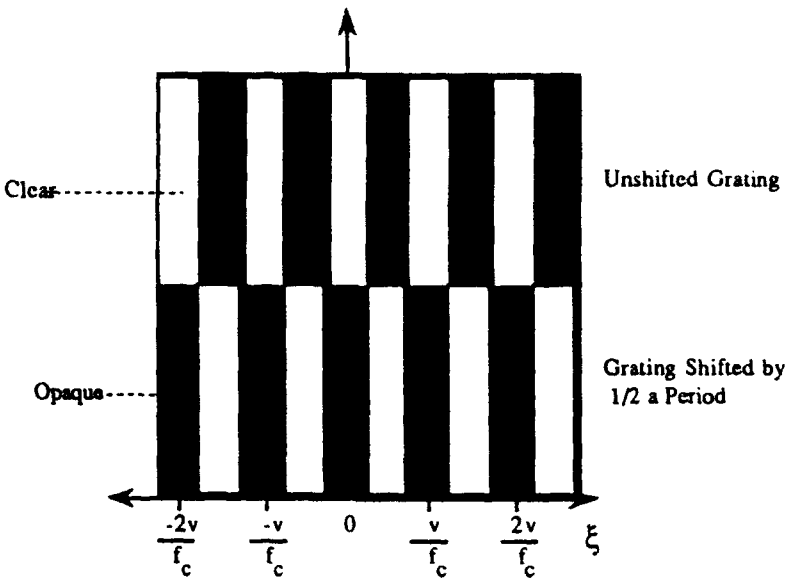


Figure 5. Grating used for encoding bi-polar signals.

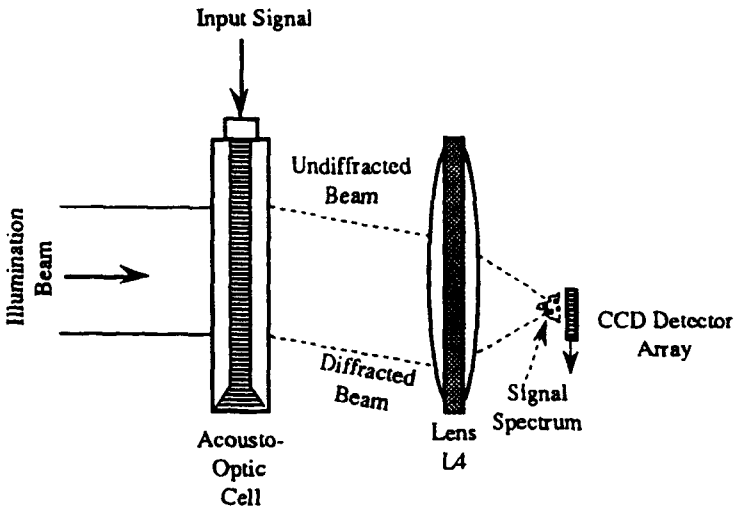


Figure 6 Schematic of acousto-optic spectrum analyzer. The undiffracted beam has been removed for clarity.

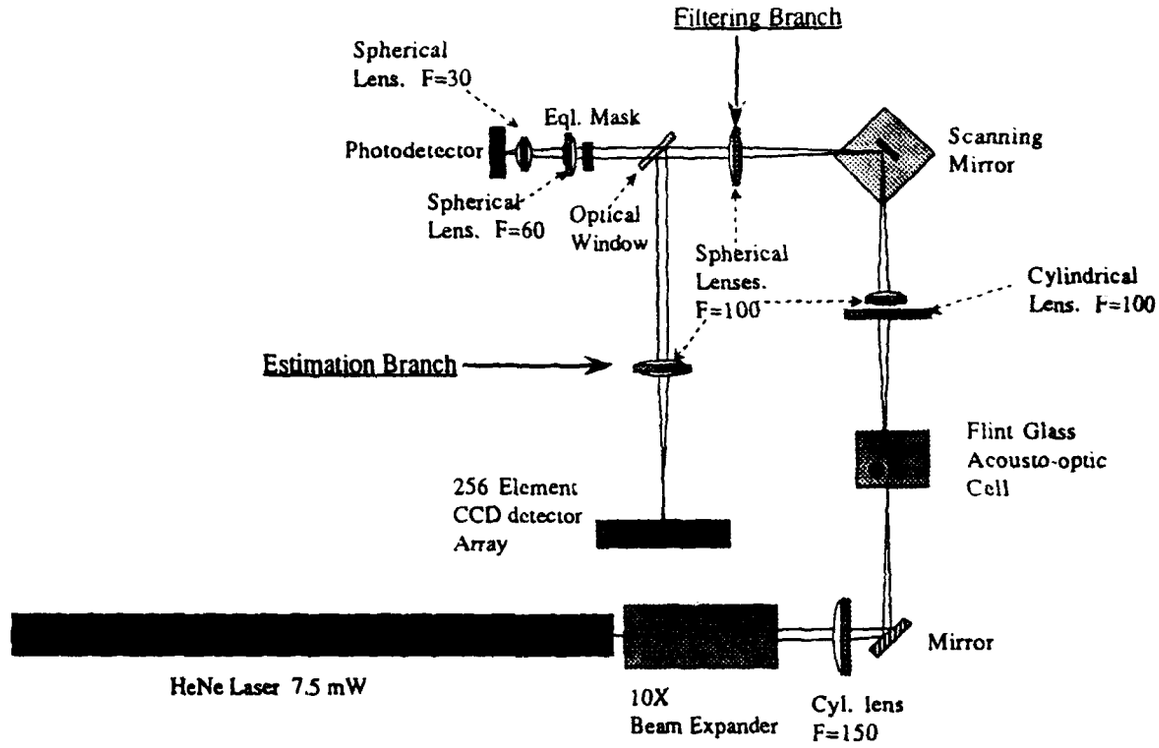


Figure 7. Optical layout of the experimental acousto-optic equalizer.

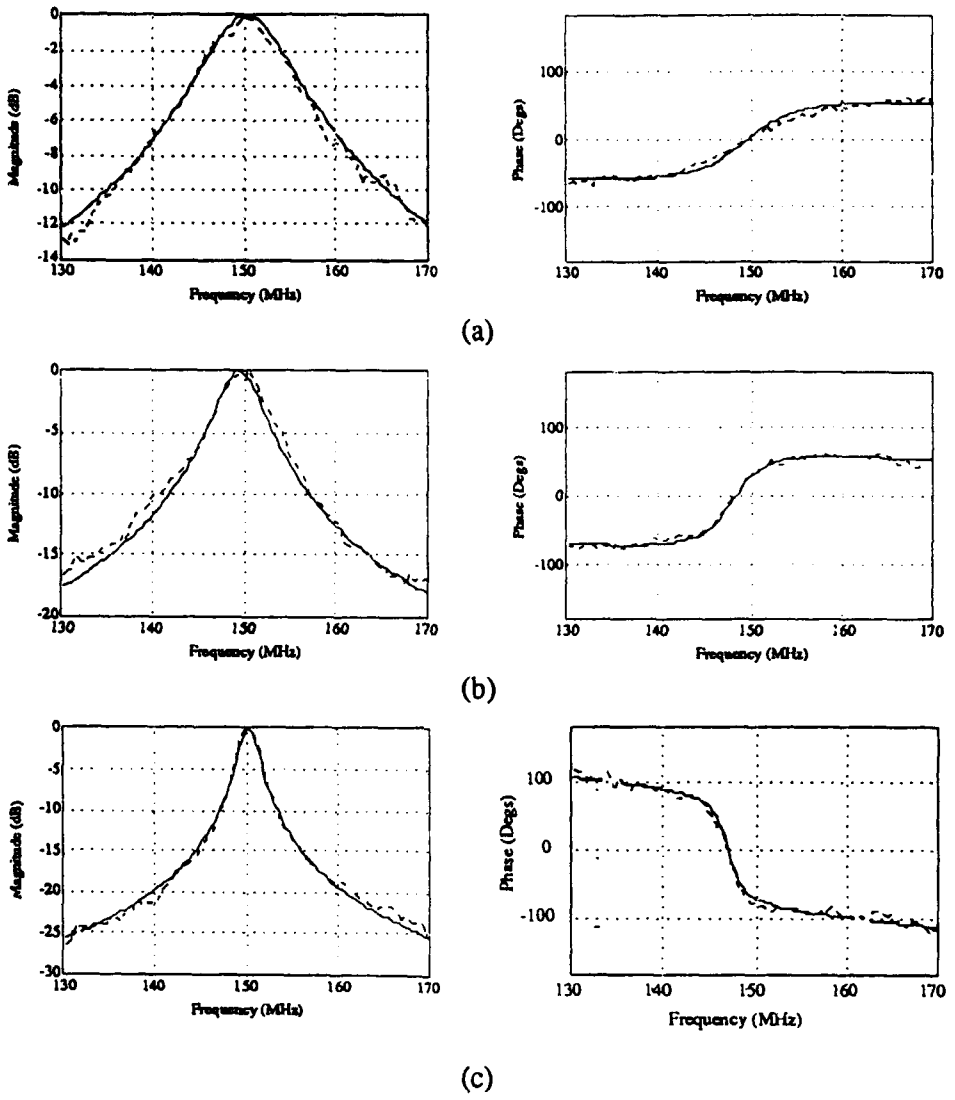
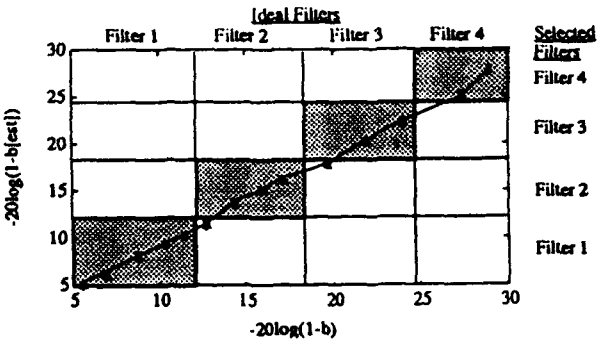


Figure 8. Measured (- -) and theoretical (—) magnitude and phase response for three of the eight equalization filters. C

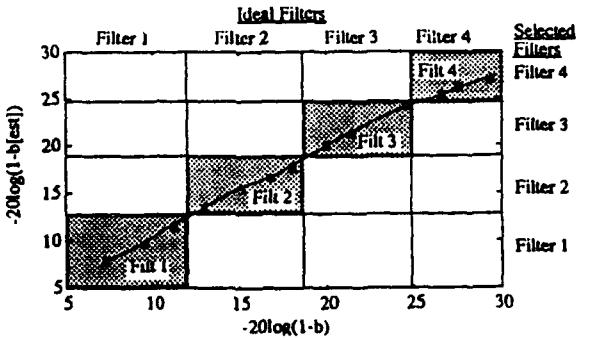
a) b=0.820

b) b=0.905

c) b=1.047



(a)



(b)

Figure 9. Plot of actual and estimated notch depth. Shaded regions correspond to channels estimates that select the best filter from the library.

a) f_0 = -4.7 MHz

b) f_0 = +0.8 MHz.

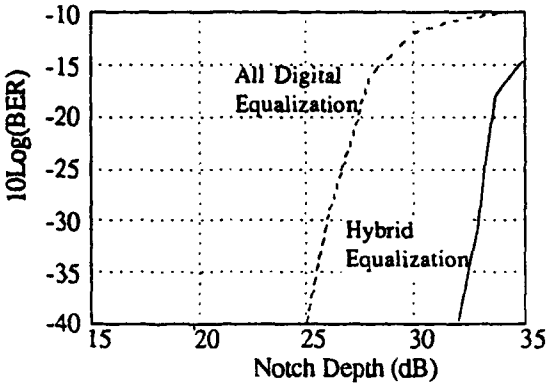


Figure 10. Static Bit Error Rate Measurements for 256-Level Quadrature Amplitude Modulation. The notch location f_0 is 3 MHz.

CAPTIONS LIST

Figure 1. Schematic of a microwave digital radio with acousto-optic equalizer.

Figure 2. Magnitude and phase plots for Rummler's channel model.

Figure 3. Mean square error output from an 11 tap equalizer. The notch location is 4 MHz. The lower curve is the smallest error than can be obtained from any linear processor.

Figure 4. Top view two acousto-optic architectures for filtering electrical input signal. Both architectures use homodyne detection.

a) Time domain filtering

b) Fourier domain filtering

Figure 5. Grating used for encoding bi-polar signals.

Figure 6 Schematic of acousto-optic spectrum analyzer. The undiffracted beam has been removed for clarity.

Figure 7. Optical layout of the experimental acousto-optic equalizer.g

Figure 8. Measured (- -) and theoretical (—) magnitude and phase response for three of the eight equalization filters.

a) b=0.820

b) b=0.905

c) b=1.047

Figure 9. Plot of actual and estimated notch depth. Shaded regions correspond to channels estimates that select the best filter from the library.

a) f_0 = -4.7 MHz

b) f_0 = +0.8 MHz.

Figure 10. Static Bit Error Rate Measurements for 256-Level Quadrature Amplitude Modulation. The notch location f_0 is 3 MHz.

REFERENCES

1. M. Kohiyama and O. Kurita , "Future Trends in Microwave Digital Radio: a View from Asia", IEEE Comm. Mag., Vol. 25, No. 2, 41-45 (1987).
2. C. Anderson and A. VanderLugt, "Hybrid acousto-optic and digital equalization for microwave digital radio channels," Optics Letters, 1182-1184 (1990).
3. Jon W. Bayless, Arthur A. Collins, and Robert D. Pedersen, "The Specification and Design of Bandlimited Digital Radio Systems," IEEE Trans. on Communications, Vol. COM-27, NO. 12, 1763-1770 (1979).
4. J. K. Chamberlain, F. M. Clayton, H. Sari, and P. Vandamme, "Receiver Techniques for Microwave Digital Radio," IEEE Communications Magazine, 43-54 (1986).
5. E. Johnson, "An adaptive IF Equalizer for Digital Transmission," Intern. Conf. Comm., 13.6.1-13.6.4 (1982).
6. C. L. Ruthroff, "Multiple-path fading on line-of-sight microwave radio systems as a function of path length and frequency," Bell System Tech. Journal, **50**, 2375-2398 (1971).
7. P. Balaban, "Statistical Model for Amplitude and Delay of Selective Fading," AT&T Tech. Jour., 2525-2529 (1985).
8. W. D. Rummler, "A New Selective Fading Model: Application to Propagation Data," Bell Sys. Tech. Jour., Vol. 58, No. 5, 1037-1071 (1979).
9. S. Qureshi, "Adaptive Equalization", Proc. of the IEEE, Vol. 73, No. 9, 1349-1387 (1985).

11. C. Anderson, Hybrid Acousto-optic and Digital Equalization for Microwave Digital Radios, PhD Dissertation, North Carolina State University, (1991).

12. A. Rosenthal, "Applications of Ultrasonic Light Modulators to Signal Recording, Display, Analysis, and Communications," IEEE Trans. Ultrasonic Eng., Vol. UE-8, 1 (1961).

13. L. Slobodin, "Optical Correlation Techniques", Proc. IEEE, Vol. 51, 1782-1799 (1963).

14. E. Felstead, "A simplified Coherent Optical Correlator," Appl. Optics, Vol.7, 105-114 (1968).

15. A. VanderLugt, Optical Signal Processing, (John Wiley & Sons, New York, 1992), Chapters 7-11.

16. T. Karnowski and A. VanderLugt, "Generalized Filtering in Acousto-optic Systems using Area Modulation," Appl. Optics, Vol. 30, No. 17, 2344-2353 (1991).

17. D. Hecht, "Spectrum Analysis using Acousto-optic Devices", in SPIE Proc. on Acousto-optics, (Vol. 90., 1976) p. 148.

18. A. Oppenheim and R. Schafer, Digital Signal Processing, (Prentice Hall, Englewood Cliffs, New Jersey, 1975).

19. A. Papoulis, Probability, Random Variables, and Stochastic Processes, (McGraw Hill, New York, 1965).

20. J. Thomas, An Introduction to Statistical Communication Theory, (Wiley, New York, 1969).

APPENDIX D

SIGNAL DISTORTION IN AN ADAPTIVE EXCISION SYSTEM

Reprinted from Optical Engineering

Volume 31, Pages 606-613, March 1992

Signal distortion in an adaptive excision system

Reeder N. Ward*
Anthony VanderLugt, FELLOW SPIE
North Carolina State University
Electrical and Computer Engineering Department
Daniels Hall, Box 7911
Raleigh, North Carolina 27695-7911

Abstract. A narrowband signal excision system uses spatial light modulators in the spatial frequency domain to notch energy caused by narrowband interferers. We review the signal excision process and discuss the optical system developed to monitor the noise spectrum. A postdetection system searches for narrowband interferers in wideband spectra and provides control signals to implement the required notch; experimental results of the distortion imposed on short pulse signals are given.

Subject terms: signal processing; distortion; excision systems.

Optical Engineering 31(3), 606-613 (March 1992).

1 Introduction

As the spectral environment becomes more congested, wide-bandwidth signals are frequently corrupted by narrowband interference. Because an optically generated Fourier transform is physically accessible, we can use spatial light modulators (SLMs) to block the energy from the interference signals while allowing the spectral energy from the wideband signals to pass through the system, thereby improving the signal-to-noise ratio (SNR) at the output.

In the early 1980s Erickson[1,2] performed analyses and experiments dealing with signal excision using spatial filters in the spatial frequency plane. Lee et al.[3] reported on experiments in which they used a threshold detector array to control SLMs in the spatial frequency plane. Roth[4] reported on the use of a photodetector array to remove interfering signals and Anderson et al.[5] demonstrated excision using a GaAs photodetector array. Brandstetter and Grieve[6] built an optical notching filter that uses a recursive technique to obtain increased notch depths.

In this paper we describe an adaptive system in which the response time of the feedback loop is fast enough to track changes in the rf environment for many applications and show how the shapes of short pulses are distorted when narrowband interferers are excised. In Sec. 2 we describe a signal excision system using SLMs in the spatial frequency domain to notch the energy caused by narrowband interferers. In Sec. 3 we describe the physical layout of a signal excision system and discuss the optical modifications implemented to monitor the noise spectrum. In Secs. 4 and 5 we present experimental results produced by the system, including the distortion effects of the excision process on short pulses.

*Current affiliation: Harris Government Systems Sector, Melbourne, Florida 32901.

Paper 11061 received June 17, 1991; revised manuscript received Sept. 18, 1991; accepted for publication Oct. 7, 1991.
 0091-3286/92/$2.00

2 Background for Signal Excision

The spectra of wideband signals are spread over a large region of the spatial frequency domain. The spectra of narrowband interference signals, on the other hand, occupy only isolated, small regions of the spatial frequency domain. A spatial filter that removes the energy associated with the narrowband interference signal improves the SNR at the output of the optical system, significantly aiding in the recovery of the signal. The signal waveform may, however, become distorted in the process, depending on the frequency of the narrowband interferer and the spectral content of the signal.

Figure 1 shows an interferometer in which an acousto-optic cell at plane P_3, driven by an electrical signal $f(t)$, modulates light in both space and time. Although the signal to be processed is usually an rf signal that is translated into the bandpass of the acousto-optic cell, we elect to discuss the basic concepts in terms of baseband signals and interferers that are modulated onto a carrier frequency f_c; this analysis more nearly parallels that needed to describe the experimental results. The drive signal to the cell is $f(t)=q(t)\cos(2\pi f_c t)$, where

$$q(t)=s(t)+\sum_{j=1}^{N} A_j \cos(2\pi f_j t+\phi_j) \tag{1}$$

is the sum of the desired baseband signal $s(t)$ and N narrowband-interference signals that have arbitrary magnitudes, frequencies, and phases.

The positive diffracted order produced by the acousto-optic cell is represented by

$$f_+(x,t)=a(x)q\left(t-\frac{T}{2}-\frac{x}{v}\right)\exp\left[j2\pi f_c\left(t-\frac{T}{2}-\frac{x}{v}\right)\right]. \tag{2}$$

where $a(x)$ is the aperture weighting function, T is the time duration of the acousto-optic cell, and v is the acoustic

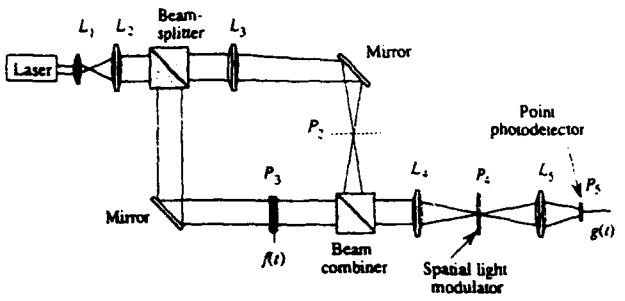


Fig. 1 Heterodyne acousto-optic system for notched filtering.

velocity. The Fourier transform of the signal beam is located at plane P_4:

$$F_+(\alpha,t)=\int_{-\infty}^{\infty} a(x)q\left(t-\frac{T}{2}-\frac{x}{v}\right) \times \exp\left[j2\pi f_c\left(t-\frac{T}{2}-\frac{x}{v}\right)\right]\exp(j\pi\alpha x)\,dx \quad , \quad (3)$$

where the time-dependent Fourier transform is expressed in terms of a spatial frequency α. We remove the factors that are not a function of x from the integral to find that

$$F_+(\alpha,t)=\exp\left[j2\pi f_c\left(t-\frac{T}{2}\right)\right]\int_{-\infty}^{\infty} a(x)q\left(t-\frac{T}{2}-\frac{x}{v}\right) \times \exp[j2\pi(\alpha-\alpha_c)x]\,dx \quad . \quad (4)$$

This result shows that $F_+(\alpha,t)$ is the Fourier transform of the aperture-weighted baseband signal, centered at a spatial frequency α_c.

Light in the upper branch of the interferometer passes through lens L_3 and becomes a point source at plane P_2. A point source located a distance x_0 from the optical axis at plane P_2 is represented by $r(x)=\text{sinc}[(x-x_0)/d_0]$, where d_0 is the distance from the centroid of the sinc function to its first zero. The Fourier transform of the reference signal, after a small rotation of the beam combiner to center the light at α_c, becomes

$$R(\alpha)=\text{rect}[(\alpha-\alpha_c)d_0]\exp(j2\pi\alpha x_0) \quad , \quad (5)$$

where the exponential term is generated by the off-axis position of the source.

Heterodyne detection can be achieved by using a photodetector immediately after the SLM at plane P_4 to integrate the intensity due to the sum of the Fourier transforms of the signal and reference beams:

$$g(t)=\int_{-\infty}^{\infty} |F_+(\alpha,t)+R(\alpha)|^2|H(\alpha)|^2P(\alpha)\,d\alpha \quad , \quad (6)$$

where $H(\alpha)$ is the spatial frequency response of the SLM used to notch the narrowband interference signals and $P(\alpha)$ is the response of the photodetector. An alternative method of detection is to use a photodetector at the image plane of the reference and signal functions, as shown in Fig. 1. The filtered signal beam at plane P_5 is

$$f_+(u,t)=\int_{-\infty}^{\infty} a(x)f_+(x,t)h(x-u)\,dx \quad . \quad (7)$$

where u is the spatial coordinate at plane P_5 and $h(u)$ is the impulse response of the filter function $H(\alpha)$. The reference beam also passes through the filter function $H(\alpha)$ and, at plant P_5, becomes

$$r(u)=\int_{-\infty}^{\infty} r(x)h(x-u)\,dx \quad . \quad (8)$$

For image plane detection, the output signal is

$$g(t)=\int_{-\infty}^{\infty} |f_+(u,t)+r(u)|^2\,p(u)\,du \quad , \quad (9)$$

where $p(u)$ is the response of the photodetector at plane P_5. We expand the integrand of Eq. (9) and find that the output photocurrent consists of three terms:

$$g_1(t)=\int_{-\infty}^{\infty} |r(u)|^2p(u)\,du \quad , \quad (10)$$

$$g_2(t)=\int_{-\infty}^{\infty} |f_+(u,t)|^2p(u)\,du \quad , \quad (11)$$

$$g_3(t)=2\,\text{Re}\left[\int_{-\infty}^{\infty} f_+(u,t)r^*(u)p(u)\,du\right] \quad . \quad (12)$$

Since $g_1(t)$ is not a function of time, its spectrum is concentrated at zero temporal frequency. It is relatively straightforward to show that the spectral content of $g_2(t)$ is also concentrated at zero frequency. Since the temporal spectrum of $g_3(t)$ is centered at f_c, the baseband terms can be removed by a bandpass filter, leaving only the cross-product term as given by Eq. (12). We substitute Eqs. (7) and (8) into Eq. (12) to find that

$$g_3(t)=2\,\text{Re}\left\{\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} a(x)f_+(x,t)h(x-u)\,dx\right] \times\left[\int_{-\infty}^{\infty} r^*(y)h^*(y-u)\,dy\right]p(u)\,du\right\} \quad . \quad (13)$$

where y is a dummy variable.

The general result from Eq. (13) cannot be reduced further without specifying the factors in the integrand. Suppose, as an example, that we consider the simple case for which $H(\alpha)=1$ over the range of spatial frequencies oc-

cupied by the signal. Since this filter in no way alters the signal or its spectrum, we can replace $h(u)$ by a delta function so that Eq. (13) reduces to

$$g_3(t) = 2\ \mathrm{Re}\left[\int_{-\infty}^{\infty} a(u) f_+(u,t) r^*(u) p(u)\ \mathrm{d}u\right] . \tag{14}$$

We substitute $r^*(u)$ into Eq. (14) to find that

$$g_3(t) = 2\ \mathrm{Re}\left\{\int_{-\infty}^{\infty} a(u) f_+(u,t) \times \mathrm{sinc}[u - x_0)/d_0] p(u)\ \mathrm{d}u\right\} . \tag{15}$$

Since the reference function $r(u)$ is a stationary probe at plane P_5, the integral is in the form of a temporal convolution. Furthermore, since the value of d_0 was set so that $R(\alpha)$ is uniform, the probe behaves as a delta function so that Eq. (15) becomes

$$g_3(t) = 2a(x_0)p(x_0)s(t-\tau)\ \cos[2\pi f_c(t-\tau)] , \tag{16}$$

where $\tau = T/2 + x_0/v$. Equation (16) shows that the input signal is completely recovered except for a time delay τ when the filter function is uniform over all frequencies. For the general case where $H(\alpha) \neq 1$, the signal may be distorted as we show in Sec. 5.

3 The Presort Processor

The presort processor, located at the Photonics Center at Rome Air Development Center, is an optical signal processor which excises narrowband interference from wideband signals. This system, made available to us for implementing certain modifications and for subsequent experimental work, is diagrammed in Fig. 2. The presort processor uses a prism to expand light in the direction of acoustic propagation and illuminates the acousto-optic cell with a Gaussian aperture weighting function. The time-bandwidth product and the Gaussian weighting of the TeO_2 acousto-optic cell provides 200 resolvable frequencies at the spatial frequency plane. The acousto-optic cell receives the electrical signal $f(t)$, which includes the signal $s(t)$ and narrowband interferers. Light in the upper branch of the interferometer passes through lens L_3 to create a point source located at plane P_2 whose distance from the beam combiner is equal to the distance between the acousto-optic cell at plane P_3 and the beam combiner.

Ideally, a single SLM would be used in the spatial frequency plane to excise unwanted signals. A wide variety of devices, such as those based on liquid-crystal technology, magneto-optic technology, or deformable-mirror technology, could be used. The requirements for the presort processor were that the SLM must provide switching times of less than 1 μs and extinction ratios of at least 30 dB. Multichannel acousto-optic devices were chosen to satisfy these requirements, but such devices are not well suited to providing a 100% duty cycle due to crosstalk constraints for adjacent channels. Two such devices are therefore needed, one in each of two spatial frequency planes created by the

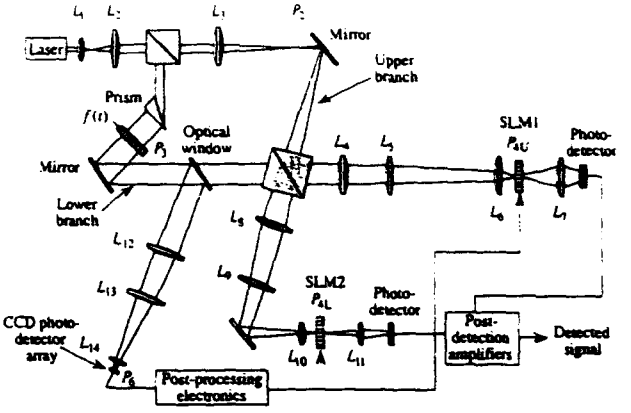


Fig. 2 Modified presort processor.

presort processor. Figure 2 shows that the beam combiner at the output of the interferometer provides the necessary Fourier transforms at planes P_{4U} and P_{4L} where the multichannel spatial light modulators SLM1 and SLM2 are located.

Since both optical segments after the beam combiner operate identically, we describe just the straight segment in detail. Spherical lenses L_4 and L_6 form a telescopic system to display the Fourier transform of the signal $f(t)$ at plane P_{4U}, which contains the SLM that blocks or passes various frequency components of the signal. The SLMs used in this system each have 100 channels, arranged with a 50% duty cycle, resulting in half the spectrum being covered by SLM1 in the straight segment while the other half of the spectrum is covered by SLM2 in the folded segment. The width of each channel is equivalent to a 1-MHz-frequency interval. If a narrowband interferer is present, the appropriate acousto-optic channels are turned off so that light corresponding to that interferer is no longer incident on the photodetectors. Analyses have shown that at least three adjacent SLM channels must be switched to provide a 25-dB notch depth for the Gaussian incident illumination. The photodetector output from the folded segment is then added to the photodetector output of the straight segment so that, in effect, the two SLMs cover the entire bandwidth. No phase errors are generated by path length differences between the two segments after the beam combiner because the signal and reference beams experience the same path lengths in both cases.

3.1 *Modifications to the Presort Processor*

Figure 2 also shows how we modified the optical portion of the presort processor. To provide for adaptive operation, we implemented a power spectrum analyzer that detects narrowband interferers. An optical window, inserted at a 20-deg angle after the acousto-optic cell, reflects 4% of the light. Lenses L_{12}, L_{13}, and L_{14} create the Fourier transform of the reflected signal beam at plane P_6, while also creating an image of the acousto-optic cell line illumination function in the orthogonal direction at the CCD array.

The beam passing through the optical window is displaced by the splitting action of the optical window. For a 6-mm-thick window, inserted at a nominal 20-deg angle, the beam displacement of 0.68 mm merely affects the rel-

ative position of the probe relative to the acousto-optic cell, as characterized by Eq. (15), and has no significant effect on the performance of the system. Of greater significance is that a second surface reflection from the optical window produces a ghost beam, parallel to the first surface reflection, which induces a fringe pattern on the spectrum. The frequency of the fringe pattern is a function of the displacement between the signals produced by the first and second surface reflections: the contrast of the fringe pattern is highly dependent on the magnitude of the second surface reflection. Even though the second surface reflectivity was measured at only 0.17%, the fringe contrast varied by 3.6 dB. Since the fringe pattern is fixed in space and time, it can be compensated for by the use of a look-up table. The need for compensation may be eliminated altogether in some applications because the magnitude of the interferers generally are about 30 dB above the signal spectrum and do not typically need to be measured with a high degree of accuracy. Wedge in the optical window, caused by the nonparallel surfaces of the optical window, shifts the spectrum of the signal at planes P_{4U} and P_{4L}. These shifts are easily compensated for when the system is calibrated.

3.2 *Electronic Modifications to the Presort Processor*

Figure 3 is a block diagram of the postprocessing electronics that we designed and built to implement the adaptation process. The power spectrum of the signal, detected by the CCD array, is sequentially clocked at a 1-MHz rate into an 8-bit analog-to-digital converter that converts the spectrum to a digital number stream. The digital numbers are compared to a threshold that is adaptively set by the postprocessing electronics; alternatively, the threshold can be set by the operator. If a digital number is greater than the threshold, the circuit decides that a narrowband interferer is present. The locations of the pixels for which the spectrum exceeds threshold are identified by the threshold circuit and clocked into the excising routine. The excising routine uses a look-up table format to provide the location information required to switch the appropriate SLM channels.

The adaptation mode of operation involves driving the postprocessing electronics to a steady-state solution. Suppose that switches one and two shown in the block diagram of Fig. 3 are closed, while switch three is open. In this mode of operation, the operator inputs the number of narrowband interferers to be excised and the threshold search feedback loop sets a threshold level so that the required number of narrowband interferers are excised.

Figure 4 illustrates how the threshold is set when multiple-narrowband interferers are present. Suppose that we require the number of excised narrowband interferers to be no more than three. The threshold search operates on the condition of whether the number of narrowband interferers is greater than or equal to the preset number of narrowband interferers. The first time the spectrum is read out by the CCD array, the system automatically sets the threshold to its largest value. The number of narrowband interferers exceeding threshold, which may be equal to zero, is then compared to the preset number of interferers. The second time the array is clocked out the postprocessing electronics sets the threshold to the next step below the initial value; there are 16 equally spaced steps across the range of an 8-bit word.

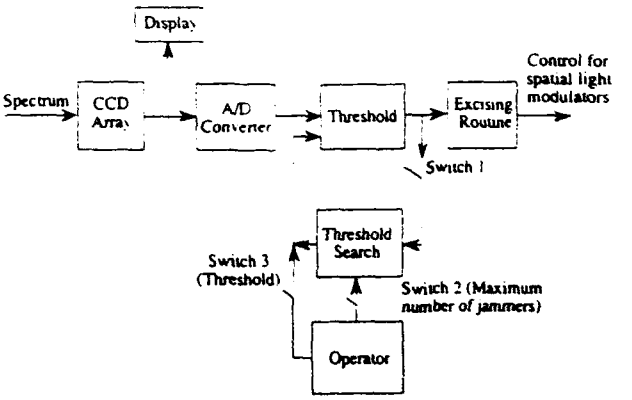


Fig. 3 Block diagram of the postdetection electronics.

The number of narrowband interferers exceeding this threshold are again counted and compared to the preset number so that the appropriate SLM channels can be switched to remove those narrowband interferers that exceed the threshold. This adaptive process is repeated until the preset number of narrowband interferers to be excised is reached. The threshold remains fixed until some narrowband interferers disappear or appear in the spectrum, or a new preset number is chosen.

For experimental purposes, we included an operator interaction mode in which switches one and two in the diagram shown in Fig. 3 are open, while switch three is closed. The operator reduces or increases the threshold, depending on how many narrowband interferers are to be excised. We implemented a visual display to enhance the operation of the system by showing the power spectrum of the signal and the threshold level of the postprocessing electronics. The spectrum of the signal is tapped directly off the CCD photodetector array and drives channel one of an oscilloscope. The threshold, which appears as a horizontal line across the display, drives channel two of the oscilloscope. In this mode of operation, the operator, in effect, completes the feedback loop. Figure 4 shows that the system has adjusted the threshold so that the three strongest interferers exceed threshold; the appropriate SLMs needed to excise these interferers can therefore be activated.

The excising algorithm is a versatile part of the postprocessing electronics. The CCD photodetector array provides four CCD photodetector elements to sample the spatial frequencies corresponding to one SLM channel; because there are 200 channels, we use 800 of the elements available in a 1024-element array. The excising algorithm, stored in electrically programmable read-only memories, searches for interferers that have a width of one frequency resolution channel. If the interferer is located directly on a SLM channel, that channel plus the neighboring two channels are switched; no more than three channels are ever switched for one interferer. If the interferer is principally located on two SLM channels, those two SLM channels are switched plus the SLM channel that is the nearest neighbor to the channel on which the most energy of the interferer is incident.

In dynamic environments, narrowband interferers are not necessarily stationary in amplitude or frequency, but they are more likely to appear and disappear throughout the spec-

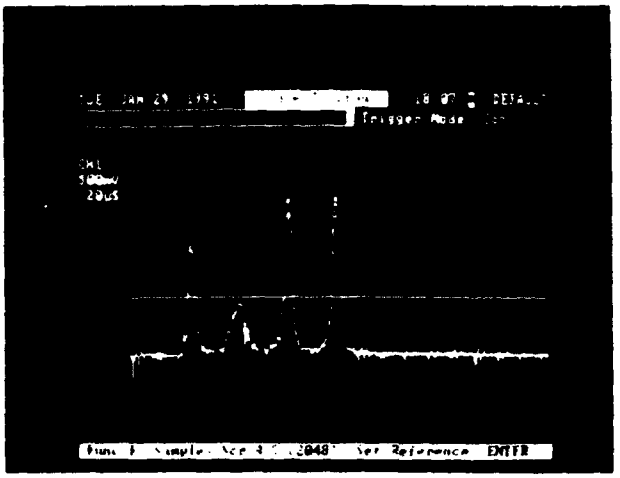


Fig. 4 Spectrum with narrowband interferers.

trum in a random fashion. We used a relatively slow CCD array that has a single output video line and a 1 ms per frame readout rate. The postprocessing electronics require 5.4 ms to process this information and to provide the addresses needed to notch the interferers. This response time can be shortened by a factor of about 10 by using CCD arrays that have multiple output video lines with faster readout rates and by using faster electronic components.

4 Experimental Results of Excising Narrowband Interference

To illustrate the narrowband interference excision process, we used a 100-ns duration pulse train with a repetition period of 5 μs as a representative wideband signal and modulated it onto a carrier frequency $f_c = 350$ MHz. We then introduced a narrowband interference signal, passed the composite signal through the presort processor, excised the interference signal, and displayed the pulse after it had been demodulated to baseband. For these experiments the narrowband interference was a single frequency f_j. The drive signal to the acousto-optic cell is therefore

$$f(t) = s(t)\cos(2\pi f_c t) + A_j \cos(2\pi f_j t + \phi_j) \quad , \tag{17}$$

where $s(t)$ is the baseband pulse signal, A_j is the magnitude, and ϕ_j is the phase of the interferer in the rf band.

Although the optical system has a SNR greater than 25 dB over a 200-MHz band, the rf mixers limit the overall system performance to a much lower figure. Since triggering the oscilloscope to display the waveform was difficult, we elected to overdrive the local oscillator input, thus producing a larger signal level as well as a larger noise level at the output. We subsequently averaged the output signal to reduce the noise level. Since the averaging is equivalent to coherent detection, the SNR attained in the following set of experiments is larger than the actual SNR at the output of the system.

The reference output pulse shown in Fig. 5(a) was added to a narrowband interferer at 307 MHz to produce the composite signal shown in Fig. 5(b); the pulse is completely obscured in the composite signal. In the frequency domain the pulse signal produces a spectrum centered at 350 MHz, resembling a sinc function with a 20-MHz-wide mainlobe. Figure 5(c) shows that the spectrum of the narrowband interferer, located in the center of the second sidelobe of the pulse, is 30 dB above the noise floor. After the SLM channels corresponding to the frequency of the interferer were switched, the output of the system was demodulated to produce the output pulse shown in Fig. 5(d). The output pulse has a shape similar to that of the reference pulse, except that the rise-time overshoot, apparent in the reference pulse, is partially removed when we excise the narrowband interferer. This example illustrates qualitatively that the presort processor is capable of removing narrowband interferers without seriously distorting the pulse shape.

There are various methods for quantitatively measuring signal distortion. We chose the root-mean-square error criterion:

$$\text{rms error} = \left\{ \frac{1}{T_r} \int_0^{T_r} [b(t) - d(t)]^2 \, dt \right\}^{1/2} \quad , \tag{18}$$

where the undistorted pulse, at the output of the system in the absence of excision, is indicated by $b(t)$, the distorted signal is indicated by $d(t)$, and T_r is the pulse repetition period.

Since most of the pulse signal energy is in the mainlobe of its corresponding sinc function in the frequency domain, we expect the greatest pulse shape distortion to occur when the narrowband interferer is inside the mainlobe of the sinc function. Suppose that we vary the frequency of the narrowband interferer so that it passes through the mainlobe from low to high frequencies, distorting the spectrum of the pulse as the interferer is excised. Figure 6 shows the rms error as a function of the frequency of the interferer. The rms error is small when the interferer frequency is located near the low-frequency edge of the mainlobe because the mainlobe has a low magnitude in this region. As we increment the frequency of the narrowband interferer through the mainlobe, the rms error increases until the frequency reaches 332 MHz, which is approximately the center frequency of the pulse used in this experiment. The rms error is largest when the interferer is centered in the mainlobe of the signal because the SLM channels remove a considerable amount of signal energy in addition to excising the interferer. The error then decreases as the frequency of the interferer increases until the first null of the sinc function is reached at 342 MHz.

5 Pulse Experiments

When we excise narrowband interferers, part of the energy of the intended signal is also removed. We now explore the nature of the pulse distortion caused by removing some of its frequency components. Fourier series analysis shows that all repetitive signals consist of weighted discrete frequency components. The dc value and the fundamental frequency reside in the mainlobe of the sinc function due to an isolated pulse, while the higher frequency components reside in the sidelobes of the sinc function. Therefore we expect that the removal of the mainlobe frequency components will have the greatest affect on the overall shape of the pulse, whereas the removal of higher frequency components will affect mostly pulse ripple, pulse rise time, and pulse fall time.

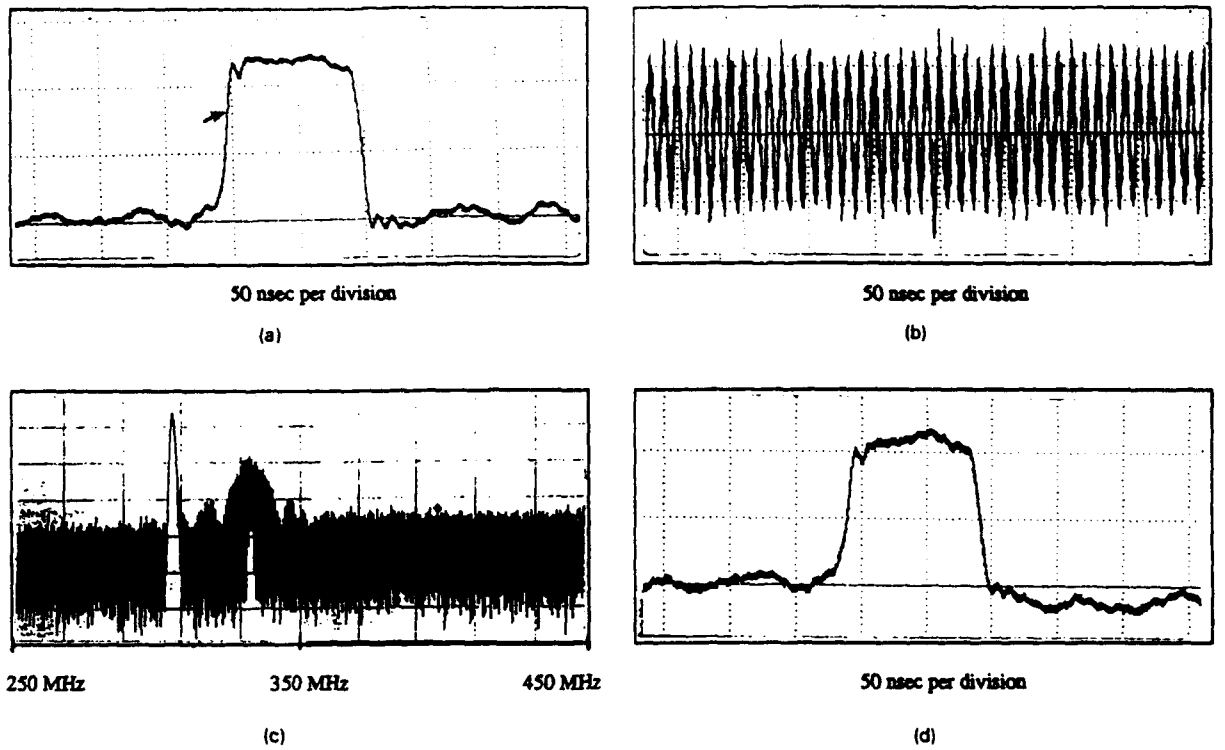


Fig. 5 Excision of narrowband interferer: (a) reference pulse (50 ns/div), (b) pulse plus a cw interferer at the system output, (c) spectrum of pulse and interferer (20 MHz/div), and (d) output pulse.

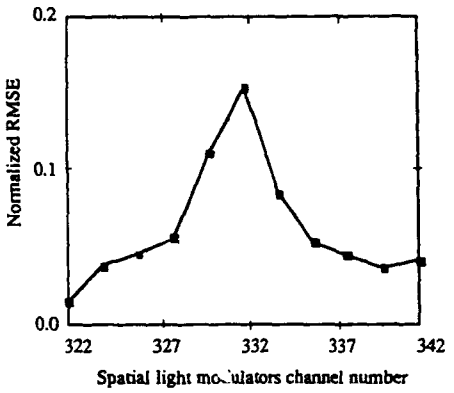


Fig. 6 Root-mean-square error as a function of the interferer frequency.

We begin by removing frequency components symmetrically about the center frequency of a 78-ns pulse signal; this procedure simulates a situation in which the interferer is introduced in the baseband signal and allows us to study the effects of notching multiple interferers on pulse distortion. As an example, suppose that the frequency of the baseband interferer is $f_j = 18$ MHz. When the signal is brought to the IF center frequency of 350 MHz, the interferer energy is concentrated at $f_c \pm f_j$, or at 332 and 368 MHz. Since the first zeros of the spectrum of the 78-ns pulse occur at $f_c \pm 12.8$ MHz, we find that the interferer energy is concentrated in the first sidelobes of the sinc function associated with the signal spectrum. When these frequency components of the pulse are removed, the edges of the pulse become rounded, as the simulation in Fig. 7(a) and the experimental results of Fig. 7(b) show. As we remove frequency components in the second sidelobe and beyond, the edges on the pulse remain rounded and ripple is evident on the top of the pulse due to the removal of the higher frequency components.

Figure 8 shows the result of a computer simulation of the rms error as a function of the frequency of the narrowband interferer. The rms error begins at 0.50 and increases to 0.68 at 352 MHz. The increase in error is due to the removal of the central frequency components, which requires that only three SLM channels be switched off, whereas the removal of the first set of symmetric frequency components requires that six adjacent SLM channels, centered in the mainlobe, be switched. Since the symmetrical notches are at the highest magnitude portion of the pulse spectrum, the rms error is large. The rms error drops rapidly as the excision positions move away from the central frequency, following the general shape of the sinc^2 spectrum produced by the pulse signal.

Figure 8 also shows the experimental results of symmetrically removing frequency components from the spectrum of a 78-ns pulse. The rms error, at the center frequency, begins at 0.29 and increases to 0.37 at 352 MHz. The experimental rms error is generally smaller than the computer simulation results because infinite notch depths were used in the computer simulations. Other reasons for a departure of the experimental results from the simulation results are that the local oscillator frequency tended to drift as the experiment data were collected, and the SNR of the system

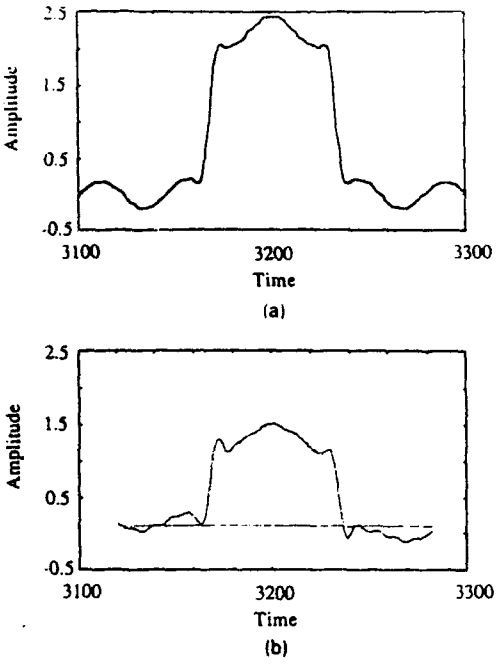


Fig. 7 Pulse shape distortion due to symmetric excision of a baseband interferer: (a) computer simulation and (b) experimental results.

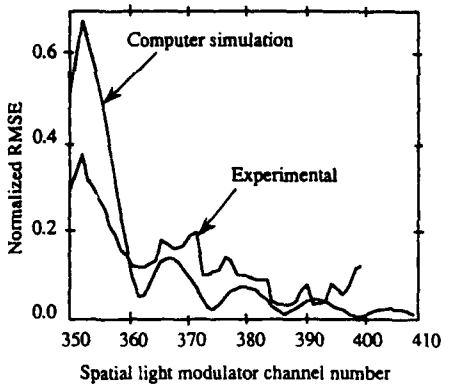


Fig. 8 Root-mean-square error for a 78-ns pulse.

limited the accuracy of the data when the higher frequencies of the pulse were removed.

The effects of excising narrowband interferers on the pulse shape are also a function of the pulse duration. The least distortion is expected when the pulse width is small because the pulse spectrum is then spread over a large portion of the frequency domain. The removal of energy corresponding to the notches needed to excise a few interferers should therefore have little effect on the pulse shape. However, when the pulse width is large, the pulse spectrum is narrow so that a given notch width represents a large fraction of the signal spectral components. To illustrate this effect, we repeated the frequency removal experiments using a 468-ns pulse, whose mainlobe is only 4.3 MHz wide, as the input signal. A pulse of this duration represents a fairly severe test because the cell duration is only 2 μs. The spectrum of a 468-ns-wide pulse is therefore nearly the same as that of a narrowband interferer.

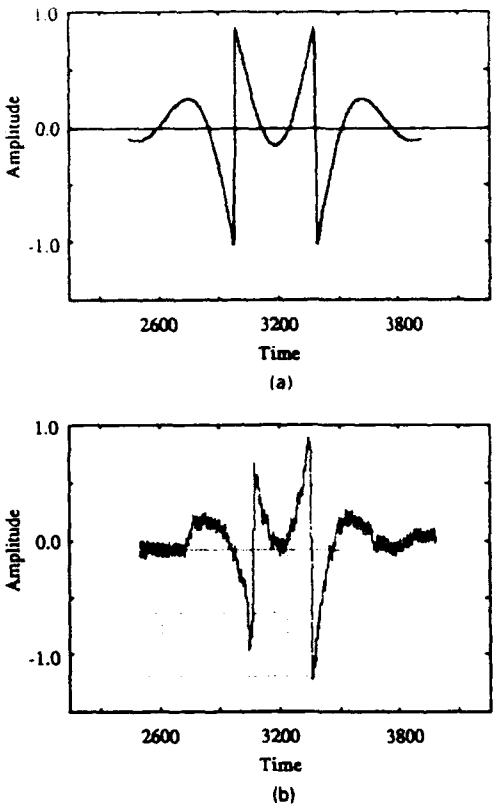


Fig. 9 Removal of central frequency components of the 468-ns pulse: (a) computer simulation and (b) experimental results.

Figures 9(a) and 9(b) illustrate the pulse distortion when we switch the three central SLM channels, thus producing the largest rms error. By notching the dc component and a significant amount of the fundamental, we produce a result similar to a differentiation of the pulse. The experimental results agree with the computer simulations; the discrepancy between the two plots is caused by the removal of slightly less than the complete mainlobe of the sinc function in the experimental results.

Figure 10 shows the computer simulation and experimental results of the rms error for the narrow pulse as a function of SLM channel position. The simulations predict that the rms error should be largest when we switch SLM channels centered at 350 MHz because the notch, which is 3 MHz wide, removes most of the mainlobe energy. By a similar line of reasoning, we expect the rms error to decline quickly as the frequency of the interferer moves away from the center frequency of the pulse. We note that the shape of the rms error of the experimental results is similar to the computer simulations. An exception occurs for somewhat higher frequency values when the interferer is in the sidelobes of the signal spectrum, because the simulations did not model the noise of the system.

6 Conclusions

Frequency plane excision is one method for removing narrowband interference from wideband signals. We modified a presort processor, which uses frequency plane excision, to provide control signals to the SLMs that perform the notching of the interference. We added a narrowband in-

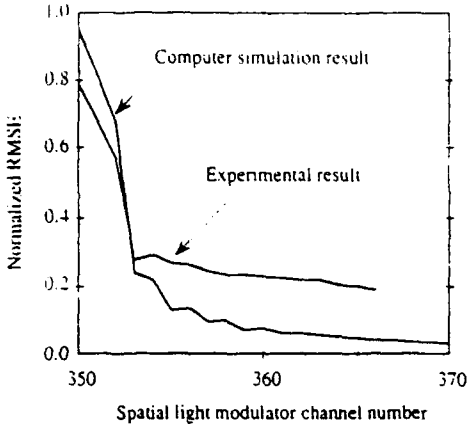


Fig. 10 Root-mean-square error for the 468-ns pulse.

terferer to a short pulse, performed the excision, and analyzed the output pulse shape. We studied the distortion of short pulses produced by the removal of various frequency components of pulses having various widths and analyzed the resulting distortion at baseband. These experiments showed that the distortion is most severe when the excision is performed in the mainlobe of the sinc function, while the distortion decreases rapidly when the excision is in the sidelobes of the pulse. The response time could be improved from its present value of about 5 ms by increasing the CCD array readout rate and the speed of the postprocessing electronics.

Acknowledgment

This work was supported by grants from the U.S. Army Research Office and from Rome Air Development Center.

References

1. J. Erickson, "Linear acousto-optic filtering with heterodyne and frequency-plane control," Ph.D. Thesis, Stanford University Electrical Engineering Dept. (June 1981).
2. J. Erickson, "Optical excisor performance limits versus improved signal detection," *Proc. SPIE* **639**, 232–239 (1986).
3. J. N. Lee, N. J. Berg, M. W. Casseday, and P. S. Brody, "High speed adaptive filtering and reconstruction of broad-band signals using acousto-optic techniques," *Proc. 1980 Ultrasonics Symp.* pp 488–492 (1980).
4. P. J. Roth, "Optical excision in the frequency plane," *Proc. SPIE* 352, 17–23 (1982).
5. G. W. Anderson, F. J. Kub, R. L. Grant, N. A. Papanicolaou, J. A. Modolo, and D. E. Brown, "Programmable frequency excision and adaptive filtering with a GaAs/AlGaAs/GaAs heterojunction photoconductor array," *Opt. Eng.* **29**(10), 1243–1248 (1990).
6. R. W. Brandstetter and P. G. Grieve, "Excision of interference from radio signals by means of a recursive optical notching filter," *Opt. Eng.* **29**(7), 804–815 (1990).

Reeder N. Ward received a BSEE degree from West Virginia University in 1988. He received the MSEE degree from North Carolina State University in 1991. While at North Carolina State University, he modified and performed experiments on an acousto-optic signal processor located at the Photonics Center at Rome Air Development Center. Presently he is employed at the Harris Corporation in Melbourne, Florida.

Anthony VanderLugt received BSEE and MSEE degrees from the University of Michigan, Ann Arbor, and Ph.D. and D.Sc. degrees from the University of Reading, England. He was a member of the research staff at the Institute of Science and Technology at the University of Michigan from 1959 to 1969 where he worked on optical spatial filtering techniques, Fourier optics, and applications of holography. In 1969 he joined the Harris Corporation as director of research at the Electro-Optics Center in Ann Arbor. From 1973 to 1979 he managed research and development groups at the Government Systems Division of Harris in Melbourne, Florida. From 1979 to 1986 he was a senior scientist in the Advanced Technology Department where his major area of interest was optical signal processing. In 1986 he joined the faculty at North Carolina State University as professor of electrical and computer engineering; he teaches courses on optical signal processing and continues his research work in this field. VanderLugt is a Fellow of the OSA and SPIE and a member of IEEE, Sigma Xi, Tau Beta Pi, and Eta Kappa Nu. He was the American editor for *Optica* from 1970 to 1974 and a member of the advisory board from 1974 to 1979. He is currently a topical editor for *Applied Optics*.

APPENDIX E

MULTICHANNEL ACOUSTO-OPTIC CROSSBAR SWITCH

Reprinted from Applied Optics

Vol. 30, pp. 4245-4256 10 October 1991

Multichannel acousto-optic crossbar switch

Dan Owen Harris

We analyze an acousto-optic crossbar switch architecture that can be used to implement an $N \times N$ point-to-point switch with just N hardware complexity. In our analysis, we determine that insertion loss and cross talk are minimized if we place the output ports in the diffraction far field of the acousto-optic cell. Using this result, we develop an optimum switch design based on Fourier optics: a Fourier transform lens is used both to scale the output beams for efficient coupling to the output ports and to provide a necessary optical fan-in from input to output ports. We demonstrate the performance of switch configurations using single-mode fiber input ports in conjunction with single-mode fiber, multimode fiber, and photodiode output ports.

Key words: Crossbar switch, photonic switch, acousto-optics.

I. Introduction

Switching networks are an important part of a more general data communications technology. As we increase the power of switching networks, communications systems that service a large number of input and output terminals become more flexible and, in effect, more useful. In recent years, the need has increased for switching systems that serve a larger number of high-capacity terminals, and the limitations of electronic switching techniques have 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, exhibiting excellent information processing and carrying capabilities that alleviate some of the problems encountered with presently used electronic switches.

An $N \times N$ point-to-point crossbar switch based on the deflection of optical beams can be constructed with just N deflectors. Deflecting architectures also provide the potential for small insertion loss and low cross-talk levels, even for large N. Given this high level of performance, deflecting architectures appear to be one of the best means of implementing photonic switches. Acousto-optic technology is a powerful and relatively mature technology that can be used to deflect optical beams. Switching architectures that are based on acousto-optic deflection have been proposed[1–3] and demonstrated.[4,5]

In Section II we briefly review the basic acousto-optic switching architecture that we have proposed,[3] while Section III contains descriptions of relevant performance parameters. We begin Section IV with an analysis of the best-case diffraction character for the acousto-optic switch; the results of this analysis indicate that Fourier domain diffraction produces optimum performance, so we dedicate the remainder of Section IV to the analytical characterization of a Fourier domain switch. In Section V we discuss the practical performance capabilities of state-of-the-art multichannel acousto-optic cell technology and then demonstrate switch performance for a particular acousto-optic cell in Section VI.

II. Architecture

A conceptual drawing of the basic acousto-optic switching architecture is show in Fig. 1. In this configuration, light is delivered to the switch by input ports arranged in a linear array. Light from each input port is collimated along the z axis, and each collimated beam is incident at the Bragg angle on one channel of a multichannel acousto-optic cell. For a point-to-point crossbar switch, the acoustic waves are created by the application of monotone rf signals to each of the piezoelectric transducers; these acoustic waves propagate along the x axis in each channel. The acoustic wave deflects a large portion of the incident light at an angle in the x–z plane that is proportional to the rf addressing frequency. Therefore, to access a given output port, we tune the rf signal to provide the proper deflection angle. Light reaches the output

The author is with the Department of Electrical and Computer Engineering, North Carolina State University, Daniels Hall, Raleigh, North Carolina 27695-7911.

Received 12 September 1990.
0003-6935/91/294245-12$05.00/0.

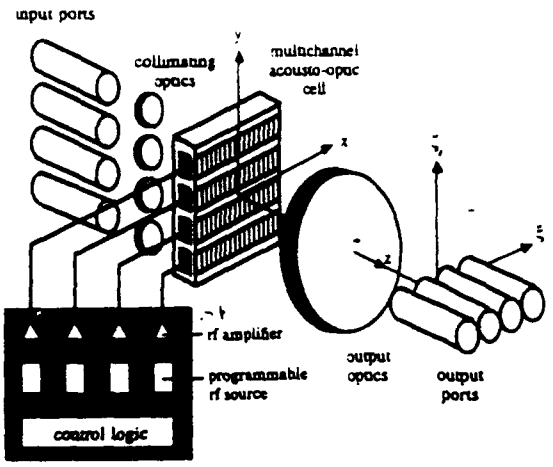


Fig. 1. Basic configuration of an acousto-optic photonic switch.

ports by passing through an optical system that directs the optical power from each vertical input port position onto the horizontal output axis.

In our point-to-point architecture, we stipulate that the rf signals are to be supplied to each channel by a single programmable rf source. Since only one acoustic channel and one rf source are required per input port and there is no path contention, this point-to-point switch is nonblocking with N hardware complexity. As N becomes large, the architecture can provide a greatly reduced complexity compared with conventional switching technologies, which require $O(N^{3/2})$ cross points for a nonblocking point-to-point switch.[6] Our approach is similar to the strategy taken by Wilson *et al.*[4] but in sharp contrast to the design proposed by Stephens *et al.*,[5] which uses N fixed oscillators. When fixed oscillators are used, they must be connected to the acousto-optic cell transducers through some type of electrical switching fabric, nullifying the hardware complexity advantage of the deflecting architecture. The architecture used by Stephens *et al.* has N^2 hardware complexity but does permit arbitrary signal fanout (multicasting). We have shown that the programmable source approach can be used to implement arbitrary signal fanout with just $O(N \log N)$ hardware complexity.[7]

III. Switch Performance

A. Insertion Loss

Losses in an acousto-optic switch are caused primarily by reflections, acousto-optic cell diffraction efficiency, and coupling inefficiencies at the output ports. Reflection losses at most optical interfaces do not exceed 0.2 dB, and these losses can be reduced further through the use of antireflection coatings. Maximum acousto-optic Bragg diffraction efficiencies for cells with moderate rf bandwidths ($W \approx 100$ MHz) range from 50 to 90%, while the maximum diffraction efficiencies for larger bandwidth cells ($W \approx 1000$ MHz) are ~10%.[8]

Coupling efficiency at the output ports is highly application dependent. When single-mode fibers are used to supply light at the input and collect light at the output of an $N \times N$ crossbar switch, the limited numerical aperture (NA) of the output fiber will cause the average coupling efficiency at the output ports to be no better than $O(1/N)$. This fan-in loss is present with any type of crossbar interconnection between similar waveguides and can be explained using arguments based on conservation of radiance.[9]

To overcome the dependence of coupling efficiency on N, we can use a photodiode to collect the light at the output port. Since a photodiode has a nearly unlimited NA, the coupling efficiency is identical for almost any approach angle. A convenient way to use photodiodes to collect light is to incorporate them into a multielement array. In some cases, however, we may find that cross talk among photodiode array elements is prohibitively high. In such cases, we may use large-core, large-numerical-aperture multimode fibers to collect light at the output ports and then carry the optical energy to discrete photodiodes.

An additional factor that can significantly affect output port coupling efficiency is the amplitude profile of the acoustic wave. Acoustic-wave propagation is subject to the same type of diffraction effect that occurs in optical propagation,[10] and the nonuniformity of the acoustic-beam profile distorts the output spot and tends to decrease coupling efficiency.

B. Cross Talk

In an acousto-optic photonic switch, leakage of optical power into inappropriate output ports is due to optical and acoustical as well as electrical effects. Optical cross talk is caused by the diffraction-induced spreading and eventual overlapping of light beams along the output axis. This diffractive spreading depends on the input beam profile as well as characteristics of the optical system between the acousto-optic cell and the output axis.

Acoustic cross talk occurs because the acoustic waves diffract and spread into adjacent channels. In essence, acoustic energy from a primary channel leaks into adjacent channels, deflecting some of the light in these adjacent channels to the output port that corresponds to the primary channel. Most multichannel acousto-optic cells are designed so that acoustic cross-talk levels are at least 30–40 dB below the primary acoustic-wave levels.[11,12]

Electrical cross talk originates from the crosscoupling of rf signals among the transducers of the multichannel cell. The primary applied signal and cross-coupled rf energy to a transducer produce a multifrequency electrical input to that channel. The presence of these multiple frequencies generates cross talk at other ports that correspond to the extraneous frequencies present in the channel and also at ports that represent frequencies that are linear combinations of the primary frequency and the extraneous frequencies. The latter cross talk is produced by intermodulation beams that result from multiple deflections within the channel.[13] As noted above, in applications where photodiode arrays are used, electri-

cal cross coupling among array elements can be a significant source of cross talk.

Assuming the incoherent addition of light that originates from the different input ports in the switch, we express the signal-to-cross-talk ratio (SCR) for a direct detection system as

$$\mathrm{SCR} = \frac{P_{i,k}}{\sum_{i \neq k} P_{i,k}}, \tag{1}$$

where $P_{i,k}$ represents the optical power collected at output port i that is intended for port k. Cross talk at a given output port from the various optical, acoustical, and electrical effects will also originate from different input ports, which implies that cross talk from these three effects will add incoherently. Therefore we analyze the three sources of cross talk separately and calculate the overall SCR, using the equation

$$\mathrm{SCR} = \left[\frac{1}{\mathrm{SCR}_o} + \frac{1}{\mathrm{SCR}_a} + \frac{1}{\mathrm{SCR}_e}\right]^{-1}, \tag{2}$$

where SCR_o, SCR_a, and SCR_e are the SCR's resulting from optical, acoustical, and electrical phenomena, respectively.

C. Signal Bandwidth

The fundamental cause of distortion in the acousto-optic switch is dispersion within the optical elements. For a switch that uses standard lenses and free-space propagation, the primary source of dispersion-related distortion is the acousto-optic cell itself. For operating wavelengths in the single-mode fiber-optic range of 1300–1550 nm, we find that typical acousto-optic cells will support signal bandwidths in the terahertz range.[7]

D. Reconfiguration Time

There are two basic components to reconfiguration time: the setup time required for the rf sources and the more fundamental acousto-optic cell reconfiguration time. The reconfiguration time of the acousto-optic cell is equivalent to the transit time, i.e., the time required for one point on the acoustic wave to traverse the full diameter of the input beam. The minimum switch reconfiguration time is therefore given as the sum of the rf source reconfiguration and acousto-optic cell transit times and is typically of the order of microseconds.

IV. Optimal Diffraction Character

We examine the effect of optical diffraction on switch performance and, based on our findings, propose a switch design that minimizes performance degradation resulting from these effects. We also derive design equations that identify trade-off relationships among the number of ports, transit time, and various acousto-optic cell and optical system parameters.

A. Minimal Diffraction-Degraded Performance

Both coupling efficiency and optical cross talk are dependent on the optical diffraction between the acousto-optic cell and the output axis. In analyzing the dependence of coupling efficiency and cross talk on diffraction, we initially model the system with no optical elements between the input and output ports; in this manner, we can determine the optimal free-space diffraction character and then design our optical system to produce this diffraction character.

To model light collimated from a single-mode fiber input port, we use a class of truncated Gaussian aperture functions that have amplitudes given by

$$a(x) = \mathrm{rect}\left[\frac{x}{L}\right] \exp\left[-2A\left(\frac{x}{L}\right)^2\right], \tag{3}$$

where L is the width of the collimated input beam aperture along the x axis in Fig. 1 and A determines the intensity of the beam at the two truncation points. Note that the beams must be truncated due to the finite extent of elements in the optical system. We use an efficient numerical technique based on fast Fourier transform methods to compute the diffraction patterns[7] and calculate the coupling efficiency and optical SCR's for the case of a photodiode array at the output ports. We chose to represent the output ports by photodiode array elements because such an analysis provides insight into the general effect of diffraction on loss and cross talk while allowing us to calculate coupling efficiency and cross talk by a relatively simple trapezoidal numerical integration[14] of the appropriate intensity distribution over the extent of each of the output ports.

Additional parameters necessary for the diffraction model are shown diagrammatically in Fig. 2. To make our analysis more general, we normalize the relevant parameters as follows: $w_n = w/\xi_0$, where $\xi_0 = z\lambda/L$ is the width of a Rayleigh resolution element along the output axis; $s_n = s/w$ is the reciprocal of the output port duty cycle; and $z_n = z/D_m$, where $D_m = L^2/1.25\lambda$ is the z-axis transition point from near-field to far-field diffraction for a rectangular aperture.[15] Accordingly, the product $w_n s_n$ represents the number of resolvable

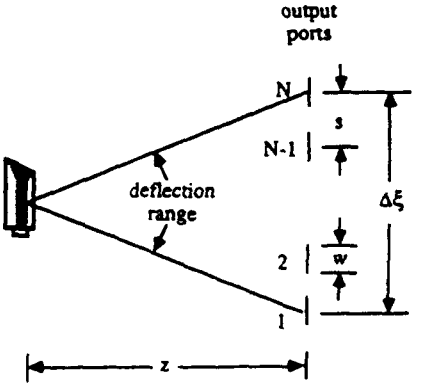


Fig. 2. Limited deflection range of an acousto-optic cell and containment of the output ports within it.

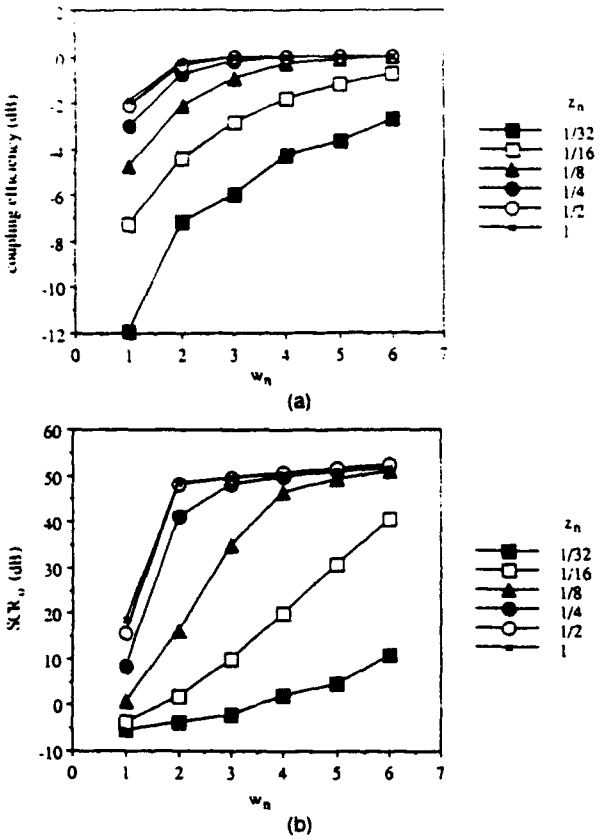


Fig. 3. Plots of calculated (a) coupling efficiency and (b) SCR_o versus the normalized port width for various values of the normalized propagation distance. In these plots, the normalized port separation $s_n = 2$.

elements contained within the output port separation.

In Figs. 3(a) and 3(b), we plot coupling efficiency and worst-case SCR_o versus w_n at the center port of a three-output-port switch for various values of z_n. By using three ports, we consider only a nearest-neighbor contribution for optical cross talk. In this example, we use $A = 8$ and $s_n = 2$; calculations using other values of A and s_n produced similar results.

From the figure we see that when z_n is small (i.e., when we are in the near-field or Fresnel diffraction domain), both the coupling efficiency and SCR for a fixed normalized port width increase as we increase the propagation distance. This occurs when the output axis is close to the cell, because the output port width and separation will be much less than the aperture width. Therefore the output port will not be wide enough to capture the output light efficiently, while the proximity of the ports implies that portions of the light from the main lobes of beams deflected to the two exterior ports will fall upon the center port, yielding high cross-talk levels. The main-lobe width remains fairly constant in the Fresnel domain; thus increasing the normalized propagation distance within the Fresnel domain will increase both the actual port width and separation, resulting in increased collection efficiency and lower cross talk.

As we increase z_n further, we find that the coupling efficiency and SCR_o for each w_n settle to steady-state maximum values; this coupling efficiency and cross-talk character is typical of far-field or Fourier domain diffraction. In the Fourier domain, the basic signature of the diffraction pattern remains the same, but the output axis scale of the pattern increases with z_n at the same rate as the actual port width and separation. Since the diffraction pattern signature is fixed and the collection aperture, port separation, and diffraction pattern scale all increase at the same rate, we do not expect any change in coupling efficiency or optical cross talk as we increase z_n.

In a second diffraction analysis, we determine the best value of A for our system. Figure 4 shows plots of coupling efficiency and optical SCR versus normalized port width for Fourier domain diffraction using Gaussian apertures with $A = 2$, 4, 8, and 16. The output port configuration and duty cycle are identical to those used in the propagation distance analysis. From Fig. 4(a), we see that in general the coupling efficiency decreases as A increases. In Fig. 4(b), however, we observe that the maximum SCR_o increases as we increase A.

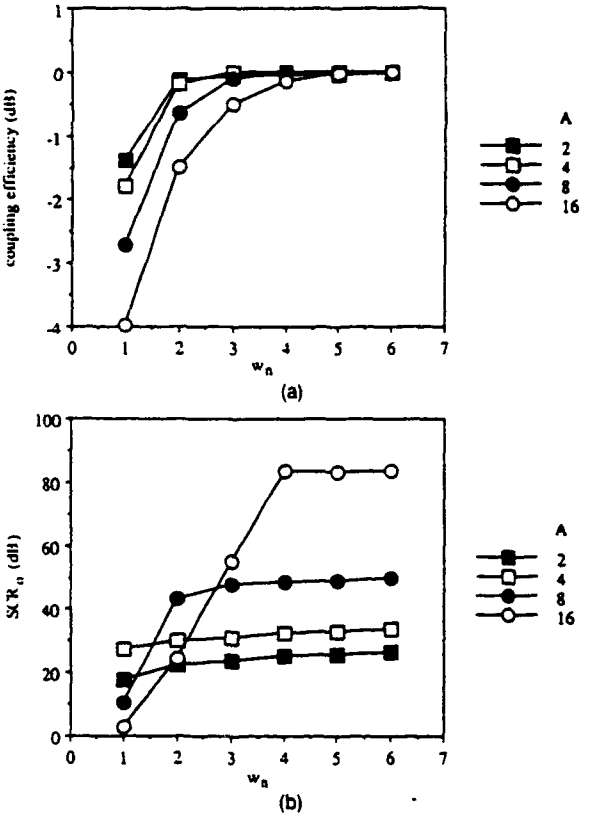


Fig. 4. Plots of calculated (a) coupling efficiency and (b) SCR_o versus the normalized port width for various values of aperture weighting parameter A. In these plots, the normalized port separation $s_n = 2$.

To keep the optical cross talk below typical acoustic cross-talk levels, we require SCR_o to be well in excess of 30 dB. Given the data in Fig. 4(b) we see that, to satisfy this optical cross-talk specification, we should truncate our input beam at intensity levels corresponding to values of $A = 8$ or $A = 16$. Suppose that we desire a coupling efficiency of no less than -0.5 dB (which is negligible relative to the best-case diffraction efficiency losses of 1–3 dB); Fig. 4(a) indicates that we require $w_n \geq 2$ to achieve this objective regardless of the value of A. The SCR_o plot in Fig. 4(b) shows that the highest SCR_o at this minimum value of $w_n = 2$ is produced by an aperture with $A = 8$ and that the value of SCR_o at this point exceeds 40 dB. Although we could obtain even better coupling efficiency and SCR_o by using $A = 16$ and $w_n = 4$, the normalized output port width is larger than the minimal value of $w_n = 2$. As we verify empirically in Subsection IV.B, increasing w_n allows fewer ports to be placed within the deflection range for our fixed s_n. Since the coupling loss and SCR are acceptable for $A = 8$ and $w_n = 2$, a truncation parameter of $A = 8$ is preferred.

The results presented above suggest that we obtain the best combined coupling efficiency and cross-talk performance if we use Fourier domain diffraction and a Gaussian aperture truncation parameter of $A = 8$. By using a configuration of this type, we can expect to obtain coupling efficiencies (neglecting NA effects) that are better than -1 dB, with worst-case SCR_o exceeding 40 dB.

B. Fourier Domain Optical Cross-Talk

To understand better the nature of optical cross talk for our chosen diffraction domain and aperture truncation, we calculated values of optical cross talk in the Fourier domain using $A = 8$ for various combinations of output port count and normalized output port width and separation. Some of the more important results are summarized in Table I.

First, we observe that the cross-talk levels remain fairly constant for a given value of the product $w_n s_n$, which, as stated above, represents the number of resolution elements contained within the port separation. Since the number of resolution elements contained within the entire deflection range, and in effect the information capacity, is equal to the time–bandwidth product (TW) of the acousto-optic cell system, we find that

$$N = \frac{TW}{w_n s_{n,\,SCR_o}} \tag{4}$$

where $w_n s_{n,\,SCR_o}$ is the number of resolution elements in the output port separation required to maintain the desired SCR_o. Equation (4) is intuitively satisfying, since the maximum number of output ports is proportional to the information capacity of the system.

Next, we find that as $w_n s_n$ increases from 1 to 4, the SCR's increase dramatically, ultimately reaching values above 40 dB. The improvement in the optical SCR gained by increasing $w_n s_n$ beyond 4, however, is much less pronounced. Since increasing $w_n s_n$ beyond 4 affords little additional cross-talk advantage while serving to increase the required TW needed for a given number of output ports, we conclude that the optimum value of $w_n s_n$ is ~ 4.

Finally, we see that the cross-talk levels calculated for $N = 3$ are similar to those calculated for $N = 301$. This means that for any value of N most of the optical cross talk comes from the nearest-neighbor output ports. Therefore

$$SCR_o \approx \frac{P_{n,n}}{P_{n,n-1} + P_{n,n+1}} \tag{5}$$

provides an accurate order of magnitude approximation for optical cross-talk levels in the switch, implying that the SCR_o is essentially independent of N. Additionally, expression (5) legitimizes the generality of results given in Figs. 3 and 4, where we consider only nearest-neighbor optical cross-talk contributions.

Table I. Calculated Signal-to-Cross-Talk Ratios (in decibels) for Various Numbers of Output Ports and Normalized Values of Port Width and Separation

		S_n			
w_n	N	1	2	4	8
1	3	5.0	23.0	48.4	53.1
	301	5.0	22.9	46.9	51.5
2	3	22.4	48.1	52.6	—
	301	22.3	46.5	50.9	—
4	3	46.8	50.9	—	—
	301	44.9	49.1	—	—
8	3	49.1	—	—	—
	301	46.9	—	—	—

C. Fourier Domain Switch Design

We can induce Fourier diffraction at a reduced propagation distance from the acousto-optic cell by using a simple lens.[16] We can also scale the output beams to match the physical size of the output ports using the same lens. In the following sections, we derive equations that govern the design of a switch that uses a Fourier transform lens system in the output optics.

1. *Output Port Capacity*

Figure 5 contains a functional diagram for the deflection plane of a Fourier domain acousto-optic switch. The two lenses and acousto-optic cell combine to produce a spatially offset image of the input device on the output axis, where the magnitude of the magnification is $M_\xi = F_2/F_1$.[17] Also, the width of a Rayleigh resolution element along the ξ axis is $\xi_0 = \lambda F_2/L$, and

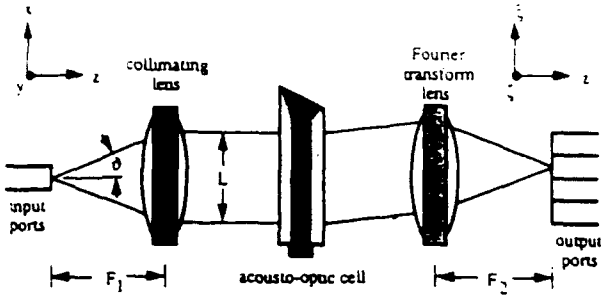


Fig. 5. Top view of a Fourier domain acousto-optic switch.

the overall deflection range is given by

$$\Delta\xi = \frac{\lambda W F_2}{v}, \tag{6}$$

where v is the velocity of sound in the acousto-optic cell.[18] From Fig. 2 we see that the number of ports that fit within the deflection range is $N = \Delta\xi/s + 1$. We solve this expression for $\Delta\xi$ and substitute the result into Eq. (6) to find that

$$F_2 = \frac{v(N-1)s}{\lambda W}. \tag{7}$$

The collimating lens focal length F_1 is dictated by the desired aperture width and the numerical aperture (NA) of the input port device. For the purposes of this analysis, it is convenient to quantify the NA of our input port as sin ϑ, where ϑ is the angle at which the far-field intensity from the input port has decreased to e^{-2} of the maximum. Since we desire $A = 8$ truncation, we need to find the e^{-8} intensity point on the input beam. The distance to the e^{-8} point on a Gaussian beam is exactly twice as far from the maximum intensity as the e^{-2} point; thus the expression for the collimating lens focal length becomes

$$F_1 = \frac{L}{4 \sin \vartheta}. \tag{8}$$

We rearrange Eq. (8) to find that the ratio $L/(4 \sin \vartheta F_1)$ is equal to unity. Multiplying $\Delta\xi/s$ in our expression for N by this ratio, noting that $T = L/v$, and making use of the expression for $\Delta\xi$ from Eq. (6) and the expression for ξ-axis magnification, we can write the maximum number of output ports as

$$N = \mathrm{TW}\frac{\lambda M_\xi}{4s \sin \vartheta} + 1. \tag{9}$$

Here we see that as N grows large, the number of output ports is approximately proportional to the TW of the acousto-optic cell system. It is interesting that this is similar to the dependence we found in Eq. (4). Furthermore, using the information presented above, we can show that

$$\frac{4s \sin \vartheta}{\lambda M_\xi} = \frac{s}{\xi_0} \equiv w_n s_n, \tag{10}$$

which implies that, if s is the minimum port separation required to maintain a given SCR, and $N \gg 1$, Eq. (9) is in fact an analytical restatement of the empirically derived Eq. (4).

2. *Optical System*

The general optical system we propose for implementing a Fourier domain acousto-optic switch is shown in Fig. 6. Here we use Fourier transform optics not only to provide the proper diffraction character along the ξ axis but also to focus the collimated beams in the ζ direction onto the ξ axis. As with the rf source selection, our approach is similar to that taken by Wilson *et al.*[4] but significantly different from the system demonstrated by Stephens *et al.*[5] In the Stephens switch, N fibers are used for each output port—one for each input port in an $N \times N$ switch. Each of the N^2 output fibers requires its own gradient-index rod lens to focus the deflected light onto the fiber core, while the N output fibers representing a specific output port are connected by a guided-wave power combiner.

In our optical system, we first collimate the beam with the spherical lens L_1 so that the e^{-2} diameter of the input beam equals the transducer height H (this is sufficient for most of the input light to interact with the sound column). H is therefore related to the input side of the optical system by $H = 2F_1 \sin \vartheta$. We may be required to use a pair of cylindrical lenses (L_2 and L_3) between the collimating lenses and acousto-optic cell to expand the beam in the x direction. We choose the ratio of cylindrical lens focal lengths as

$$\frac{F_3}{F_2} = \frac{L}{2H}, \tag{11}$$

where, as before, L is the e^{-8} aperture width required to provide the necessary TW.

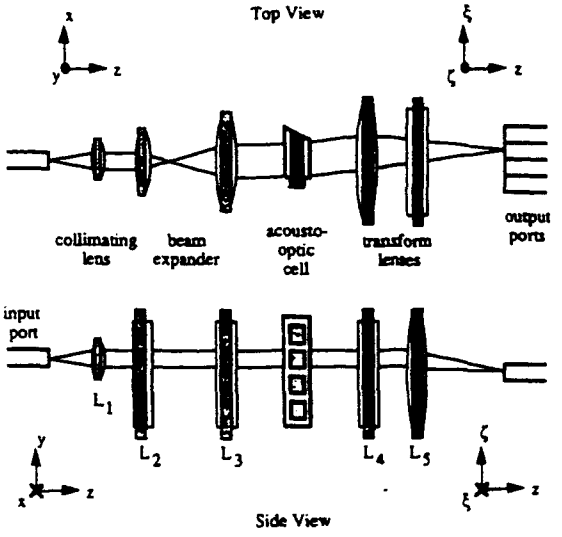


Fig. 6. General form of the optical system used in a Fourier domain acousto-optic switch. Note that only one input port is shown in the side view. This is done only to avoid clutter in the diagram. The remaining input ports are implied.

For the Fourier transform lens system, we may need two separate lenses to optimize the switch performance. The cylindrical lens in the x direction, L_4, provides the proper spot size in the ξ direction for efficient coupling to the output ports. The required focal length of the transform lens in the x direction is

$$F_4 = \frac{F_3}{F_2} F_1 M_\xi, \tag{12}$$

where M_ξ represents the ratio of the output-to-input port spot size that is needed for efficient coupling along the ξ axis. The focal length of L_5, the cylindrical lens in the y direction, is given by

$$F_5 = F_1 M_\zeta, \tag{13}$$

where M_ζ is the magnitude of the optimum spot size magnification along the ζ axis.

To determine the optimum value of M_ζ for single-mode fiber input-to-output ports, we first note that the sine of the worst-case angle of arrival to the output ports is half of the ratio of the overall input port array height to the transform lens focal length F_5. Using the fact that M_ζ is equal to F_5/F_1, we find that the worst-case angle of arrival to the output port is governed by

$$\sin \theta_{max} = \frac{\sin \vartheta (N-1)\gamma}{M_\zeta}, \tag{14}$$

where γ is the ratio of the center-to-center transducer spacing to the transducer height.

We represent the ζ dependence of the output beam amplitude as a Gaussian spot with the e^{-2} beam radius $M_\zeta \omega_0$, where ω_0 is the mode-field radius of both the input and output port fibers. We then account for an oblique approach angle θ by multiplying the Gaussian amplitude distribution by the phasor $\exp(-j2\pi\zeta \sin\theta/\lambda)$. To avoid NA selective losses, M_ζ is typically much larger than unity. Using overlap integrals,[19] we find that the fan-in efficiency for a given incidence angle is approximated by

$$\rho_f = \frac{2}{M_\zeta} \exp\left[-\left(\frac{\omega_0^2 k_0^2 \sin^2\theta}{2}\right)\right] \tag{15}$$

for $M_\zeta \gg 1$.

Expression (15) shows that the fan-in efficiency at the output ports decreases with increasing approach angle, implying that the worst-case fan-in efficiency occurs for the port with the largest approach angle. Substituting the expression for the worst-case approach angle from Eq. (14) into expression (15), we find that the poorest fan-in efficiency in our system is

$$\rho_f|_{\theta_{max}} = \frac{2}{M_\zeta} \exp\left[-\left(\frac{\omega_0^2 k_0^2 NA^2 \gamma^2 N^2}{2M_\zeta^2}\right)\right]. \tag{16}$$

The value of M_ζ that maximizes the worst-case fan-in efficiency is given by

$$M_\zeta = \omega_0 k_0 NA \gamma N \tag{17}$$

When we use this optimum value of M_ζ with most commercially available single-mode fibers,[20] fan-in efficiency is approximately $1/\gamma N$ for interior input ports, falling to $0.62/\gamma N$ at the largest approach angle. Here we effectively introduce coupling loss from spot size mismatch in the ζ direction to avoid NA selective exclusion of the exterior input ports.

As mentioned in Subsection III.A, we can use either multimode fiber or a photodiode array element to collect light at the output axis to increase fan-in efficiency. The multimode optical fibers we use have circular cores, and a good many photodiode elements are square. Therefore the optimum optical system design for these output port devices yields circularly symmetric output spots, where $M_\zeta = M_\xi$. Note that we must make sure that the NA of any multimode fiber that we use at the output ports is sufficient to accommodate the maximum approach angle given by Eq. (14).

An alternative means of increasing fan-in efficiency in a single-mode fiber switch design would be to incorporate a second reconfigurable deflector into the output optics. In this manner, the incoming light beams from each input port could be deflected along a single optical axis. Since all light approaches the output ports from a common direction, NA effects would be neutralized, allowing fan-in efficiency to approach unity regardless of the number of input ports.

V. Multichannel Acousto-Optic Cell Technology

Currently, two of the more powerful multichannel acousto-optic cell designs for switching applications are based on the longitudinal acoustic mode in TeO_2 and the shear acoustic mode in GaP. We briefly discuss the characteristics of specific multichannel cells that have been developed using each of these materials and the resulting performance of a switch that could be implemented with each technology.

The first cell is a 32-channel device based on the longitudinal TeO_2 configuration.[12,21] This device has an rf bandwidth of 200 MHz and a maximum TW of 1000. Using Eq. (4) and the optimum value of $\omega_a s_a = 4$, we find that this cell would support up to 250 output ports. At the full capacity of TW = 1000, the transit time will be ~5 μs. The maximum diffraction efficiency in any one channel of this cell is ~30% at λ = 633 nm for 1 W of applied rf power. To avoid performance degradation owing to thermal loading, however, we limit the maximum total power that can be applied to the cell at any one time to 10 W. This translates to maximum diffraction efficiencies of about 10% for a 32-input port switch. Acoustic and electrical SCR for the cell are better than 30 dB for switching applications.

A 64-channel cell constructed from GaP (Ref. 11) has a TW of 200. Thus we could fit about 50 output

translation stage. The fiber is moved along the ξ axis to emulate each output port in a multiport array.

Our photodiode array is composed of silicon p-i-n diodes that measure 500 × 500 μm and are arranged linearly with 550-μm center-to-center spacings. Electrical cross talk among the nearest-neighbor elements is specified to be less than 2%.

The rf sources that we use are based on direct digital frequency synthesizers. The output of each synthesizer is a monotone signal with a programmable frequency ranging from 0 to 110 MHz. The spurious noise for the synthesizer is specified to be 40 dB below the signal level over the entire frequency band, while the reconfiguration time is ~ 125 ns.

B. Single-Mode Fiber Output Ports

To the author's knowledge, these experiments represent the first reported demonstrations of free-space switches that use single-mode fiber at both the input and output ports. For the 4 × 4 single-mode system, we use an objective with a 7.2-mm focal length to collimate the light from the input fiber, while the two cylindrical lenses used to expand the beam in the x direction have focal lengths of 19 and 150 mm. The combined effect of these three lenses produces a collimated elliptical beam with an e^{-2} width of 9.09 mm in the x direction and a 1.15-mm e^{-2} height in the y direction. We align the input ports to the four innermost channels of the eight-channel acousto-optic cell and use a single spherical achromatic doublet with a focal length of 88.9 mm for the Fourier transform lens. This produces spots at the output plane that are elliptically shaped, having the same proportionality as the input beams, but the major and minor axes are rotated 90°. Note that the relatively large spot size in the ζ direction not only provides optimum fan-in efficiency as per Eq. (17) but also relaxes requirements for the ζ-axis alignment of the optical system; ξ-axis alignment is not critical because our rf sources are in effect continuously tunable, and the exact ξ-axis location of the deflected spot can be adjusted by slightly changing the rf addressing frequency. For the experiments described here, we chose nominal rf addressing frequencies of 59.7, 73.2, 86.7, and 100.2 MHz for output ports A–D, respectively, which represents an output port separation of nearly 180 μm.

The measured losses for the 4 × 4 single-mode switching element are given in Table II. Total insertion losses ranged from 15 to 19 dB. These results include 2 dB of reflective and absorptive loss in the optical system, a large portion of which is due to 1 dB of loss associated with the collimating objective. After taking diffraction efficiency into account, we found that the coupling loss for a deflected beam was ~ 2–2.5 dB in excess of that measured for the undeflected beam. This corresponds well to theoretically calculated acoustic-beam profile-induced losses of 1.8–2.0 dB.[14]

Table II. Measured Insertion Loss (in decibels) for a 4 × 4 Single-Mode Switching Element

Channel	Port A	Port B	Port C	Port D	Fan-in Loss
3	19.0	17.5	17.5	18.5	12.0
4	16.8	15.3	15.3	16.4	10.0
5	16.7	15.3	15.3	16.5	9.9
6	19.0	17.4	17.5	18.7	11.9
Diffraction efficiency loss	2.7	1.5	1.6	2.6	——

The major cause of insertion loss in this system was optical fan-in. The theoretical values for fanin loss calculated from expression (15) were 9.0 dB for channels 4 and 5 and 11.9 for channels 3 and 6, which basically match the measured values in Table II. The transform lens system could be made more efficient by increasing the duty cycle of the input beams along the y axis. At present, the ratio of input beam separation to height is 2.5, but in the limit where this ratio is reduced to unity (as could be approached with a material like GaP),[12] the theoretical coupling efficiencies are 5.8 and 7.5 dB for the interior and exterior channels, respectively.

Average cross-talk levels for the 4 × 4 single-mode switching element, calculated from the measured data for each of the 24 interconnection permutations, are compared with signal levels in Fig. 7. Overall, SCR's ranged from 36 to 43 dB. We determined that the most significant causes of cross-talk in the single-mode switch are extraneous beams resulting from acoustic waves produced by cross-coupled electrical energy between transducers as well as intermodulation from the interaction of the primary and cross-coupled acoustic waves.

We measured the reconfiguration time for the 4 × 4 single-mode switch to be 5.31 μs. This measured value is close to the theoretical prediction that includes 125 ns for synthesizer reconfiguration, 4.33 μs of transit time for an e^{-2} diameter of 18.2 mm, plus 475 ns of delay due to 2 mm of space between the transducer and the e^{-3} intensity point of the collimated beam.

The optical system for the 8 × 8 single-mode fiber switching element is basically the same as that used for the 4 × 4 version. The only difference is that the

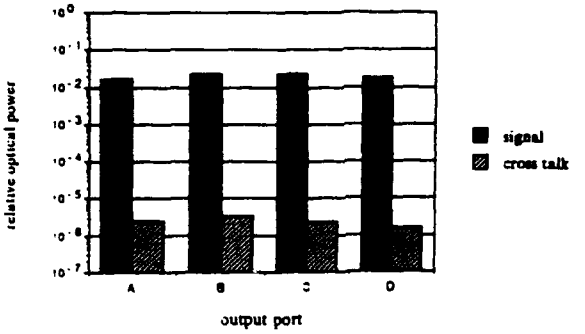


Fig. 7. Measured average output signal and cross-talk levels relative to the input signal for 4 × 4 single-mode fiber to the single-mode fiber switch.

Fourier transform lens for the 8 × 8 switch has a focal length of 200 mm. The rf addressing frequencies are 63, 68, 73, 78, 83, 88, 93, and 98 MHz for output ports A–H, respectively, resulting in an output port separation of nearly 150 μm.

The interconnection permutation we chose to characterize has the input port at channel 1 connected to output port H, 2 to B, 3 to C, 4 to E, 5 to A, 6 to G, 7 to F, and 8 to D. This particular permutation contains interconnections between various combinations of internally and externally located input and output ports. It also has some adjacent addressing frequencies applied to adjacent channels as well as some adjacent addressing frequencies applied to well-separated channels. This permutation produces a wide range of losses and cross-talk levels, providing a good indication of the upper and lower limits of 8 × 8 switch performance.

The measured signal and cross-talk levels for the representative interconnection permutation are given in Fig. 8. Insertion losses ranged from 21 to 25 dB, and SCR's were between 35 and 42 dB. On average, insertion losses were higher, while SCR's and reconfiguration time were the same as for the 4 × 4 switch.

The primary difference in the 8 × 8 and 4 × 4 single-mode switches is related to fan-in loss. In the 8 × 8 case, the measured fan-in loss was within 1.5 dB of the 14.6-dB theoretical value at the two interior channels (4 and 5), while being within 0.5 dB of the theoretical value of 17.9 dB at the exterior channels (1 and 8). These numbers represent a fan-in loss increase of ~6 dB compared to the 4 × 4 switch, which is basically the difference in measured insertion losses between the two configurations. Unfortunately, the fan-in loss experienced here was above optimum because of design limitations imposed by the cell aperture and rf bandwidth.[7] If we could increase either the TW of the cell or the output port duty cycle, we could theoretically achieve fan-in losses of ~11.5 dB for the interior channels and 15 dB for the exterior channels. This represents an increase of only 2–3 dB in overall insertion loss compared to the 4 × 4 switch.

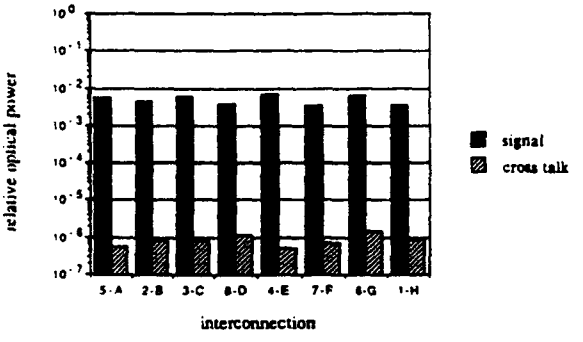


Fig. 8. Measured output signal and cross-talk levels relative to the input signal for a representative interconnection permutation of an 8 × 8 single-mode fiber to single-mode fiber switch

C. Multimode Fiber Output Ports

For multimode fiber output ports, we use a 9-mm focal length objective as a collimating lens and an achromatic transform lens with an 88.9-mm focal length. Here no x-axis beam expansion is necessary, and the resulting circular collimated beam has an e^{-2} diameter of 1.44 mm. The rf addressing frequencies are identical to those used in the 4 × 4 single-mode switch: 59.7, 73.2, 86.7, and 100.2 MHz for ports A–D, respectively, resulting in an output port separation of nearly 180 μm.

The measured insertion losses for the multimode fiber output port configuration are given in Table III. These losses ranged from 5 to 7 dB, which represents significant improvement compared with the single-mode experiment. The reflection and absorption loss of the system inclusive of the transform lens is only 1.3 dB, with the majority of that due to a 1-dB loss from the collimating objective; this is comparable with the loss in the single-mode system. Diffraction efficiency losses ranged from 2.3 to 3.5 dB, which are higher than for the single-mode output ports primarily because the multimode configuration has a larger y-direction collimated beam diameter, meaning that a smaller percentage of the incident light interacts with the acoustic wave.

The measured coupling loss to the fiber for the undeflected light beam was ~0.7 dB, close to the theoretical Fresnel reflection limit of 0.35 dB. The excess coupling losses for the deflected beams were ~1–1.5 dB, which corresponds well to calculated losses of 1.0–1.3 dB from acoustic-wave-induced output spot distortion. From Table III, we see that losses are fairly independent of the acoustic channel, which implies that NA selective losses are negligible. This insignificant NA selective loss is the primary reason for the reduction in insertion loss that results from replacing the single-mode fiber with multimode fiber at the output ports.

The average signal and cross-talk levels for a 4 × 4 switching element are given in Fig. 9. The SCR's ranged from about 31 to 35 dB, which is considerably worse than for the single-mode switches. The change

Table III. Measured Insertion Loss for Multimode Output Port Configuration (decibels)

	Port			
Channel	A	B	C	D
1	6.7	5.4	5.7	7.3
2	6.7	5.4	5.5	7.0
3	6.6	5.3	5.5	6.8
4	6.7	5.2	5.3	6.3
5	6.4	5.1	5.2	6.3
6	6.5	5.3	5.3	6.6
7	6.5	5.3	5.4	6.7
8	6.6	5.3	5.4	6.6
Diffraction efficiency loss	3.5	2.3	2.5	3.5

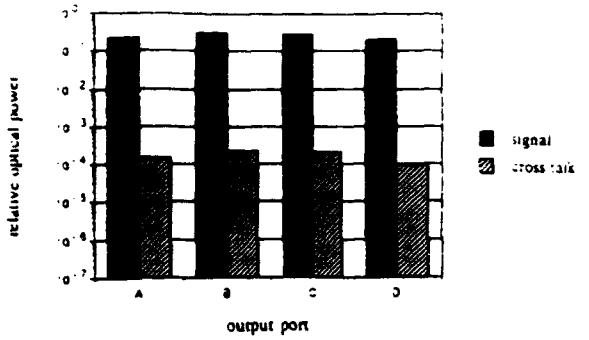


Fig. 9. Measured average output signal and cross-talk levels relative to the input signal for the 4 × 4 single-mode fiber to multimode fiber switch.

is due primarily to a relative rise in optical cross-talk levels because of the increase in the output port duty cycle. The electrical and acoustic cross-talk levels were practically the same as those for the 4 × 4 single-mode element. Note that the optical cross-talk levels are somewhat higher than what we expect from the analyses in Sections IV. The discrepancy exists because our analyses do not account for anomalous scattering effects that are present in the experimental optical system.

We measured a reconfiguration time of 1.03 μs for the multimode fiber output port switch, which coincides almost exactly with the theoretical prediction based on 125 ns of frequency synthesizer setup time, 676 ns of transit time, and 238 ns of delay corrsponding to 1 mm of space between the tranducer and the e^{-3} intensity point of the collimated beam.

These multimode output port experiments most closely resemble the two previous acousto-optic switch demonstrations we have cited. Insertion losses for our switch are markedly improved compared with those for both of the other switches. Stephens *et al.*[5] had average insertion losses ranging from 13 to 18 dB for a 4 × 4 switch between single-mode fiber input ports and 200-μm core diameter multimode fiber output ports. This is comparable with the performance we achieved in our single-mode-to-single-mode switch. In the demonstration by Wilson *et al.*,[4] an 8 × 8 switch between single-mode and 62.5-μm core multimode fiber yielded overall losses ranging from 17 to 40 dB. We should mention that the use of translation stages at the input and output ports represents a significant difference in the present configuration and the two systems demonstrated previously. However, no insertion loss advantage was gained at the output ports from using translation stages, since the only degree of freedom is in the ξ direction, and any ξ-axis misalignments that may occur in a multifiber system can be counteracted by rf tuning. Since we also only have one degree of freedom at the input ports, the only real advantage that the translation stage affords over the other systems is exact alignment to the acoustic channel in the y direction. Alignment in this direction, however, is not nearly as critical as in the x direction and should provide an advantage of no more than 1 or 2 dB in terms of overall loss.

Cross-talk levels for all three switch implementations were of the order of 30 dB below signal, but there are differences in reconfiguration times. The Stephens implementation had a measured reconfiguration time of ~1.2 μs dominated by a 1.1-μs delay due to 5-mm separation between the collimated beam and the transducer. Although the transit time for their configuration was only ~150 ns (as we define it), the rf bandwidth was 120 MHz, yielding a TW that is only 30% less than that used in our switch. This suggests that our design can achieve similar transit times for comparable rf bandwidths. Their rf source switching time of 10 ns was, however, significantly less than the 125 ns required for our frequency synthesizers. This reduction is advantageous when rf reconfiguration time is dominant (typically for high rf bandwidths). In the Wilson demonstration, voltage-controlled oscillators were used as programmable rf sources, and the source setup dominated the overall reconfiguration time of 10 μs. The switching time of their system could be reduced by using a different rf source.

D. Photodiode Array Output Ports

For the photodiode array output ports, we use a 7.2-mm focal length objective (to produce a collimated circular beam with a 1.15-mm diameter) in conjunction with a 300-mm focal length achromatic transform lens. The rf addressing frequencies are 51.5, 73.8, 86.1, and 98.4 MHz for ports A, B, C, and D, respectively, which provides the desired output port separation of 550 μm.

The relative signal and cross-talk levels for the photodiode array output configuration are given in Fig. 10. Overall, insertion losses ranged from ~3 to 4 dB with no evidence of NA selective loss. The losses here are lower than those for the multimode fiber output ports, because the diode collection areas are large enough that acoustic-wave profile beam spreading does not cause appreciable excess loss and also

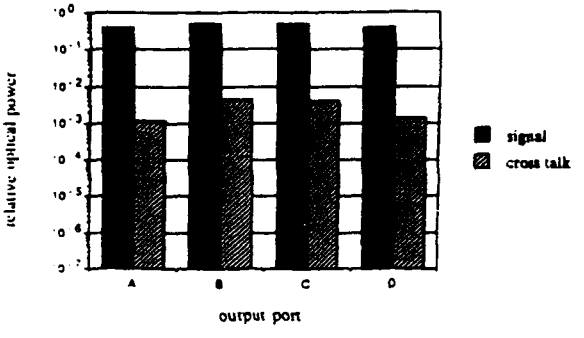


Fig. 10. Measured average output signal and cross-talk levels relative to the input signal for the 4 × 4 single-mode fiber to photodiode array switch.

because diffraction efficiencies are slightly higher owing to a reduction in the collimated beam diameter.

Average SCR's for the photodiode array configuration ranged from 20 to 25 dB, indicating that crosstalk levels are significantly higher than for the multimode fiber output ports. The reason for this increase is the dominant electrical cross talk among nearest-neighbor elements in the photodiode array.

VII. Conclusion

Deflecting photonic switch architectures are attractive because they can be used to construct an $N \times N$ crossbar switch with just N deflecting elements. We have determined that the best-case loss and cross-talk performance for an acousto-optic deflecting switch is achieved using far-field diffraction. Such a switch can be constructed by using a Fourier transform lens between the acousto-optic cell and the devices that collect the light along the output axis. With the Fourier domain switch, the transform optics provide the proper spot size for optimum coupling efficiency at the output ports as well as focusing the light from the vertical input port array onto the horizontal output axis.

We have experimentally verified the performance of several Fourier domain multichannel acousto-optic switch configurations. We achieved insertion losses ranging from 3 to 7 dB for connections from an array of eight single-mode fiber input ports to output ports consisting of either multimode fibers or a photodiode array. When single-mode fiber output ports are used, insertion losses jump to 15–19 dB for a 4×4 switch and 21–25 dB for an 8×8 switch. These increased losses are due primarily to fan-in loss, which could potentially be eliminated by placing a secondary deflector into the output optics of the switch. SCR's were well in excess of 30 dB for both the 4×4 and 8×8 single-mode fiber switches, while hovering around 30 dB for the multimode output ports. Cross-talk for our particular photodiode array configuration was only of the order of 20 dB below signal because of electrical crosstalk among the array elements. Reconfiguration times for the experimental switches were ~ 1 μs for the multimode output port configuration, increasing to 5 μs for the single-mode fiber output port designs.

This research was supported by the U.S. Army Research Office and BellSouth Enterprises, Inc.

References

1. A. VanderLugt, "Fresnel transforms and optical computing," in *Optical and Hybrid Computing*, H. H. Szu, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **634**, 51–56 (1986).
2. A. A. Sawchuk, B. K. Jenkins, C. S. Raghavendra, and A. Varma, "Optical crossbar networks," Computer **20**(6), 50–60 (1987).
3. D. O. Harris and A. VanderLugt, "Acousto-optic photonic switch," Opt. Lett. **14**, 1177–1179 (1989).
4. M. L. Wilson, D. L. Fleming, and F. R. Dropps, "A fiber-optic matrix switchboard using acousto optic Bragg cells," in *Components for Fiber Applications III and Coherent Lightwave Communications*, P. M. Kopera and H. R. Sunak, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **988**, 56–62 (1988).
5. W. E. Stephens, P. C. Huang, T. C. Banwell, L. A. Reith, and S. S. Cheng, "Demonstration of a photonic space switch utilizing acousto-optic elements," Opt. Eng. **29**, 183–191 (1990).
6. V. E. Benes, *Mathematical Theory of Connecting Networks and Telephone Traffic* (Academic, New York, 1965).
7. D. O. Harris, "Acousto-optic photonic switching," Ph.D. dissertation, North Carolina State University (University Microfilms, Ann Arbor, Mich., 1990).
8. A. Korpel, "Acousto-optics—a review of the fundamentals," Proc. IEEE **69**, 48–53 (1981).
9. J. W. Goodman, "Fan-in and fan-out with optical interconnections," Opt. Acta **32**, 1489–1496 (1985).
10. W. T. Maloney, G. Meltz, and R. L. Gravel, "Optical probing of the Fresnel and Fraunhofer regions of a rectangular acoustic transducer," IEEE Trans. Sonics Ultrason. **SU-15**, 167–172 (1968).
11. W. R. Beaudet, M. Popek, and D. R. Pape, "Advances in multi-channel Bragg cell technology," in *Advances in Optical Information Processing II*; D. R. Pape, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **639**, 28–33 (1986).
12. M. Amano and E. Roos, "32 channel acousto-optic Bragg cell for optical computing," in *Acousto-Optic, Electro-Optic, and Magneto-Optic Devices and Applications*, J. A. Lucero, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **753**, 37–42 (1987).
13. D. L. Hecht, "Multifrequency acoustooptic diffraction," IEEE Trans. Sonics Ultrason **SU-24**, 7–18 (1977).
14. P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, New York, 1972).
15. A. VanderLugt, "Fresnel transforms and Bragg cell processors," Appl. Opt. **24**, 3846–3859 (1985).
16. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, Calif., 1968).
17. E. Hecht and A. Zajac, *Optics* (Addison-Wesley, Menlo Park, Calif., 1979).
18. A. VanderLugt, "Bragg cell diffraction patterns," Appl. Opt. **21**, 1092–1100 (1982).
19. C. M. Miller, S. C. Mettler, and I. A. White, *Optical Fiber Splices and Connectors* (Marcel Dekker, New York, 1986).
20. E.g., Corning, Inc., product information sheet for PI-108 SMF single-mode fiber.
21. S.-C. Lin, "Crosstalk characteristics of multichannel acousto-optic Bragg cells," in *Advances in Optical Information Processing III*, D. R. Pape, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **936**, 76–84 (1988).
22. I. C. Chang, "Acousto-optic devices: material issues," in *Spatial Light Modulators, and Applications I*, U. Efron, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **465**, 55–65 (1984).

APPENDIX F

MULTICHANNEL ACOUSTO-OPTIC CROSSBAR SWITCH WITH ARBITRARY SIGNAL FAN-OUT

Reprinted from Applied Optics

Volume 31, Pages 1684-1686, 10 April 1992

function is flowing along the trajectory. In figurative terms this process can also be regarded as a reproduction of the picture of the flight of the wave train that is emitted by the laser.

References

1. N. Abramson, "Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena," Appl. Opt. **22**, 215–232 (1983); N. Abramson, "Light-in-flight recording 2: compensation for the limited speed of the light used for observation," Appl. Opt. **23**, 1481–1492 (1984); "Light-in-flight recording 3: compensation for optical relativistic effects," Appl. Opt. **23**, 4007–4014 (1984).
2. D. I. Staselko, Yu. N. Denisyuk, and A. G. Smirnov, "Holographic registration of a picture of temporal coherence of a wave train of a pulse radiation source," Opt. Spectrosc. **26**, 413–420 (1969).
3. Yu. N. Denisyuk, D. I. Staselko, and R. R. Herke, "On the effect of the time and spatial coherence of radiation source on the image produced by a hologram," presented at the International Symposium on Holography Applications, July 1970; Nouv. Rev. Opt. Appl. **1**, Suppl. 2, 4–5 (1970).

Multichannel acousto-optic crossbar switch with arbitrary signal fan-out

Dan Owen Harris and A. VanderLugt

The authors are with the Department of Electrical and Computer Engineering, North Carolina State University, Daniels Hall, Raleigh, North Carolina 27695-7911.

Received 29 August 1990.
0003-6935/92/111684-03$05.00/0

We propose an N × N acousto-optic switch architecture capable of arbitrary signal fan-out with O(N log N) hardware complexity. We also investigate the impact of signal fan-out on loss and cross talk.

Key words: *Crossbar switch, photonic switch, broadcast network, acousto optics.*

We have previously described the operation and performance of an acousto-optic photonic switching architecture.[1,2] In our architecture we rely on Bragg diffraction of light by monotone acoustic waves in a multichannel acousto-optic cell. Light that passes through an acoustic channel is efficiently deflected at an angle that is proportional to an applied radio frequency (rf), so that we steer light to an appropriate output port through proper selection of the rf. Simultaneous interconnections between a single input port and several output ports, or signal fan-out, are obtained by superimposing two or more acoustic waves within a single acoustic channel to provide the necessary multiple beam deflection. In this Technical Note we introduce an acousto-optic crossbar switching architecture that has arbitrary signal fan-out capability; we also point out ways in which the performance of this architecture deviates from our point-to-point crossbar switch and present experimental results that illustrate some of these differences.

As an example of the basic architecture, Fig. 1 shows a configuration for implementing a 4 × 4 switch with arbitrary fan-out capability. The top four acoustic channels of the cell have one programmable monotone rf source, and each of the channels is aligned to one of the input ports; these channels provide point-to-point deflection of the input beams to one of the four primary output ports (A–D), or to one of two auxiliary output ports. To implement signal fan-out from one of the input ports, we first deflect the light from that input to an auxiliary output port, which feeds the signal back to one of the two fan-out channels. These two fan-out channels have more than one programmable rf driver attached to their transducers, so we can create multiple tone acoustic waves to effect the desired signal fan-out. Note that either a photodiode or an optical fiber can be used to collect light at the auxiliary output ports: the output of the photodiode would be used to drive a laser source at the input of the fan-out channel, while the fiber can directly feed the optical signal back to the fan-out channel.

We construct an $N \times N$ crossbar switch with arbitrary signal fan-out by using the following algorithm. To our $N \times N$ point-to-point crossbar, we first add $\mathrm{int}(N/2)$ acoustic channels with two monotone programmable rf drivers and $\mathrm{int}(N/2)$ auxiliary output ports that are used to direct light to each of the new fan-out channels, where $\mathrm{int}(x)$ designates the integer nearest to but not exceeding x. These additional acoustic channels and output ports allow the switch to provide interconnection permutations that require fan-out to two output ports. To increase our capability to include fan-out to three output ports, we simply add a monotone programmable rf driver to each of the bottom $\mathrm{int}(N/3)$ fan-out channels. After three-port fan-out is implemented, we can further enhance our capability to four-port fan-out by adding another rf source to each of the bottom $\mathrm{int}(N/4)$ fan-out channels. To achieve arbitrary signal fan-out capability, we continue this process of adding a rf source to the bottom $\mathrm{int}(N/i)$ fan-out channels, for values of $i = 5, \ldots, N$.

After we have completed adding rf sources to the fan-out channels as prescribed in the previous paragraph, we are left with an $N \times N$ crossbar switch that requires $\mathrm{int}(3N/2)$ acoustic channels and output ports and has a rf source count given by

$$N_{RF} = \mathrm{int}\left(\frac{N}{2}\right) + \sum_{i=1}^{N} \mathrm{int}\left(\frac{N}{i}\right). \tag{1}$$

Since $\mathrm{int}(N) \le N$, we can approximate Eq. (1) by using

$$N_{RF} \sim N\left(\frac{1}{2} + \sum_{i=1}^{N} \frac{1}{i}\right). \tag{2}$$

which slightly overestimates the number of rf sources. The well-known series expansion for $\log_e(N)$ is written as[3]

$$\log_e(N) = \sum_{i=1}^{\infty} \frac{1}{i}\left(\frac{N-1}{N}\right)^i. \tag{3}$$

As N grows large, we find that the series expansion for $\log_e(N)$ and the summation in approximation (2) approach equality. Using this result for large N, we find that

$$N_{RF} \approx N\left(\frac{1}{2} + \log_e N\right). \tag{4}$$

Approximation (4) states that the required number of programmable monotone rf signal sources is $O(N \log N)$. Therefore, since the required number of acoustic channels

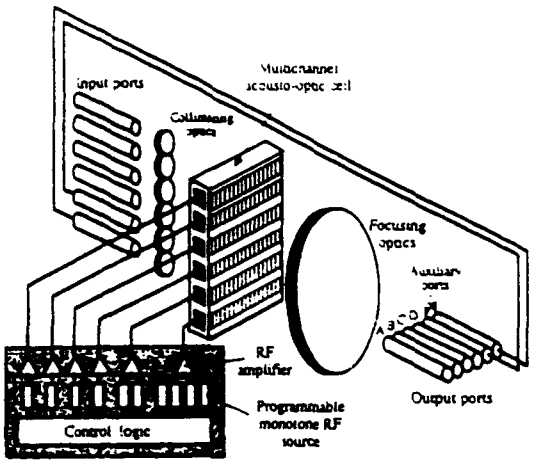


Fig. 1. Acousto-optic photonic switch architecture for arbitrary signal fan-out. The hardware complexity of this architecture approaches $O(N \log N)$.

is only $O(N)$, we can use our approach to implement a crossbar switch that is capable of arbitrary signal fan-out with just $O(N \log N)$ hardware components, well below the typical N^2 hardware complexity that is required to implement an equally powerful crossbar by using other types of electrical or optical switching architecture (for examples of N^2 complexity architectures, see Refs. 4–6).

The operation and performance of our fan-out architecture is similar to that of its point-to-point counterpart,[1,2] with a few exceptions. First, splitting an optical signal evenly over M output ports imposes a $1/M$ reduction in the power delivered to each port; also, when a fiber provides direct feedback to the fan-out channels, the optical signal must pass through the acousto-optic system twice, so insertion loss—excluding splitting loss—is at least twofold that of our point-to-point architecture. Next we note that reconfiguration time can also be longer for the fan-out architecture if we do not have a controller that can simultaneously reconfigure the necessary rf sources. If we must access one rf source at a time, the reconfiguration time will increase with M; however, the M dependence of overall reconfiguration time for our architecture is similar to that of any other fan-out switch. Finally, the nonlinear interaction of multiple frequencies in the fan-out channels causes additional signal losses and cross talk, the extent of which we detail in the remainder of this Technical Note.

Nonlinear effects that occur in multifrequency acousto-optic diffraction reduce diffraction efficiency below what is expected when power splitting alone is considered; this reduction in diffraction efficiency increases the overall insertion loss of the switch. When two equally strong acoustic waves are present in a single channel, the Bragg diffraction efficiency for each of the deflected beams is given by[7]

$$\eta' = \eta |J_1(\Delta\varphi)|^2. \quad (5)$$

In Eq. (5) η is the maximum monotone Bragg diffraction efficiency at the specified acoustic frequency, $J_1(\cdot)$ is the first-order Bessel function of the first kind, and $\Delta\varphi = \gamma(P_{RF})^{1/2}$, where P_{RF} is the RF power per frequency applied to the acousto-optic cell transducer and γ is constant for a given acousto-optic cell and fixed operating wavelength.[7]

Multifrequency acousto-optic diffraction also gives rise to extraneous deflected beams that result from nonlinear intermodulation effects.[7,8] In our switching application, the beams that can affect performance are typically those due to third-order intermodulation products. The deflection angles associated with these intermodulation products are linear combinations of the angles that correspond to the constituent acoustic waves. In switch designs where the output ports are equally spaced, some of the intermodulation beams deflect light directly to inappropriate output ports, which contributes to cross talk at those ports. The Hecht limit for the third-order intermodulation product power that results from two equal-strength monotone acoustic waves is written as[7]

$$I = \eta J_3(\Delta\varphi)^2, \quad (6)$$

where $J_3(\cdot)$ is the third-order Bessel function of the first kind and η is now the maximum monotone Bragg diffraction efficiency for the frequency corresponding to the intermodulation product.

We experimentally verified switch performance degradation that results from nonlinear acousto-optic interaction. In our experiment we characterize the diffraction efficiencies and intermodulation cross talk that occur from the application of two equally powerful monotone RF signals to an acoustic channel in a 4 × 4 switch within a single-mode fiber network. The apparatus we used in this experiment is shown in Fig. 2. Light at 633 nm, supplied to the switch by a single-mode fiber, is collimated by a 7.2-mm focal-length microscope objective. To achieve enough acousto-optic cell time–bandwidth product, the input beam is expanded in the x direction with a pair of 19- and 150-mm focal-length cylindrical lenses in a confocal arrangement. The resulting collimated elliptical beam is incident on the acousto-optic cell at the Bragg angle. The cell uses the longitudinal mode in TeO_2 and has a center frequency of 80 MHz and a bandwidth of 40 MHz; at 633 nm, $\gamma = 0.123/(\text{mW})^{1/2}$ for our cell. Light deflected from the cell is focused onto the output axis by a spherical lens having a focal length of 88.9 mm. We place a single-mode fiber on a translation stage along the output axis to collect the light at each of four possible output port positions.

In our switch design, we specify addressing frequencies of 59.7, 73.2, 86.7, and 100.2 MHz for output ports A–D, respectively; this results in a center-to-center spacing of 180 μm along the output axis. In our experiment we apply a two-tone RF signal to the cell with equal-strength frequency components at 73.2 and 86.7 MHz; this is the signal we would apply to access ports B and C simultaneously. Besides the two primary deflected beams, application of this RF signal also creates intermodulation beams that correspond to 59.7 and 100.2 MHz; light from these beams results in cross talk at ports A and D.

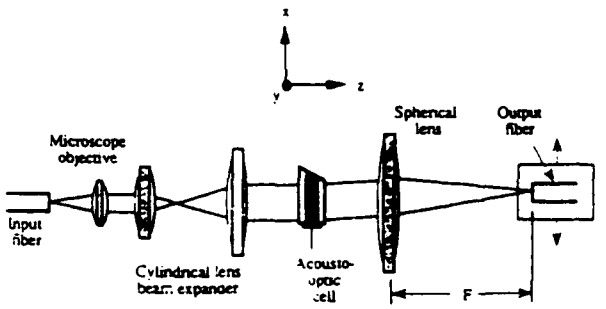


Fig. 2. Top view of the apparatus that was used in the two-tone diffraction efficiency and intermodulation experiments. Here, F is the focal length of the spherical lens, and Bragg incidence to the acousto-optic cell is implied.

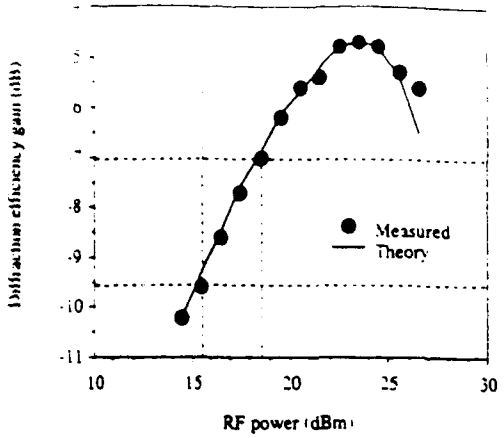


Fig. 3. Average measured diffraction efficiencies versus applied RF power for beams deflected to ports B and C by a two-tone RF signal with constituent frequencies of equal strength. Diffraction efficiency values are referenced to the maximum achievable diffraction efficiency for a monotone RF signal.

The results of diffraction efficiency and intermodulation level measurements are shown in Figs. 3 and 4. In Fig. 3 we plot the average of the two diffraction efficiencies for the beams deflected to ports B and C versus applied RF power and the corresponding theoretical predictions; these values are quoted in decibels and are referenced to the maximum diffraction efficiency that can be achieved when a single monotone signal is applied to the cell. Note that the measured efficiencies conform to theory to within a few tenths of a decibel. The largest efficiency is ~35% of our reference level, which occurs when our two-tone signal contains 23.5 dBm of RF power at each of the two frequencies. Figure 4 shows similar measurements and theoretical calculations for the average intermodulation cross-talk power at the two extreme ports A and D. The levels are quoted in decibels and are again referenced to the maximum diffraction efficiency achieved when a single monotone RF signal is applied to the transducer. In many instances, the measured intermodulation levels are ~3 dB above the Hecht limit; the intermodulation power in excess of our theory is most likely due to acoustic nonlinearities.[9]

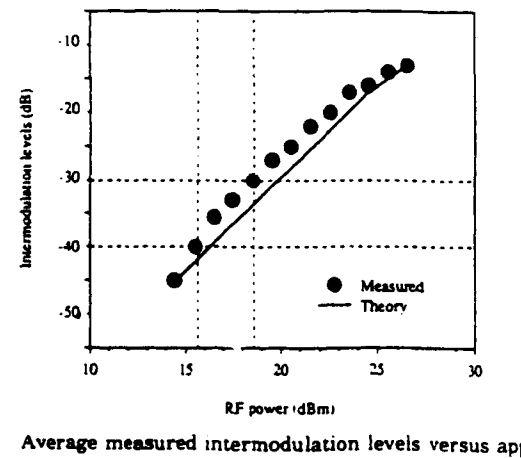


Fig. 4. Average measured intermodulation levels versus applied RF power; these intermodulation beams appear at ports A and D and result from the applied RF signals that were used to collect the data in Fig. 3. Intermodulation levels are referenced to the maximum achievable diffraction efficiency for a monotone RF signal

From Fig. 4 we find that as we increase the applied RF power, intermodulation beam levels quickly exceed the 30-dB cross-talk levels that we achieved in the point-to-point configurations. For an applied RF power of 18.5 dBm or less, the graphs of diffraction efficiency gain and intermodulation levels in Figs. 3 and 4, respectively, show that the intermodulation beams are at least 30 dB below the maximum monotone diffraction efficiency, while the two-tone diffraction efficiency is reduced by more than 7 dB relative to the monotone case. To maintain a 30-dB intermodulation signal-to-cross-talk ratio for a point to point interconnection at port A or D, our multicasting diffraction efficiency at ports B and C must therefore be at least 7 dB less than for the point-to-point case.

To determine the reduction in diffraction efficiency that is necessary to maintain a 30-dB intermodulation signal-to-cross-talk ratio for multicast interconnections, we must find the levels of applied RF power at which the ratio of diffraction efficiency gain to intermodulation level is equal to 30 dB or more. Using an iterative process, we find that the intermodulation beams are at least 30 dB below the two-tone diffraction efficiency at RF powers of < 15.5 dBm. Combining this information with the typical 2–3-dB diffraction efficiency losses we encounter in point-to-point configurations,[1,2] we find that, to maintain a 30-dB signal-to-cross-talk ratio for all interconnections in our multicasting application, the worst-case diffraction efficiency losses are 12 dB or greater than those of a switch without fan-out capability.

In conclusion, we have shown that a crossbar switch with arbitrary fan-out capability can be implemented by using acousto-optic technology with only $O(N \log N)$ hardware complexity. However, insertion loss and cross-talk characteristics of fan-out interconnections are degraded somewhat compared to a point-to-point switch. We note that it may be possible to alleviate increased cross talk by spacing the output ports at irregular intervals along the output axis, but further research is needed to determine the practicality of this cross-talk reduction strategy.

This work was supported by the U.S. Army Research Office and by BellSouth Enterprises, Inc.

Dan Owen Harris is currently with Siecor Corporation, 489 Siecor Park-HE, Hickory, N.C. 28603-0489.

References

1. D. O. Harris and A. VanderLugt, "Acousto-optic photonic switch," Opt. Lett. **14**, 1177–1179 (1989).
2. D. O. Harris, "Multichannel acousto-optic crossbar switch," Appl. Opt. **30**, 4245–4256 (1991).
3. W. H. Beyer, *Standard Mathematical Tables*, 27th ed. (CRC Press, Boca Raton, Fla., 1984).
4. B. E. Briley, *Introduction to Telephone Switching* (Addison-Wesley, Reading, Mass., 1983).
5. A. R. Dias, R. F. Kalman, J. W. Goodman, and A. A. Sawchuk, "Fiber-optic crossbar switch with broadcast capability," Opt. Eng. **27**, 955–960 (1988).
6. W. E. Stephens, P. C. Huang, T. C. Banwell, L. A. Reith, and S. S. Cheng, "Demonstration of a photonic space switch utilizing acousto-optic elements," Opt. Eng. **29**, 183–191 (1990).
7. D. L. Hecht, "Multifrequency acoustooptic diffraction," IEEE Trans. Sonics Ultrason. **SU-24**, 7–18 (1977).
8. I. C. Chang, "Acoustooptic devices and applications," IEEE Trans. Sonics Ultrason. **SU-23**, 2–22 (1976).
9. G. Elston, "Intermodulation products in acoustooptic signal processing systems," in *IEEE 1985 Ultrasonics Symposium Proceedings* (Institute of Electrical and Electronics Engineers, New York, 1985), Vol. 1, pp. 391–397

APPENDIX G

SAMPLING OF FRESNEL TRANSFORMS

Reprinted from Proceeding of the SPIE

Advances in Optical Information Processing IV
Volume 1296, Pages 246-252, April 1990

Sampling of Fresnel transforms

A. VanderLugt

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

Abstract

Domains other than the time or frequency domains arises naturally in coherently illuminated optical systems that produce Fourier transforms. We extend the well known result that N samples are sufficient to represent the information content in the object, image, and Fourier planes to include any Fresnel plane of a coherently illuminated optical system, provided that we use a specified sampling technique.

1. Introduction

In an optical Fourier transform system there is a continuum of Fresnel domains available for implementing processing operations between the object plane and the Fourier transform plane. 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 Fourier transform. The question arises as to whether N samples are sufficient to accurately measure the intensity in any of the Fresnel transform planes.

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 quantities in an optical system is that the sampling distance d_0 for a bandlimited object having a cutoff spatial frequency α_{co} is $d_0=1/(2\alpha_{co})$. If the object has length L, the number of samples required is $N=L/d_0$.

Consider the Fourier transform system of Figure 1. An object f(x) in the front focal plane P_1 of the lens is illuminated by a plane wave of coherent light. In the object plane P_1 all information

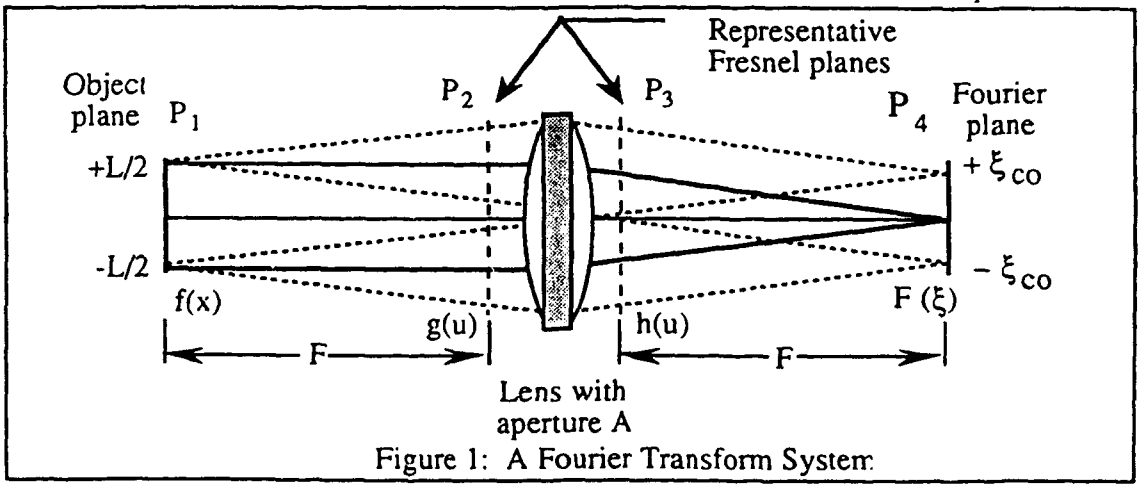


Figure 1: A Fourier Transform System

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 showed 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.[1] A related study showed that the packing density for a generalized imaging system is highest in either the object plane P_1 or the Fourier plane P_4, depending on the cutoff frequency α_{co}.[2] 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.

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 α_f in some Fresnel planes is at least as large as α_{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:

$$r(u) = \int_{-\infty}^{\infty} d(x) e^{-j\frac{\pi}{\lambda D}(u-x)^2}\, dx, \qquad (1)$$

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 $d(x)=\mathrm{sinc}(x/d_0)$ is that

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

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

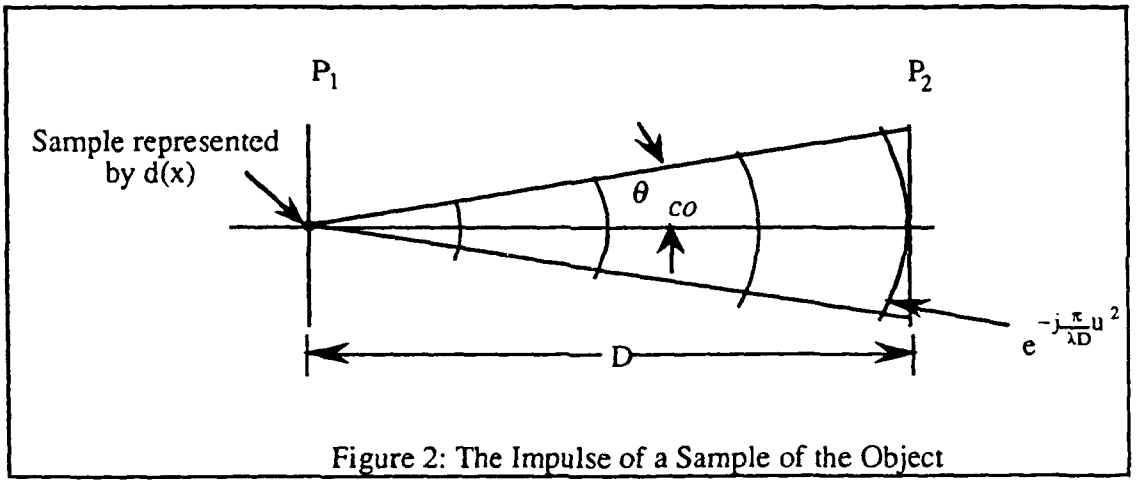


Figure 2: The Impulse of a Sample of the Object

the physical cutoff angle θ_{co} and the cutoff spatial frequency α_{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 (2). 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 $1 \le n \le (N-1)$, and centered on the optical axis at x=0.The object distribution is therefore

$$f(x) = d(x - nd_0/2) + d(x + nd_0/2), \qquad (3)$$

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

$$I(u) = 2[1 + \cos(2\pi n d_0 u / \lambda D)]. \quad (4)$$

From (4) we find that the spatial frequency α_f of the intensity, in a general Fresnel plane, due to two samples in the object plane is

$$\alpha_f = \frac{n d_0}{\lambda D}. \quad (5)$$

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 α_f is produced only by two closely spaced samples. Light rays from widely spaced samples do not overlap because θ_{co} has a finite value. 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, 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 $\alpha_f=L/\lambda F=\alpha_{co}$.

3. The Optimum Sampling Distribution for a Maximum Capacity System

As we have seen, the distribution of spatial frequencies throughout out the Fresnel plane is 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. 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 $|u| \approx d_0$.

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

$$\alpha_f(u) = \frac{L}{\lambda F}\left[1 - \frac{|u|}{L}\right]. \quad (6)$$

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

$$c(u) = 1 + \cos\left[\frac{\pi}{\lambda F}(L-u)^2\right] \quad (7)$$

as shown in Figure 3a. 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 (6).

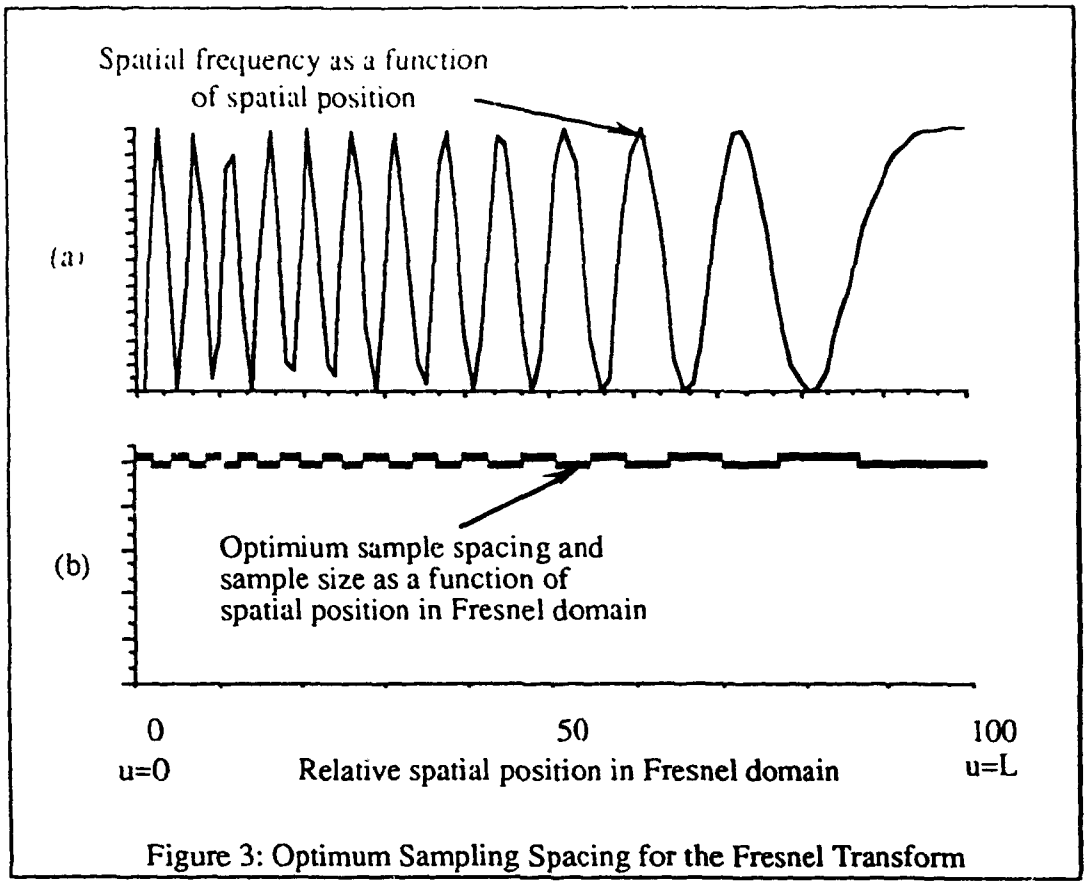


Figure 3: Optimum Sampling Spacing for the Fresnel Transform

The optimum sampling of the chirp function is found by applying the Nyquist sampling criterion that requires two samples per highest frequency. For the chirp function (7), 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 (7).

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 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 4 which 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 α_{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.

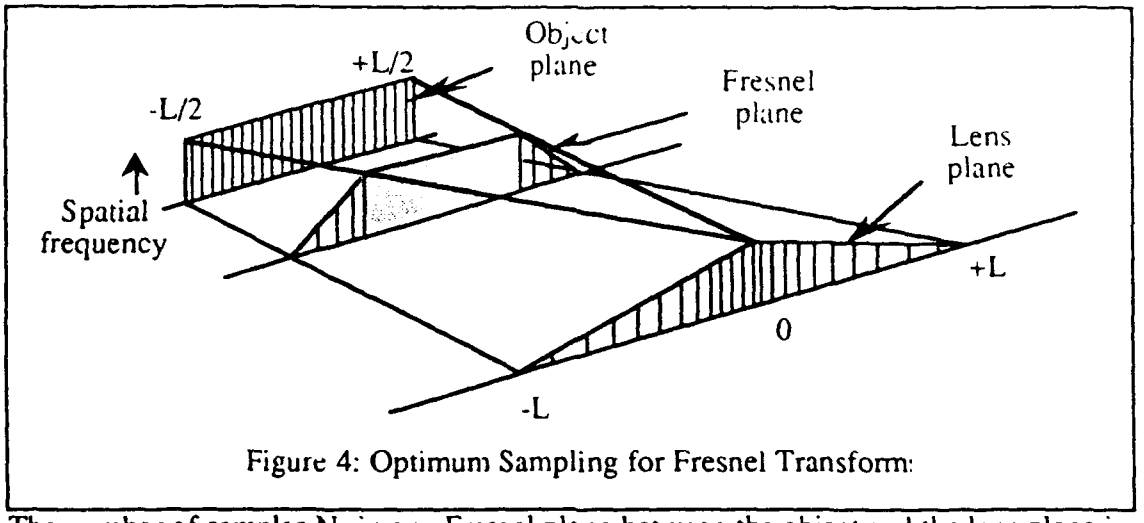


Figure 4: Optimum Sampling for Fresnel Transform

The number of samples N_f in any Fresnel plane between the object and the lens plane is found with the aid of Figure 4 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:

$$N_f = \frac{L - 2\theta_{co} D}{d_0} + \frac{4\theta_{co} D}{2d_0} = \frac{L}{d_0} = N, \qquad (8)$$

where $2d_0$ is the sample spacing for the average spatial frequency in the nonuniformly sampled region. From (8) 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 3, 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. Such a system is shown in Figure 5a for the situation where the magnification M =-2. The ray trace for the two edge samples as as shown in Figure 5a 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 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. In this region, the highest spatial frequency α_f, as given by the hyperbolic function (5), is now less than α_{co} because these Fresnel planes are beyond the distance where the marginal rays first cross. 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.

From the scissors diagram in Figure 5a, 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

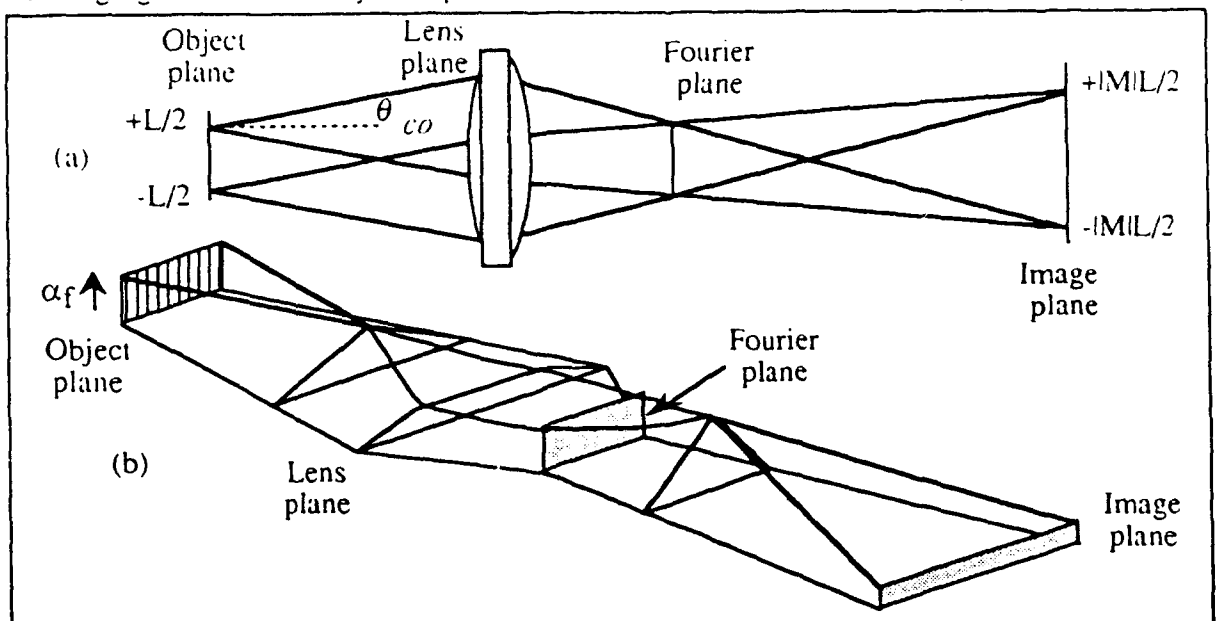


Figure 5: High Bandwidth Imaging System

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 uniform at

$$d_F = \frac{2\xi_{co}}{N} = \frac{2\lambda F \alpha_{co}}{N} = \frac{\lambda F}{L}, \qquad (9)$$

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. 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 second triangular space/frequency section, which corresponds to the second crossover plane for the rays in the scissors diagram. 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. In a more detailed analyses, we prove that the number of samples required in each section of Figure 5b is equal to N, as confirmed by the fact that the areas of each of the sections are equal.[3]

5. Summary and Conclusions

We have 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 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.

6. Acknowledgments

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

7. References

1. A. VanderLugt, "Design relationships for holographic memories," Appl. Opt., Vol. 12, p. 1675, (1973)

2. A. VanderLugt, "Packing density in holographic systems," Appl. Opt., Vol. 14, p. 1081 (1975)

3. A. VanderLugt, "Optimum Sampling of Fresnel Transforms," Submitted to Applied Optics

APPENDIX H

OPTIMUM SAMPLING OF FRESNEL TRANSFORMS

Reprinted from Applied Optics

Volume 29, Pages 3352-3361, 10 August 1990

A reprint from Applied Optics

Optimum sampling of Fresnel transforms

A. VanderLugt

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 intensity signal in any Fresnel plane of a coherently illuminated optical system with exactly N samples, provided that we use a specified nonuniform sampling technique.

I. 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 after using the Fourier transform operation to display the frequency content of the signal. Correlation is extensively used in signal detection and can be implemented in either the time or 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.[1]

In optical processing, we associate the spatial and spatial frequency domains of a coherently illuminated optical system with the time and temporal frequency domains. In addition to these two domains, an optical system has a continuum of Fresnel transform domains between the object plane and Fourier transform plane. The properties of the Fresnel transform of a signal is a unique mixture of the properties of the space and frequency distributions of the signal. The Fresnel transform resembles one or the other of these distributions depending on the location of the Fresnel plane.

The sampling theorem is an important signal processing tool for characterizing the information content of signals or the capacity of systems. For example, Toraldo[2] used the sampling theorem to determine the degrees of freedom in an image and the capacity of an optical channel, as did Linfoot[3] in his study of optical images. O'Neill and Walther[4,5] used sampling theory to study the importance of phase in optical systems, and Barakat,[6] at about the same time, applied sampling to optical diffraction theory. 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 1970s.[7,8] Marks *et al.*[9] subsequently applied similar analyses to space variant systems.

The required sample spacing for the image of an object is a function of the system magnification. 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.[10] Hence, if N samples are sufficient to represent the object accurately, we know that N samples, whose size and spacing are scaled by the magnification, are also sufficient to represent the image accurately. We show that N samples are sufficient to represent the Fresnel transform at any plane, but only in the condition that the sample spacings are nonuniformly distributed. The details of the sampling distribution are dependent on the bandwidth of the object and the particular Fresnel plane being sampled.

In Sec. II we use a simple optical system to illustrate the basic principles for sampling Fresnel transforms. In Sec. III we develop the optimum sampling distribution when the information capacity of the system is maximized; such a system images an object at an infinite conjugate plane. In Sec. IV we remove this restriction and derive the sampling procedure for the completely general case of a finite conjugate imaging system. In Sec. V we find the maximum spatial frequency at any Fresnel transform plane in a generalized imaging system.

II. Fourier Transform System

In a coherently illuminated optical system we deal with amplitude distributions which are not unfortu-

The author is with North Carolina State University, Electrical & Computer Engineering Department, Raleigh, North Carolina 27695-7911.

Received 27 October 1989.
0003-6935/90/233352-10$02.00/0.

nately directly observable or measurable. The sampling spacings and spatial frequencies are always associated with the intensities of optical signals, since it is the intensity that is observed and measured. Although we use amplitude distributions to calculate the Fresnel transform, all spatial frequencies and sample spacings discussed in this paper are those associated with the intensity of the Fresnel transform.

From the sampling theorem, we know that a signal is accurately represented by $N = 2TW$ samples, where T is the time duration of the signal and W is its bandwidth. The required sample spacing is, therefore, $T_0 = T/N = 1/(2W)$. The corresponding notion in an optical system is that the sample spacing d_0 for a bandlimited object $f(x)$, having a cutoff spatial frequency α_{co}, is $d_0 = 1/(2\alpha_{co})$. The number of samples required for an object of length L is $N = L/d_0$. The required number of samples for a 2-D object is $N = N_x N_y$, where N_x and N_y are the number of samples required in the x- and y-directions.

A. Fourier Domain

Consider the Fourier transform system, shown in Fig. 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[11]:

$$g(u) = \int_{-\infty}^{\infty} f(x) \exp\left[-j \frac{\pi}{\lambda F}(u - x)^2\right] dx, \tag{1}$$

where F is the distance from plane P_1 to plane P_2 and the object is limited to 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:

$$h(u) = g(u) \exp\left(j \frac{\pi}{\lambda F} u^2\right). \tag{2}$$

A second Fresnel transform similar to Eq. (1) is applied to $h(u)$ and produces the light distribution at plane P_4:

$$F(\xi) = \int_{-\infty}^{\infty} h(u) \exp\left[-j \frac{\pi}{\lambda F}(\xi - u)^2\right] du. \tag{3}$$

We now substitute Eq. (2) and Eq. (1) into Eq. (3) to obtain the Fourier transform relationship[12]

$$F(\xi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x) \exp\left[-j \frac{\pi}{\lambda F}(u - x)^2\right] \times \exp\left(j \frac{\pi}{\lambda F} u^2\right) \exp\left[-j \frac{\pi}{\lambda F}(\xi - u)^2\right] dx\,du$$

$$= \int_{-\infty}^{\infty} f(x) \exp\left(j \frac{2\pi}{\lambda F} \xi x\right) dx. \tag{4}$$

Although the configuration shown in Fig. 1 is the simplest to analyze, the Fourier transform exists in a much wider range of conditions, such as when plane P_1 is at a different position relative to the lens or when using divergent or convergent illumination.[12]

In the object plane P_1, all information is confined to the region defined by the aperture L. In the Fourier plane P_4, all information passes through an aperture $2\xi_{co}$. A previous study showed 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.[7] A related study showed that the packing density, expressed in bits per unit area, for a generalized imaging system is highest in either plane P_1 or P_4, depending on the cutoff frequency α_{co}.[8] For the special case in which 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. Since the lens aperture A, as shown in Fig. 1, must be large enough to transmit both the undiffracted and diffracted light, we find that $A = 2L$. Hence the spatial extent of the Fresnel transform at planes P_2 and P_3 is twice that of either the object or Fourier transform.

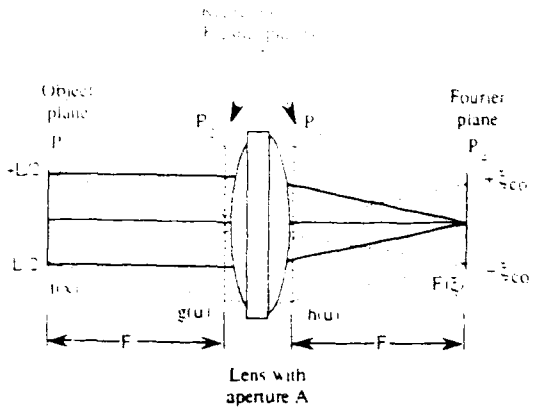


Fig. 1. Fourier transform system.

B. 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 required number of samples remained the same. Unfortunately, this is not the case because the spatial frequency α_f in some Fresnel planes is at least as large as α_{co}, and the spatial extent of the Fresnel transform is larger than that of the object. To illustrate this point, consider the general Fresnel transform $r(u)$, due to one sample of the object, valid in the region between planes P_1 and P_2, as shown in Fig. 2:

$$r(u) = \int_{-\infty}^{\infty} d(x) \exp\left[-j \frac{\pi}{\lambda D}(u - x)^2\right] dx, \tag{5}$$

where $d(x)$ is the object sampling 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 good working approximation for a sampling function $d(x) = \mathrm{sinc}(x/d_0)$ is that[11]

$$r(u) = \begin{cases} \exp\left(-j \frac{\pi}{\lambda D} u^2\right); & |u| \le \theta_{co} D, \\ 0; & \text{else.} \end{cases} \tag{6}$$

The energy due to the sampling function is, therefore,

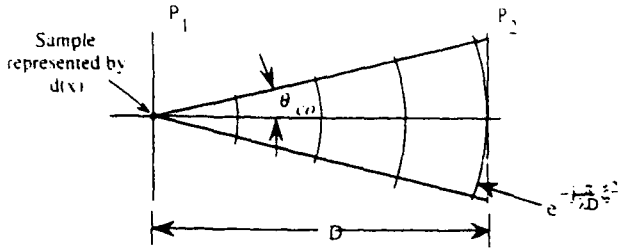


Fig. 2. Impulse of a sample of the object.

contained primarily within a cone whose apex angle is $2\theta_{co}$, as shown in Fig. 2. The physical cutoff angle θ_{co} is determined by the rays that pass through the cutoff spatial frequency positions $\pm\xi_{co}$ at the Fourier plane of Fig. 1.

An equally valid method for determining the spatial frequencies at an arbitrary Fresnel plane is to use the angular spectrum associated with $f(x)$. In that method, we decompose $f(x)$ into a set of orthogonal plane waves that propagate into free space. These plane waves interfere to establish the spatial frequencies appropriate for the Fresnel plane in question. In this paper, we elect to decompose $f(x)$ into a set of orthogonal sampling functions.

Suppose that the object $f(x)$ contains N samples, each spaced a distance d_0 from its neighbors. Consider the light diffracted by a pair of samples in plane P_1, which are separated by nd_0, where $1 \le n \le (N-1)$ and which are centered on the optical axis at $x = 0$. This two-sample object is, therefore, represented by

$$f(x) = d(x - nd_0/2) + d(x + nd_0/2), \tag{7}$$

and the intensity at a plane an arbitrary distance D from the object plane P_1 is found by substituting Eq. (7) into Eq. (5) to find that

$$I(u) = \left|\int_{-\infty}^{\infty} f(x) \exp\left[-j\frac{\pi}{\lambda D}(u-x)^2\right]dx\right|^2$$

$$= \left|\int_{-\infty}^{\infty} [d(x - nd_0/2) + d(x + nd_0/2)] \times \exp\left[-j\frac{\pi}{\lambda D}(u-x)^2\right]dx\right|^2. \tag{8}$$

Since the sampling function for a bandlimited system behaves as a δ-function, we use the sifting theorem to find that

$$I(u) = \left|\exp\left[-j\frac{\pi}{\lambda D}(u - nd_0/2)^2\right] + \exp\left[-j\frac{\pi}{\lambda D}(u + nd_0/2)^2\right]\right|^2$$

$$= 2[1 + \cos(2\pi nd_0 u/\lambda D)]. \tag{9}$$

The energy in $I(u)$ is concentrated in the region $|u| \le (\theta_{co}D - nd_0)$ as we more fully explore in Sec. III. From Eq. (9) we find that the spatial frequency α_f of the intensity, in a general Fresnel plane, due to two samples in the object plane is

$$\alpha_f = \frac{nd_0}{\lambda D}. \tag{10}$$

This result is subject to the constraint that $nd_0 \le L$, so that the samples lie within the region that defines the object and 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, interference is produced only by two closely spaced samples; light rays from widely spaced samples do not overlap due to the bandlimited nature of $f(x)$ and the finite values of θ_{co}. 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 edge samples of the object in Fig. 3(a), we note that light from these two samples overlap only when $D = F$ so that the maximum spatial frequency, at all Fresnel planes between P_1 and P_2, is the cutoff frequency $\alpha_{co} = L/\lambda F$.

From this same ray diagram, we see that the Fresnel diffraction pattern at the lens plane, therefore, has twice the spatial extent as the object. As we just showed, the highest spatial frequency at the Fresnel plane is the same as that for the object. We conclude, therefore, that we need at least $2N$ samples to measure $I(u)$ accurately, if the samples are uniformly spaced. For 2-D objects, we require 4 times the number of samples which increases the computational burden on digital postprocessing operations.

III. 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. We continue by analyzing the maximum capacity system

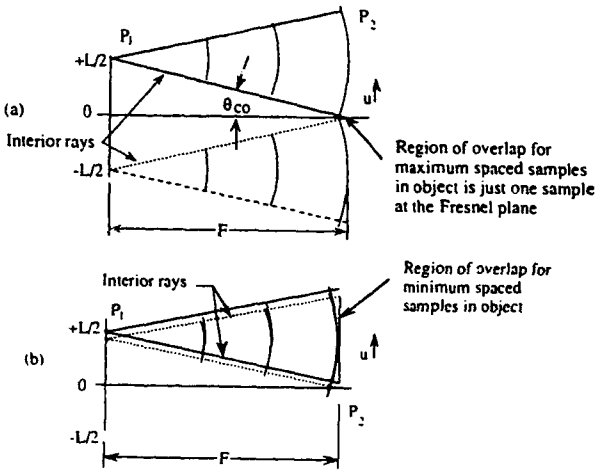


Fig. 3. Sample spacing to produce (a) maximum frequency but minimum overlap and (b) minimum frequency but maximum overlap.

described in Sec. II. The spatial frequency at any position in the Fresnel plane is most easily found with the aid of a graphic 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 Fig. 1. From Eq. (10) 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, with the restriction that the light from the two samples overlaps at the observation plane.

A. Distribution of Spatial Frequencies in a Fresnel Domain

As noted above, the maximum separation between samples 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, in accordance with Eq. (10), where the interior rays from the two samples cross. The region of overlap at a Fresnel plane is found from Fig. 3(a) and from straightforward geometrical calculations to have the minimum value of $R = L - (N-1)d_0 = d_0$.

From Fig. 3(b), 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 Eq. (10), of $\alpha_f = d_0/\lambda F$. The amount of overlap at the Fresnel plane is, however, at a maximum value; the region of overlap is $R = L - d_0$. As we see from Fig. 3, the region of overlap in the Fresnel plane P_2 is always centrally located opposite the midpoint of the pair of samples. Samples having large separations produce higher spatial frequencies at the Fresnel plane, but the samples are constrained to a small range of positions near the optical axis in plane P_1. The region of overlap in plane P_2 decreases as the sample separation increases, vanishing when the separation approaches L as shown in Fig. 3(a). On the other hand, closely spaced samples can occur anywhere in plane P_1 so that the region of overlap in plane P_2 is large. As a result of these considerations, we find that low spatial frequencies are present throughout the $|u| \le L$ region in plane P_2, but high spatial frequencies exist only near the optical axis in a region for which $|u| \le d_0$.

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 where the highest frequency is localized, we sample the object uniformly. A similar argument holds for the Fourier plane so that the information in the Fourier plane is also uniformly sampled. However, from the ray diagram associated with Fig. 3 for the spatial frequencies in the lens plane, we see that all spatial frequencies are present in the Fresnel plane only in the vicinity of $u = 0$ and that the spatial frequencies at any other position in the Fresnel plane decrease to zero as $|u| \to L$.

The distribution of spatial frequencies as a function of the variable u in the Fresnel plane is found by a simple extension of the graphic 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

$$\alpha_f(u) = \frac{L}{\lambda F}\left(1 - \frac{|u|}{L}\right). \tag{11}$$

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

B. Optimum Sample Spacing Distribution

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 for the Fresnel transform at the lens plane by the spatial frequency distribution associated with a chirp function

$$c(u) = 1 + \cos\left[\frac{\pi}{\lambda F}(L-u)^2\right], \tag{12}$$

as shown in Fig. 4(a). The chirp function has its maximum frequency at $u = 0$, its minimum frequency at $u = L$, and the rate of change is linear as required by Eq. (11). The spatial frequency distribution of the chirp as a function of the variable u is identical to that of the Fresnel transform for an arbitrary object as established by Eq. (11). To find the required sample spacings, we apply the Nyquist sampling criterion that requires two samples per cycle of the spatial frequency at each position in the Fresnel plane. For the chirp function (12), the optimum sample spacings are shown in Fig. 4(b), where alternate samples are staggered in the vertical direction for clarity. As desired, the sample spacing is small near the optical axis where the spatial frequencies are high in the Fresnel transform, and the sample spacing increases linearly as we move away from the optical axis, as required by Eq. (11).

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 π and solving the relationship

$$\frac{\pi}{\lambda F}(L-u)^2 = \frac{\pi L^2}{\lambda F} - \pi \tag{13}$$

for u:

$$u = L - \sqrt{L^2 - \lambda F}, \tag{14}$$

which is the value of the minimum sample spacing $d_{f\min}$. We can rewrite Eq. (14) as

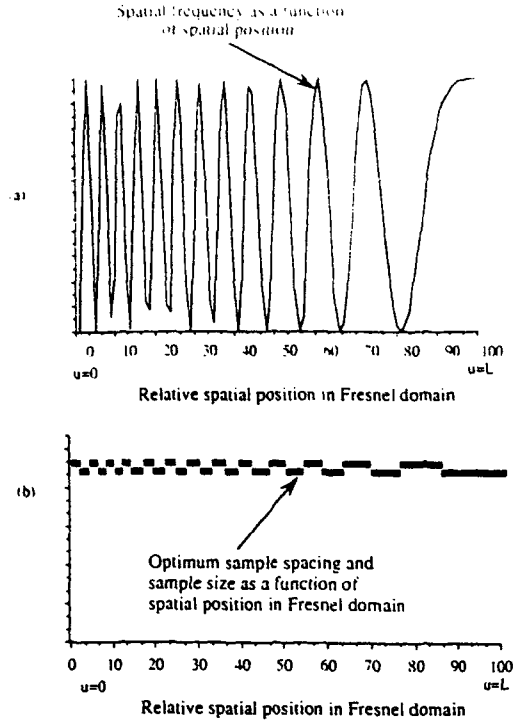


Fig. 4. Optimum sampling spacing for the Fresnel transform: (a) the chirp function representation of the spatial frequency distribution; (b) the sample spacing at the lens plane.

$$u = L - L\sqrt{1 - \frac{\lambda F}{L^2}}$$

$$= L\left(1 - \sqrt{1 - \frac{1}{L\alpha_{co}}}\right)$$

$$= L\left(1 - \sqrt{1 - \frac{2}{N}}\right) \cong \frac{L}{N} \cong d_0. \tag{15}$$

Thus the minimum sampling spacing in any Fresnel plane between P_1 and P is the same as that for the object. As an aside, this result is characteristic of all systems that use Fresnel transforms. It is consistent, for example, with the fact that a synthetic aperture radar antenna of length d_0 is just sufficient to sample the chirp radar return produced by a target whose size is d_0.

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 π and solving the relationship

$$\frac{\pi}{\lambda F}(L - u)^2 = \pi \tag{16}$$

for u:

$$u = L - \sqrt{\lambda F}. \tag{17}$$

The maximum sample spacing $d_{f\max}$ is found by subtracting the value of u given by Eq. (17) 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 whose focal length is F.

C. Example

As an example of nonuniform sample spacing, suppose that we have an object whose length is $L = 100$ mm and whose cutoff frequency is $\alpha_{co} = 0.25$ cycles/mm. For this object, 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 still α_{co}, but the length of the Fresnel plane is $2L$ compared with L for the object plane.

To illustrate the optimum sampling distribution for the Fresnel transform, we set the parameter $\lambda F = 400$ and find, from Eqs. (14) and (15), that $d_0 = d_{f\min} = 2$ mm and that $d_{f\max} = 20$ mm in agreement with the results shown in Fig. 4(b). 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 cycles/mm, we find that $N = 50$. The number of samples at the Fresnel planes P_2 and P_3 in Fig. 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.

D. Space/Frequency Diagrams

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 Fig. 1. First, from the ray diagram of Fig. 3, we see that the maximum frequency is the same at all planes because at least one pair of samples in the object plane produces the cutoff spatial frequency α_{co} at each Fresnel plane. 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 toward P_2. On the other hand, the total spatial extent of the Fresnel transform also increases linearly as D increases. These features are illustrated diagrammatically in Fig. 5 for all the Fresnel planes betweeen P_1 and P_2.

The diagram in Fig. 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. The object must be regularly sampled with sample spacing d_0, because the maximum spatial frequency α_{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 spatial frequency gradually goes to

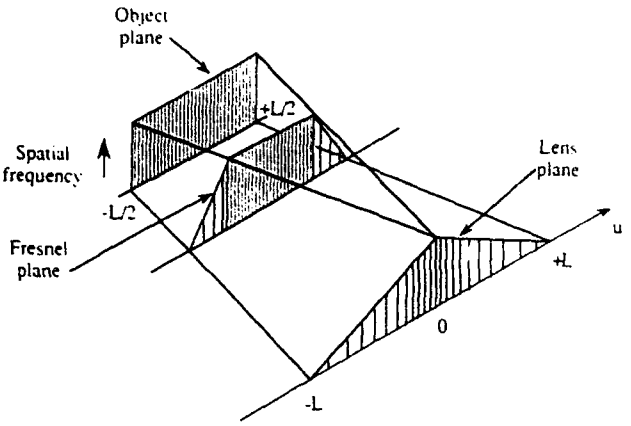


Fig. 5. Optimum sampling for Fresnel transforms. The height gives the value of the maximum spatial frequency as a function of position.

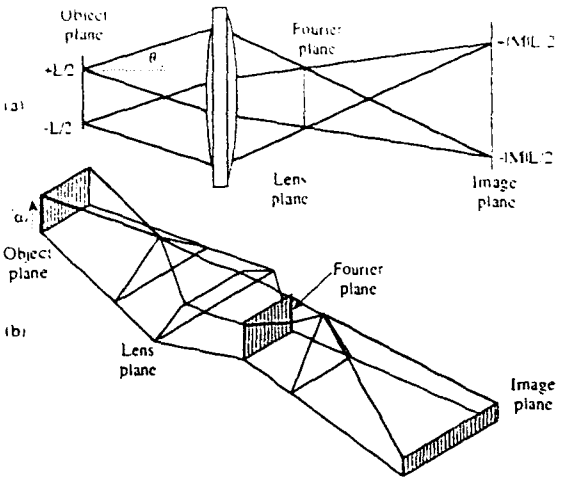


Fig. 6. High bandwidth imaging system: (a) the scissors diagram; (b) the space/frequency distribution.

zero are nonuniformly sampled. The optimum sampling within the two end regions is found by appending chirp functions with appropriate chirp rates to the central region. The height of the space/frequency diagram is proportional to the required sample density at a given position in an arbitrary Fresnel plane.

The required number of samples N_f in any Fresnel plane between the object and lens plane is found with the aid of Fig. 5. The number of samples is the sum of the uniformly spaced samples in the region for which $|u| \le (L/2 - \theta_{co}D)$ and the nonuniformly spaced samples in the outer regions of the trapezoid:

$$N_f = \frac{L - 2\theta_{co}D}{d_0} + \frac{4\theta_{co}D}{2d_0} = \frac{L}{d_0} = N, \tag{18}$$

where $2d_0$ is the sample spacing for the average spatial frequency in the nonuniformly sampled region. From Eq. (18) we find that the number of samples, when the optimum sampling technique is used, is the same in all planes between the object and 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 Fourier plane is the mirror image of that shown in Fig. 5, 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 nonuniform sampling procedure is used and provided that the Fourier transform system is structured for maximum information capacity.

IV. Sampling Strategy for an Unconstrained Imaging System

We now extend these sampling concepts to all Fresnel planes in an imaging system having finite magnification. A general imaging system without the maximum capacity constraint evolves from the Fourier transform system of Fig. 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 Fig. 6(a) for the situation where the magnification $M = 2$.

We begin the analysis of the general sampling strategy 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 Fig. 2. The ray trace for the two edge samples as shown in Fig. 6(a) 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 Fig. 6(b) 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 in Sec. IV.B, followed by a similar analysis for those planes on the image side of the lens in Sec. IV.B.

A. 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 Fig. 1. For finite imaging conditions, the lens is located farther away from the object plane so that the light from each object sample continues to diverge between this section and lens plane. To maintain an optimum sampling distribution in this region, we must modify the sample spacing strategy developed so far. We begin by noting that the highest spatial frequency α_f, as given by the hyperbolic function (10), is now $<\alpha_{co}$ because these Fresnel planes are beyond the distance where the interior rays first cross. The angle subtended by the two edge samples of the object, therefore, decreases below that needed to produce the cutoff spatial frequency. The section of the space/frequency distribution is still trapezoidal, however, by virtue of the arguments given above, but the maximum frequency in the central part of the trapezoid is smaller than the cutoff frequency.

A typical section of the space/frequency distribution

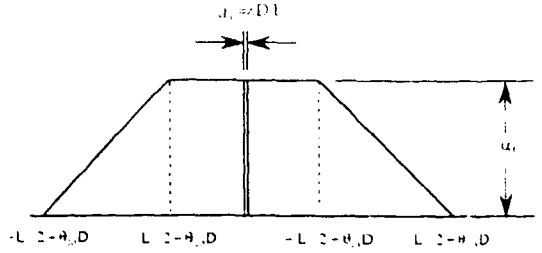


Fig. 7. General trapezoidal region.

in this region is shown in Fig. 7. The number of samples N_f in a Fresnel plane between the triangular section and 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 spatial samples in the outer regions of the trapezoid:

$$N_f = \frac{2(-L/2 \times \theta_{co}D)}{d_f} + \frac{2L}{2d_f} = \frac{2\theta_{co}D}{d_f}, \tag{19}$$

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 α_f is $d_f = \lambda D/L$ in Eq. (19) 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 in the most general imaging condition.

B. Image Side of the Lens

From the scissors diagram in Fig. 6(a), 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. 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 each 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

$$d_F = \frac{2\xi_{co}}{N} = \frac{2\lambda F\alpha_{co}}{N} = \frac{\lambda F}{L}, \tag{20}$$

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 for which the object and Fourier plane have the same size so that, in general, $d_F \neq d_0$. The sample spacing in the Fourier plane may be $\gtrless d_0$, depending on the geometry of the system, as we explore further in Sec. V.

As we progress from the Fourier plane to the image plane, the space/frequency sections remain trapezoidal, while the maximum spatial frequency at the optical axis 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 second 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 $|M|d_0$ as required by geometrical optics.

V. Maximum Spatial Frequencies

We have seen that the optimum sample spacing in any Fresnel plane is nonuniform, generally combined with some regions of uniform sampling. The question arises as to which planes require the most dense sample spacings. To develop the answer, consider the scissors and space/frequency diagrams given in Fig. 8, which is the same system configuration as given in Fig. 6, except that the object bandwidth is much narrower. The sample size and sample spacing at the object plane are larger than those for the wide bandwidth system, leading to a smaller diffraction angle θ_{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 α_{co}, where α_{co} for the narrow bandwidth system is less that the cutoff frequency for the wide bandwidth systems discussed above. As in the wide 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 sample cross, leading to the first triangular space/frequency section shown in Fig. 8(b). Between

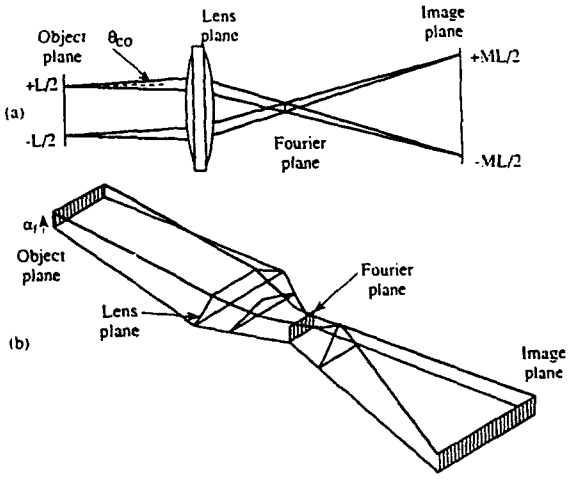


Fig. 8. Low bandwidth imaging system: (a) the scissors diagram; (b) the space/frequency distribution.

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 wide bandwidth system, light propagating to the edge samples at the image does 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 at the image plane 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 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 $|M|L$, 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 $\alpha_{co}/|M|$. Note that the space/frequency diagram for the narrow bandwidth object as shown in Fig. 8(b) is completely nested within that of the wideband object as shown in Fig. 6(b).

A. Fresnel Plane Containing the Highest Spatial Frequency

We now sharpen the qualitative results obtained so far to find the planes for which the spatial frequency is at a maximum and to quantify the magnitude of this frequency. 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 Sec. IV, the maximum spatial frequency must, therefore, 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 Eq. (11), it is useful to plot the value of the spatial frequency that occurs at the axis. Consider the scissors diagram in the upper part of Fig. 9 for a wide bandwidth object (solid rays) and a narrow bandwidth object (dotted rays). As an example, consider a system which has an object length $L = 1$ mm and is configured to provide a magnification $M = 2$. If the focal length of the lens is $F = 8$ mm, the distance from the object plane to the lens plane is $(1 - 1/M)F = 12$ mm, and the distance from the lens to the image plane is $(1 - M)F = 24$ 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 Fig. 9. In this case the cutoff frequency at the object plane is $\alpha_{co} = 400$ cycles/mm, and the maximum frequency remains at this value until we reach the plane at which the marginal rays first cross. From this

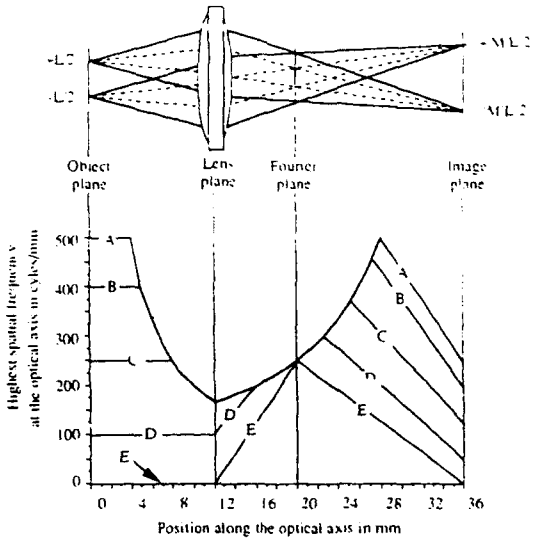


Fig. 9. Maximum frequency as a function of position.

plane to the lens plane the maximum frequency decreases following the hyperbolic curve

$$\alpha_f = \frac{L}{\lambda D}, \tag{21}$$

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

$$\alpha_f = \frac{-ML}{\lambda D}, \tag{22}$$

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 narrower 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 lie closer to the lens so that the maximum frequency path spends less time on the hyperbolas. At a somewhat narrower 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 discussed in Sec. IV.C. 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 hyper-

bola. This phenomenon 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 Eq. (6) between the object and 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 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 Eq. (20).

B. Four Planes that may Contain the Maximum Frequency

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 wider 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 Eq. (22). 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

$$D = \frac{-ML}{\frac{L}{F} - \frac{2\theta_{co}}{M}}, \tag{23}$$

which, when substituted into Eq. (22), yields the result that the shift occurs when

$$\alpha_{co} \geq \frac{ML}{(1 + M)\lambda F}. \tag{24}$$

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

The maximum spatial frequency must occur in one or more of four possible planes: the object plane; Fourier plane; image plane; or second crossover plane where the space/frequency diagram has a triangular section. 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. As a note in passing, it is possible for the maximum frequency to be the same in three of the four planes but not at all four planes simultaneously.

C. Smallest Maximum Spatial Frequency

It is quite simple to find the plane containing the minimum spatial frequency at the optical axis. For narrow 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

$$\alpha_{co} \geq \frac{-M^2L}{(M - 1)\lambda F}. \tag{25}$$

The minimum frequency occurs at the lens plane for all object bandwidths that satisfy (25).

An analysis similar to that used in connection with Fig. 9 applies to the case when the magnification is less than one. If $M = -\frac{1}{2}$, both the scissors diagrams and space/frequency diagrams are quite different, however, from those given in Fig. 8. We cannot, therefore, simply read the space/frequency diagrams from right to left. The reason is that 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 when we use collimated illumination, 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 formulas and calculations needed to describe this even more general case are straightforward and not given here.

VI. Summary and Conclusions

Geometrical optics ensures that the information content in the object and image planes of an optical system are equal; N samples are sufficient to measure accurately 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 spatial frequency may occur in at least one, but not more than three, of four planes in a generalized imaging system: object plane; Fourier plane; second crossover plane; or image plane.

This nonuniform sampling may have application to showing how to space the elements of arrays such as discrete element phased arrays in the microwave region. The nonuniform sampling procedure reduces

the required number of samples by up to a factor of 4, for the 2-D case, leading to less computations in applications such as image restoration. It is also a useful aid in understanding how the space and spatial frequency information is localized in Fresnel planes.

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

References

1. Special Issue, Proc. IEEE **69** (1981).
2. G. Toraldo di Francia, "Resolving Power and Information," J. Opt. Soc. Am. **45**, 497–501 (1955); "Capacity of an Optical Channel in the Presence of Noise," Opt. Acta **2**, 5–8 (1955).
3. E. H. Linfoot, "Information Theory and Optical Images," J. Opt. Soc. Am. **45**, 808–819 (1955).
4. E. L. O'Neill and A. Walther, "The Question of Phase in Image Formation," Opt. Acta **10**, 33–40 (1963).
5. A. Walther, "The Question of Phase Retrieval in Optics," Opt. Acta **10**, 41–49 (1963).
6. R. Barakat, "Application of the Sampling Theorem to Optical Diffraction Theory," J. Opt. Soc. Am. **54**, 920–930 (1964).
7. A. VanderLugt, "Design Relationships for Holographic Memories," Appl. Opt. **12**, 1675–1685 (1973).
8. A. VanderLugt, "Packing Density in Holographic Systems," Appl. Opt. **14**, 1081–1087 (1975).
9. R. J. Marks III, J. F. Walkup, and M. O. Hagler, "A Sampling Theorem for Space-Variant Systems," J. Opt. Soc. Am. **66**, 918–921 (1976).
10. B. D. Guenther, *Modern Optics* (Wiley, New York, 1990).
11. A. VanderLugt, "Fresnel Transforms and Bragg Cell Processors," Appl. Opt. **24**, 3846–3859 (1985).
12. A. VanderLugt, "Operational Notation for the Analysis and Synthesis of Optical Data Processing Systems," Proc. IEEE **54**, 1055–1063 (1966).

APPENDIX I

SHORT PULSE DETECTION BY ACOUSTO-OPTIC PROCESSING

Reprinted from Proceeding of the SPIE

Advances in Optical Information Processing V
Volume 1704, Pages 146-152, April 1992

Short pulse detection by acousto-optic processing

A. VanderLugt

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

ABSTRACT

In this paper we show how the Fresnel transform helps to detect time-delayed short pulse signals. A short pulse disperses spatially as it propagates through free space according to the Fresnel transform. A point photodetector produces a signal whose temporal frequency is directly proportional to the time delay between the pulses. When this signal is fed to an acousto-optic spectrum analyzer, the frequency domain represents a time delay domain.

1. THE FRESNEL TRANSFORM

Consider the beam combining part of the heterodyne acousto-optic system shown in Figure 1. Light diffracted by the two acousto-optic cells propagates into free space to produce a Fresnel transform at plane P_3. A short pulse $s_1(t)$ is defined as a pulse whose duration is significantly less than the duration T=L/v of the acousto-optic cell, where L is the length of the cell and v is the acoustic velocity. For a cell whose time-bandwidth product is TW, the smallest resolvable time slot is T_0=T/N=1/2W, where N=2TW is the number of resolvable elements and W is the bandwidth of the cell. A short pulse, then, consists of a few samples of duration T_0; such a pulse drives the acousto-optic cell in the lower branch of the interferometer.

The delayed signal $s_2(t)$ drives the acousto-optic cell in the upper branch of the interferometer. To simplify the discussion, we use the Raman-Nath mode of illumination at wavelength λ. The Fresnel transform for the signal in the lower branch is[1]

$$F_1(\eta,t) = \int_{-\infty}^{\infty} b(x) f_1(t - \tfrac{T}{2} - \tfrac{x}{v}) e^{-j(\pi/\lambda D)(\eta - x)^2} dx, \qquad (1)$$

where b(x) is an aperture weighting function, $f_1(t)=s_1(t)\cos(2\pi f_c t)$, and η is the spatial coordinate in plane P_3. Similarly, the Fresnel transform for the signal in the upper branch is

$$F_2(\eta,t) = \int_{-\infty}^{\infty} b(x) f_2(t - \tfrac{T}{2} - \tfrac{x}{v}) e^{-j(\pi/\lambda D)(\eta - x)^2} dx, \qquad (2)$$

where $f_2(t)=s_2(t)\cos(2\pi f_c t)=s_1(t-\tau)\cos[2\pi f_c(t-\tau)]$. At plane P_3 these two terms add coherently and a small photodetector is placed at η_0 in plane P_3 to detect the light intensity

 0-8194-0869-7/92/$4.00

$I(\eta,t)=|F_1(\eta,t)+F_2(\eta,t)|^2$. The detected output signal is

$$g(t)=\int_{-\infty}^{\infty}|F_1(\eta,t)+F_2(\eta,t)|^2\,H(\eta-\eta_0)d\eta, \tag{3}$$

where $H(\eta-\eta_o)$ describes the geometry and sensitivity of the photodetector. When we expand the integrand of (3) into its component parts, we obtain three terms:

$$g_1(t)=\int_{-\infty}^{\infty}|F_1(\eta,t)|^2\,H(\eta-\eta_0)d\eta, \tag{4}$$

$$g_2(t)=\int_{-\infty}^{\infty}|F_2(\eta,t)|^2\,H(\eta-\eta_0)d\eta, \tag{5}$$

and

$$g_3(t)=2\,\mathrm{Re}\{\int_{-\infty}^{\infty}F_1(\eta,t)F_2^*(\eta,t)H(\eta-\eta_0)d\eta\}. \tag{6}$$

The cross-product term as given by (6) generally contains the information of interest; the other two terms represent unwanted signals that must be removed in the post-processing operation.

It has been shown that the optimum distance between the acousto-optic cell the Fresnel transform plane satisfies the relationship that[1]

$$\frac{v^2}{\lambda D}=\frac{W}{T}, \tag{7}$$

in which the geometrical parameters on the left are related to the signal parameters on the right. If D is smaller than the optimum value, the signal duration is less than T because of insufficient dispersion; if D is larger than the optimum value, the signal amplitude is reduced while the signal duration is still limited to T seconds by virtue of the duration of pulse in the cell.

The edge rays associated with the pulse are shown by solid lines in Figure 3 and while those associated with the delayed pulse are shown dotted. The cross-product signal $g_3(t)$ as given by (6) does not begin until $t=\tau$ and this signal ends at $t=T$. The length of overlap at plane P_2 is therefore $L_\tau=v(T-\tau)$. Diffraction effects extend the time of overlap somewhat, but the strongest contribution to the cross-product output occurs for $\tau\le t\le T$. The sum of $g_1(t)$ and $g_2(t)$ exists, however, for $0\le t\le T+\tau$.

From (1) and (2) we see that the impulse responses at plane P_3 has the form

$$F_1(\eta,t)=b(vt-L/2)e^{-j(\pi/\lambda D)[\eta-v(t-T/2)]^2};\ 0\le t\le T, \tag{8}$$

and

$$F_2(\eta,t)=b(vt-v\tau-L/2)e^{-j(\pi/\lambda D)[\eta-v(t-\tau-T/2)]^2};\ \tau\le t\le T+\tau. \tag{9}$$

We substitute (8) and (9) into (3) and perform the integration by letting $H(\eta-\eta_o)=\delta(\eta-\eta_o)$ represent a point photodetector; we also use (7) to express result for the crossproduct term as

$$g_3(t) = 2b(vt - L/2)b(vt - v\tau - L/2)\cos(2\pi W\tau t/T + \phi_0);\ \tau \le t \le T, \qquad (10)$$

where ϕ_0 is an unimportant phase. The spatial frequency generated by the Fresnel transforms of two pulses therefore creates a temporal frequency that is linearly proportional to the time delay between the pulses. As the input pulses move under the aperture function b(x), the spatial modulation is converted to temporal modulation as shown by (8), (9), and (10). The most important result from (10) is that the output of the photodetector is a cosine whose temporal frequency is linearly proportional to τ:

$$f = \frac{v^2}{\lambda D}\tau = \frac{W}{T}\tau. \qquad (11)$$

The range of temporal frequencies is $0 \le f \le W$ as the time delay ranges over the interval $0 \le \tau \le T$.

2. The Fourier Transform of the Fresnel Transform

The output of the Fresnel preprocessing system as given by (3) is used to drive an acousto-o c spectrum analyzer. In the spectrum analyzer, the cell length is L=vT, where v is the acoustic velocity of the cell; the time bandwidth product is TW. Since the output g(t) from the Fresnel transform preprocessor is a baseband signal, we create a new signal $r(t)=g(t)\cos(2\pi f_c t)$ that drives the acousto-optic cell of the spectrum analyzer. The output of the spectrum analyzer is[2]

$$G(\alpha,t) = \int_{-\infty}^{\infty} a(x)g(t - \tfrac{T}{2} - \tfrac{x}{v})e^{j2\pi f_c(t-\frac{T}{2}-\frac{x}{v})}e^{j2\pi\alpha x}dx, \qquad (12)$$

where a(x) is an aperture weighting function. The range of integration is set by either the aperture weighting function a(x) for the spectrum analyzer or the duration of g(t). We are most interested in the cross product term $g_3(t)$ given by (6) because it contains the desired information about the time delays between the two signals. We concentrate on the example given by (10) for two impulse signals.

We set a(x)=b(x)=rect(x/L) for convenience, substitute $g_3(t)$ into (12), and retain only the positive diffracted order to find that

$$G_3(\alpha,t) = 2\,\mathrm{Re}\{\int_{-\infty}^{\infty} \mathrm{rect}(x/L_\tau)e^{j2\pi(f_c \pm \frac{W\tau}{T})(t-\frac{T}{2}-\frac{x}{v})}e^{j2\pi\alpha x}dx\}. \qquad (13)$$

As a result of the integration the photodetector output is proportional to

$$|G_3(\alpha,t)|^2 = \mathrm{sinc}^2[(\alpha - \alpha_c + W\tau/T)L_\tau] + \mathrm{sinc}^2[(\alpha - \alpha_c - W\tau/T)L_\tau], \qquad (14)$$

where we assume that the maximum overlap length L_τ is smaller than L. Equation (14) reveals that the output of the spectrum analyzer is, for this example, a sinc-function located at the spatial

frequencies $\alpha_c+W\tau/T$ and $\alpha_c-W\tau/T$.

3. Calculations of the Fresnel and Fourier Transforms

We further illustrate the theory by calculating the Fresnel transform of some short pulses for specific time delays. Consider a system in which the acousto-optic cell has a time duration of T=20μs, a bandwidth of W=50Mhz, and an acoustic velocity of v=0.617Km/s. The length of the cell is L=12.34mm so that the optimum distance between planes P_1 and P_3 in Figure 1 is $D=L^2/\lambda TW$=305mm for λ=500nm. A pulse of 20ns duration spreads light over a distance L in plane P_3, thus satisfying the condition needed to produce an output signal with the longest possible duration.

The diffraction patterns at plane P_3 are calculated using the exact form of the Fresnel-Kirchhoff equation. We model the pulse as a rectangle in plane P_1; its shape is not strongly dependent on the aperture function b(x) because the pulse is much shorter than the aperture. In Figure 3, we show the intensity $I(\eta_0,t)$ at the output of the Fresnel transform preprocessor for a time delay between pulses of τ=400ns. As the pulses move through the cell, the envelope takes on the shape of the Gaussian aperture function as indicated by (10). At t=τ, the first instant when both pulses are in the acousto-optic cell, one edge of the transform is positioned over the detector. At t=T/2, the transform is centered at the detector; at t=T-τ, the other edge of the transform is positioned over the detector. This moving spatial Fresnel transform is therefore converted to a time signal $g_3(t)$ as indicated by (10). The modulation is 100% because the pulses are so close that they sample nearly the same value of b(x) as they move through the cell. The interference pattern consists of a cosine function on a bias; the spatial frequency is α=f/v=1.62cycles/mm as calculated from (11), yielding about 20 fringes in the interference pattern.

This signal is then spectrum analyzed to measure the spatial frequency and to estimate the time delay. The total aperture function from (10) is approximately a Gaussian function whose intensity is $1/e^3$ at the aperture edges. Figure 4 shows the output of the Fourier transform of $g_3(t)$ at the instant when it is centered in the acousto-optic cell of the spectrum analyzer. In this calculation we ignore the influence of the spatial carrier frequency α_c so that the transform is shifted to baseband. As expected, the Fourier transform consists of the response to the aperture function at α=0 and α=1.62cycles/mm; we show only the positive spatial frequencies. The relationship between the time delay and the spatial frequency is $\tau=\lambda D\alpha/v$ so that the peak response is at τ=400ns as expected. The signal drops by 3dB at 400±5ns so that the time delay can be measured with a high degree of accuracy. The accuracy of the estimate is a function of the number of elements in the photodetector array and the noise in the system.

In Figure 5 we show the Fresnel transform of two pulses with the time delay increased to 8μs which is 40% of the cell length and near the upper limit of useful operation. Here we see that no interference is evident until t=8μs, and the detector output initially follows the envelope of the Gaussian aperture function since only one pulse is in the cell. The second pulse then enters the cell and generates the spatial frequency associated with the interference pattern; the envelope is the product of Gaussian functions separated by the spatial equivalent of 8μs (5mm). The modulation continues until t=T and then the intensity reverts to the Gaussian envelope until the second pulse leaves the cell at time T+τ. The percentage modulation varies in time as the pulses evolve because they are spaced far enough apart so they sample different values of b(x) as they propagate. When they are centered in the cell, the amplitudes are equal because b(–x)=b(x) and the modulation is 100% as shown at t=14μsec.The duration of g(t) is 36μsec; we show only the central 20 μsec of the signal in Figure 5. The spatial frequency in this case is 31.8cycles/mm which is beyond the resolution of the plotter; aliasing is therefore evident in the plot.

The spectrum of this signal is shown in Figure 6; it is similar to that described above, except that the mainlobe width of the signal at 31.8cycles/mm (corresponding to τ=8μsec) is somewhat

greater because the time duration of the modulation in $g_3(t)$ is less and the magnitude is reduced for the same reason. Nevertheless, the signal drops 3dB at 8μsec±6.25ns, showing that the absolute accuracy in measuring the time delay is not significantly greater that in the previous example. The maximum time delay that can be supported is application dependent; this result illustrates the trend in the response.

4. ACKNOWLEDGMENTS

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

5. REFERENCES

1. A. VanderLugt, "Fresnel Transforms and Bragg Cell Processors," Appl. Opt., Vol. 24, p. 3846 (1985)
2. A. VanderLugt, *Optical Signal Processing*, p. 399, Wiley, 1992

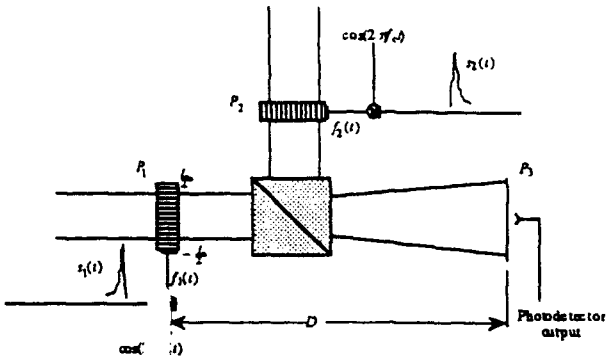


Figure 1: Fresnel transform

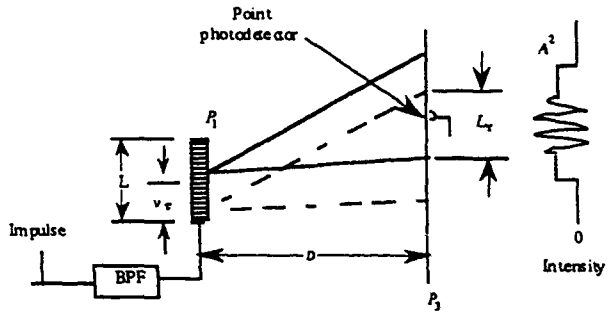


Figure 2: Geometrical interpretation of Fresnel transform

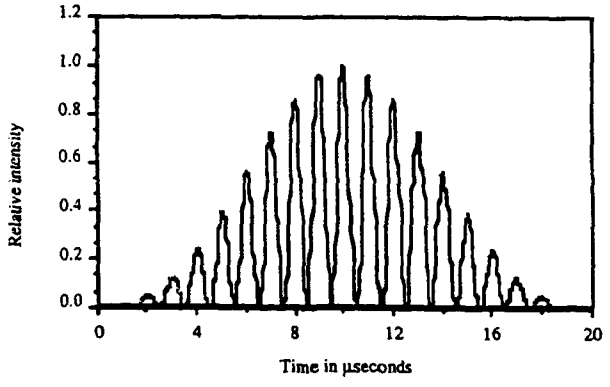


Figure 3: Temporal response of the Fresnel preprocessor to two short pulses 0.4μsec apart

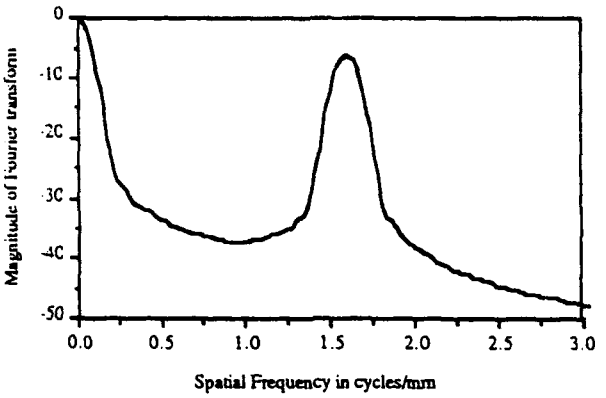


Figure 4: Fourier transform of signal produced by the Fresnel preprocessor(1.59cycles/mm=0.4μsec).

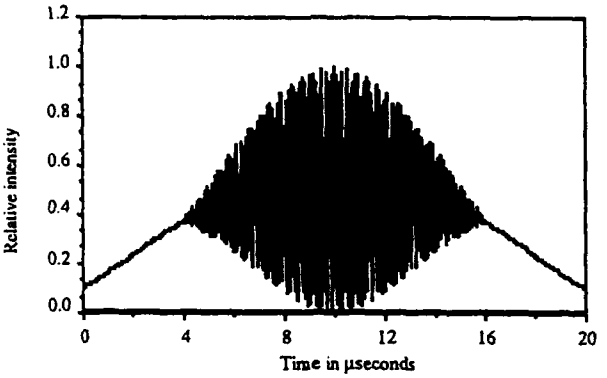


Figure 5: Temporal response of the Fresnel preprocessor to two short pulses 8μsec apart

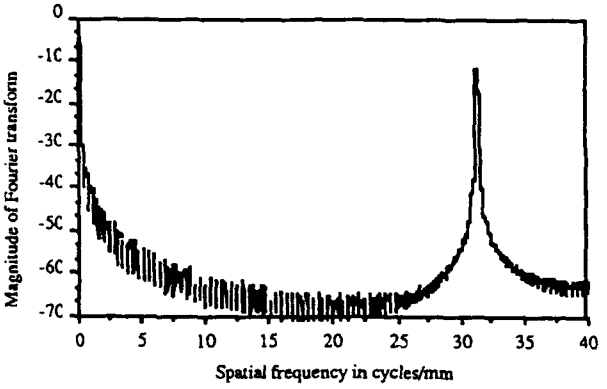


Figure 6: Fourier transform of signal produced by Fresnel preprocessor (31.8cycles/mm=8μsec)

APPENDIX J

TIME DELAY DETECTION OF SHORT PULSES BY FRESNEL AND FOURIER TRANSFORMATIONS

Reprinted from Applied Optics

Volume 32, Pages 3761-3771, 10 July 1993

Time-delay detection of short pulses by Fresnel and Fourier transformations

A. VanderLugt, C. S. Anderson, and P. J. W. Melsa

The Fresnel transform of a signal, when used as a preprocessing step to a subsequent spectrum analysis, helps to detect the time delay between short-pulse signals. Light from the short pulses disperses spatially as it propagates through free space to create an interference pattern. As the pulses move through an acousto-optic cell, a point photodetector produces a signal whose temporal frequency is directly proportional to the time delay between the pulses. When this signal is fed to a spectrum analyzer, the frequency domain represents a time-delay domain.

1. Introduction

In many applications the time delay between two signals is useful for determining the source of information. For example, in ranging systems the pulse and its echo constitute the two signals; the pulse may be repeated at known time intervals to extend the unambiguous range or to improve the measurement accuracy. In applications such as detecting exotic emitters, the pulses generally have different times of arrival at two or more antenna elements; in this case, we may use a dualchannel or a multichannel acousto-optic cell to process the signals and to sort them according to their angles of arrival. In other applications in which only a single pulse is received, a known time difference can be induced between the primary signal and its replica so that a single-channel acousto-optic cell is sufficient; in this case, we are interested in pulse detection and simply use the time delay as an aid for doing so.

Correlation is often used as a processing technique to estimate the time delays of signals that have large time–bandwidth products.[1–4] Alternatively, spectrum analysis in the form of a cepstrum may be used to measure time delays.[5] In both cases, the signals generally have a long duration so that the time–bandwidth product is large, resulting in a high signal-to-noise ratio for the processed signal at the output of the system. Short-pulse signals, however, are more difficult to detect because, although the bandwidth of the signals may be high, the short time duration yields a small time–bandwidth product. Correlation is ineffective for detecting short signals having a small time–bandwidth product because the correlation gain is small and the output signals tend to resemble the input signals. Spectrum analysis tends to be ineffective because the pulse energy is spread over a wide band of frequencies in the Fourier domain, leading to a low signal-to-noise ratio in that domain. Neither time-domain nor frequency-domain processing therefore provides the desired detection operation.

2. Background

We briefly review, for comparison purposes, how optical Fourier-transform techniques have been used to estimate time delay between two signals. As an example, consider the application in which an rf receiver consists of two antenna elements whose base-line distance is D_{12}; the received signals are $s_1(t)$ and $s_2(t)$. Suppose that a source is located at an angle θ with respect to the boresight of the array. The rf energy arrives at antenna element 2 with time delay $\tau = (D_{12}/c)\sin\theta$ with respect to that at antenna element 1 so that $s_2(t) = s_1(t - \tau)$.

We measure the angle of arrival by estimating the time delay between the two signals. Suppose that we produce the Fourier transform of these received signals in a heterodyne acousto-optic spectrum analyzer, as shown in Fig. 1.[6] The final lens in the system produces the Fourier transform of the light distribution created by the acousto-optic cells in planes P_1 and P_2. The Fourier transform of the first

A. VanderLugt is with the Department of Electrical and Computer Engineering, Box 7911, Daniels Hall, North Carolina State University, Raleigh, North Carolina 27695. C. S. Anderson is with Dynetics, Inc., P.O. Drawer B, Huntsville, Alabama 35814-5050. P. J. W. Melsa is with Tellabs Research Center, 3702 North Main Street, Mishawaka, Indiana 46545.

Received 26 May 1992.
0003-6935/93/203761-11$06.00/0.

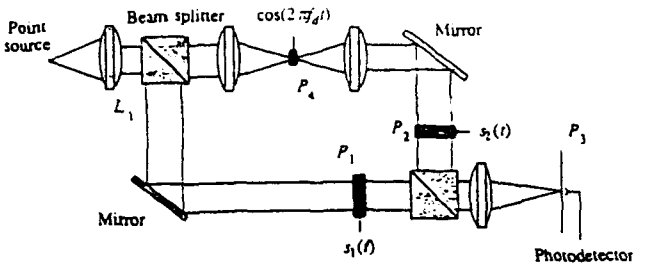


Fig. 1. Heterodyne acousto-optic Fourier-transform system.

input signal is

$$F_1(\alpha, t) = \int_{-\infty}^{\infty} a(x) f_1\left(t - \frac{T}{2} - \frac{x}{v}\right) \exp(j2\pi\alpha x)\mathrm{d}x, \quad (1)$$

where α is a spatial frequency and v is the velocity of the acoustic wave. The aperture weighting function $a(x)$ accounts for the illumination profile and truncation effects caused by the acousto-optic cell. We ignor an exponential function that describes the Bragg-angle illumination, and we assume that the rf signal is centered at f_c, the midpoint of the bandpass of the acousto-optic cell. For $s_2(t) = s_1(t - \tau)$ we have

$$F_2(\alpha, t) = F_1(\alpha, t)\exp(j2\pi f_d t)\exp(j2\pi\alpha v\tau), \quad (2)$$

so that the spectrum $F_2(\alpha, t)$ is identical to $F_1(\alpha, t)$ except for a phasor indicating that the modulator at plane P_4 has shifted the temporal frequency of light in the upper branch of the interferometer by f_d and a phase term $\phi(\alpha) = 2\pi\alpha v\tau$ that contains the time-delay information used to calculate the angle of arrival. The intensity at plane P_3 is the magnitude squared of the sum of $F_1(\alpha, t)$ and $F_2(\alpha, t)$:

$$I(\alpha, t) = |F_1(\alpha, t)|^2[1 + \cos[2\pi f_d t + \phi(\alpha, t)]]. \quad (3)$$

We demodulate $I(\alpha, t)$ in time and extract the phase information by comparing the phase of the detected frequency with that of a reference frequency.

Jernigan proposed driving a single-channel acousto-optic cell with the sum of $s(t)$ and $s(t - \tau)$.[7] In this case, the temporal offset frequency is zero, so that the cepstrum is

$$I(\alpha, t) = |F_1(\alpha, t)|^2[1 + \cos(2\pi\alpha v\tau)], \quad (4)$$

and the time delay is encoded as a spatial frequency in the α direction. Jernigan suggested scanning the intensity at the Fourier plane with a vidicon and spectrum analyzing the resulting temporal signal by using the fast-Fourier-transform algorithm. In this fashion, frequency bins correspond to time delays.

Each of the techniques discussed has drawbacks; the first requires a quadrature demodulation scheme and has a limited ambiguous range on the time delay between pulses because the phase is measured modulo 2π. The second technique requires scanning the intensity in the Fourier plane, which is basically a batch processing operation, followed by a fast Fourier transform; these operations become computationally intensive for real-time applications. The performance of this technique is sharply reduced by the presence of cw signals. We show in Section 3 how a Fresnel transform can be used to preprocess signals, before they are fed to a spectrum analyzer, to improve the detection of the time delay associated with short-pulse signals.

3. Fresnel Transform

In a coherently illuminated optical system a continuum of Fresnel transforms exists between the space (time) plane and the Fourier (frequency) plane of an optical system that uses an acousto-optic cell as the input spatial light modulator. The Fresnel transform is therefore a mixed time–frequency function whose characteristics resemble those of the time or the frequency function, depending on the plane of observation. At the midpoint between the two planes the characteristics are balanced between those of the time and the frequency functions.

Light diffracted by the two short pulses disperses spatially as it propagates through free space according to the Fresnel transform to create a sinusoidal interference pattern. As the short pulses travel through an acousto-optic cell at velocity v, so too does the interference pattern travel at velocity v. A point photodetector at an appropriate Fresnel-transform plane therefore produces a signal whose temporal frequency is directly proportional to the time delay between the pulses. When the output signal from the Fresnel transform preprocessor is used to drive a spectrum analyzer, the frequency domain is mapped into a time-delay domain.

We treat the general case in which the pulses may be generated by a laser range finder, a time-domain reflectometer, echoes in a communication system, a dual-antenna collection system, or similar systems. In all cases the shape of the pulse is affected by the signal source, the transmission medium, and the frequency response of the acousto-optic cell. We initially assume that these processes are nondispersive so that delayed pulses are represented as $s_2(t) = as_1(t - \tau)$, where a is an attenuation factor and τ is the delay. Without loss of generality, we set $a = 1$.

We begin our discussion of the basic ideas by modifying the last part of the heterodyne acousto-optic cell configuration, as shown in Fig. 2. This system is similar to that shown in Fig. 1, except that we remove the light modulator from plane P_4 and the final lens in the system; the light distribution at plane P_3 then becomes a Fresnel transform. A short pulse $s_1(t)$ is defined as a pulse whose duration is significantly less than the duration $T = L/v$ of the acousto-optic cell; such a pulse drives the acousto-optic cell in the lower branch of the interferometer. The delayed signal $s_2(t)$ drives the acousto-optic cell in the upper branch of the interferometer.

To simplify the discussion, we use the Raman–

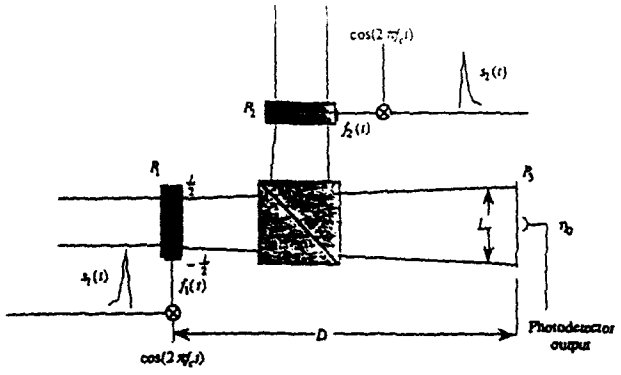


Fig. 2. Fresnel transforms of short pulses.

Nath mode and illuminate the acousto-optic cells normal to their entrance faces with collimated light of wavelength λ. Light diffracted by the signals in each of the acousto-optic cells propagates a distance D into free space, as shown in Fig. 2. The Fresnel transform for the signal in the lower branch is[8]

$$F_1(\eta, t) = \int_{-\infty}^{\infty} b(x) f_1\left(t - \frac{T}{2} - \frac{x}{v}\right) \exp[-j(\pi/\lambda D)(\eta - x)^2] dx, \tag{5}$$

where $f_1(t) = s_1(t)\cos(2\pi f_c t)$, $b(x)$ is an aperture weighting function that may be different from that used in Eq. (1) and η is the coordinate in plane P_3. Similarly, the Fresnel transform for the signal in the upper branch is

$$F_2(\eta, t) = \int_{-\infty}^{\infty} b(x) f_2\left(t - \frac{T}{2} - \frac{x}{v}\right) \exp[-j(\pi/\lambda D)(\eta - x)^2] dx, \tag{6}$$

where $f_2(t) = s_2(t)\cos(2\pi f_c t) = s_1(t - \tau)\cos[2\pi f_c(t - \tau)]$. At plane P_3 these two terms add coherently in the region L_τ where they overlap.

A small photodetector is placed at η_0 in plane P_3 to detect the light intensity; the detected output signal is

$$g(t) = \int_{-\infty}^{\infty} |F_1(\eta, t) + F_2(\eta, t)|^2 H(\eta - \eta_0) d\eta, \tag{7}$$

where $H(\eta - \eta_0)$ describes the geometry and sensitivity of the photodetector. When we expand the integrand of Eq. (7) into its component parts, we obtain three output terms:

$$g_1(t) = \int_{-\infty}^{\infty} |F_1(\eta, t)|^2 H(\eta - \eta_0) d\eta, \tag{8}$$

$$g_2(t) = \int_{-\infty}^{\infty} |F_2(\eta, t)|^2 H(\eta - \eta_0) d\eta, \tag{9}$$

$$g_3(t) = \Re\left\{\int_{-\infty}^{\infty} F_1(\eta, t) F_2^*(\eta, t) H(\eta - \eta_0) d\eta\right\}. \tag{10}$$

The cross-product term as obtained by Eq. (10) generally contains the information of interest; the other two terms represent unwanted signals whose effects must be removed by postprocessing operations, as discussed in Section 6.

Because all planes to the right of the beam combiner are Fresnel planes, the question arises as to which is the optimum plane at which to detect the light. The optimum plane is found by considering the diffraction of light owing to cw signals and to short pulses separately. Suppose that an applied signal has a lowest cw frequency f_1, which diffracts light at an angle θ_1, as shown by the solid rays in Fig. 3(a). The highest cw frequency f_2 diffracts light at an angle θ_2, as shown by the dashed rays in Fig. 3(a). If the distance D between planes P_1 and P_3 is too large, a point photodetector at the Fresnel plane P_3 will not detect light over the full frequency range of the signal.

From the diagram we see that the intersection of the upper ray from the lowest frequency and the lower ray from the highest frequency occurs at a distance above the optical axis equal to

$$\eta_0 = -\frac{L}{2} + \theta_2 D = \frac{L}{2} + \theta_1 D. \tag{11}$$

Because the relationship between the diffracted angle

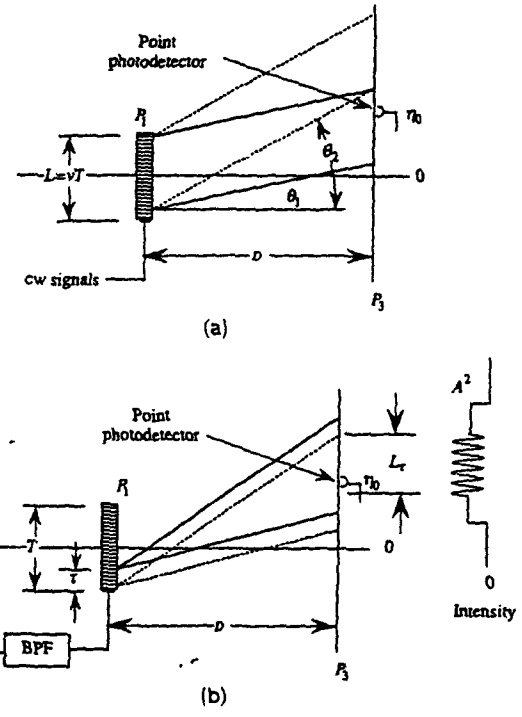


Fig. 3 Geometrical interpretation of Fresnel diffraction: (a) cw signal representation, (b) short-pulse representation.

and the applied frequency is $\theta = \lambda\alpha = f/v$, we find that

$$D \le \frac{vL}{\lambda(f_2 - f_1)} \le \frac{L^2}{\lambda TW}, \tag{12}$$

where $W = f_2 - f_1$ is the rf bandwidth of the acousto-optic cell. Energy from all cw frequency components of the signal are therefore detected only if the distance D satisfies inequality (12).

For a short pulse, light is diffracted within a cone whose upper and lower rays are shown as solid lines in Fig. 3(b) and whose ray angles are θ_2 and θ_1. We express $\Delta\theta = \theta_2 - \theta_1$ in terms of the pulse length L_0 and find that $\Delta\theta = \lambda/L_0$.[9] This cone of rays covers an interval $\Delta\theta D$ at plane P_3 and, as the short pulse travels through the cell, so too does the diffracted light travel with velocity v at plane P_3. Energy from a pulse delayed by a time interval τ produces a similar cone of rays, as shown by the dashed lines in Fig. 3(b). For interference between any two pulses, the cone of rays must overlap, and for any given time delay, we want to maximize the time for which the photodetector senses the interference. By tracing the rays associated with two pulses separated by the maximum time T, we find a second inequality for the distance D:

$$D \ge \frac{vL}{\lambda(f_2 - f_1)} \ge \frac{L^2}{\lambda TW}. \tag{13}$$

Inequalities (12) and (13) are satisfied only when the equality holds for both. By rearranging the parameters, we find that the optimum distance between the acousto-optic cell and the Fresnel-transform plane satisfies the relationship that[8]

$$\frac{v^2}{\lambda D} = \frac{W}{T}, \tag{14}$$

in which the geometrical parameters of the system on the left-hand side of the equality are related to the signal parameters on the right.

From the ray diagrams associated with the two pulses in Fig. 3(b), we find that the cross-product signal $g_3(t)$ as obtained by Eq. (10) does not begin until $t = \tau$ and cannot last beyond the time that the first pulse leaves the cell at $t \doteq T$. In general, the time interval is obtained by $T_\tau = vL_\tau$, where the length of overlap L_τ at plane P_3 is obtained by the amount of pulse dispersion minus the pulse delay:

$$L_\tau = \Delta\theta D - v\tau. \tag{15}$$

When the pulse width is at its minimum value of $L_0 = L/TW$, L_τ reaches its maximum value of $L_\tau = v(T - \tau)$; diffraction effects extend somewhat beyond the region suggested by the cones in Fig. 3(b), but the strongest contribution to the cross-product output occurs for the value of L_τ obtained by Eq. (15). In terms of time durations, the sum of $g_1(t)$ and $g_2(t)$ exists for a time interval $0 \le t \le (T + \tau)$, while the cross-product term exists only during some fraction of the time interval $\tau \le t \le (T - \tau)$ as determined by the overlap region L_τ and the velocity v.

4. Fourier Transform of the Fresnel Transform

We analyze the frequency content of the output of the Fresnel preprocessing system as obtained by Eq. (7) to extract the time-delay information. This spectrum analysis may be implemented either digitally or optically; the choice depends on the computational intensity for the application at hand. Here we discuss performing the Fourier transform optically as might be required, for example, in applications in which the time of arrival of the signal is unknown and we wish to perform an instantaneous Fourier transform over a sliding window for all time. For generality, and to assist in finding the optimum processing conditions, we assign different parameters to the acousto-optic cells in the two systems. In the spectrum analyzer, the cell length is $L_1 = v_1T_1$, where v_1 is the acoustic velocity of the cell; the time–bandwidth product is T_1W_1. Because the output $g(t)$ from the Fresnel-transform preprocessor is a baseband signal, we create a new rf signal $r(t) = g(t)\cos(2\pi f_c t)$ that derives the acousto-optic cell of the spectrum analyzer. The output of the spectrum analyzer is

$$G(\alpha, t) = \int_{-\infty}^{\infty} a(u) g\left(t - \frac{T_1}{2} - \frac{u}{v_1}\right) \times \exp\left[j2\pi f_c\left(t - \frac{T_1}{2} - \frac{u}{v_1}\right)\right] \exp(j2\pi\alpha u)\,du. \tag{16}$$

The range of integration is set by either $a(u)$ or the duration of $g(t)$. We are most interested in the cross-product term $g_3(t)$ obtained by Eq. (10) because it contains the desired information about the time delays between the wo signals. We concentrate first on this signal component and treat $g_1(t)$ and $g_2(t)$ as obtained by Eqs. (8) and (9) in Subsection 6.C.

We substitute $g_3(t)$ into Eq. (16) to find that

$$G_3(\alpha, t) = 2\,\mathrm{Re}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(u) F_1\left(\eta, t - \frac{T_1}{2} - \frac{u}{v_1}\right) \times F_2^*\left(\eta, t - \frac{T_1}{2} - \frac{u}{v_1}\right) H(\eta - \eta_0) \times \exp\left[j2\pi f_c\left(t - \frac{T_1}{2}\right)\right] \times \exp[j2\pi(\alpha - \alpha_c)u]\,d\eta\,du\right\}, \tag{17}$$

where α_c is the spatial frequency equivalent of the temporal carrier frequency f_c. The integration on η can be performed directly by replacing the point-photodetector response $H(\eta - \eta_0)$ by $\delta(\eta - \eta_0)$ and by applying the sifting theorem to obtain

$$G_3(\alpha, t) = 2\,\mathrm{Re}\left\{\int_{-\infty}^{\infty} a(u)F_1\left(\eta_0, t - \frac{T_1}{2} - \frac{u}{v_1}\right) \times F_2^*\left(\eta_0, t - \frac{T_1}{2} - \frac{u}{v_1}\right)\exp\left[j2\pi f_c\left(t - \frac{T_1}{2}\right)\right] \times \exp[j2\pi(\alpha - \alpha_c)u]\,du\right\}. \quad (18)$$

Equation (18) is the central result of this analysis.

To gain insight into the behavior of $G_3(\alpha, t)$, consider the response of the system to the ultimate short pulse: an idealized impulse so that $s_1(t)$ becomes $\delta(t)$. Although we can work with Eq. (18) directly, it is instructive to return to Eq. (10) so that we find $g_3(t)$ as an intermediate result. We use Eqs. (5) and (6) to find that

$$F_1(\eta, t) = \int_{-\infty}^{\infty} b(x)\delta\left(t - \frac{T}{2} - \frac{x}{v}\right)\exp[-j(\pi/\lambda D)(\eta - x)^2]\,dx,$$

$$= b(vt - L/2)\exp[-j(\pi/\lambda D)(\eta - vt + L/2)^2]. \quad (19)$$

In a similar fashion

$$F_2(\eta, t) = b(vt - v\tau - L/2) \times \exp[-j(\pi/\lambda D)(\eta - vt + v\tau + L/2)^2], \quad (20)$$

which is the same as Eq. (19) except for the effects of the delay τ in the aperture weighting function and in the exponential. We use Eqs. (19) and (20) in Eq. (10) to calculate the cross-product term:

$$g_3(t) = 2\,\mathrm{Re}\left\{\exp[j[2\pi W\tau t/T + \phi_0(\tau)]] \times \int_{-\infty}^{\infty} b(vt - L/2)b(vt - v\tau - L/2)H(\eta - \eta_0) \times \exp(-j2\pi v\tau\eta/\lambda D)\,d\eta\right\}, \quad \tau \le t \le T, \quad (21)$$

where $\phi_0(\tau) = (\pi/\lambda D)(Lv\tau - v^2\tau^2)$ and where we use relationship (14), which optimizes the system performance. As before, we represent the photodetector response $H(\eta - \eta_0)$ by $\delta(\eta - \eta_0)$ and use the sifting theorem so that

$$g_3(t) = 2b(vt - L/2)b(vt - v\tau - L/2) \times \cos[2\pi(W\tau t/T) + \phi_1(\tau)], \quad \tau \le t \le T, \quad (22)$$

where

$$\phi_1(\tau) = (\pi/\lambda D)[(L - 2\eta_0)v\tau - v^2\tau^2] \quad (23)$$

is an unimportant phase term. From Eq. (22) we see that the spatial aperture function $b(x)$ of the Fresnel-transform system imparts modulation on the temporal carrier frequency. For example, the envelope values for the leading pulse are $b(-L/2)$, $b(0)$, and $b(L/2)$ at the corresponding times $t = 0$, $t = T/2$, and $t = T$; the delayed pulse experiences the same modulation profile at a later time. The most important result from Eq. (22) is that the output of the photodetector at η_0 is a modulated carrier whose temporal frequency is linearly proportional to τ:

$$f = \frac{W}{T}\tau. \quad (24)$$

The range of temporal frequencies is therefore $0 \le f \le W$ as the time delay ranges over the interval $0 \le \tau \le T$. An interesting aspect of the Fresnel preprocessor is that the required detector bandwidth is a function only of the relative time delay, not the pulse width.

From Eq. (22), the sign of τ is ambiguous; positive and negative values of τ produce the same temporal frequency. In applications such as range finding the delay is known to be positive. But in an emitter-sorting application this ambiguity means that we cannot determine on which side of the boresight the emitter is located. We can overcome this difficulty in the system shown in Fig. 1 by translating one of the cells a distance $L_d = v\tau_d$, where τ_d is the largest expected time delay. Although this offset reduces the available range of time delays that are estimated, the offset τ_d is generally small for emitter sorting unless the base-line distance is large. For example, the maximum τ_d occurs at the end-fire direction; even for antennas located 300 m apart, the delay offset is only 1 μs, which is a small effect on the total delay range available when the cell in the Fresnel preprocessor is made from TeO_2 operating in the slow shear mode. Alternatively, if the outputs of the two antenna elements are summed and used to drive a single acousto-optic cell, one of the outputs can be delayed by a known amount to change the apparent boresight angle.

Because the frequency of the signal $g_3(t)$ is a linear function of the time delay between the short pulses, we can display the time delays by feeding $g_3(t)$ to a spectrum analyzer. We retain just the positive frequencies and set $a(x) = \mathrm{rect}(x/L)$ and $b(x) = \mathrm{rect}(x/L)$ for convenience so that Eq. (16) becomes

$$G_3(\alpha, t) = \mathrm{sinc}[(\alpha - \alpha_c \mp W\tau/L)L_\tau] \times \cos[2\pi(f_c + W\tau/T)t + \phi_1(\tau)], \quad \tau \le t \le (T - \tau). \quad (25)$$

Equation (25) reveals that the output of the spectrum analyzer is, for this example, a sinc function located

at the spatial frequency $\alpha_c \pm W\tau_c/L$ with an associated temporal frequency of $f_c + W\tau_c/T$. The $\pm$ sign in Eq. (25) arises because $g_3(t)$ is a baseband signal so that the spectrum is redundant about f_c.

From Eq. (25) we see that the first zero of the sinc function is at a spatial frequency $1/L_\tau$ with respect to that of the centroid. Because $L_\tau = v(T - \tau)$, we see that the sinc function at the output of the spectrum analyzer has its minimum width when $\tau = 0$. For longer time delays the amount of overlap of the two beams from the Fresnel transform system decreases, which, in turn, causes the time duration of $g_3(t)$ to decrease. The result is that the response in the Fourier plane of the spectrum analyzer is both broader and weaker. As $\tau \to T$, the spatial duration L_τ of $g_3(t)$ tends toward zero so that the frequency resolution deteriorates.

The fact that the response is broad does not necessarily mean that we cannot accurately find its centroid to measure τ; the fact that the response is weak is more serious because we may reach a point at which the signal-to-noise ratio is inadequate. A limit on the range of delays to $|\tau| \leq T/2$ ensures that the response to the shortest signal in the spectrum analyzer is not more than 3 dB down relative to that produced by the longest signal. In practice, this limitation may not be severe because the time delay associated with an acousto-optic cell represents a considerable amount of free-space propagation time owing to the large velocity ratio of c/v. For example, an acousto-optic cell with $T = 25\mu s$ can handle a dual-antenna system with a base-line distance of $D_{12} \approx 4{,}000$ m; this is considerably longer than those normally used in practice.

In this analysis we thought that advantages might accrue if the parameters associated with the acousto-optic cells in the Fresnel- and Fourier-transform systems are different. We address this question here. From Eqs. (13) and (15) we conclude that the maximum signal duration produced by the Fresnel preprocessing system is T so that the time durations of the cells in the two systems are equal. From Eq. (24) we conclude that the maximum bandwidth produced by the Fresnel preprocessing system is W so that the bandwidths of the cells in the two systems are also equal; the requirements on the acousto-optic cells are therefore identical.

5. Simulations and Experimental Results of the Fresnel and Fourier Transforms

We illustrate the theory by simulating the Fresnel transform of short pulses with specific time delays. Consider a system in which the acousto-optic cell has a time duration $T = 20$ μs, a bandwidth $W = 50$ Mhz, and an acoustic velocity $v = 0.617$ km/s. The length of the cell is $L = 12.34$ mm and the optimum distance between planes P_1 and P_3 in Fig. 3 is $D = L^2/\lambda TW = 305$ mm for $\lambda = 500$ nm. The shortest possible pulse of 20-ns duration therefore spreads light over a distance $L_\tau = L$ in plane P_3, thus satisfying the condition needed to produce an output signal with the longest possible duration.

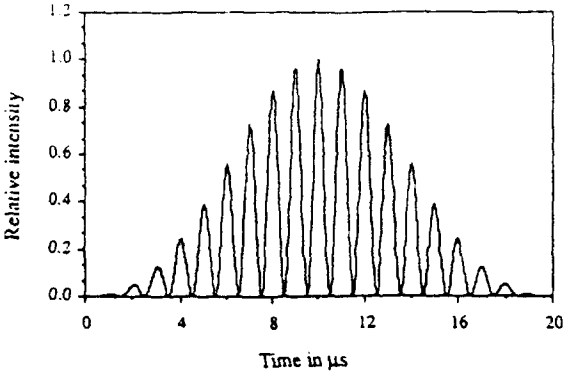


Fig. 4. Temporal response of the Fresnel preprocessor to two short pulses 0.4 μs apart.

We calculated the diffraction patterns at plane P_3 using the exact form of the Fresnel–Kirchhoff equation; the pulse is a rectangular function in plane P_1, and the aperture function $b(x)$ is Gaussian with its amplitude at the edges of the cell being $1/e^{1.5}$. In Fig. 4 we show the intensity $I(\eta_0, t)$ at the output of the Fresnel-transform preprocessor for a time delay between pulses of $\tau = 400$ ns. As the pulses move through the cell, the envelope takes on the shape of the Gaussian aperture function as indicated by Eq. (22). At $t = \tau$, the first instant when both pulses are in the acousto-optic cell, one edge of the Fresnel transform is positioned over the detector. At $t = T/2$ the transform is centered at the detector; at $t = T - \tau$ the other edge of the transform is positioned over the detector. This moving spatial Fresnel transform is therefore converted to a time signal $g_3(t)$, as indicated by Eq. (22). The modulation is nearly 100% throughout because the pulses are so close that they sample nearly the same value of $b(x)$ as they move through the cell. The interference pattern consists of a cosine function on a bias; the spatial frequency is $\alpha = f/v = 1.62$ cycles/mm as calculated from Eq. (24), yielding ~ 20 fringes in the interference pattern.

This signal is then spectrum analyzed to measure the spatial frequency and to estimate the time delay. Figure 5 shows the output of the Fourier transform of $g_3(t)$ at the instant when it is centered in the acousto-optic cell of the spectrum analyzer. The total aperture function used in Eq. (18) at this time instant is approximately a Gaussian function whose intensity is $1/e^{4.5}$ at the aperture edges. In this calculation we ignore the influence of the spatial carrier frequency α_c so that the transform is shifted to baseband. As expected, the Fourier transform consists of the response to the aperture function at $\alpha = 0$ and $\alpha = 1.62$ cycles/mm; we show only the positive spatial frequencies in Fig. 5. Because the relationship between the time delay and the spatial frequency is $\tau = \lambda\alpha D/v$, the peak response is at $\tau = 400$ ns, as expected. The signal drops by 3 dB at ± 10 ns relative to the peak so

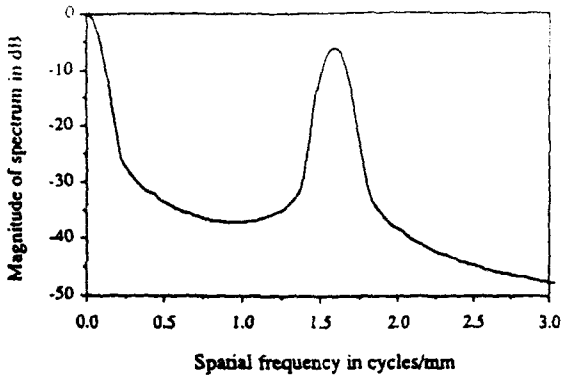


Fig. 5. Fourier transform of the signal produced by the Fresnel preprocessor (. ' ? cycles/mm = 0.4 μs).

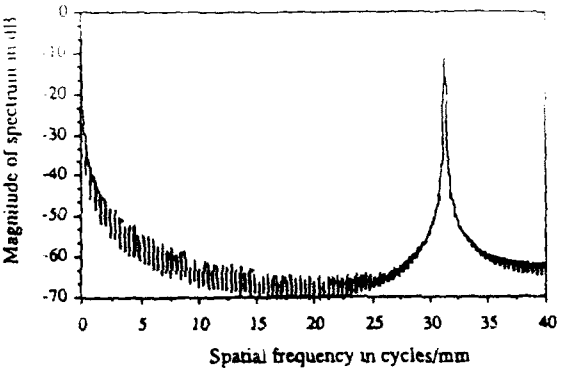


Fig. 7. Fourier transform of signal produced by Fresnel preprocessor (32.4 cycles/mm = 8 μs).

that the time delay can be measured with a reasonably high degree of accuracy. The accuracy of the estimate is a function of the number of elements in the photodetector array and the noise in the system.

In Fig. 6 we show the Fresnel transform of two pulses with the time delay increased to 8 μs, which is 40% of the cell length and near the upper limit of useful operation. The total duration of $g(t)$ is 36 μs; we show only the central 20 μs of the signal in Fig. 6. We see that the detector output follows the Gaussian aperture function for the first 12 μs because only one pulse is in the cell. The second pulse then enters the cell to generate the spatial frequency associated with the interference pattern; the envelope is the product of Gaussian functions separated by the spatial equivalent of 8 μs (5 mm). The modulation continues until t = 24 μs, and then the intensity reverts to the Gaussian envelope until the second pulse leaves the cell. The percentage modulation varies in time as the pulses evolve because they are spaced far enough apart so that they sample different values of $b(x)$ as they propagate. When they are centered in the cell, the amplitudes are equal because $b(-x) = b(x)$ and the modulation is 100%, as shown at t = 18 μs. The spatial frequency in this case is 32.4 cycles/mm, which is beyond the resolution of the graphics program and the plotter; aliasing is therefore evident in the plot.

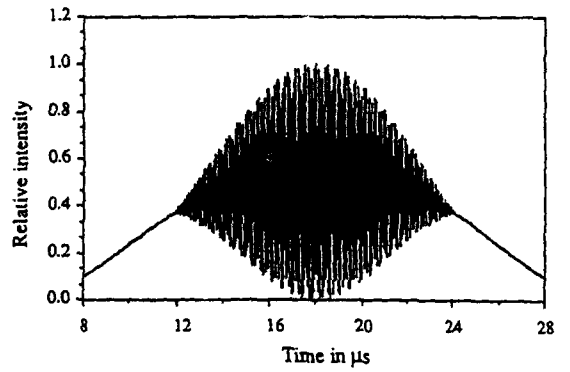
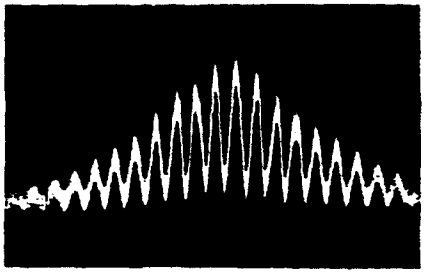


Fig. 6. Temporal response of the Fresnel preprocessor to two short pulses 8 μs apart.

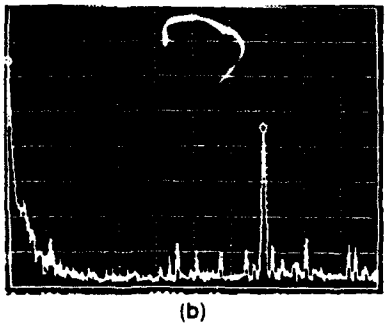
The spectrum of this signal is shown in Fig. 7; it is similar to that described above, except that the main-lobe width of the signal at 32.4 cycles/mm (corresponding to τ = 8 μs) is somewhat greater and the magnitude is reduced because the time duration of the modulation in $g_3(t)$ is less. The broadening of the main lobe of the diffracted light relative to that shown in Fig. 5 is obscured by the change of scale in Fig. 7. From the data we find that the signal drops 3 dB at ±12.5 ns relative to the centroid of the main lobe, showing that the absolute accuracy in measuring the time delay is less than in the previous example. The maximum time delay that can be supported is application dependent; this result illustrates the trend in the response.

To test the theory and to validate the simulations, we created a short pulse of 25-ns duration and delayed it by τ = 500 ns. The acousto-optic cell material is TeO_2, operated in the slow shear mode. The cell has a bandwidth of W = 40 MHz, and we illuminated only the central T = 10 μs of the cell with the Gaussian beam from a He–Ne laser. The acoustic velocity for the acousto-optic cell is v = 0.617 km/s so that the distance from the cell to the optimum Fresnel plane, as obtained by Eq. (14) for λ = 632.8 nm, is D = 150 mm. According to Eq. (24), the output of the Fresnel preprocessor should generate a temporal frequency of $f = W(\tau/T)$ = 2 MHz with an associated spatial frequency of $\alpha = f/v$ = 3.25 cycles/mm. The edge rays for this pulse form an angle $\Delta\theta$ = 0.102 rad so that the amount of spatial overlap, as obtained by Eq. (15), is L_τ = 6.17 mm. Finally, the number of fringes that we expect under the envelope of the interference pattern as obtained in Eq. (22) is $\alpha L_\tau = 20$.

Figure 8(a) shows the experimental result for these conditions. We see that the output of the Fresnel preprocessor consists of approximately 20 well-modulated fringes within a Gaussian envelope that exists for 10 μs. Since the two pulses are closely spaced relative to the width of the Gaussian illumination, the experimental results are similar to those



(a)



(b)

Fig. 8. Experimental results: (a) temporal response of the Fresnel preprocessor, (b) Fourier transform of the signal produced by the Fresnel preprocessor.

shown in Fig. 4, in which the modulation extends to the ends of the Gaussian illumination profile. Figure 8(b) displays the output of an electronic spectrum analyzer when it is driven by the signal shown in Fig. 8(a). To obtain a more accurate estimate of the temporal frequency, we repeated the drive signal to the Fresnel preprocessor so that the resolution bandwidth of the spectrum analyzer could be set to 10 kHz. Figure 8(b) shows the dc signal that is the sum of $g_1(t)$ and $g_2(t)$ and a strong cw frequency at 1.924 MHz; this frequency is close to the calculated value of 2 MHz, with the discrepancy probably caused by a small error in setting the distance D. The experimental results therefore illustrate all the features predicted by the analysis.

6. Bias Terms and Self-Noise

So far we have assumed that the short pulses are isolated, without the presence of cw signals. In this section we consider the interaction of the Fresnel transforms of short pulses and those of cw signals that may be present, as well as various other sources of signal-generated noise.

A. Continuous-Wave Signals

Suppose that a single cw signal is present in the received signal along with short pulses. We first consider the interaction of the Fresnel transform of a cw signal with its delayed replica. The cw signal is of the form $\cos(2\pi f_j t)$ and, if we consider just the positive diffracted order, the intensity at the Fresnel diffraction plane can be expressed as

$$I(\eta, t) = \left|\exp\left[j2\pi f_j\left(t - \frac{T}{2} - \frac{\eta}{v}\right)\right] + \exp\left[j2\pi f_j\left(t - \tau - \frac{T}{2} - \frac{\eta}{v}\right)\right]\right|^2 = 2[1 + \cos(2\pi f_j \tau)]. \quad (26)$$

In this calculation we assume that the two signals have unit amplitude and that we can safely ignore the effects of diffraction because the photodetector is within the geometrical shadow region of the edges of the acousto-optic cell. We note from Eq. (26) that the intensity pattern is not a function of either space or time; this interference pattern therefore does not contribute noise to the output of the system because its spectrum is located at $f = 0$ and its energy is removed from the output signal by a bandpass filter.

B. Continuous-Wave and Short-Pulse Signals

Next, we consider the interference of a cw signal and the Fresnel response to an impulse. The impulse response is found from Eq. (5), for the condition that $b(x) = 1$, as

$$\int_{-\infty}^{\infty} \delta\left(t - \frac{T}{2} - \frac{x}{v}\right)\exp[-j(\pi/\lambda D)(\eta - x)^2]dx = \exp\left[-j(\pi/\lambda D)\left(\eta - \frac{vT}{2} - vt + \frac{L}{2}\right)^2\right]. \quad (27)$$

The intensity at the Fresnel plane owing to the sum of the cw signal and the response to an impulse is

$$I(\eta, t) = \left|\exp\left[j2\pi f_j\left(t - \frac{T}{2} - \frac{\eta}{v}\right)\right] + \exp\left[-j\frac{\pi}{\lambda D}\left(\eta - \frac{vT}{2} - vt + \frac{L}{2}\right)^2\right]\right|^2 = 2\left\{1 + \cos\left[2\pi f_j\left(t - \frac{T}{2} - \frac{\eta}{v}\right) + \frac{\pi}{\lambda D}\left(\eta - vt + \frac{L}{2}\right)^2\right]\right\}, \quad 0 \le t \le T, \quad (28)$$

which, when we integrate the intensity over the point detector located at η_0, produces a chirp signal:

$$g(t) = 2\left\{1 + \cos\left[2\pi(f_j - 2v\eta_0)t + \frac{2\pi W}{T}t^2 + \phi_0\right]\right\}, \quad (29)$$

where ϕ_0 is an unimportant phase.

In addition to the noise contribution obtained by Eq. (29), we encounter other contributions owing to combinations of the cw signal and the delayed short pulse or owing to the short pulse and the delayed cw

signal. All these combinations, however, have the same general form as that of Eq. (29). The starting frequency of the chirps is dependent on the frequency of the cw signal and the photodetector position; the maximum frequency range is W. Not all these frequency components necessarily fall into the range of frequencies that represents the desired time-delay range in the spectrum analyzer and, because the energy in these chirp signals tends to be spread over all frequencies, this source of noise is rather easily removed by the inherent thresholding operation at the output of the spectrum analyzer.

C. Bias-Term Noise

So far we have ignored the contributions owing to the bias terms $g_1(t)$ and $g_2(t)$ from the Fresnel-transform preprocessor. The temporal spectra of these terms are at baseband and are collected along with the desired cross-product signal. We therefore must translate $g_3(t)$ to an intermediate frequency f_d so that it is separable from the bias terms. To calculate the value of f_d introduced by the acousto-optic cell at plane P_4 in Fig. 1, we find the temporal Fourier transform of $g_1(t)$; the second bias term $g_2(t)$ has the same spectral content as does $g_1(t)$. The required transform is

$$G_1(f) = \int_{-\infty}^{\infty} g_1(t)\exp(-j2\pi ft)\mathrm{d}t, \tag{30}$$

where

$$g_1(t) = \int_{-\infty}^{\infty} |F_1(\eta, t)|^2 H(\eta - \eta_0)\mathrm{d}\eta. \tag{31}$$

Normally we would substitute Eq. (5) into Eq. (31) and then substitute Eq. (31) into Eq. (30). When we integrate Eq. (31) over a small photodetector for which $H(\eta - \eta_0) \approx \delta(\eta - \eta_0)$, we find that $g_1(t) = |F_1(\eta_0, t)|^2$ and we see that Eq. (5) provides little insight for a general signal $f_1(t)$ that contains the pulse we wish to detect.

The analysis based on simple geometrical optics described in Section 3 is now extended to assist us here because we are interested only in the total frequency range of the spectral function $G_1(f)$, not the exact form of $g_1(t)$ or $G_1(f)$. We decompose the band-limited signal $f_1(t)$ as a set of weighted cw components so that the light amplitude leaving the acousto-optic cell is represented by

$$f_1\left(t - \frac{T}{2} - \frac{x}{v}\right) = \sum_{n=N_1}^{N_2} c_n \exp\left[j2\pi n f_0\left(t - \frac{T}{2} - \frac{x}{v}\right)\right], \tag{32}$$

where c_n is the amplitude of the nth frequency component, $f_0 = 1/T$ is the smallest resolvable frequency, and $N_2 - N_1 + 1 = TW$ is the number of frequencies contained in $f_1(t)$. The geometrical interpretation of Eq. (32) is that the light associated with the nth frequency component is a plane wave traveling at an angle $\theta_n = \lambda n f_0 / v$ with respect to the optical axis and oscillating at a temporal frequency $f_n = nf_0$.

In the absence of diffraction each plane wave has the same length L as the acousto-optic cell so that, after it propagates to plane P_3 of Fig. 3(a), the signal is represented as

$$F_1(\eta, t) = \sum_{n=N_1}^{N_2} c_n \,\mathrm{rect}[(\eta - \theta_n D)/L] \times \exp[j2\pi n f_0(t - T/2)] \tag{33}$$

We use Eq. (33) in Eq. (31), perform the sifting operation to get $F_1(\eta_0, t)$ and use the result in Eq. (30):

$$G_1(f) = \int_{-\infty}^{\infty} \sum_{n=N_1}^{N_2} \sum_{n=N_1}^{N_2} c_n c_m \times \mathrm{rect}[(\eta_0 - \theta_n D)/L]\mathrm{rect}[(\eta_0 - \theta_m D)/L] \times \exp[j2\pi(n - m)f_0(t - T/2)]\exp(-j2\pi ft)\mathrm{d}t. \tag{34}$$

The product of the rect functions at the point-photodetector position in plane P_3 of Fig. 3(b) is unity for all n and m. Hence we find that

$$G_1(f) = \sum_{n=N_1}^{N_2} \sum_{n=N_1}^{N_2} \delta[f - (n - m)f_0], \tag{35}$$

which shows that the temporal frequency content of $G_1(f)$ exists for $0 \le |f| \le 2W$.

We have already shown that the spectral content $G_3(f)$ of the desired cross-product term $g_3(t)$ also exists for $0 \le |f| \le W$ if we permit the time delay to range over the full time interval T. Hence the offset frequency must satisfy the condition that $f_d \ge (2W + W/2) = 2.5W$ so that the cross-product term can be separated from the bias terms by bandpass filtering. In some special cases in which the time delay is known to be bounded, the offset may be reduced further, as shown by Eq. (24); in no event can f_d be less than $2W$.

D. Self-Noise

A more severe, and less easily eliminated, source of noise at the output of the spectrum analyzer is due to the presence of cw signals having different frequencies in the cross-product term. The contribution of the cw components at plane P_3 owing to $f_1(t)$ is still obtained by Eq. (33), but the contribution from the upper branch is modified to become

$$F_2(\eta, t) = \sum_{m=N_1}^{N_2} c_m \,\mathrm{rect}[(\eta - \theta_m D)/L] \times \exp[j2\pi m f_0(t - T/2)]\exp(j2\pi f_d t). \tag{36}$$

As a result, we find that $G_3(f)$ owing to cw components is

$$G_3(t) = \int_{-\infty}^{\infty} \sum_{n=N_1}^{N_2} \sum_{n=N_1}^{N_2} c_n c_m \times \mathrm{rect}[(\eta_0 - \theta_m D)/L]\mathrm{rect}[(\eta_0 - \theta_m D)/L] \times \exp[j2\pi(n - m)f_0(t - T/2)] \times \exp[j2\pi(f - f_d)t]dt, \quad (37)$$

and the temporal spectrum becomes

$$G_3(f) = \sum_{n=N_1}^{N_2} \sum_{n=N_1}^{N_2} \delta[f - f_d - (n - m)f_0]. \quad (38)$$

We see that although the offset frequency f_d permits us to separate the cross-product term from the bias terms, $G_3(f)$ itself still contains energy from the cw components in the received signal. This represents a form of self-noise; we now consider methods to remove this source of noise and thereby improve the signal-to-noise ratio at the output of the Fourier-transform system.

7. Spectral Subtraction

Because the self-noise is in the same frequency band as the signal of interest, we need to find a method other than filtering to separate the signal and the noise. From Eqs. (24) and (14) we see that the frequency corresponding to a given τ is a function of D, which is the distance from the acousto-optic cell to the plane of the detector in the Fresnel-transform preprocessor. The optimum value for D, in terms of maximizing the time duration of the detected signal, occurs when Eq. (14) is satisfied. Suppose that we place a photodetector at a somewhat different value of D and at a different spatial position η_1. From Eqs. (34) and (35) we see that the spectrum $G_1(f)$ owing to the cw signals is not a function of η, provided that our representation of the Fresnel diffraction associated with a plane wave is reasonably accurate. A similar argument applies to $G_2(f)$ as well as to the self-noise generated by the cw components of $g_3(t)$.

We now have two outputs for which the cw components produce the same spectrum, while the signals owing to the short pulses have known relationships according to Eq. (19). For example, if the second detector is placed at a distance $D_2 = 0.9D_1$, the cw frequency produced by the time delay between short pulses will be $f_2 = 1.1f_1$ so that it is located at a different spatial frequency at the output of the spectrum analyzer.

We sketch the basics of spectral subtraction in Fig. 9. The spectrum $G_3(\eta_0, f)$ for the first detector shows the Fresnel transform of the cw signals within the cell as well as the response to the signal owing to the two short pulses at a frequency f corresponding to the time delay τ. The signal-to-noise ratio is -9.6 dB so that the desired response is buried in the self-noise (i.e., the desired signal has unit intensity while the variance of the self-noise is 9.6 dB). The second spectrum $G_3(\eta_1, f)$ shows the response from the second detector with a somewhat different self-noise response; in this case the signal-to-noise ratio is -9.4 dB and the desired signal corresponding to the same time delay τ_i occurs at $f_2 = 1.1f_1$. Because these two spectra represent integration for a time period of $\sim$T s, the spectra owing to the self-noise should be nearly identical. When these two spectra are subtracted, the signals owing to the short pulses are enhanced as shown in Fig. 9; the subtraction does not eliminate the self-noise completely owing to the small differences in self-noise intensity at the two Fresnel planes. A fairly simple post-processing operation should be able to sort the delays even when several pulses are present in the cell at the same time.

This spectral subtraction scheme may also be useful to detect short pulses in a noisy background. Suppose that a short pulse is obscured by strong cw frequency components. By introducing an artificial, but known, time delay, we can predict the temporal frequencies that a short pulse will produce at the output of the Fourier transform system. In this case, we do not need to sample the entire output of the spectrum analyzer but just the two frequencies whose values are known from the known induced time delay. After subtraction of the spectral responses, the results are thresholded to yield the detection of the short-pulse signal.

8. Summary and Conclusions

We have shown that the Fresnel-transform preprocessor has the intrinsic property of converting short pulses into longer signals so that the interference between the short pulse and its delayed replica produces a time-gated cw signal. This signal is then spectrum analyzed so that the spatial frequency domain becomes, in effect, a time-delay domain; the time delay can be measured directly from the location of the spectral content of the time-gated cw signal. The self-noise generated by the remainder of the input signal to the Fresnel preprocessor can be significantly reduced by using a second photodetector positioned so that its response to short pulses changes

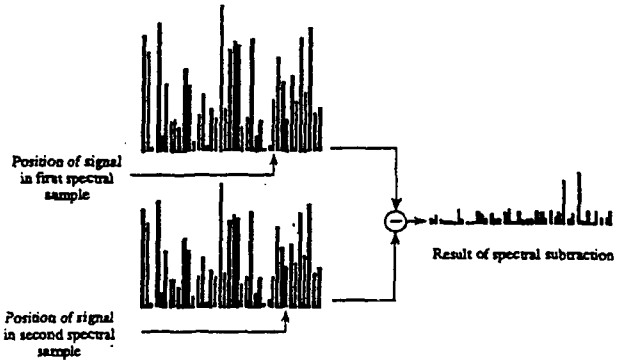


Fig. 9. Spectral subtraction to remove self-noise.

in a prescribed way while its response to the rest of the input signal is largely unaltered. The difference between the spectra of these two signals reveals the time delay between the pulses.

This research was performed under a grant from the U.S. Army Research Office.

References

1. J. C. Hassab and R. E. Boucher, "Optimum estimation of time delay by a generalized correlator," IEEE Trans. Acoust. Speech Signal Process. **ASSP-27**, 373–380 (1979).
2. W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," IEEE Trans. Inf. Theory **IT-19**, 608–614 (1973).
3. C. H. Knapp and G. C. Carter, "The centralized correlation method for estimation of time delay," IEEE Trans. Acoust. Speech Signal Process. **ASSP-24**, 320–327 (1976).
4. V. H. MacDonald and P. M. Schultheiss, "Optimum passive bearing estimation," J. Acoust. Soc. Am. **46**, 37–43 (1969).
5. D. G. Childers, E. P. Skinner, and R. C. Kemerait, "The cepstrum: a guide to processing," Proc. IEEE **65**, 1428–1443 (1977).
6. A. VanderLugt, "Use of decimated photodetector arrays in spectrum analysis," Appl. Opt. **27**, 2061–2070 (1988).
7. J. L. Jernigan, "Optical autocorrelation signal processor," U.S. patent 4,198,634 (15 April 1980).
8. A. VanderLugt, "Fresnel transforms and Bragg cell processors," Appl. Opt. **24**, 3846–3859 (1985).
9. A. VanderLugt, *Optical Signal Processing* (Wiley, New York, 1992), Chap. 3, p. 84.

APPENDIX K

DETECTION OF SHORT PULSES BY FRESNEL AND FOURIER TRANSFORMATIONS

Accepted for publication by Applied Optics.

Detection of Short Pulses by Fresnel Preprocessing

by

Michael J. Wardlaw

NSWC/DL, Code F44

Dahlgren, Va. 22448-5000

A. VanderLugt

Electrical and Computer Engineering Department

North Carolina State University

Box 7911 Daniels Hall

Raleigh, North Carolina 27695

919-515-5192

Abstract

Short pulses create unique Fresnel diffraction effects as they propagate through an acousto-optic signal processing system that produce an interference pattern. By following the detection of the interferometric pattern with spectral analysis, we extract information about pulse width, pulse separation, and relative time of arrival. Laboratory models of the Fresnel transform processing system and an optical spectrum analyzer were built and tested to verify the basic concepts. Pulses as short as the theoretical limit of 20 nanoseconds for this system and separated by as little as 60 nanoseconds or as much as 17 microseconds have detected and measured. Short pulses have also been detected in the presence of cw signals by using a frequency-variant matched filtering operation.

1.0 Introduction

In a previous paper, we developed the theory for using a Fresnel preprocessing technique to measure the time delay between two short-pulse signals and gave some preliminary results.[1] In this paper we describe extensive simulations and experimental results for this technique, including the development of techniques to detect short pulses in the presence of cw signals. This detection scheme recognizes that the Fresnel transform of a short pulse is dependent on the plane at which it is measured, whereas the response to cw signals are independent of the position of the detection plane. The two types of signals are distinguished by subtracting the Fresnel transforms from two different planes and by using frequency-variant matched filtering. Calculations of the signal-to-noise ratio, false alarm rates and detection rates are given for both linear and nonlinear thresholding operations. A pulse generation, detection, enhancement, and measurement system is used to test the applicability of this technology to real time signal processing problems.

Before presenting the experimental results, we summarize the fundamental concepts developed in Reference 1. In Figure 1 we show an acousto-optic cell of length L driven by two short pulses separated by a time duration τ. A short pulse s(t) is defined as a pulse whose duration is significantly less than the duration T=L/v of the acousto-optic cell, where L is the length of the acousto-optic cell and v is the velocity of the acoustic wave. The Fresnel transform shows that light diffracted by the two short pulses disperses spatially within an included angle $2\theta_{co}$ as it propagates through free space to create a sinusoidal interference pattern at plane P_2. As the short pulses travel through the cell at velocity v, so too does the interference pattern travel at velocity v. A point photodetector at a Fresnel transform plane P_2 therefore produces a signal whose temporal frequency is directly proportional to the time delay between the pulses. When the output signal from the Fresnel transform preprocessor is used to drive a spectrum analyzer, the frequency domain is mapped into a time-delay domain.

To simplify the theory, we use the Raman-Nath mode of illumination with collimated light of wavelength λ. Light diffracted by the signal in the cell propagates a distance D into free space as shown in Figure 1. The Fresnel transform at plane P_2 for the first pulse in the cell is[2]

$$F_1(\eta,t) = \int_{-\infty}^{\infty} b(x) f_1(t - \tfrac{T}{2} - \tfrac{x}{v}) e^{-j(\pi/\lambda D)(\eta - x)^2} dx, \qquad (1)$$

where $f_1(t)=s(t)\cos(2\pi f_c t)$, $s(t)$ is the baseband representation of the pulse, f_c is the midband frequency of the cell, $b(x)$ is the illumination function, and η is the coordinate at plane P_2. Similarly, the Fresnel transform at plane P_2 for the delayed signal is

$$F_2(\eta,t) = \int_{-\infty}^{\infty} b(x) f_2(t - \tfrac{T}{2} - \tfrac{x}{v}) e^{-j(\pi/\lambda D)(\eta - x)^2} dx, \qquad (2)$$

where $f_2(t)=s(t-\tau)\cos[2\pi f_c(t-\tau)]$. These two terms add coherently at plane P_3 in the region L_τ where they overlap to produce the sinusoidal interference signal.

A small photodetector, placed at η_o in plane P_2, detects the light intensity to produce the output signal

$$g(t) = \int_{-\infty}^{\infty} |F_1(\eta,t) + F_2(\eta,t)|^2 H(\eta - \eta_0) d\eta, \qquad (3)$$

where $H(\eta-\eta_o)$ describes the geometry and sensitivity of the photodetector. When we expand the integrand of (3) into its component parts, we obtain three output terms:

$$g_1(t) = \int_{-\infty}^{\infty} |F_1(\eta,t)|^2 H(\eta - \eta_0) d\eta, \qquad (4)$$

$$g_2(t) = \int_{-\infty}^{\infty} |F_2(\eta,t)|^2 H(\eta - \eta_0) d\eta, \qquad (5)$$

and

$$g_3(t) = 2\,\mathrm{Re}\{\int_{-\infty}^{\infty} F_1(\eta,t) F_2^*(\eta,t) H(\eta - \eta_0) d\eta\}. \qquad (6)$$

The cross-product term as given by (6) generally contains the information of interest; the other two terms represent unwanted signals whose effects must be removed by post-processing operations.

For a short pulse, light is diffracted mainly within a cone whose semi-angle is θ_{co}, where $\theta_{co}=\lambda/L_0$ and L_0 is *the minimum* width of the pulse. This cone of rays covers an interval $2\theta_{co}D$ at plane P_2 and energy from a pulse delayed by a time interval τ produces a similar cone of rays as shown in Figure 1. The distance to the optimum Fresnel diffraction plane is given by the relationship that[1]

$$\frac{v^2}{\lambda D}=\frac{W}{T}, \qquad (7)$$

where $W=f_2-f_1$ is the rf bandwidth of the acousto-optic cell. The length of overlap L_τ at plane P_3 is given by the amount of pulse dispersion minus the pulse delay:

$$L_\tau = 2\theta_{co}D - v\tau. \qquad (8)$$

When the pulse width is at its minimum value of $L_0=L/TW$, L_τ reaches its maximum value of $L_\tau=v(T-\tau)$ as can be seen by rearranging (7) and substituting it into (8). Diffraction effects extend somewhat beyond the region suggested by the cones in Figure 1; but the strongest contribution to the cross-product output occurs for the value of L_τ given by (8). In terms of a time duration, the cross-product term exists only during some fraction of the time interval $\tau\le t\le(T-\tau)$ as determined by the overlap region L_τ and by the velocity v.

The output of the Fresnel preprocessor, for a short pulse, is approximated by [1]

$$g(t)=|b(vt-L/2)|^2+|b(vt-v\tau-L/2)|^2$$

$$+2b(vt-L/2)b(vt-v\tau-L/2)\cos[2\pi(W\tau t/T)]; \quad \tau\le t\le T, \qquad (9)$$

where we ignore unimportant phase terms. The result (9) shows that the output of the photodetector is a modulated sinusoid whose temporal frequency is linearly proportional to τ:

$$f = \frac{W}{T}\tau. \qquad (10)$$

The range of temporal frequencies is therefore $0 \le f \le W$ as the time delay ranges over the interval $0 \le \tau \le T$. When g(t) is fed to a spectrum analyzer, the frequency domain is mapped into a time delay domain.

2.0 Experimental Design and Implementation

The experimental design and implementation is based on an input/output data acquisition and control philosophy. Functions requiring multiple configurations are generally controlled via a Macintosh® computer. In many cases the control interface also provided a means to collect data from the test instruments. The data acquisition and control software for the computer controlled functions are implemented using LabVIEW®2. This software package provides many of the subfunction utilities needed to design and implement control over an IEEE-488 bus, as well as to control the discrete and analog I/O interfaces installed within the computer.

The experimental system is segmented into three sections from an electrical viewpoint. The first section includes the computer and its peripherals. The second section includes those equipments directly controllable via one of the two data acquisition and control interfaces (oscilloscope, spectrum analyzer, arbitrary waveform generator, and local oscillators). The third section includes all the electronic equipment used to support and condition the signals both being generated and collected (amplifiers, local oscillator, cables, and so forth).

From an optical viewpoint the system is divided into two distinct subsystems, the Fresnel preprocessor and the spectral analysis subsystem. Details of these two subsystems are given in the following sections.

2.1 Fresnel Preprocessor

Figure 2 shows the layout of the Fresnel preprocessor. The acousto-optic cell interaction material is TeO_2, operated in the slow-shear mode, and has a bandwidth of W=50MHz, an access time T=40us, and a propagation velocity v=0.617mm/μs. For these parameters we compute the length of the cell as L=vT=24.7mm so that the optimum distance D between the acousto-optic cell and the point detector is found from (7) as 357mm. The light source is a GaAlAs, thermoelectric cooled laser diode; its output power output is 150mw at a wavelength of 853nm. Spherical lenses L_1 and L_2 expand the beam while cylindrical lens C_3 produces a line focus at the acousto-optic cell. Cylindrical lenses C_4 and C_5 image the line focus at the primary Fresnel plane, while the Fresnel transform is created in the direction parallel to the acousto-optic cell by allowing light to propagate into free space to plane P_2. The detector consists of a 200μm diameter, high-speed silicon avalanche photodiode, with its associated circuitry and mounting hardware. Since the physical size of the detector is too large to detect the interference fringe pattern at the primary Fresnel plane with good response, we use lens L_6 to magnify the spatial fringe pattern at plane P_2 to match the size of the detector which is located at plane P_3.

2.2 Spectrum Analysis Subsystem

The spectrum analysis subsystem can be implemented either electronically or optically; we demonstrate both techniques. The electronic spectrum analyzer is a conventional laboratory instrument. The layout of the optical spectral analysis system built for this application is shown in Figure 3. The light source is a HeNe laser with an output power level of 4mw at a wavelength of 632.8nm. Lenses L_1 and L_2 expand the beam while cylindrical lens C_3 produces a line focus at the acousto-optic cell. Lens L_5 produces the Fourier transform of the signal at the output plane in the horizontal direction while the lens combination C_4 and L_5 image the line focus from the cell onto the photodetector array. The parameters of the acousto-optic cell are the same as for the one used in the Fresnel preprocessor. The photodetector is a 128 x 8 tapped CCPD photodiode linear array. The frequency span covered by this spectrum analyzer arrangement is 33.3MHz, using 467 photodetector elements from the array, with a frequency spacing of 71.KHz; this frequency span

was adequate for our experimental purposes.

2.3 Data Acquisition and Control

Data acquisition and control is centralized within a Macintosh® FX computer. Two specialized I/O cards are used; one card provides an IEEE-488 interface capability while the other provides both digital and analog I/O capability. The IEEE-488 interface is connected to an arbitrary function generator, a spectrum analyzer, and a programmable oscilloscope. The digital I/O interface is connected to two digital direct-synthesizer boards which produce the cw signals needed to test system performance.

Flexibility of the data acquisition and control system is useful for generating arbitrary waveforms, particularly arbitrary pulses. We created several virtual instrument data acquisition and control programs utilizing a LabView®2 development environment on the computer. The virtual instruments are designed to communicate with the arbitrary function generator by means of a callable subfunction. This design allowed other higher-level virtual instruments, independent of the arbitrary function generator hardware communication constraints, to create the actual waveform needed for any particular test in terms of pulse width, pulse separation, and delay time.

3.0 Simulations and Experimental Results

The basic design of this system allows for a straightforward simulation of the entire system. An example of the output of the simulation program is given in Figure 4. In the upper part of Figure 4 we show the simulated output of the Fresnel preprocessor for a pair of pulses whose widths are 40ns and whose separation is 500ns. The output consists of an interference fringe pattern described analytically by (9), including the effects of the Gaussian illumination b(x). The time scale ranges from 0-2000 units, with the largest value corresponding to the largest possible duration of the signal as given by (8). In the time interval denoted by 0–350 units, neither pulse is illuminated and the output is zero. At about 350 units, the first pulse is illuminated, but there is no evidence of fringes because both pulses must be illuminated to produce interference fringes. At about 400 units, both pulses are illuminated, but the fringe contrast is low because the the

second signal is illuminated by a Gaussian function whose amplitude is low at the edges of the acousto-optic cell. As the two pulses move through the cell, the contrast increases and attains its maximum value when they are centered under the Gaussian illumination (at 1000 units). As the pulses continue through the acousto-optic cell, the modulation of the fringes decays because the output of the system is symmetric about the central spatial position. In the lower part of Figure 4 we show the one sided spectrum of the simulated signal. From (10) we calculate that the temporal frequency should be f=1.25MHz as confirmed by the position of the spectral line at 1.25MHz in the simulation result.

The upper part of Figure 5 shows the experimental output produced by the Fresnel preprocessor for the same parameters used in the simulation. The correspondence of the two results is generally good; the experimental results exhibit, of course, noise associated with the detector and its circuitry, and have a somewhat different envelope due to the fact that the pulses do not have perfectly sharp rise and fall times due to the finite bandwidth of the acousto-optic cell. In the lower part of Figure 5, we show the two-sided spectrum of the experimentally produced signal. This result is the output of a virtual spectrum analyzer implemented using LabVIEW®2, which captured the data from an electronic spectrum analyzer. Unless otherwise noted, all spectral data presented here use the settings as shown on the virtual instrument shown in Figure 5. The dc content of the signal is located at a relative position of 200 and we see spectral lines located at 200±25. Since the frequency scale is 20MHz/200 = 100kHz per division, the frequency of the detected signal from the Fresnel preprocessor is at 1.25MHz which confirm both the theoretical and simulation calculations.

3.1 Pulse Separation versus Frequency and Signal-to-Noise Ratio

Table 1 shows the calculated frequency and the measured frequency for various pulse separations; the pulse width is fixed at 100ns, with the pulse separation ranging from 1μs to 16μs. We see that the measured frequencies track the calculated frequencies to within an average error of about 1% and verifies the relationship between pulse separation and frequency as given by (10).

The signal-to-noise ratio is the factor limiting the useful separation between the pulses; this data is also included in Table 1. The noise in this instance is generated by the photodetector and its associated circuitry. The signal-to-noise ratio changes as a function of pulse separation for two reasons. First, as the pulse separation increases, the time duration of the sinusoidal part of the output signal decreases as indicated by (8) and (9); this shorter signal contains less energy and produces a lower signal-to-noise ratio. Second, as the pulse separation increases so that the value of τ increases, the product b(vt-L/2)b(vt-vt-L/2) of the two Gaussian illumination terms as given by (9) becomes smaller.

There is also a strong relationship between the peak signal-to-noise ratio and the pulse width. As the pulse width is increases, for a fixed pulse separation, the peak signal-to-noise ratio increases because the pulse intercepts more of the laser power. Very wide pulses, however, also imply larger minimum pulse separations and eventually lead to lower signal-to-noise ratios as discussed above in connection with Table 1. Throughout this experimental work, the signal-to-noise ratio was obtained from a spectrum analyzer for which the analysis bandwidth was set at 10kHz. This means that we integrated the results from 2-3 successive sets of pulses. As a result, the signal-to-noise ratio shown is higher than the actual value at the output of the optical system by about 4dB.

3.2 Frequency-Variant Processing

We need to distinguish the Fresnel transforms of short pulses from those produced by cw signals, such as narrowband radio sources. When two or more cw signals propagate through the acousto–optic cell, the difference frequency generated after square-law detection lies within the operating bandwidth of the detector in the Fresnel preprocessor. If the difference frequency is in the band of frequencies generated by pulses, the response to cw signals becomes a noise source that must be removed.

The Fresnel preprocessor is a space-variant processor because the output is a function of the detector position within a given Fresnel plane as well as being a function of the propagation distance D as shown by (1) or (2). Recall that light diffracted by a short pulse creates a

diverging wave front as shown in Figure 1, while light diffracted by a cw signal propagates as a plane wave whose angle with respect to the optical axis is a function of the cw frequency. The significance of this space-variant feature of Fresnel transforms is that the frequency detected due to the interference of plane wavefronts does not change as a function of distance D, whereas the frequency detected due to diverging wavefronts does change as a function of the distance D.

In our experiments we detected the Fresnel transform at two planes separated by a distance ΔD. The distance from the acousto-optic cell to the detector plane is conveniently altered without moving any of the components in the Fresnel preprocessor by inserting a Pechan prism between cylindrical lenses C_4 and C_5 in the diagram of Figure 2. This prism is non-deviating and forms an erect image, while introducing a fixed optical path difference to alter the value of D.

We distinguish cw signals from short pulse signals by taking measurements at two different values of D and subtracting the spectrum of one measurement from that of the other. Since spectral contributions from the cw signals are essentially the same for the two cases, they are removed by the subtraction process. The spectral contribution from the pulses is retained, however, because there is a known shift in the frequency of the sinusoidal interference pattern. The ability to distinguish and remove cw signals, through the space-variant property of the Fresnel transform requires, a frequency-variant operation on the output of the spectrum analyzer; this is a direct result of the fact that we are detecting the signal in a Fresnel domain as opposed to a space domain or to a Fourier domain.

3.2.1 Simulation of Frequency-Variant Processing

We performed a simulation to illustrate the basic space-variant properties of the Fresnel preprocessor and the frequency-variant post-detection operations. Figure 6a shows a possible spectrum over a frequency range of 20MHz, corresponding to the frequency range used experimentally. The spectrum consists of a sequence of cw signals, with the output produced by a pair of short pulses being located at 12MHz. Zero-mean random noise was added to this spectrum to simulate receiver noise; the signal cannot be seen in Figure 6a because the signal-to-noise ratio for this output is –8.3dB. A second spectrum, simulating that obtained with

the Pechan prism in the system to produce a new Fresnel detection plane, is shown in Figure 6b. The spectrum due to the cw signals is the same as that shown in Figure 6a, except that a new zero-mean random noise was used to simulate noise from the second photodetector. The simulated change in the distance D between measurements is 10% so that the signal due to the short pulses is now located at 13.2MHz; the signal-to-noise ratio for this case is -8.9dB.

In Figure 6c we show the results of subtracting the second spectral output from that of the first. As expected, the cw part of the spectrum, except for the noise, cancels because it is not a function of the distance D; energy due to the short pulses, however, does not cancel because the change in the distance D causes a frequency shift as predicted by (7) and (9). Evidence of the short pulses is shown by a positive spectral line at 12MHz and a negative spectral line, due to the subtraction process, at 13.2MHz, along with the residual spectrum from the zero-mean noise.

The most important result of spectral subtraction is that the signature of the short pulse response becomes a positive/negative doublet, with the unique property that the frequency span between the spectral lines of the positive/negative doublet is linearly related to the midband frequency. From (9) and (7) we find that the frequency of the output of the Fresnel preprocessor is

$$f = \frac{v^2}{\lambda D}\tau, \tag{11}$$

so that the differential in the frequency span as a function of D is

$$\Delta f = \frac{v^2}{\lambda}\left[\frac{-\Delta D}{D^2}\right]\tau = \frac{-\Delta D}{D}\frac{v^2}{\lambda D}\tau = \frac{-\Delta D}{D}f. \tag{12}$$

The frequency difference between the spectral lines of the positive/negative doublet, for fixed values of D and ΔD, is therefore a linear function of the frequency. Any processing applied to the data must therefore have a frequency variant response, because the positive/negative doublet

spectral lines are spaced close together at the low temporal frequencies of the output of the Fresnel preprocessor and far apart at high temporal frequencies.

Two processing algorithms were applied to the simulated data. The first is a nonlinear dual threshold technique implemented by applying a positive and a negative threshold directly to the difference data shown in Figure 6c; the threshold level is based on the standard deviation of the difference data. The next step is to apply an AND logic operation which requires that if a positive spectral line at a frequency f exceeds threshold then a negative spectral line at $f+\Delta f$ must also exceed threshold. We expected that this AND operation would help discriminate against occasional isolated spectral lines due to noise that exceed threshold. We refer to this entire processing operation as the dual threshold operation.

The second algorithm is a frequency-variant matched filtering operation applied to the data shown in Figure 6c. The frequency-variant matched filter consists of a positive/negative doublet whose frequency spacing is the linear function of the frequency given by (12). In Figure 6d we show the result of the frequency variant matched filtering operation. A threshold based on the standard deviation of the data shown in Figure 6d is then applied to output of the matched filter.

We tested the dual threshold and the matched filtering techniques against simulated data of the type given in Figure 6 containing 20 signals distributed among 1000 data samples. The first step is to determine the proper threshold level; this level is generally a function of the costs associated with missed signals, false alarms, probability of detection, and so forth. To determine the false alarm rate, we set the signal level equal to zero so that the output consisted of only noise. We then applied the two algorithms to the data and measured the false alarm rate.

The data shown in Figure 7a reveals that, since the residual noise has zero mean, the false alarm rate for the matched filter result is near the expected value of 0.5 when the threshold is set at zero. The false alarm rate for the dual threshold technique is 0.25 as expected because the logical AND operation requires the conjunction of two events, each of whose probability is equal to 1/2. The probability of false alarms decreases for both techniques as the threshold level increases. At threshold levels of $1.7\sigma_d$ for the dual threshold case and $2.4\sigma_m$ for the matched

filter case, where σ_d and σ_m are the standard deviations for the respective data sets, the measured probability of false alarm is $2(10^{-5})$. Based on the results of this simulation, as shown in Figure 7a, we conclude that the dual threshold technique is more robust with respect to false alarm rate.

The other important performance criterion, however, is the probability of signal detection. We set the threshold at the values noted above so that the probability of false alarm was the same for both techniques. We then increased the signal level, repeating the simulation 10 times at each threshold level; the averaged results are shown in Figure 7b. The probability of detection increases as the signal level increases for both techniques; either technique provides a probability of detection approaching one for signals whose levels exceeds 1.1. The output signal-to-noise ratio at this signal level is +9.2dB for the dual-threshold case and about 3dB more for the matched filter case. The input signal-to-noise ratio was -8.1 db for both cases; most of the processing gain is due to the subtraction of the cw signals. Although the matched filter produces a relatively small increase in processing gain because the space bandwidth product of the positive/negative doublet signal is only equal to two; the 3dB additional gain is sufficient to improve its performance significantly. From Figure 7b we see that the matched filter technique performs uniformly better than the dual-threshold technique. The matched filter is therefore the baseline technique in the experimental work described in the next section.

3.2.2 Experimental Results of Frequency Variant Processing

We experimentally tested the space variant property for the Fresnel transform by modifying the Fresnel preprocessor so that the Fresnel plane distance can quickly and repeatability be changed from one value to another. Instead of moving the detector to a new longitudinal position, we inserted a Pechan prism in the optical path to increase its effective optical path length by approximately 35mm, which is of the order of 10%. We measured the frequencies produced by pulses as described in Section 3.1, but this time with and without the prism, to get the results shown in Table 2. As predicted, the frequencies from the Fresnel preprocessor are shifted to lower values when the prism is in the Fresnel preprocessor because the optical path is

longer.

The input signals used in this test are two cw signals generated by local oscillators and two pulses produced by the arbitrary function generator. The spectrum of the output from the Fresnel preprocessor is recorded in two data files: one file contains the spectrum of the output signal obtained without the prism, as shown in Figure 8a, and the other file contains the spectrum of the output signal obtained with the prism, as shown in Figure 8b. The spectrum is offset so that the dc component occurs at the relative frequency position denoted by 200; the spectra are all symmetric about this frequency, as expected. Since the frequency span is ±20MHz, one unit on the horizontal scale represents 100kHz.The primary difference in the two spectra is that the spectral frequency component due to the cw signal does not change in frequency when the path length changes, whereas the spectral component due to the short pulses does change in frequency as the path length changes, as predicted by the theory.

The cw difference frequency produced spectral lines at about ±8MHz, while the short pulses produced the higher frequency components at about ±10MHz. The two types of signals are difficult to distinguish in this display because the Fresnel transform converts the short pulse energy into a cw signal whose frequency is a measure of the pulse separation while the cw signal remains a cw signal. Since the cw signal exists for all time, while that of the short pulses exists only during the time period produced by the overlap of the pulses, the spectral lines for the short pulses are somewhat broader. The same general comments apply to the data collected when the Fresnel plane is moved so that the optical path increased by 35mm as shown in Figure 8b. The positions of the spectral lines due to the cw signal have not changed, while those of the short pulses have moved to frequencies that are about 10% lower.

Spectral subtraction as described in Section 3.2.2 was applied to these experimental results to remove the effects of the cw signal and to enhance the detection of the short pulse information. The approach is to convert the logarithmic spectral scale, as produced by the spectrum analyzer, to a linear scale before they are subtracted. The results of this type of linear subtraction is shown in Figure 9a. In this case, the cw signals are removed and the spectral lines

due to the short pulses form a positive/negative doublet. These two spectral lines are at frequencies that differ by 10% due to the change in the distance of the Fresnel detection plane.

We note that the negative portion of the doublet has a smaller magnitude than the positive portion. There are two reasons for this difference: the prism transmittance is not 100% and the expanding light amplitude from the short pulses decreases as the distance to the Fresnel plane increases as suggested by Figure 1. We did not attempt to implement any form of gain control or equalization to correct this problem. Also, the dc frequency component is not as cleanly removed as the cw components; the dc component is much stronger than any other spectral component (it is saturated in the data shown in Figure 8) and small changes in the signal levels between the two detection planes can lead to a relatively large difference after spectral subtraction. The dc frequency can, of course, be masked from the output if necessary.

The application of a matched filter to the difference spectrum results in the output shown in Figure 9b. The generation of this result can be envisioned by convolving the data shown in Figure 9a with a negative/positive doublet, the convolution taking place by sliding the data over the filter from left to right. In Figure 9b, the earliest output of the filter is at the 400 position on the scale. The first negative response is due to the multiplication of the negative part of the matched filter with the positive spectral line. The strongest positive output is at 300, the position where the negative and positive part of the filter and the data overlap exactly. A weaker, negative response is due to the overlap of the smaller negative spectral line from the data overlaps with the positive part of the matched filter doublet.

Since the difference spectrum near the 200 position in Figure 9a is a single spectral line resulting from incomplete subtraction of the dc spectral component, the output in Figure 9b gives the impulse response of the matched filter (with the expected reversal of coordinates).The output near the 100 position on the negative frequency scale is redundant, but the result shown here illustrates that the response is the negative of that shown for the positive frequencies near the 300 position. The signal is detected by setting a threshold in the positive frequency region from 200 to 400.

In this experimental illustration of matched filtering, we did not implement the frequency-variant feature of the matched filter, as can be deduced by considering the response to the dc component, because the input data set originated from the electronic spectrum analyzer. If the spectrum were computed optically, with the results stored on a CCD array whose contents are converted to a video signal as demonstrated in Section 3.5, the frequency-variant requirement of the matched filter could be easily implemented by changing the spacing of the positive/negative matched filter doublet as the data is read out from the array.

3.3 Multiple Sets of Pulses

In some applications multiple sets of pulses may be present within the acousto-optic cell at the same time. To test the performance of the system under this condition, we generated five pulses within a time interval equal to about 40% of the cell transit time. By recognizing that fringe patterns will be generated from each pulse pair, we can predict that the number of combinations, or pulse pairs, available to generate frequencies at the output of the Fresnel preprocessor and, therefore, at the output of the spectrum analyzer. For our scenario, five pulses, taken two at a time, yield 10 possible frequencies.

The values of each of these frequencies are given in Table 3 for the combinations denoted by AB, BC, AC, CD, BD, AD, DE, CE, BE, and AE, where AB denotes the separation between pulses A and B. As before, we note a good correlation between the frequencies predicted and those measured. Figure 10 shows the spectrum of the output of the Fresnel preprocessor and the ten frequencies produced by the pulses are clearly evident, with frequency spacings as given in Table 3. The signal-to-noise ratio drops as a function of pulse separation as explained before, and the signal-to-noise ratio for the larger separations given by AE, BE, and CE are probably too low to be useful.

3.4 Optical Spectrum Analyzer Performance

As an alternative to using a conventional electronic spectrum analyze, we built an optical spectrum analyzer to measure the spectrum of the output from the Fresnel preprocessor as described in Section 2.2 and shown in Figure 3. We used a linear charge coupled photodiode

array containing 1024, 13μm x 13μm detectors to detect the spectrum. The photodetector array is read out at a video rate that converts photodetector position to time and is displayed on an oscilloscope. Each element on the photodetector array corresponds to a particular frequency. If the array is read out and displayed on an oscilloscope, the conversion factor from time to frequency is 0.156MHz/μs.

We fed the output from the Fresnel preprocessor to both the electronic and optical spectrum analyzers for comparison and show the results in Figure 11. The optical spectrum analyzer shows evidence of fixed pattern noise caused by clocking out the results from the photodetector array. This noise source is usually removed, in an operational system, by subtracting it from the output to recover the true data. The analysis bandwidths of both spectrum analyzers were set at 3KHz to obtain these results.

The optical spectrum analyzer, although perhaps not needed in all applications, is ideally suited for implementing the frequency-variant matched filtering described in Section 3.2 because it calculates the Fourier transform continuously and because it can implement the filter as the data is transmitted over the video output line .

4.0 Summary and Conclusions

We detected and distinguished pulses whose durations are as short as 20ns, which is at the theoretical limit for the 50MHz bandwidth of the experimental system, and separated by as little as 60 nanoseconds or as much as 17 microseconds. As the pulse duration increases, the time duration of the output of the Fresnel preprocessor decreases and its spectrum broadens. We enhanced the frequencies generated due to short pulse interactions while simultaneously de-emphasizing the frequencies generated by cw waveforms. We also tested the interactions among multiple pairs of pulses. Again, the frequencies that were measured correlated well with our developed theory.

We showed that two detector positions allowed us to separate short pulses from cw signals by using a frequency-variant matched filtering operation. Future work might consider how an array of photodetector elements, with each element at a different Fresnel plane should increase

the discriminatory power of the system. Another improvement would be to implement a refined version of the frequency-variant matched filter to more nearly match the signal waveforms and to implement the processing in real-time as the video data is clocked out of the photodetector array of the optical spectrum processing subsystem.

5. Acknowledgment

This research was performed under a grant from the U.S. Army Research Office.

References

1. A. VanderLugt, C. S. Anderson, and P. J. W. Melsa, "Time Delay Detection of Short Pulses by Fresnel and Fourier Transformations," Appl. Opt., **32**, xxxx-xxxx (1993) in Press

2. A. VanderLugt, "Fresnel Transforms and Bragg Cell Processors," Appl. Opt., **24**, 3846-3859 (1985)

Figure Caption List

Figure 1: Diffraction caused by two short pulses

Figure 2: Fresnel transform preprocessor

Figure 3: Optical spectrum analyzer

Figure 4: Results of the Fresnel processing using a simulation program

Figure 5: Experimental results verifying the simulation

Figure 6: Simulation of spectral subtraction to detect short pulses

Figure 7:Simulation results for the dual threshold and the matched filter approaches;

(a) False alarm rates, and (b) Probability of detection

Figure 8: Spectrum for two short pulses and two cw signals:

(a) without Pechan prism, (b) with Pechan prism

Figure 9: Spectral subtraction: (a) Logarithmic subtraction, (b) Linear subtraction

Figure 10: Spectrum of the output of the interferometric system when five pulses are present

Figure 11: Comparison of electronic and optical spectrum analyzers:

(a) Electronic spectrum analyzer, (b) Optical spectrum analyzer

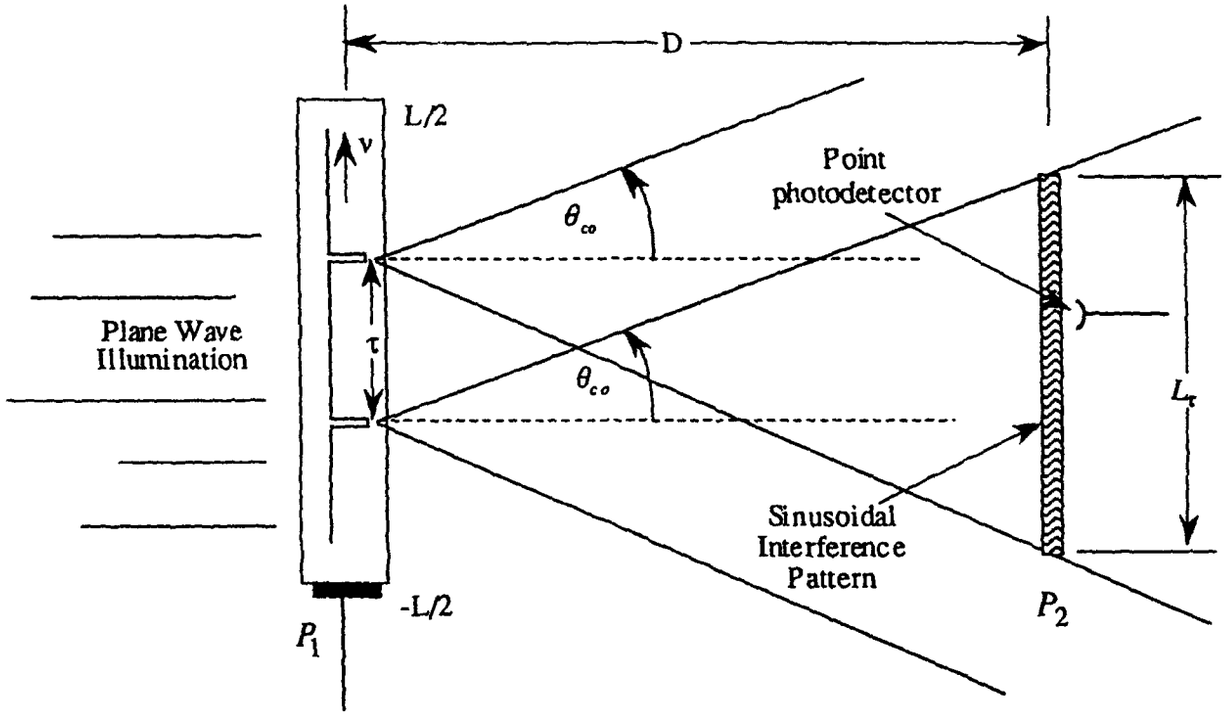


Figure 1: Diffraction caused by two short pulses

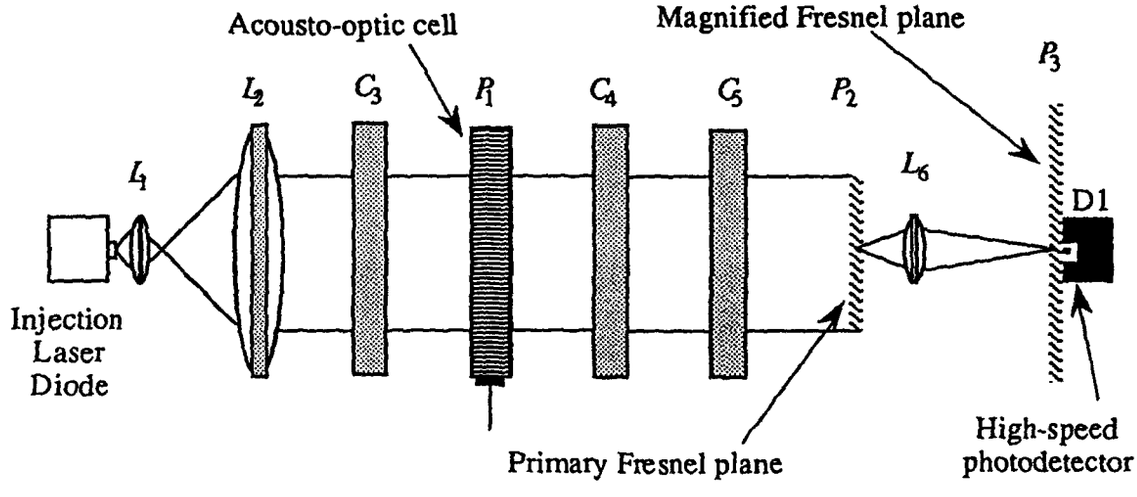


Figure 2: Fresnel preprocessor

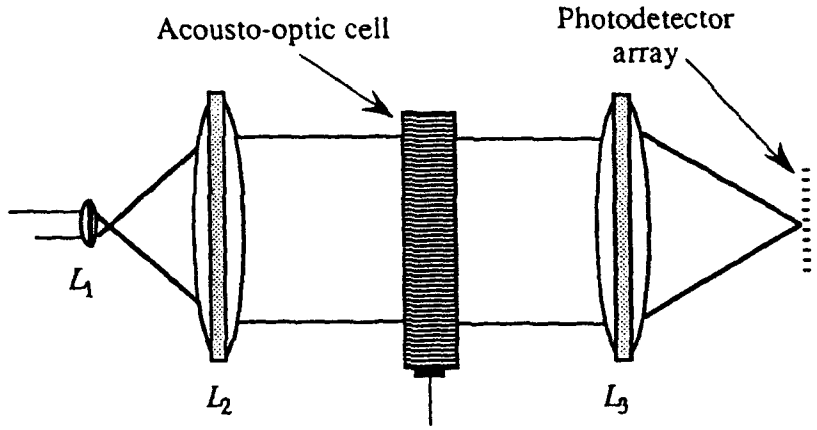


Figure 3: Optical spectrum analyzer

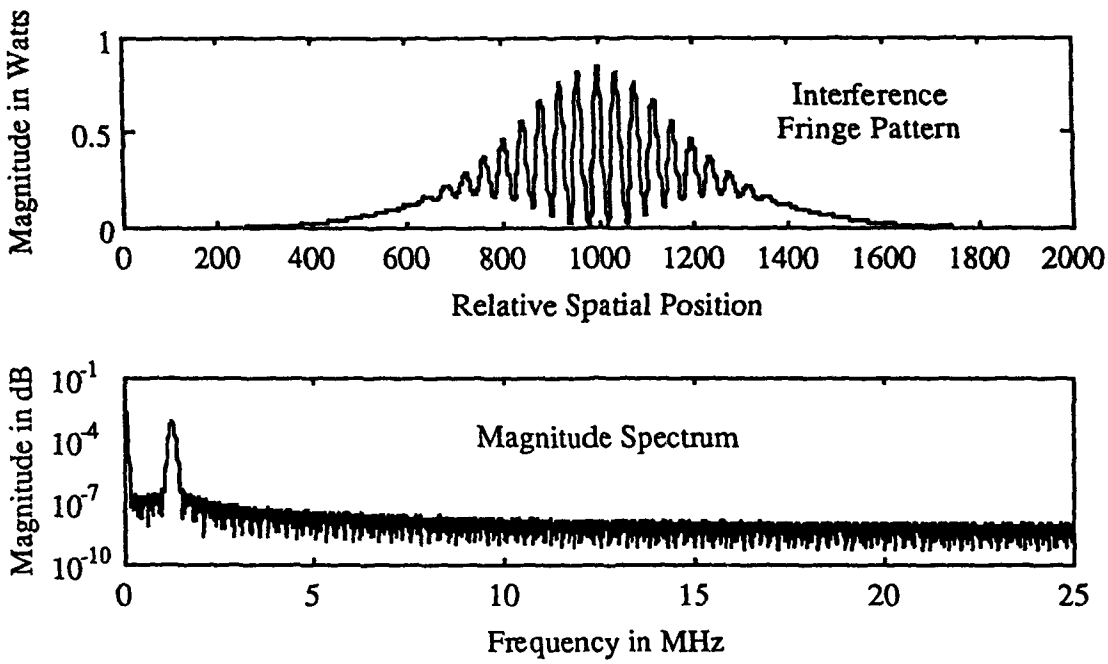
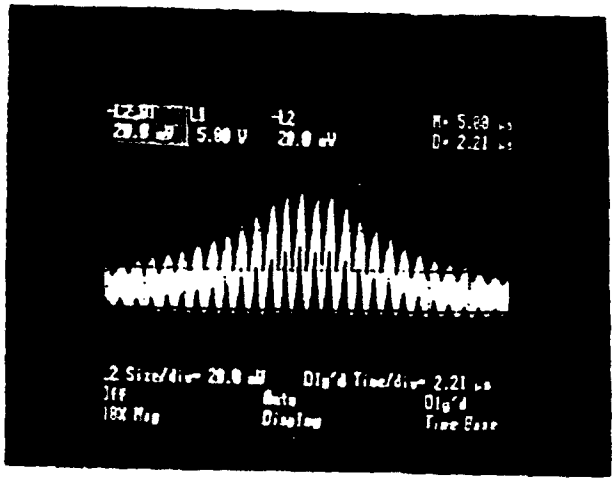


Figure 4: Results of the Fresnel processing using a simulation program



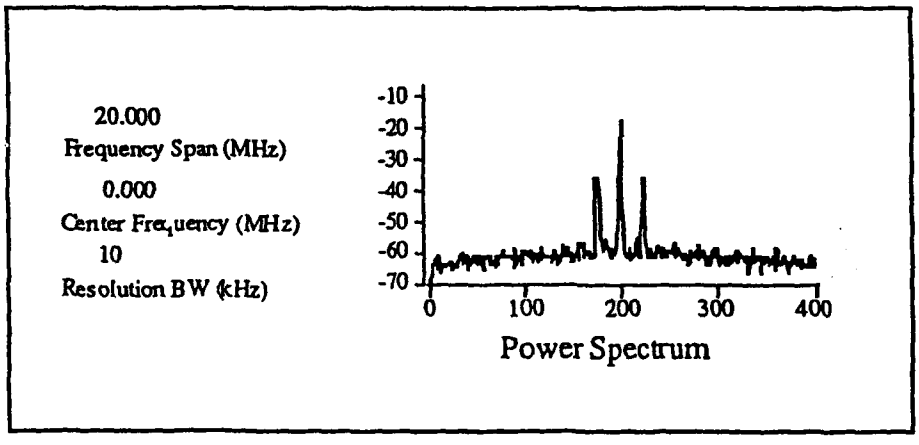


Figure 5: Experimental results verifying the simulation

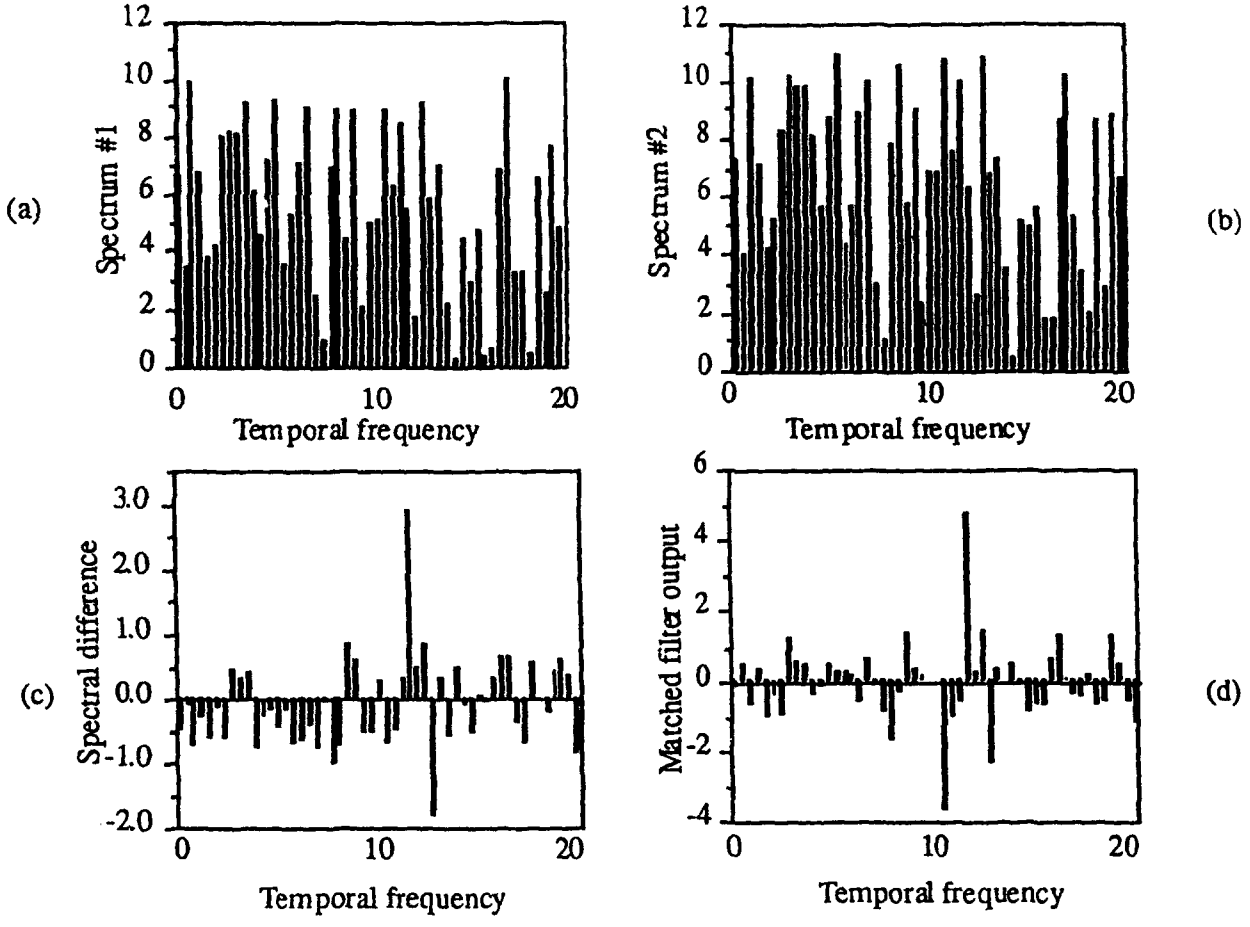


Figure 6: Simulation of spectral subtraction to detect short pulses:
(a) Data from normal mode, (b) Data with Pechan prism,
(c) Linear spectral subtraction, and (d) Matched filter output

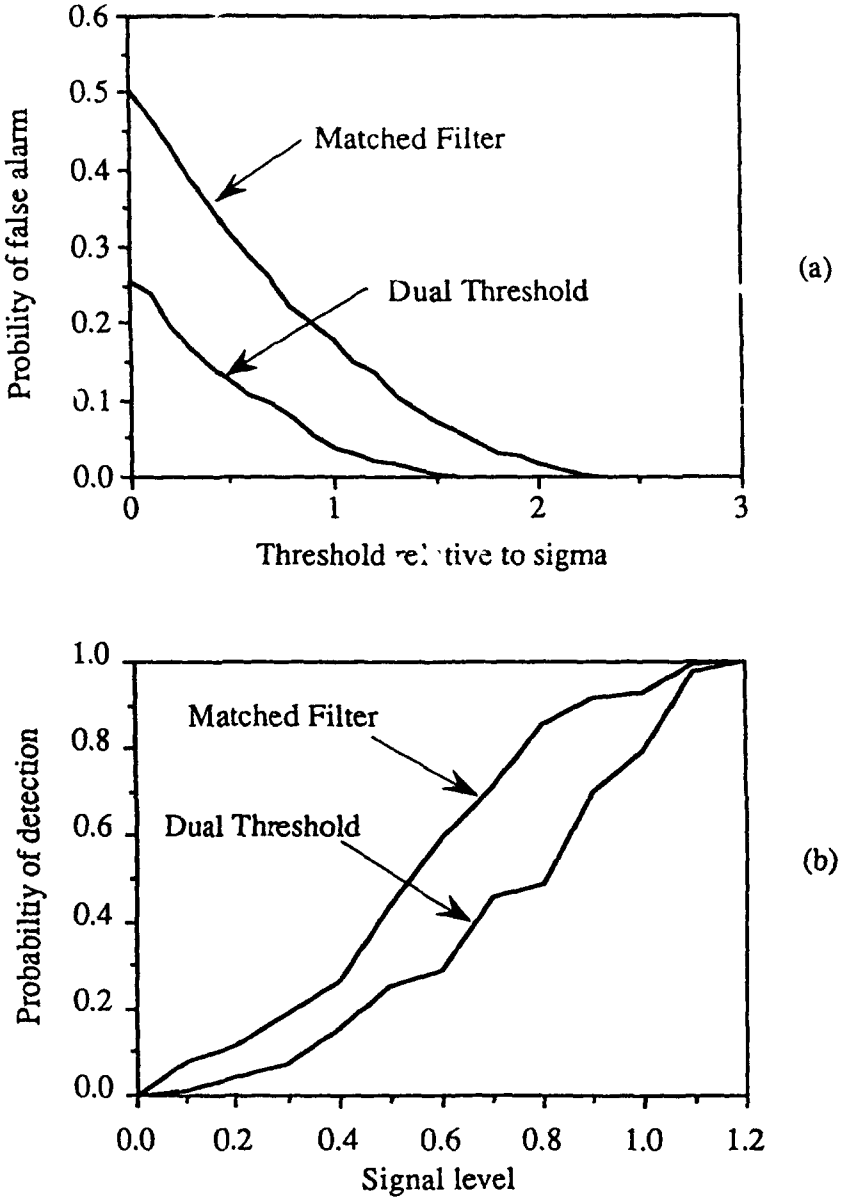


Figure 7: Simulation results for the dual threshold and the matched filter approaches; (a) False alarm rates, and (b) Probability of detection

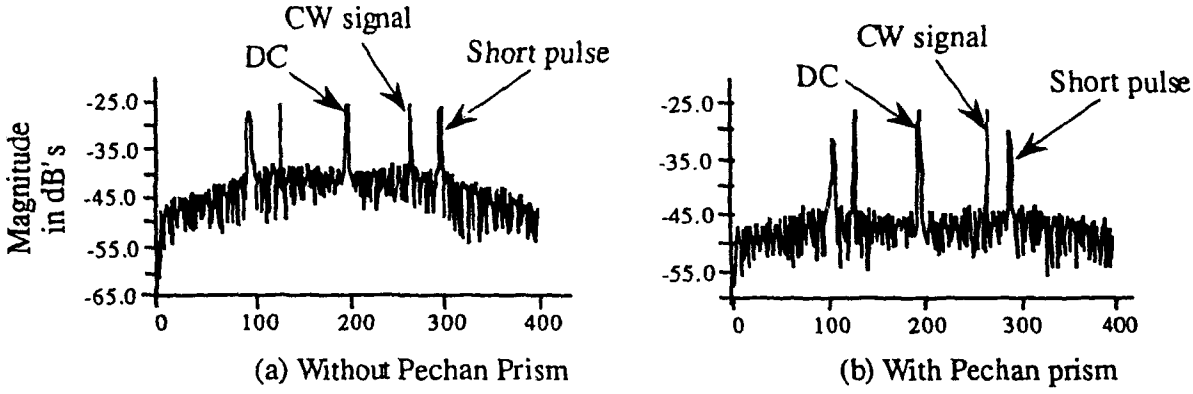


Figure 8: Spectrum for two short pulses and two cw signals

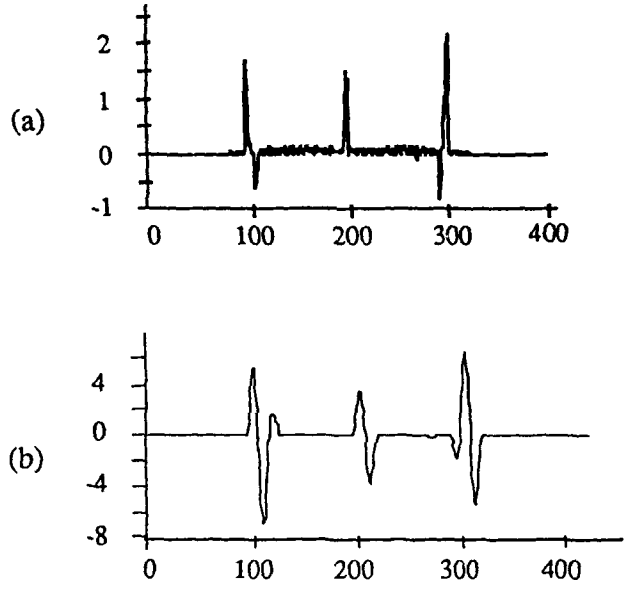


Figure 9: Subtraction of spectra in Figure 7:
(a) Spectral difference and (b) Matched filter ouptue

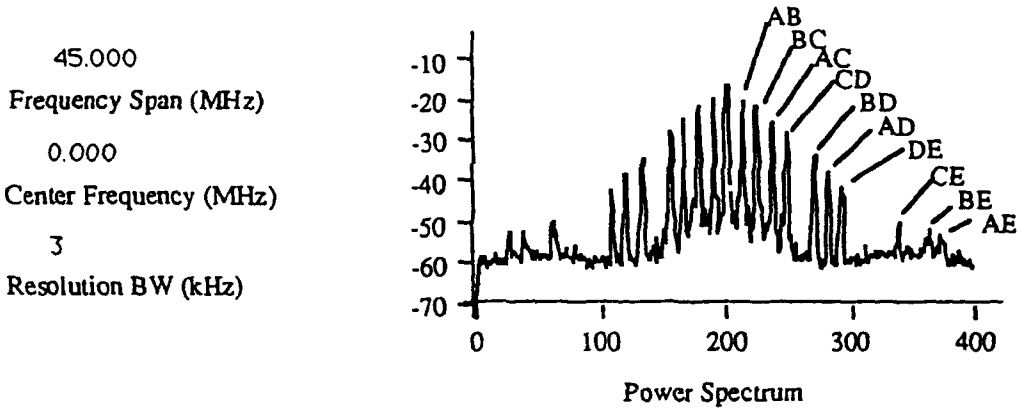


Figure 10: Spectrum of the output of the interferometric system when five pulses are present

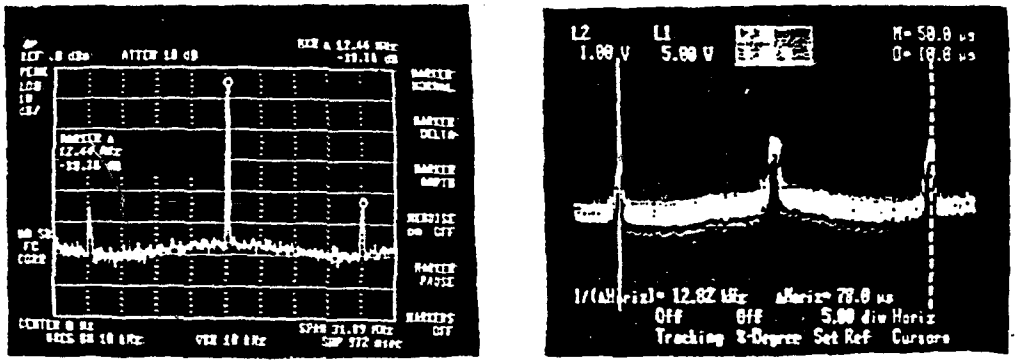


Figure 11: Comparison of electronic and optical spectrum analysis:
(a) electronic spectrum analyzer, (b) optical spectrum analyzer

Pulse separation (microseconds)	Calculated frequency (MHz)	Measured frequency (MHz)	SNR (dB)
1	1.25	1.3	41
2	2.5	2.58	40
4	5	5.03	34
6	7.5	7.58	27
8	10	10.13	19
10	12.5	12.68	14
12	15	15.15	12
14	17.5	17.63	9
16	20	20.1	3

Table 1: Calculated and measured frequency, along with SNR versus pulse separation.

Pulse separation (microseconds)	Measured frequency-without prism (MHz)	Measured frequency-with prism (MHz)	Average SNR (dB)
1	1.31	1.13	37
2	2.56	2.25	34.5
3	3.81	3.44	32
4	5.01	4.63	29
5	6.31	5.75	26.5
6	7.56	6.88	24
7	8.81	8.06	20
8	10.1	9.25	16
9	11.3	10.5	14

Table 2: Measured frequency for system with and without prism

Pulse pair	Pulse separation (microseconds)	Calculated frequency (MHz)	Measured frequency (MHz)	SNR (dB)
AB	1	1.25	1.25	38
BC	2	2.5	2.56	35
AC	3	3.75	3.75	33
CD	4	5	5.06	31
BD	6	7.5	7.56	25
AD	7.5	8.75	8.81	20
DE	8	10	10.13	18
CE	12	15	15.13	12
BE	14	17.5	17.69	10
AE	15	18.75	18.88	8

Table 3: Calculated and measured frequency, along with SNR, for multiple pulses.

APPENDIX L

DESIGN RELATIONSHIPS FOR ACOUSTO-OPTIC SCANNING SYSTEMS

Reprinted from Applied Optics.

Volume 31, pp. 4058-4068, 10 July 1992

a reprint from Applied Optics

Design relationships for acousto-optic scanning systems

A. VanderLugt and A. M. Bardos

We develop the general scanning relationships of an acousto-optic system by using both a purely geometric-optics and a physical-optics approach; each approach provides useful insights into the scanning relationships. The diffraction approach reveals that there are four basic scanning configurations: a long or short chirp scanner, either aperture or repetition-rate limited. The throughput rate for a scanner is always maximized if we use the short-chirp-scanning, repetition-rate-limited mode of operation. The maximum rate may be achieved with other configurations, but at the expense of a decrease in some of the other performance parameters. Examples are given of how these design relationships are used.

I. Introduction

When acousto-optic cells are driven by a sinusoidal signal, light is deflected at an angle proportional to the frequency of the signal and a lens can be used to focus the deflected light wave to a point. Scanning of the light at the focal plane is achieved by controlling the frequency of the applied signal. Although the basic concepts of acousto-optic scanning are documented,[1-11] the design of an acousto-optic scanner often represents a compromise in performance among several parameters such as spot size, number of samples per scan line, scan duty cycle, sample rate, and throughput rate. In this paper we describe a basic scanning system in which an acousto-optic cell is used to scan light along a line and develop the relationships that govern its performance. Although we concentrate on line-scanning applications, the results are also useful in signal-processing applications that use chirp signals.

In Section II, we review the basic parameters associated with acousto-optic cells in terms of the random-access scanning mode of operation. In Section III we develop the general scanning relationships from a purely geometric viewpoint, whereas the same results are developed in Section IV, using a physical-optics approach; each approach provides useful insights into the scanning relationships. In particular, the diffraction view reveals that there are four basic scanning configurations: a long or short chirp scan, either aperture or repetition-rate limited. The throughput rate for a scanner is always maximized if we use the short-scan, repetition-rate-limited mode of operation. The maximum rate may be achieved in other configurations, but at the expense of a decrease in some of the other performance parameters. Examples of how these design relationships are used are given in Section 5. In Section 6 we consider the effects of using a truncated Gaussian illumination instead of the uniform illumination that was used throughout our analyses to obtain closed-form solutions for the design relationships.

II. Background

Acousto-optic cells diffract light at an angle that is, to a good approximation, a linear function of the input frequency. Figure 1 shows the connections among the diffraction angle, the spatial frequency, the temporal frequency, and the acoustic wavelength. For a given acoustic velocity v and drive frequency f, the acoustic wavelength is $\Lambda = v/f$. Because the spatial frequency within the cell is $\alpha = 1/\Lambda$, the relationship between spatial and temporal frequencies becomes

$$\alpha = 1/\Lambda = f/v. \tag{1}$$

The diffraction angle is connected to the spatial and temporal frequencies by

$$\theta = \lambda/\Lambda = \lambda\alpha = \lambda f/v, \tag{2}$$

which nicely ties together the important parameters.

The acousto-optic cell, when driven by a cw frequency, behaves as a random-access beam deflector

A. VanderLugt is with the Department of Electrical and Computer Engineering, Daniels Hall, North Carolina State University, Raleigh, North Carolina 27695; A. M. Bardos is with Harris Government Systems Sector, Melbourne, Florida 32901.

Received August 6, 1991.
0003-6935/92/204058-11$05.00/0.

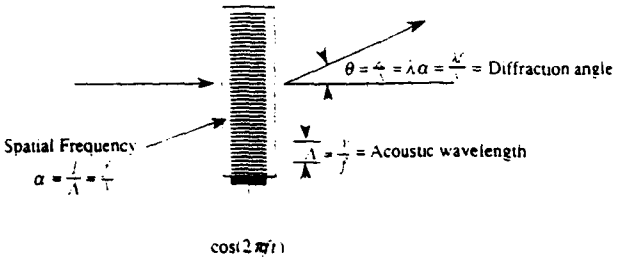


Fig. 1. Relationships among wavelength, spatial frequency, and diffraction angle.

that addresses a specific position at the focal plane of a lens. Figure 2 shows an acousto-optic cell at plane P_1 that has an acoustic velocity v and a length $L = vT$. We drive the cell with a signal $f(t) = \cos(2\pi f_k t)$ to access the kth spot position in the scan line. This signal produces a positive diffracted order whose Fourier transform is

$$F_+(\xi, t) = \int_{-\infty}^{\infty} f_+(x, t) \exp\left(j \frac{2\pi}{\lambda F} \xi x\right) dx, \qquad (3)$$

where $f_+(x, t) = \text{rect}(x/L) \exp[j2\pi f_k(t - x/v - T/2)]$, ξ is the coordinate in the Fourier plane, and F is the focal length of the lens. The Fourier transform of $f_+(x, t)$ is

$$F_+(\xi, t) = \int_{-L/2}^{L/2} \exp\left[j2\pi f_k\left(t - \frac{T}{2} - \frac{x}{v}\right)\right] \exp\left(j \frac{2\pi}{\lambda F} \xi x\right) dx$$

$$= \exp[j2\pi f_k(t - T/2)] \, \text{sinc}\left[\left(\frac{\xi}{\lambda F} - \frac{f_k}{v}\right) L\right], \qquad (4)$$

where we ignore amplitude-scaling factors. From Eq. (4) we find that the lens focuses light at the spatial position

$$\xi_k = (\lambda F/v) f_k \qquad (5)$$

at plane P_2. This result shows that the spot position is linearly proportional to the applied frequency. The first zero of the sinc function occurs at

$$\xi = d_0 = \lambda F/L. \qquad (6)$$

We use d_0 as both a measure of the spot size and as the Nyquist sampling distance at plane P_2.

We can also use cw frequencies to scan a light beam along a line in a stepwise fashion. A continuous scanning action provided by a chirp drive signal, however, provides higher line-scan rates. Figure 3 shows an acousto-optic cell driven by a signal whose frequency increases linearly from f_1 to f_2 in a time duration T_c. This frequency modulation signal, called a chirp signal, is characterized by

$$c(t) = \cos(2\pi f_1 t + \pi a t^2); \qquad 0 \le t \le T_c, \qquad (7)$$

where a is the chirp rate, expressed in hertz per second. From Eq. (7) we see that the instantaneous frequency $f_i(t)$ is

$$f_i(t) = \frac{1}{2\pi} \frac{\partial}{\partial t} (2\pi f_1 t + \pi a t^2)$$

$$= f_1 + at; \qquad 0 \le t \le T_c. \qquad (8)$$

Because the instantaneous frequency of the chirp sweeps over the bandwidth $W = f_2 - f_1$ of the acousto-optic cell in the time interval T_c, the chirp rate a is

$$a = W/T_c \qquad \text{for upchirp signals}$$

$$= -W/T_c \qquad \text{for downchirp signals.} \qquad (9)$$

The instantaneous frequency at the end of the acousto-optic cell is f_e; the frequency at the beginning of the cell is

$$f_b = f_e + aT$$

$$= f_e + WT/T_c. \qquad (10)$$

In the example shown the chirp frequency is increasing in time, which is generally called the upchirp condition; the chirp frequency may also decrease in time, which is generally called the downchirp condition.

The behavior of the scanning action produced by the cell can be explained by using the elementary diffraction theory and geometric ray tracing, or by using a diffraction integral. Each method provides useful insights into the scanning phenomena; we begin with the ray-tracing approach.

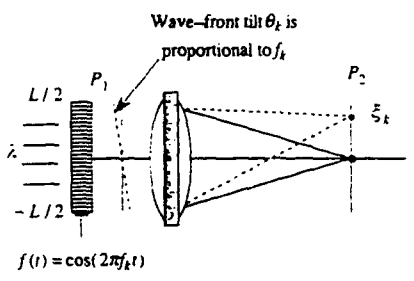


Fig. 2. Acousto-optic scanner.

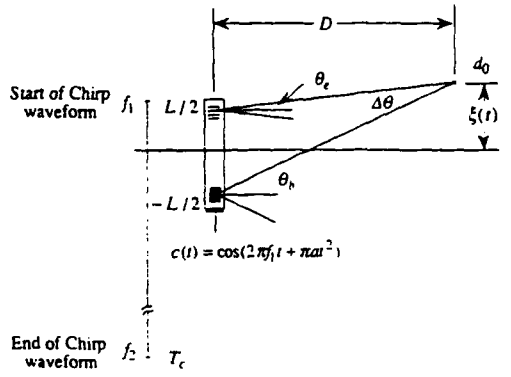


Fig. 3. Linear scanning with chirp waveform.

III. Linear Scanning with Chirp Waveforms: The Ray-Tracing Approach

The ray-tracing approach helps us to visualize where light from the acousto-optic cell is focused in the optical system and illustrates that a chirp signal in an acousto-optic cell produces a result similar to that of a laterally moving lens. Consider a ray trace for a stationary chirp segment that has just filled the acousto-optic cell, as shown in Fig. 3. The basic diffraction theory shows that the instantaneous frequency in the small region near the end of the cell produces an undiffracted waveform, indicated by a ray traveling parallel to the optical axis, along with positive and negative diffracted waveforms, indicated by rays that each make an angle θ_e with respect to the undiffracted light. The diffraction angle is related to the spatial frequency α_e and the temporal frequency f_e according to Eq. (2):

$$\theta_e = \alpha_e \lambda = \lambda f_e / v. \tag{11}$$

A similar relationship holds for the region near the beginning of the cell:

$$\theta_b = \alpha_b \lambda = \frac{\lambda f_b}{v} = \frac{\lambda (f_e + WT/T_c)}{v}. \tag{12}$$

When we trace the rays associated with the positive diffracted orders of each subregion within the cell, we find that they intersect at a distance D from the cell.

For the small-diffraction angles produced by the acousto-optic cell, the included angle between the extreme rays is

$$\Delta\theta = \theta_b - \theta_e = \lambda WT / v T_c, \tag{13}$$

so that the distance to the plane of focus is

$$D = L/\Delta\theta = v^2 T_c / \lambda W. \tag{14}$$

From Eq. (14) we find a relationship, frequently used in optical signal processing, between the chirp rate a and the radius of curvature D of the chirp wave front within the cell. By rearranging the factors, we find that

$$v^2/\lambda D = W/T_c = a. \tag{15}$$

The key geometric parameters of the acoustic signal, such as v, λ, and D, are on the left of the relationship, whereas the key drive-signal parameters, such as W and T_c are on the right.

The spot size, found from the fundamental diffraction relationship given by Eq. (3), is $d_0 = \lambda/\Delta\theta$, where $\Delta\theta$ is the angle between two converging rays:[12]

$$d_0 = \lambda/\Delta\theta = v T_c / TW. \tag{16}$$

For a given chirp duration, the spot size is inversely proportional to the time–bandwidth product of the cell.

The scanning velocity is most easily calculated by noting that the spot position, as a function of time, is

$$\xi(t) = -L/2 + D\theta_b(t) = -L/2 + D\frac{\lambda(f_e + Wt/T_c)}{v}, \tag{17}$$

where we used the general form of Eq. (12) to produce Eq. (17). The scanning velocity v_s is then

$$v_s = \frac{\partial}{\partial t}\xi(t) = D\frac{\lambda W}{v T_c}. \tag{18}$$

We use the value of D from Eq. (14) in Eq. (18) to find that

$$v_s = \frac{v^2 T_c}{\lambda W}\frac{\lambda W}{v T_c} = v. \tag{19}$$

The scanning velocity is therefore always equal to the acoustic velocity and cannot be controlled by any of the signal parameters.

The length of the scan line is equal to the product of the scan velocity and the active scanning time. Scanning begins at $t = T$ and continues until $t = T_c$, so that the active scan-time interval is $(T_c - T)$. The length of the scan line is therefore

$$L_s = v_s(T_c - T) = (T_c/T - 1)L, \tag{20}$$

so that the scan line is longer than the length of the acousto-optic cell. The number of samples in a scan line is

$$M = \frac{L_s}{d_0} = \frac{v(T_c - T)}{v T_c/TW} = \left(1 - \frac{T}{T_c}\right)TW, \tag{21}$$

so that, if $T_c \gg T$, the number of samples in a scan line approaches the time-bandwidth product of the cell.

IV. Linear Scanning with Chirp Waveforms: The Diffraction Approach

To control the scanning velocity, we must introduce a lens to the right of the acousto-optic cell. To analyze this condition we turn to the diffraction approach; in the process, we develop some new analytical tools and provide other useful insights. Furthermore, we can now more fully address the effects produced by the temporal characteristics of the chirp waveform.

In Fig. 4 we show a condition for which the acousto-optic cell aperture is small compared with the

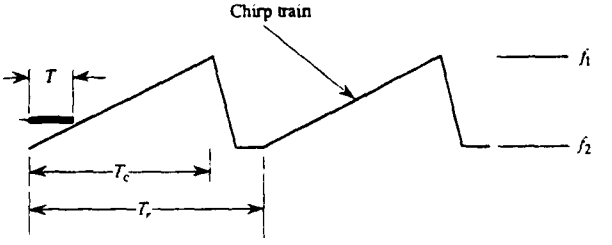


Fig. 4. Chirp train and its associated frequency–time relationship.

chirp period. The chirp duration T_c is the time between the lowest and highest frequencies of the chirp; the difference between the lowest and highest frequencies is the bandwidth $W = f_2 - f_1$ of the cell. The chirp-repetition period T_r is the time interval between a given point on one chirp segment and a similar point on the next chirp segment, e.g., the time between the highest frequencies in the two segments. The repetitive nature of the chirp train, shown in Fig. 4, is expressed by a time convolution of the chirp signal with an impulse train:

$$f(t) = c(t) * \sum_{n=-\infty}^{\infty} \delta(t - nT_r)$$

$$= \cos(2\pi f_1 t + \pi a t^2) * \sum_{n=-\infty}^{\infty} \delta(t - nT_r). \quad (22)$$

We consider the general case in which T_c may be larger than, comparable with, or even less than the aperture time T of the acousto-optic cell. We classify scanners according to two criteria: the active aperture time and the active scan time. If $T_c \geq T$, the active aperture time is governed by the time duration of the acousto-optic cell. We refer to this condition as the long-chirp scanner. If $T_c < T$, the active aperture time is governed by the duration of the chirp. We refer to this condition as the short-chirp scanner. For long-chirp scanning, the active scan time is $T_s = T_c - T$, as we noted in Section IV. For short-chirp scanning, the active scan time $T_s = T - T_c$. These two scan times can be combined to give a single active scan time of $T_s = |T - T_c|$.

Figure 4 shows that T_r must be greater than T_c when we use a voltage-controlled oscillator to generate the chip signal, because the signal does not return instantaneously from f_2 to f_1. A part of the chirp train is therefore not available for active scanning. We can, however, arrange for the chirp waveforms to overlap to an arbitrary extent by impulsing a surface acoustic-wave device that produces a chirp waveform at arbitrary repetition intervals. The active scan time is either T_s or T_r, whichever is shorter. If the active scan time is T_s, the system is aperture limited. If the active scan time is T_r, the system is repetition-rate limited. There are therefore four basic scanning configurations: a long or short chirp scan, either aperture or repetition-rate limited.

A. Long-Chirp, Aperture-Limited Scanner

The optical arrangement for a long-chirp, aperture-limited scanner is essentially the same as that shown in Fig. 2, except that the acousto-optic cell is now driven by a chirp signal represented by Eq. (7). For this exercise, we select the negative diffracted order whose amplitude just after the acousto-optic cell is

$$f_-(x, t) = \text{rect}(x/L) \exp\left\{-j\left[2\pi f_1\left(t - \frac{T}{2} - \frac{x}{v}\right) + \pi a\left(t - \frac{T}{2} - \frac{x}{v}\right)^2\right]\right\}, \quad T \leq t \leq T_c, \quad (23)$$

where the scan time starts at $t = T$ and ends when the trailing edge of the chirp segment arrives at the transducer. For the moment we assume that the lens is in contact with the acousto-optic cell; we show how to handle a finite separation later in this section. The positive lens is represented by the phase response

$$h(x) = \exp\left(j\frac{\pi}{\lambda F}x^2\right), \quad (24)$$

so that the light distribution to the right of the lens is

$$r_-(x, t) = \text{rect}(x/L) \exp\left\{-j\left[2\pi f_1\left(t - \frac{T}{2} - \frac{x}{v}\right) + \pi a\left(t - \frac{T}{2} - \frac{x}{v}\right)^2\right]\right\} \exp\left(j\frac{\pi}{\lambda F}x^2\right), \quad T \leq t \leq T_c. \quad (25)$$

The light distribution at any plane a distance D_f to the right of the lens is given by the Fresnel transform of $r_-(x, t)$:

$$F(\xi, t) = \int_{-\infty}^{\infty} r_-(x, t) \exp\left[-j\frac{\pi}{\lambda D_f}(\xi - x)^2\right]dx. \quad (26)$$

We substitute Eq. (25) into Eq. (26) to find that

$$F(\xi, t) = \int_{-\infty}^{\infty} \text{rect}(x/L) \exp\left\{-j\left[2\pi f_1\left(t - \frac{T}{2} - \frac{x}{v}\right) + \pi a\left(t - \frac{T}{2} - \frac{x}{v}\right)^2\right]\right\} \exp\left(j\frac{\pi}{\lambda F}x^2\right) \times \exp\left[-j\frac{\pi}{\lambda D_f}(\xi - x)^2\right]dx, \quad T \leq t \leq T_c. \quad (27)$$

We use Eq. (15) in Eq. (27) to find that

$$F(\xi, t) = \exp[-j2\pi f_1(t - T/2)] \times \int_{-\infty}^{\infty} \text{rect}(x/L) \exp(j2\pi f_1 x/v) \times \exp\left\{\left[-j\frac{\pi v^2}{\lambda D}\left(t - \frac{T}{2} - \frac{x}{v}\right)^2\right]\right\} \times \exp\left(j\frac{\pi}{\lambda F}x^2\right) \exp\left[-j\frac{\pi}{\lambda D_f}(\xi - x)^2\right]dx$$

$$= \exp(j\phi) \int_{-L/2}^{L/2} \exp\left\{j\frac{\pi x^2}{\lambda}\left[\frac{1}{F} - \frac{1}{D} - \frac{1}{D_f}\right]\right\} \times \exp\left\{j\frac{2\pi x}{\lambda}\left[\frac{v(t - T/2)}{D} + \frac{\xi}{D_f} + \frac{\lambda f_1}{v}\right]\right\}dx, \quad T \leq t \leq (T_c - T), \quad (28)$$

where we collect all phase factors that are not functions of x into ϕ. Note that the chirp rate $a = W/T_c$ is positive when we use the upchirp mode, as we do

here, and negative when we use the downchirp mode of modulation.

The focal position occurs where the light intensity is highest; this condition arises when the integral has its maximum value. The integral from Eq. (28) clearly has its maximum value when the integrand is set equal to one. Let us begin, however, by setting the value of the exponential that is quadratic in x equal to one. The first condition necessary to obtain focus is therefore that

$$\frac{1}{F} - \frac{1}{D} - \frac{1}{D_f} = 0 \tag{29}$$

or that

$$D_f = \frac{DF}{D - F}. \tag{30}$$

When Eq. (30) is satisfied, Eq. (28) produces the spatial light distribution at the focal plane given by

$$F(\xi, t) = \int_{-L/2}^{L/2} \exp\left\{ j \frac{2\pi x}{\lambda} \left[\frac{v(t - T/2)}{D} + \frac{\xi}{D_f} + \frac{\lambda f_1}{v} \right] \right\} dx$$

$$= L \operatorname{sinc}\left[\frac{v(t - T/2)L}{\lambda D} + \frac{\xi(D - F)L}{\lambda F D} + \frac{f_1 L}{v} \right], \quad T \le t \le T_c. \tag{31}$$

The position of this scanning spot at any instant in time is found by setting the argument of the sinc function equal to zero, which is equivalent to setting the value of the exponential in Eq. (28) that is linear in x equal to 1:

$$\xi = -\frac{\lambda f_1 DF}{v(D - F)} - \frac{v(t - T/2)F}{D - F}, \quad T \le t \le T_c. \tag{32}$$

The spot position at the beginning of scan, when $t = T$, is

$$\xi_b = -\frac{\lambda f_1 DF}{v(D - F)} - \frac{v(T/2)F}{D - F}, \tag{33}$$

whereas the spot position at the end of scan, when $t = T_c$, is

$$\xi_e = -\frac{\lambda f_1 DF}{v(D - F)} - \frac{v(T_c - T/2)F}{D - F}, \tag{34}$$

so that the length of scan is

$$L_s = |\xi_e - \xi_b| = \left| \frac{v(T_c - T)F}{D - F} \right|. \tag{35}$$

The scanning velocity is readily obtained from Eq. (32) as

$$v_s = \frac{\partial \xi}{\partial t} = -\frac{vF}{D - F}. \tag{36}$$

The scanning velocity v_s has the same or opposite direction as v, depending on the value D of the wave-front radius of curvature. If we take the velocity of the acoustic wave as positive, the rules are as follows:

1. When D is positive and greater than F, as for the case analyzed in this section, v_s is negative so that the spot moves in the negative x direction. In this case, the chirp signal in the acousto-optic cell is equivalent to a negative lens whose focal length is longer than that of the positive lens. The light therefore focuses at a plane to the right of the lens because Eq. (30) shows that D_f is positive.
2. When D is negative, the scanning spot moves in the positive x direction. In this case, the focal length of the chirp is positive, the net result, as confirmed by Eq. (30), is equivalent to that of two positive lenses working together.
3. When D is positive and less than F, the scanning velocity is positive. Therefore the spot moves in the same direction as v, but the light does not focus anywhere to the right of the lens. In this case the focal length of the chirp is equivalent to a negative lens whose focal length is shorter than that of the positive lens, and Eq. (30) confirms that D_f is negative.

The rules are illustrated in Fig. 5. The negative diffracted order satisfies the first rule. The value of D is positive, which is equivalent to stating that the focal power of the chirp is negative, so the scan plane lies to the right of the plane at which the undiffracted light is focused. As the chirp signal flows through the acousto-optic cell the ray angles increase in the negative direction, leading to a negative scan velocity. The start of the scan position as seen from Eq. (33) is negative, as is the end of the scan position as seen from Eq. (34).

The positive diffracted order satisfies the second rule. The value of D is negative, which equivalent to stating that the focal power of the chirp is positive, so the scan plane lies to the left of the plane at which the undiffracted light is focused. As the chirp signal flows through the acousto-optical cell, the ray angles increase in the positive direction, leading to a positive scan velocity. The start of the scan position as seen from Eq. (33) is positive, as is the end of the scan

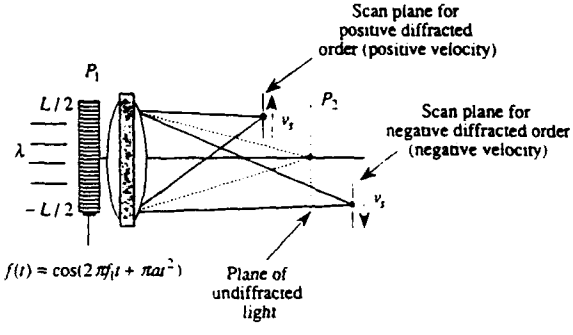


Fig. 5 Relationship of scanning action for upchirp signals.

position, as seen from Eq. (34). The positive diffracted order exists for all values of D because the equivalent focal length of two positive lenses must be positive.

Rule 3 applies to a special case for the negative diffracted order and states that the light may not focus for certain values of D. For example, as the value of D approaches F, the negative power caused by the chirp signal subtracts from the positive power of the lens, and the focal plane for the negative diffracted order recedes to infinity. When $D = F$ the two focal powers cancel exactly and the focal plane is at infinity. As we stated in the third rule, the negative diffracted order does not focus at any plane to the right of the lens if D is positive and less than F; instead, it generates a virtual scan plane.

When we drive the acousto-optic cell with a downchirp signal of the form

$$c(t) = \cos(2\pi f_2 t - \pi a t^2), \tag{37}$$

instead of with the upchirp signal as given by Eq. (7). The same general results apply except that we interchange f_1 and f_2 to account for the different starting frequency and replace D by $-D$ to account for the downchirp nature of $c(t)$. The roles of the two scan planes shown in Fig. 5 are now interchanged, so the negative diffracted order focuses to the left of the positive diffracted order. As we expected the scan velocities also have opposite signs, so the spots scan toward the optical axis instead of away from the optical axis. The choice of the upchirp or downchirp mode conveniently allows the system designer to select the scanning direction.

Equation (36) shows that we can control the scanning velocity by selecting the value of the focal length of the lens. For a desired scan velocity, the required focal length of the lens is

$$F = \frac{D}{1 - v/v_s}. \tag{38}$$

The signs of D and v_s can combine, according to the rules, only to cause the focal length of the lens to be positive. As $v \to v_s$ the lens focal length is infinite, as we discussed in Section 3.

The size of the scanning spot is obtained from Eqs. (16) and (30):

$$d_0 = \frac{\lambda}{L/D_f} = \frac{\lambda DF}{(D - F)L}. \tag{39}$$

As the chirp radius of curvature $D \to \infty$ the spot size tends to a value of $d_0 = \lambda F/L$, as we expected, because the chirp waveform then contributes no power to the system and the lens alone acts on the diffracted light.

The number of resolvable elements or samples in the scan line is

$$M = \frac{L_s}{d_0} = \left|\frac{v(T_c - T)L}{\lambda D}\right|. \tag{40}$$

We use Eq. (15) in Eq. (40) to find that

$$M = \left|1 - \frac{T}{T_c}\right| TW. \tag{41}$$

just as we found from the geometric analysis.

The scan duty cycle is defined as the ratio of the active scan time divided by the repetition period:

$$U = \frac{\min(T_c, T_r)}{T_r} = \frac{T_c - T}{T_r}. \tag{42}$$

The sample rate at which samples are recorded is given by the ratio of the scan velocity to the spot size:

$$R_s = \frac{|v_s|}{d_0} = \frac{T}{T_c} W. \tag{43}$$

The throughput rate is the average number of samples recorded per unit time and is the product of the sample rate and the scan duty cycle,

$$R_t = UR_s = \left|1 - \frac{T}{T_c}\right| \frac{T}{T_r} W. \tag{44}$$

where we have used Eqs. (42) and (43) to produce Eq. (44).

If the acousto-optic cell and the lens are not in contact, as suggested by Fig. 5, we can use the thin-lens formula to find the equivalent position of the plane at which the light is focused. If the separation between two thin lenses having powers $K_1 = 1/F_1$ and $K_2 = 1/F_2$ is d_{12}, the equivalent power of the combination is[13]

$$K_{eq} = K_1 + K_2 - d_{12}K_1K_2. \tag{45}$$

We associate the power of the chirp signal in the acousto-optic cell with K_1, so that $K_1 = 1/D$, and the lens power with K_2. The net power of the combination gives the distance to the scan plane from the acousto-optic cell as $D_f = 1/K_{eq}$.

B. Short-Chirp, Aperture-Limited Scanner

In the short-chirp scanner for which T_c is less than T, the active aperture time is limited by the chirp duration T_c. Suppose that one of the chirp signals in the chirp train is completely within the acousto-optic cell, as shown in Fig. 6. Some of the design relationships from Eq. (31) onward are modified somewhat in this case. For example, the start of the scan time is T_c and the end of the scan time is $T - T_c$, whereas the integrations are over a spatial range $L_c = vT_c$. We

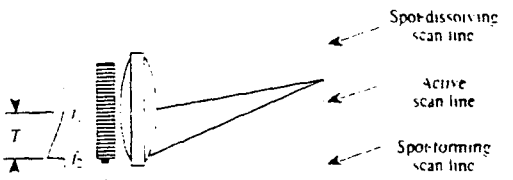


Fig. 6. High duty cycle scanner.

study the diffraction phenomenon as the chirp transitions into and out of the cell at the end of this section.

The new form of Eq. (31) is

$$F(\xi, t) = \int_{-L/2}^{L-L/2} \exp\left\{ j \frac{2\pi x}{\lambda} \left[\frac{v(t - T/2)}{D} + \frac{\xi}{D_f} + \frac{\lambda f_1}{v} \right] \right\} dx$$

$$= L_c \operatorname{sinc}\left[\frac{v(t - T/2)L_c}{\lambda D} + \frac{\xi(D - F)L_c}{\lambda F D} + \frac{f_1 L_c}{v} \right],$$

$$T_c \le t \le (T - T_c), \quad (46)$$

where we ignore unessential amplitude and phase factors. As before, the position of the scanning spot at any instant in time is found by setting the argument of the sinc function equal to zero, from which we find that

$$\xi = -\frac{\lambda f_1 DF}{v(D - F)} - \frac{v(t - T/2)F}{D - F},$$

$$T_c \le t \le (T - T_c). \quad (47)$$

The spot position of the beginning of scan, when $t = T_c$, is

$$\xi_b = -\frac{\lambda f_1 DF}{v(D - F)} - \frac{v(T_c/2)F}{D - F}, \quad (48)$$

and the spot position of the end of scan, when $t = T - T_c$, is

$$\xi_e = -\frac{\lambda f_1 DF}{v(D - F)} - \frac{v\left(\frac{T}{2} - T_c\right)F}{D - F}, \quad (49)$$

so that the length of scan is

$$L_s = |\xi_e - \xi_b| = \left| \frac{v(T - T_c)F}{D - F} \right|. \quad (50)$$

The scanning velocity is still given by Eq. (36), but the spot size is slightly different:

$$d_0 = \frac{\lambda}{L_c/D_f} = \frac{\lambda DF}{(D - F)L_c}, \quad (51)$$

which is similar to Eq. (39), except that L is replaced by L_c because the spot size is now determined by the length of the chirp, not the length of the acousto-optic cell. The number of samples in the scan line is

$$M = \frac{L_s}{d_0} = \left| \frac{v(T - T_c)L_c}{\lambda D} \right|. \quad (52)$$

We use Eq. (15) in Eq. (52) to find that the number of samples in the scan line is

$$M = \left(1 - \frac{T_c}{T}\right) TW. \quad (53)$$

Because the chirp duration is less than the cell duration, the maximum number of samples as set by the time-bandwidth product of the acousto-optic cell is not obtained.

The scan duty cycle for this configuration is

$$U = \frac{T - T_c}{T}, \quad (54)$$

and the sample rate is

$$R_s = \frac{v_s}{d_0} = W. \quad (55)$$

which is the maximum achievable sample rate. The throughput rate is

$$R_t = U \frac{|v_s|}{d_0} = \left(\frac{T - T_c}{T}\right) W, \quad (56)$$

which is obtained in a fashion similar to that used to produce Eq. (44).

In Fig. 6 we show the situation where at least one period of a chirp signal is fully in the acousto-optic cell and the scanning spots are therefore well formed. We now examine the scanning spot shape and position as the chirp segments enter and leave the cell. The light from these transition times is located in regions just before and just after the scan line. To account for the spot-forming condition, we modify the limits of integration in Eq. (28):

$$F(\xi, t) = \int_{-L/2}^{-L/2+vt} \exp\left\{ j \frac{\pi x^2}{\lambda} \left[\frac{1}{F} - \frac{1}{D} - \frac{1}{D_f} \right] \right\}$$

$$\times \exp\left\{ j \frac{2\pi x}{\lambda} \left[\frac{v(t - T/2)}{D} + \frac{\xi}{D_f} + \frac{\lambda f_1}{v} \right] \right\} dx,$$

$$0 \le t \le T_c, \quad (57)$$

which is applicable for a chirp waveform as it just enters the cell. The limits of integration show that the integral is over a small spatial region when t is small and that the region of integration increases linearly for $0 \le t \le T_c$. As before, we set the value of D_f so that the quadratic term in x is equal to unity, leaving the integral

$$F(\xi, t = \int_{-L/2}^{-L/2+vt} \exp\left\{ j \frac{2\pi x}{\lambda} \left[\frac{v(t - T/2)}{D} + \frac{\xi}{D_f} + \frac{\lambda f_1}{v} \right] \right\} dx$$

$$= vt \operatorname{sinc}\left\{ \left[\frac{v(t - T/2)}{D} + \frac{\xi}{D_f} + \frac{\lambda f_1}{v} \right] vt/\lambda \right\},$$

$$0 \le t \le T_c, \quad (58)$$

where we have ignored unimportant constants.

The behavior of this sinc function, whose argument is quadratic in the time variable, has some interesting features that are exhibited in the dashed-line region of Fig. 6 where the spot is first formed:

1. The amplitude of the sinc is small for small values of t, as we expected from a consideration of the

region of integration, and reaches a limit that is proportional to vT_r when the chirp in the cell is fully illuminated.

2. The centroid of the spot as a function of time is located at

$$\xi = -\frac{\lambda f_1 DF}{v(D-F)} - \frac{v(t - T/2)F}{D-F}, \qquad 0 \le t \le T_c, \tag{59}$$

just as in Eq. (47) but with a slightly different time interval of validity. The spot position when the chirp enters the cell is

$$\xi_b = -\frac{\lambda f_1 DF}{v(D-F)} - \frac{v(-T/2)F}{D-F}, \tag{60}$$

and its position when it is fully in the cell is

$$\xi_e = -\frac{\lambda f_1 DF}{v(D-F)} - \frac{v(T/2)F}{D-F}. \tag{61}$$

By comparing Eqs. (60) and (61) with Eqs. (48) and (49), we see that the end positions of the scanning spot produced by the chirp as it enters the cell are displaced a distance L_s below the corresponding positions of the active scan line.

3. The most interesting feature of the sinc function is that the spot size changes its size continuously as the chirp enters the cell. The spot size is determined by finding the position of the first zero of the sinc function relative to its centroid. This distance is

$$\Delta\xi = d_0 = \frac{\lambda D_f}{vt}, \qquad 0 \le t \le T_c. \tag{62}$$

From Eq. (58) we see that the sinc function is infinitely broad when $t = 0$, but its amplitude is zero. As time increases, the spot moves toward the active scanning region and its size decreases while its amplitude increases. The rate at which the spot size decreases as the centroid moves closer to the beginning of the active scan position is just sufficient to keep the light from spilling into the active scanning region prematurely. When the chirp has fully entered the cell at $t = T_c$ the spot has full resolution, and the active scanning begins as the chirp travels through the remainder of the acousto-optic cell.

As the chirp segment leaves the cell, the spot dissolves in an order that is a reversal of its formation. The spot gradually loses intensity as its broadens until it reaches the end of the spot-dissolving scan line shown in Fig. 6.

C. Long-Chirp, Repetition-Rate-Limited Scanner

To achieve a high throughput rate, we need a high scan duty cycle; Eq. (42) shows that we want the active scan time to equal the repetition period of the chirp train. Suppose that we use a surface acoustic device to generate a chirp segment whenever it is driven by an impulse function. By controlling the time of the impulses, we produce chirp segments with any desired repetition period T_r. Depending on the ratio of the chirp period to the chirp duration, one or more overlapping chirp segments may be in the cell at the same time. If the response of the cell is linear, the chirp signals do not interfere and the only effect of the overlapping chirps is to lower the useful diffraction efficiency. Because the chirp signal is on a carrier frequency, a nonlinear response from the cell produces higher-order terms that are easily eliminated by spatial filters.

In this section we assume that the chirp segments overlap so that $T_r \le T_c$, and so that $T_r > T$. The system parameter that is changed is the scan length, which through a line of analysis similar to that given in Subsection IV.A., is

$$L_s = |\xi_e - \xi_b| = \left|\frac{vT_rF}{D-F}\right|. \tag{63}$$

The spot size is still determined by the cell aperture because $T_r > T$:

$$d_0 = \frac{\lambda DF}{(D-F)L}, \tag{64}$$

so that the number of samples in a scan line is

$$M = (T_r/T_c)TW, \tag{65}$$

which achieves its maximum value when $T_r = T_c$. The scan duty cycle for this configuration is

$$U = \frac{\min(T_s, T_r)}{T_r} = \frac{T_r}{T_r} = 1, \tag{66}$$

as we expected. The sample rate and the throughput rate are equal in this configuration at

$$R_t = UR_s = \frac{|v_s|}{d_0} = \frac{T}{T_c}W. \tag{67}$$

As before, we see that the throughput rate is maximized only when $T_c = T$. To achieve this condition, we consider the final of the four basic configurations.

D. Short-Chirp, Repetition-Rate-Limited Scanner

In this configuration the chirp segments also overlap so that $T_r \le T_c$, and we assume that $T_r \le T$. The scan length, found through a line of analysis similar to that given in Subsection IV.C., is

$$L_s = |\xi_e - \xi_b| = \left|\frac{vT_rF}{D-F}\right|. \tag{68}$$

The spot size is now determined by the active scan aperture $L_r = vT_r$ so that

$$d_0 = \frac{\lambda DF}{(D-F)L_c}, \tag{69}$$

so that the number of samples in a scan line is

$$M = T_r W, \tag{70}$$

which achieves its maximum value when $T_r = T$. The scan duty cycle for this configuration is

$$U = \frac{\min(T_s, T_r)}{T_r} = \frac{T_r}{T_r} = 1. \tag{71}$$

as we expected. The sample rate and the throughput rate are equal in this configuration at

$$R_t = UR_s = \frac{|v_s|}{d_0} = W. \tag{72}$$

In this configuration, the throughput rate is maximized independently of the values of T_r or T_c, provided that the constraints necessary to implement the short-chirp, repetition-rate-limited scanner are satisfied.

E. Summary of Scanner Performance Criteria

Table 1 gives a summary of the important performance parameters of the four basic scanning configurations and is a useful aid in beginning a design. For example, some applications require a high throughput rate R_t. The maximum rate of W samples per second can always be achieved with a short-chirp scanner that is repetition-rate limited. The number of samples per line, however, is always less than TW because the highest ratio of T_r/T, for such a repetition-rate-limited scanner, is 1/2. In contrast, we can achieve TW spots per scan line with either of the long-chirp scanners, but only with a reduction in the throughput rate.

To appreciate the relationships among the design parameters more fully, we use two graphic representations. The graph of Fig. 7(a) shows the number of samples per scan line M, normalized to its maximum value of TW, as a function of the ratios T/T_c and

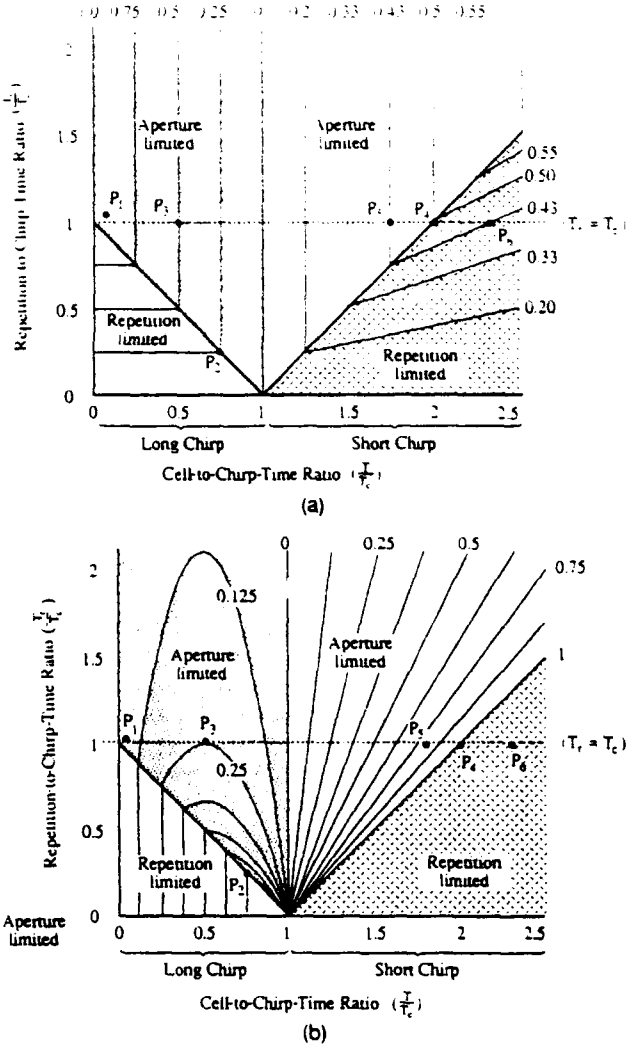


Fig. 7. Normalized plots: (a) number of samples per scan, (b) throughput rate.

Table 1. Acousto-Optic Scanner Parameters

Parameter	Long Chirp ($T_c \geq T$)		Short Chirp ($T_c < T$)	
	Aperture Limited ($T_c - T \leq T_r$)	Repetition-Rate Limited ($T_c - T > T_r$)	Aperture Limited ($T_c + T_r \geq T$)	Repetition-Rate Limited ($T_c + T_r < T$)
Active scan time, $\min(T_s, T_r)$	$T_c - T$	$T_c - T$	$T - T_c$	$T - T_c$
Active aperture time, $\min(T, T_c)$	T	T	T_c	T_c
Scan length (L_s)	$\left\|\frac{v(T_c - T)F}{D - F}\right\|$	$\left\|\frac{vT_rF}{D - F}\right\|$	$\left\|\frac{v(T - T_c)F}{D - F}\right\|$	$\left\|\frac{vT_rF}{D - F}\right\|$
Spot size (d_0)	$\frac{\lambda DF}{(D - F)L}$	$\frac{\lambda DF}{(D - F)L}$	$\frac{\lambda DF}{(D - F_c)L}$	$\frac{\lambda DF}{(D - F)L_c}$
Number of samples per line (M)	$(1 - T/T_c)TW$	$(T_r/T_c)TW$	$(1 - T_c/T)TW$	$(T_r/T)TW$
Scan duty cycle [$U = \min(T_s, T_r)/T_r$]	$(T_c - T)/T_r$	1	$(T - T_c)/T_r$	1
Sample rate (R_s)	$(T/T_c)W$	$(T/T_c)W$	W	W
Throughput rate (R_t)	$(1 - T/T_c)(T/T_r)W$	$(T/T_c)W$	$[(T - T_c)/T_r]W$	W

$T_r T_c$. The vertical line passing through $T T_r = 1$ is the dividing line between the long-chirp and short-chirp configurations. The horizontal dashed line for which $T_r = T_c$ is the boundary between those configurations in which the chirp signals do or do not overlap. The diagonal lines passing through the points (0, 1), (1, 0), and (2, 1) represent the boundaries between the repetition-rate-limited configurations (below the diagonal lines) and the aperture-limited conditions (above the diagonal lines).

The loci of the constant number of samples per scan line, normalized by TW, are shown for each of the four basic configurations. The normalized values are given along the top and right-hand sides of Fig. 7(a). We note that the largest number is obtained by using a long-chirp scanner for which the ratio T/T_c is small; the scanning may be either aperture or repetition-rate limited. When $T/T_c = 1$ the number of samples reaches its minimum value, because the active scan time is at its minimum value and only one spot can be formed in each scan line. For aperture-limited short-chirp scanners the number of samples per scan line is reciprocally related to the ratio T/T_c, whereas the lines for repetition-rate-limited short-chirp scanners have slopes whose values are equal to the normalized values themselves.

The graph of Fig. 7(b) shows the throughput rate R_t, normalized by its maximum value of W samples per second. The throughput rate is not a function of the ratio T_r/T_c for repetition-rate-limited, long-chirp scanners, as shown in the lower left-hand part of Fig. 7(b). According to Eq. (44) the normalized throughput rate follows parabolic curves when the scanner becomes aperture limited, as shown in the upper left-hand part of Fig. 7(b). Aperture-limited short-chirp scanners have throughput rates that follow straight line segments, as shown in the upper right-hand part of Fig. 7(b), whereas the normalized throughput rate is fixed at unity for all repetition-rate-limited short-chirp scanners, as shown in the lower right-hand part of Fig. 7(b).

V. Examples of an Acousto-Optic Recording System

Using the results summarized in Table 1 and Fig. 7, we provide some sample designs in this section. The parameters for these designs are given in Table 2.

A. Example 1

Suppose that we design a relatively low-performance system, such as a facsimile scanner or recorder. In this case a large number of samples per scan line is typically more important than a high throughput rate. We therefore select an acousto-optic cell, such as one made from slow shear-wave tellurium dioxide material, that has a large time-bandwidth product (for example, $T = 50$ μs and $W = 40$ MHz so that $TW = 2000$). Because the required throughput rate for a typical facsimile is well under 1 MHz, the normalized throughput rate is much less than 0.025; because the chirp period is nearly equal to the chirp duration ($T_r = T_c$), only a small data buffer is needed. This scanner–recorder configuration is represented by point P_1 in the graphs of Fig. 7.

B. Example 2

Suppose that the requirements are the same as in the first example, but we need to operate at a much higher throughput rate of 30×10^6 samples per second. If we use the same acousto-optical cell as before, the normalized throughput rate is 0.75. For long-chirp scanning, Fig. 7(b) shows that the operating point is at P_2. The architecture represented by point P_2, however, requires a fourfold overlap of chirps, thus complicating the driver and lowering diffraction efficiency. A better solution is found at point P_3 by using an 8.3-μs, 120-MHz acousto-optic cell, which operates as an aperture-limited, long-chirp scanner with a normalized throughput rate of 0.25 to provide a normalized number of samples per scan line of 0.5. The number of samples per line is the same in the two alternatives [(0.25)(50 μs)(40 MHz) = 500 as compared with (0.5)(8.3 μs)(120 MHz) = 500]. An even better solution for these requirements may be to use just 30 MHz of the 40-MHz bandwidth of the slow shear-wave cell and to operate with short-chirp scanning at point P_4, where the scan duty cycle is 100% and where the normalized number of samples is 0.48, to provide (0.48)(50 μs)(30 MHz) = 720 samples per scan line.

The graphs of Fig. 7, coupled with the data from Table 1, provide the information needed to quickly sort through the possible scanning solutions for a particular problem. For example, the two scanning configurations shown by P_5 and P_6 in Fig. 7 provide the same number of samples per scan line, but the solution at P_5 has a normalized throughput rate of only 0.875 as compared with a normalized rate of one for the solution at P_6.

Table 2. Examples of Various Scanning Configurations

Scanner Type	T (μs)	W (MHz)	T_c	T_r	Throughput Rate	Samples/Scan Line	Location in Fig. 7
Long chirp, aperture limited	50	40	$> 40\,T$	T_c	$\ll W$	$= TW$	P_1
Long chirp, aperture limited	50	40	$\frac{4}{3}T$	$\frac{1}{4}T_c$	$0.75\,W$	$0.25\,TW$	P_2
Long chirp, aperture limited	8.3	120	$2\,T$	T_c	$0.25\,W$	$0.5\,TW$	P_3
Short chirp, repetition limited	50	30	$\frac{1}{5}T$	T_c	W	$0.48\,TW$	P_4
Short chirp, aperture limited	50	40	$\frac{1}{16}T$	T_c	$0.875\,W$	$0.43\,TW$	P_5
Short chirp, repetition limited	50	40	$\frac{1}{35}T$	T_r	W	$0.43\,TW$	P_6

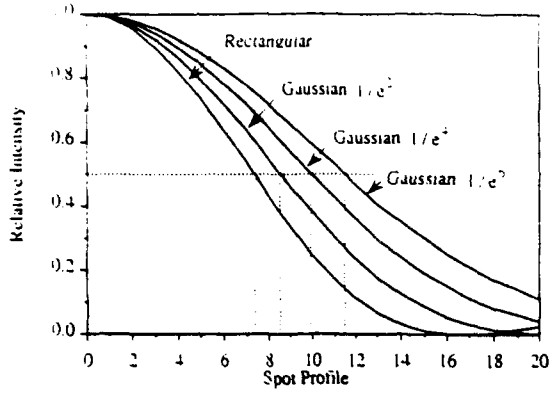


Fig. 8. Spot sizes for various illumination profiles.

VI. Other Considerations

In the analyses given so far, we have assumed uniform illumination of the acousto-optical cell so that the design relationships can be clearly stated in closed form. In practice, the acousto-optic cell is usually illuminated by a laser beam that has a Gaussian intensity weighting so that the spot size, for a given aperture, is greater than that for a uniform illuminating beam.[14] Figure 8 shows the spot sizes for a uniform illumination and for Gaussian illuminations that are truncated at the edges of the cell where the intensity drops to $1/e^2$, $1/e^4$, and $1/e^6$ of the central value. We use the half-power response of the spot distribution as a convenient measure of the spot size. The spot sizes for these Gaussian illuminations, normalized by the spot size produced by a uniform beam, have increased by a factor of G, where G is equal to 1.17 for the $1/e^2$ illumination, 1.36 for the $1/e^4$ illumination, and 1.56 for the $1/e^6$ illumination. All the relationships developed in the previous sections are still valid except that the spot size d_0 must be multiplied by G, whereas the number of samples per scan line M, the sample rate R_s, and the throughput rate R_t must all be divided by G.

Summary

The rich variety of conditions under which an acousto-optic scanner can be implemented and the interlocking nature of the performance parameters often lead to *ad hoc* solutions in which conflicts arise among the required performance specifications. In this paper we consider all possible scanning configurations and classify them into four basic types. A consistent set of design relationships for each of the scanning configurations has been developed and presented in both tabular and graphic forms from which a preliminary design is obtained. The specific solution is found by applying the appropriate design equations.

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

References

1. R. Lipnik, A. Reich, and G. A. Schoen, "Nonmechanical scanning of light using acoustic waves," Proc. IEEE **52**, 853–854 (1964).
2. J. S. Gerig and H. Montague, "A simple optical filter for chirp radar," Proc. IEEE **52**, 1753 (1964).
3. A. Korpel, R. Adler, P. Desmares, and T. M. Smith, "An ultrasonic light deflection system," IEEE J. Quantum Electron. **QE-1**, 60–61 (1965).
4. A. Korpel, R. Adler, P. Desmares, and W. Watson, "A television display using acoustic deflection and modulation of coherent light," Appl. Opt. **5**, 1667–1675 (1966).
5. E. I. Gordon, "A review of acoustooptical deflection and modulation devices," Appl. Opt. **5**, 1629–1639 (1966).
6. M. B. Schultz, M. G. Holland, and L. Davies, Jr., "Optical pulse compression using Bragg scattering by ultrasonic waves," Appl. Phys. Lett. **11**, 237–240 (1967).
7. J. H. Collins, E. G. H. Lean, and H. J. Shaw, "Pulse compression by Bragg diffraction of light with microwave sound," Appl. Phys. Lett. **11**, 240–242 (1967).
8. I. Gorog, J. D. Knox, and P. V. Goedertier, "A television-rate laser scanner," RCA Rev. **33**, 623–694 (1972).
9. L. D. Dickson, "Optical considerations for an acousto-optic deflector," Appl. Opt. **11**, 2196–2202 (1972).
10. J. H. Eveleth, "High resolution laser beam recorder with self-focusing acousto-optic scanner," U.S. patent 3,851,951 (3 December 1974).
11. M. Gottlieb, C. L. M. Ireland, and J. M. Ley, *Electro-Optic and Acousto-Optic Scanning and Deflection* (Dekker, New York, 1983).
12. A. VanderLugt, *Optical Signal Processing* (Wiley, New York, 1992), Chap. 2, p. 49.
13. Ref. 12, pp. 35–37.
14. J. Randolph and J. Morrison, "Rayleigh equivalent resolution of acousto-optic deflection cells," Appl. Opt. **10**, 1453–145450 (1971).

APPENDIX M

INTERFERENCE AND DIFFRACTION

Reprinted from IEEE Transactions on Education
Volume 35, pp.126-132 (1992)

Interference and Fresnel Diffraction

B. D. Guenther and A. VanderLugt, *Senior Member, IEEE*

Abstract—**A surprising number of traditional topics, such as the Rayleigh resolution criterion, spatial filtering, bandlimited signals, the sampling theorem, phase contrast microscopy, and white light holography can be explained using simple interference theory. These basic results are then easily extended to the N-source case to introduce diffraction theory, and facilitates the teaching of the elements of modern optics to junior- and senior-level students.**

I. Introduction

IN TEACHING applied physics and engineering students the fundamentals of physical optics, we find that they are sometimes overwhelmed by the Fresnel–Kirchhoff integral when it is used as an introduction to diffraction theory. Because students more easily understand the concept of interference fringes produced by two overlapping waves, we find that extending our discussion of interference phenomena to three sources and, ultimately, to N sources provides useful insights into diffraction phenomena and helps students visualize results produced by the Fresnel–Kirchhoff diffraction integral. In this paper, we illustrate how we blend the theories of interference and Fresnel diffraction phenomena.

We begin with the solution to the scalar wave equation because students are generally familiar with it from courses in physics, applied mathematics, mechanics, or electromagnetism. The propagating wave is a solution of the scalar wave equation:

$$\nabla^2 E = \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}. \tag{1}$$

The harmonic wave solution to the wave equation is of the form:

$$E(z,t) = E_0 \cos(\omega t - kz) \tag{2}$$

where ω is the radian frequency of light, z is the direction of propagation, and $k = 2\pi/\lambda$. Consider the interference between two waves with equal magnitudes but with different frequencies ω_1 and ω_2, propagating along the z direction. The resulting wave is obtained by a simple algebraic addition of the two waves:

$$\begin{aligned} E(z,t) &= E_0 \cos(\omega_1 t - k_1 z) + E_0 \cos(\omega_2 t - k_2 z) \\ &= \boldsymbol{E_0} \cos[(\omega t - kz) + (\Delta\omega t - \Delta k z)] \\ &\quad + E_0 \cos[(\omega t - kz) - (\Delta\omega t - \Delta k z)] \end{aligned} \tag{3}$$

where $\Delta\omega = \omega_2 - \omega_1$ and $\Delta k = k_1 - k_2$ are the difference values, while ω and k are the mean values.

Manuscript received August 1991.

B. D. Guenther is with the U.S. Army Research Office, Research Triangle Park, NC, and Duke University, Durham, NC.

A. VanderLugt is with North Carolina State University, Raleigh, NC.

IEEE Log Number 9106732.

By virtue of a trigonometric identity, we write the resultant amplitude modulated wave as:

$$E(z,t) = 2E_0 \cos(\omega t - kz)\cos(\Delta\omega t - \Delta k z). \tag{4}$$

Because the two waves are at slightly different frequencies, their relative phase difference is a function of time. The time-dependent phase causes the two waves to alternately add constructively and destructively, creating the amplitude modulation that is a periodic series of maxima due to constructive interference shown by (4). The amplitude modulation of the carrier wave is a wave of frequency $\Delta\omega$, called the beat frequency, and propagation constant Δk propagating at a velocity of:

$$u = \frac{\Delta\omega}{\Delta k}. \tag{5}$$

This velocity is called the group velocity. The modulation is on a carrier wave of frequency

$$\omega = \frac{1}{2}(\omega_1 + \omega_2) \tag{6}$$

and propagation constant k traveling at a velocity

$$v = \frac{\omega}{k}. \tag{7}$$

This velocity we recognize as the phase velocity of the carrier wave. In a nondispersive medium, the group and phase velocity are identical.

If the two waves have unequal magnitudes, (3) becomes:

$$E(z,t) = E_1 \cos(\omega_1 t - k_1 z) + E_2 \cos(\omega_2 t - k_2 z). \tag{8}$$

The intensity $I(z,t)$ is the squared magnitude of the wave $E(z,t)$:

$$I(z,t) = E(z,t)E^*(z,t). \tag{9}$$

Any observation or measurement of the intensity must account for the finite bandwidth of the detector. We illustrate this by substituting (8) into (9), to find that the observable is proportional to:

$$\begin{aligned} G(z,t) &= \langle [E_1 \cos(\omega_1 t - k_1 z) + E_2 \cos(\omega_2 t - k_2 z)] \\ &\quad \times [E_1 \cos(\omega_1 t - k_1 z) + E_2 \cos(\omega_2 t - k_2 z)]^* \rangle \\ &= \langle E_1^2 \cos^2(\omega_1 t - k_1 z) + E_2^2 \cos^2(\omega_2 t - k_2 z) \\ &\quad + 2E_1 E_2 \cos(\omega_1 t - k_1 z)\cos(\omega_2 t - k_2 z) \rangle \\ &= \left\langle E_1^2 \left\{ \frac{1}{2} + \frac{1}{2}\cos(2\omega_1 t - 2k_1 z) \right\} \right. \\ &\quad \left. + E_2^2 \left\{ \frac{1}{2} + \frac{1}{2}\cos(2\omega_2 t - 2k_2 z) \right\} \right. \end{aligned}$$

$$+ 2E_1E_2\left\{\frac{1}{2}\cos[(\omega_1+\omega_2)t-(k_1+k_2)z]\cdot\cos[(\omega_2-\omega_1)t-(k_2-k_1)z]\right\}\Big\rangle. \quad (10)$$

The bandwidths of all physical detectors are too narrow to detect the frequency of light so that the observable becomes:

$$G(z,t)=\frac{1}{2}E_1^2+\frac{1}{2}E_2^2 + \left\langle 2E_1E_2\left\{\frac{1}{2}\cos[(\omega_1+\omega_2)t-(k_1+k_2)z]\cdot\cos[(\omega_2-\omega_1)t-(k_2-k_1)z]\right\}\right\rangle. \quad (11)$$

The last term of (11) time averages to zero if the beat frequency $\Delta\omega=\omega_2-\omega_1$ is much larger than the bandwidth of the detector. In this case, the observable:

$$G(z)=\frac{1}{2}E_1^2+\frac{1}{2}E_2^2 \quad (12)$$

is simply the sum of the intensities of the individual waves and has no temporal variation. On the other hand, if the frequencies differ by less than the bandwidth of the detector, we observe the temporal fluctuation due to the envelope of the third term of (11). If the frequencies of the two waves are identical, we apply trigonometric identities to (11) to find that the observable is:

$$G(z,t)=\frac{1}{2}E_1^2+\frac{1}{2}E_2^2+E_1E_2^*. \quad (13)$$

This result is the same, to within a scaling factor, as would be obtained from (8) had we ignored the temporal oscillations of the light. Hence, when the light is monochromatic, we often express the amplitude of the sum of two complex valued waves simply as:

$$E(z)=E_1(z)+E_2(z) \quad (14)$$

where we suppress the cosinusoidal time dependence of the propagating wave. In addition to suppressing the temporal part of the electromagnetic wave, we find that the use of complex notation reduces the mathematical complexity of many problems. This notation is consistent with that used for the transfer functions for lenses, prisms, and other optical elements. As an example of this procedure, consider the interference between two plane waves which are now represented in the paraxial approximation by:

$$E(x)=e^{-jk\theta_1 x}+e^{-jk\theta_2 x} \quad (15)$$

where θ_1 and θ_2 are the angles that a ray normal to the wavefront makes with respect to the z-axis. The intensity at any plane at which these waves overlap is simply:

$$I(x)=\left|e^{-jk\theta_1 x}+e^{-jk\theta_2 x}\right|^2 = 2\{1+\cos[k(\theta_1-\theta_2)x]\} \quad (16)$$

so that the observed pattern is sinusoidal in intensity.

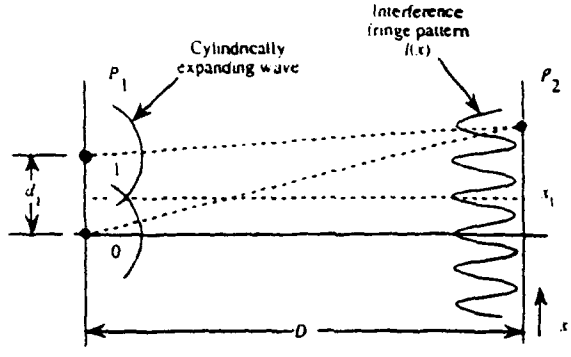


Fig. 1. One-dimensional representation of the two-source geometry.

II. Two Sources

Having illustrated how two plane waves interfere to produce a sinusoidal intensity pattern, we consider the interference between waves at a finite distance from two sources. A formulation of the problem is shown in Fig. 1. If the distance from the source plane P_1 to the observation plane P_2 is large, we approximate the electric field at plane P_2, due to a point source at the origin, by the expanding wave:

$$\frac{1}{\sqrt{r}}e^{-jkr} \quad (17)$$

propagating from plane P_1 towards plane P_2. The square root in the denominator of (17) results from solving the problem in one-dimension. For the source at the origin, we have:

$$r\doteq D\sqrt{1+(x/D)^2} \quad (18)$$

and we use the binomial expansion to obtain:

$$E(x)=E_0e^{-j(kD+\phi)}e^{-j(k/2D)x^2} \quad (19)$$

where x is the coordinate in plane P_2, and D is the distance between the planes. The magnitude E_0 of the wave is nonnegative, in accordance with the usual method of incorporating all phase information in the exponential as the factor ϕ. The first exponential factor simply indicates the number of wavelengths between planes P_1 and P_2 and is ignored in subsequent discussions. The detailed derivation of the approximation need not be given to the student initially, as it provides little additional physical insight. The basis for the approximation is given in several texts and can be covered later, if desired [1]–[3].

The field at P_2 when the two sources have equal magnitudes and are in phase is:

$$E(x)=E_0e^{-j(k/2D)x^2}+E_0e^{-j(k/2D)(x-d_1)^2} \quad (20)$$

where d_1 is the distance between the sources, and E_0 is the magnitudes at plane P_2 of the wave due to the source. The intensity at plane P_2 is:

$$I(x)=E(x)E^*(x)=|E(x)|^2 = 2E_0^2+2\,\mathrm{Re}\left[E_0^2e^{-jkx^2/2D}e^{jk(x-d_1)^2/2D}\right] \quad (21)$$

so that the normalized intensity is:

$$\frac{I(x)}{2I_0} = 1 + \cos\left[(k/D)xd_1 - (k/2D)d_1^2\right] \tag{22}$$

where $I_0 = E_0^2$. The first term of (22) is the sum of the intensities produced by the individual sources; this sum is a spatially uniform intensity, called the *bias*. The second term is a spatially varying cosine distribution, called the fringe pattern. It has a *spatial frequency* α given by the partial derivative of the phase with respect to the spatial variable x:

$$\alpha = \frac{1}{2\pi}\frac{\partial}{\partial x}\left[\frac{kxd_1}{D}\right] = \frac{d_1}{\lambda D}. \tag{23}$$

The spatial frequency of the fringe pattern, expressed in units of cycles/mm, therefore increases as the separation d_1 between the sources increases and as the observation distance D decreases. The ratio d_1/D is the angular size of the source as seen from the viewing screen.

The phase of the fringe pattern is:

$$\varphi = -\frac{\pi d_1^2}{\lambda D}. \tag{24}$$

The physical interpretation of the phase is that it specifies the relative position of the maximum intensity in the observation plane. The *principal maximum* occurs when the argument of the cosine is equal to zero, i.e., where $2xd_1 - d_1^2 = 0$. The principal maximum, therefore, occurs at $x_1 = d_1/2$, which is directly opposite the midpoint between the two sources, as shown in Fig. 1. Other intensity maxima occur when the phase is an integer multiple of 2π.

If the light from the two sources have arbitrary phases ϕ_0 and ϕ_1, the entire fringe pattern in plane P_2 is shifted according to the phase difference, $\phi_1 - \phi_0$, and the principal maximum moves to:

$$x_1 = \frac{d_1}{2} - \frac{\phi_1 - \phi_0}{kd_1/D}. \tag{25}$$

The variation of the fringe intensity is measured by the *fringe visibility*:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2E_0E_1}{E_0^2 + E_1^2}. \tag{26}$$

Maximum fringe visibility is achieved when the two sources are spatially and temporally coherent and have equal magnitudes; the visibility is not a function of the phase difference between the two sources. The two-source model, therefore, provides the basics for understanding topics such as interference, photoelectric mixing spectroscopy, and the coherence function.

III. Three Sources

The traditional approach is to proceed directly from the two-source geometry to the N-source case. We find that pausing to examine a three-source geometry provides a wealth of additional insights that helps bridge the gap between the two-source and N-source cases. We use the three-source geometry to illustrate techniques such as Zernike's method of measuring small optical path differences, phase and magnitude spatial filtering, and white-light holographic fringe formation.

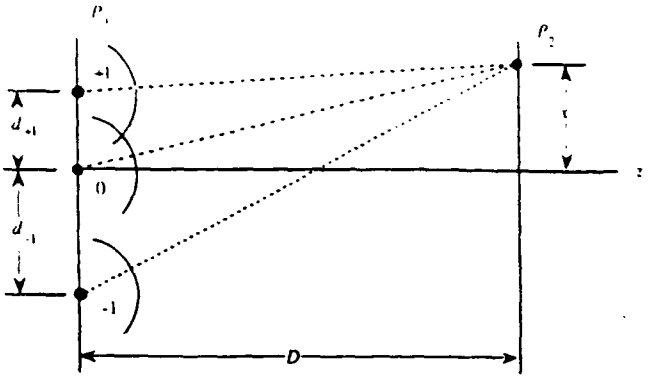


Fig. 2. One-dimensional representation of the three-source geometry.

The three-source geometry is shown in Fig. 2. To anticipate the transition to the N-source case, we subscript the sources surrounding the central source with $\pm$ indexes. The intensity at the observation plane is now the squared magnitude of the sum of three terms:

$$I(x) = \left|E_{-1}e^{-j\left[k(x+d_{-1})^2/2D+\phi_{-1}\right]} + E_0e^{-j\left[kx^2/2D+\phi_0\right]} + E_{+1}e^{-j\left[k(x-d_{+1})^2/2D+\phi_{+1}\right]}\right|^2 \tag{27}$$

where d_{+1} and d_{-1} are the distances between the sources. An expansion of this equation yields the central result:

$$\begin{aligned} I(x) = {} & \left(E_{-1}^2 + E_0^2 + E_{+1}^2\right) \\ & + 2E_0E_{+1}\cos\left[kxd_{+1}/D - kd_{+1}^2/2D - \phi_{+1} + \phi_0\right] \\ & + 2E_{-1}E_{+1}\cos\left[kx(d_{+1} + d_{-1})/D \right. \\ & \left. - k\left(d_{+1}^2 - d_{-1}^2\right)/2D + \phi_{-1} - \phi_{+1}\right] \\ & + 2E_{-1}E_0\cos\left[kxd_{-1}/D + kd_{-1}^2/2D + \phi_{-1} - \phi_0\right]. \end{aligned} \tag{28}$$

If we set any one of the three source magnitudes to zero, the result reduces to (22) as expected.

We now show how the variation of the phase of the individual sources modifies the intensity. For simplicity, we assume that $d_{+1} = d_{-1} = d$ and $E_{+1} = E_1 = E_0$. The relative intensity at plane P_2 is then:

$$\begin{aligned} \frac{I(x)}{I_0} = {} & 3 + 2\cos\left[kxd/D - kd^2/2D - \phi_{+1} + \phi_0\right] \\ & + 2\cos\left[kxd/D + kd^2/2D + \phi_{-1} - \phi_0\right] \\ & + 2\cos\left[2kxd/D + \phi_{-1} - \phi_{+1}\right]. \end{aligned} \tag{29}$$

The first term of (29) is a bias of uniform intensity. The second and third terms describe a fringe pattern produced by interference between adjacent sources; we call this the *fundamental fringe pattern*. The last term we label the *harmonic fringe pattern*; it contains twice the spatial frequency of the fundamental fringe because the outer sources are twice as far apart as adjacent sources. A periodic intensity variation along the z direction can be understood by finding the maximum

and minimum values of (29) at various planes along the z-axis. This aids in the development of a physical understanding of Fresnel zones.

Zernike used three equally spaced slits to form a high-accuracy interferometer for precisely measuring the phase of an unknown sample covering the central slit [4]. The device was initially designed as a tool for measuring the phase shifts produced by thin layers of materials used in phase-contrast microscopes. In Zernike's interferometer, the slit spacings are equal and the phases of the outer slits are the same: $\phi_{+1} = \phi_{-1}$, so that $\phi_{+1} - \phi_0 = \phi_{-1} - \phi_0 = \phi$. To maximize the fringe visibility, the central slit is usually made twice as wide as the outer slits. Using these parameters in (29) yields a relative intensity at plane P_2 of:

$$\frac{I(x)}{I_0} = 3 + 4\cos\left[kd^2/2D + \phi\right]\cos[kxd/D] + \cos[2kxd/D]. \tag{30}$$

A fundamental fringe pattern occurs in a z-plane for which $\left|\cos\left(kd^2/2D + \phi\right)\right| = 1$, or when $kd^2/2D + \phi = n\pi$, where n is an integer. The harmonic fringe pattern occurs whenever $\cos\left(kd^2/2D + \phi\right) = 0$, or when $kd^2/2D + \phi = (2n+1)\pi/2$. We first locate the fundamental and harmonic planes when the phases of the waves passing through the three slits are identical so that $\phi = 0$.

$$D_{00} = \frac{d^2}{n\lambda}$$
$$D_{10} = \frac{d^2}{\left(n + \frac{1}{2}\right)\lambda} \tag{31}$$

where D_{00} and D_{10} are the distances from the slit plane to the fundamental and harmonic planes when the phase difference is zero. The fringe patterns in these two planes are shown in Fig. 3(a).

The next step is to place the test sample phase material over the central slit. The visibility of the fringe patterns are now modified; we illustrate the harmonic fringe pattern in Fig. 3(b) for a $\lambda/100$ phase shift. We see that a relatively small phase change in the material produces an easily discernible variation in the envelope of the fringe pattern. The new locations of the fundamental and harmonic planes, where the fringe envelopes are flat as shown in Fig. 3(a), are given by:

$$D_0 = \frac{d^2}{\left(n - \frac{\phi}{\pi}\right)\lambda}$$
$$D_1 = \frac{d^2}{\left(n + \frac{1}{2} - \frac{\phi}{\pi}\right)\lambda}. \tag{32}$$

The unknown phase ϕ of the material is calculated by using the positions of the four planes:

$$\frac{2\lambda\phi}{\pi d^2} = \left[\frac{1}{D_{10}} - \frac{1}{D_1}\right] + \left[\frac{1}{D_{00}} - \frac{1}{D_0}\right]. \tag{33}$$

The accuracy of this three-slit interferometer is based simply on the ability to locate planes having intensity fringes whose envelopes are flat. Mean errors as low as $\pm 0.0017\lambda$ in the measurement of the phase of the samples have been reported [4], [5].

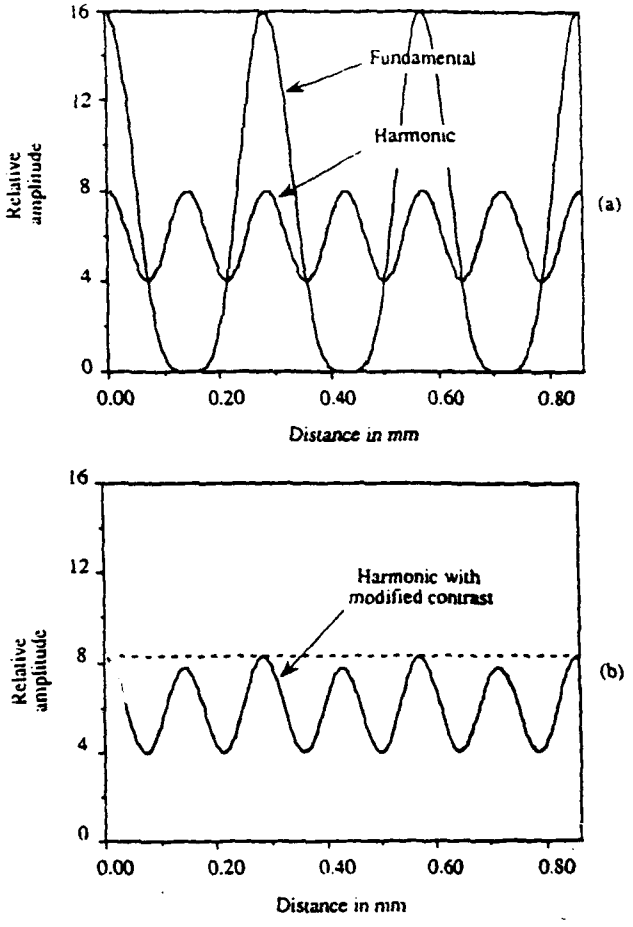


Fig. 3. Interference fringes: (a) The fundamental fringe pattern located at plane D_{00} and the harmonic fringe pattern located at plane D_{10}. (b) The harmonic fringe pattern, showing reduced contrast caused by a $\lambda/100$ phase shift in the material.

IV. Diffraction by the Source

So far, we have ignored the shape of the diffracting sources. It is convenient, as a transition to diffraction theory and signal processing application using the Fourier transform, to consider a source whose amplitude distribution is:

$$g(x) = \frac{\sin(\pi x/d_0)}{\pi x/d_0} \equiv \operatorname{sinc}(x/d_0) \tag{34}$$

where d_0 is the distance to the first zero of the sinc function. The Fourier transform of this single-point object is created by a lens L of focal length F, as shown in Fig. 4, and is defined as:

$$G(\xi) = \int_{-\infty}^{\infty} g(x)e^{jk\xi x/F}dx \tag{35}$$

where ξ is the physical coordinate in plane P_2. We define a spatial frequency variable $\alpha = \xi/\lambda F$ and associate this variable with the spatial frequency generated by the interference of two waves as given by (16). We substitute (34) into (35)

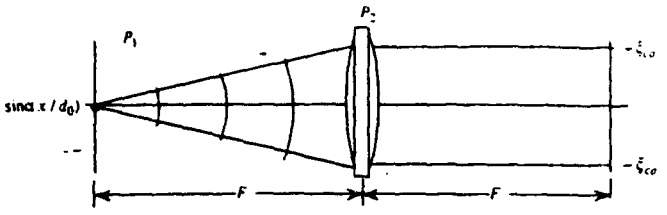


Fig. 4. The Fourier transform of a source whose amplitude in plane P_1 is $\text{sinc}(x/d_0)$.

and calculate the Fourier transform of the object as:

$$G(\xi) = \int_{-\infty}^{\infty} \text{sinc}(x/d_0) e^{jk\xi x F}\, dx = \text{rect}(\xi d_0/\lambda F) \tag{36}$$

where we ignore scaling constants. The rect function is defined as:

$$\text{rect}(x) = \begin{cases} 1; & |x| \le \frac{1}{2} \\ 0; & \text{else.} \end{cases} \tag{37}$$

From (36) and (37), we find that the light distribution in the Fourier plane is a plane wave that is sharply cutoff at:

$$\xi_{co} = \frac{\lambda F}{2d_0} \tag{38}$$

or at $\alpha_{co} = 1/(2d_0)$ which is called the *cutoff* spatial frequency.

A. Bandlimited Signals and the Sampling Theorem

Our approach to diffraction theory offers an opportunity to introduce students to the basics of sampling theory. The Whittaker–Shannon theorem for accurately representing a bandlimited signal by a sequence of sample functions states that the highest spatial frequency present must be sampled at least twice per cycle. The relationship that $d_0 = 1/(2\alpha_{co})$, where d_0 is the distance between the samples, fulfills this requirement. The limited extent of the frequency distribution $G(\alpha)$ shows that $\text{sinc}(x/d_0)$ is a *bandlimited* signal of bandwidth α_{co}. This suggests that we can represent an arbitrary bandlimited signal $g(x)$ as a sum of weighted *sample functions* of the form given by (34):

$$g(x) = \text{rect}(x/L) \sum_{n=-\infty}^{\infty} a_n e^{j\phi_n} \text{sinc}[(x - nd_0)/d_0] \tag{39}$$

where the a_n and ϕ_n are magnitude and phase coefficients, the spacing between the sample functions is d_0, and the rect function defines the length of the signal. The signal representation in (39) is equivalent to extending the three-source result given in Section III to an $N = L/d_0$ sample representation of a one-dimensional signal, in which each sample is equivalent to a source having magnitude a_n. By virtue of the shift theorem, the Fourier transform of an arbitrary object consisting of many weighted sample functions, is simply the sum of a set of weighted plane waves:

$$G(\alpha) = \text{rect}(\alpha/2\alpha_{co}) \sum_{n=-\infty}^{\infty} \alpha_n e^{j(2\pi n d_0 \alpha + \phi_n)} \tag{40}$$

where $\text{rect}(\alpha/2\alpha_{co})$ shows that the transform is bounded the frequency interval $|\alpha| \le \alpha_{co}$. These plane waves interfere in the Fourier plane in a fashion similar to that discussed in Section II. Thus, $G(\alpha)$ can be considered either as a summation of the primitive interference patterns produced by plane waves from the individual samples in $g(x)$ or as the diffraction produced by $g(x)$ taken as a whole. In this way, we show that interference and diffraction are closely related and that the distinction between them is largely arbitrary. By simple ray tracing, we find that the light from all samples of $g(x)$ in plane P_1 pass through an aperture in the Fourier plane bounded by $|\alpha| \le \alpha_{co}$. The entire signal, as well as each sample, is therefore strictly bandlimited [6].

An alternative representation of the signal is by means of weighted cosine functions in which the signal is represented, after using the Euler formula, by:

$$g(x) = \text{rect}(x/L) \sum_{n=-\infty}^{\infty} b_n e^{j(2\pi n \alpha_0 x + \phi_n)} + \text{c.c.}, \tag{41}$$

where b_n and ϕ_n are the weights, α_0 is the minimum resolvable spatial frequency, and c.c. represents the complex conjugate. The similarity of (41) to (40) suggests that the Fourier transform of $g(x)$ has the same form as (39):

$$G(\alpha) = \text{rect}(\alpha/2\alpha_{co}) \sum_{n=-\infty}^{\infty} b_n e^{j\phi_n} \text{sinc}[(\alpha - n\alpha_0)L] \tag{42}$$

where we again ignore scaling factors.

These two basic ways of representing signals in the space and frequency planes are useful in discussing a wide range of topics in optics. It emphasizes that all planes in an optical system are essentially space planes, and which one we prefer to call the frequency plane is largely a matter of choice.

B. Rayleigh Resolution Criterion

We show students how the Rayleigh resolution criterion can be easily developed from (41) and (42). In this case, we associate the input signal length L with the width of the (one-dimensional) entrance pupil of the telescope. According to Rayleigh's criterion, the angular resolution of the telescope is $\phi_0 = \lambda/L$. The image of a star, from conventional diffraction theory, has an intensity $\text{sinc}^2(\xi L/\lambda F)$ so that two stars are resolved if they are separated by a distance $\xi_0 = \lambda F/L$ at the focal plane [7]. When we associate the aperture in plane P_2 of Fig. 4 with the entrance pupil of a telescope and identify plane P_1 as the image plane, we also find that $d_0 = \phi_0 F = \lambda F/L$ in accordance with (39). Additional insights can be gained by noting that the cutoff diffraction angle θ_{co} in Fig. 4 is $\theta_{co} = \tan^{-1}(\xi_{co}/F) = \tan^{-1}(\lambda\alpha_{co}) \cong \lambda\alpha_{co}$ for small angles. Thus, the maximum half-angle through which light can be scattered by a bandlimited signal $g(x)$ having sample spacing d_0 is simply $\lambda/2d_0$.

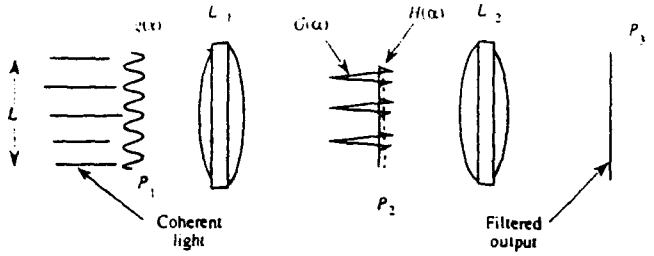


Fig. 5. Spatial filtering system. The spatial filter $H(\alpha)$ is located in plane P_2.

C. Abbe's Resolution Criterion and Spatial Filtering

We use the three-source geometry to illustrate the basic principles of optical spatial filtering [8], [9]. Suppose that the object has a relative intensity $I(x) = 3/2 + 2\cos(2\pi\alpha_1 x) + 1/2\cos(2\pi\alpha_1 x)$ with an associated amplitude $g(x) = [1 + \cos(2\pi\alpha_1 x)]$, defined for $|x| \leq L/2$, and is illuminated at plane P_1 by coherent light of wavelength λ, as illustrated in Fig. 5. The light amplitude at the Fourier plane P_2 is, from (35), equal to:

$$G(\alpha) = \text{sinc}[(\alpha + \alpha_1)L] + \frac{1}{2}\text{sinc}(\alpha L) + \frac{1}{2}\text{sinc}[(\alpha - \alpha_1)L] \tag{43}$$

which consists of sinc functions centered at $\alpha = -\alpha_1$, 0, and $+\alpha_1$. The connection between $G(\alpha)$, viewed as a three-source signal, and the signal $g(x)$ is similar to the interference patterns discussed in Section III. In terms of signal processing, we consider $G(\alpha)$ the spectrum of $g(x)$.

The spectrum $G(\alpha)$ can be modified by a *spatial filter* whose complex-valued transmittance is $H(\alpha)$. Abbe, with his theory of image formation by a coherently illuminated microscope, showed how false detail can arise when the spatial frequency spectrum of an object is altered [10]. As an example, suppose that we apply a spatial filter $H(\alpha)$ of the form:

$$H(\alpha) = \begin{cases} 1, & |\alpha| \leq \alpha_1/2; \\ 0, & \text{elsewhere.} \end{cases} \tag{44}$$

This filter allows only the sinc function from (43) located at $\alpha = 0$ to pass; it is called a *lowpass filter*. The filtered output is created at plane P_3 by lens L_2 in Fig. 5, which provides a second Fourier transform, and has intensity $I(x) = 1$ and there is no evidence of a spatial fringe pattern. As a second example, suppose that the filter $H(\alpha)$ in plane P_2 of Fig. 5 passes only two of the sinc functions from (43). If one of the outer sources is removed by the filter, the result from (43) is that:

$$I(x) = \frac{5}{4} + \cos(2\pi\alpha_1 x + \partial) \tag{45}$$

where γ is the residual phase. The image is now significantly altered relative to the object; the fundamental frequency has reduced fringe visibility and the harmonic frequency has been eliminated altogether. Finally, if we remove just the central order from the spectrum of (43), we find that the intensity at the output is:

$$I(x) = \frac{1}{2}[1 + \cos(4\pi\alpha_1 x)] \tag{46}$$

revealing that the fundamental frequency is absent in the filtered image.

The spatial filter may also change the phase of the sources in plane P_2. Suppose that the outer sinc functions are modified by a filter of the form:

$$H(\alpha) = \begin{cases} e^{i\phi_{+1}}, & \text{near } \alpha = \alpha_1 \\ 1, & \text{near } \alpha = 0 \\ e^{i\phi_{-1}}, & \text{near } \alpha = -\alpha_1. \end{cases} \tag{47}$$

The intensity distribution at plane P_3 is now highly dependent on the values of ϕ_{+1} and ϕ_{-1}. If $\phi_{+1} = -\phi_{-1}$, the result is equivalent to multiplying $G(\alpha)$ by a linear phase function so that the fringe structure in plane P_3 is simply shifted. If, however $\phi_{+1} \neq \phi_{-1}$, we need to examine (29) more closely to determine the nature of the output. The third and fourth terms can be combined to form the expression:

$$4\left\{\cos\left[kd^2/2D + \frac{\phi_{+1} + \phi_{-1}}{2} - \phi_0\right]\right\} \cdot \cos\left[2kxa/D - \frac{\phi_{+1} - \phi_{-1}}{2}\right]. \tag{48}$$

The fringe pattern described by this term has the same spatial frequency as the fringe pattern observed from two sources with a separation of d. Here, however, we encounter a curious situation in which the amplitude of the fringe pattern is modulated by a cosinusoidal function which equals zero whenever:

$$kd^2/2D + \frac{\phi_{+1} + \phi_{-1}}{2} - \phi_0 = \pm\frac{\pi}{2}. \tag{49}$$

Without loss of generality, we set $\phi_0 = 0$ and associate the focal distance F with D. The requirement for a zero amplitude fundamental fringe pattern is now that:

$$\frac{2\pi d^2}{\lambda F} = |\pi - \phi_{+1} - \phi_{-1}|. \tag{50}$$

It is clear that the fundamental can be completely eliminated at the output when ϕ_{+1} and ϕ_{-1} satisfy (50). As a note in passing, it was this spatial filtering homework problem that reminded us of Zernike's three-slit interferometer and stimulated our more general study of three-source phenomena.

E. Phase Contrast Microscopy [11]

We connect Zernike's phase contrast microscopy with spatial filtering and, at the same time, introduce additional insight into the N-source case. Zernike noted that the transmittance of a phase object can be represented in one-dimensional notation as $\exp[j\phi(x)]$. If the maximum phase deviation of the object is small so that $\phi(x) \ll 1$, we can use a power series expansion to show that $\exp[j\phi(x)] \approx 1 + j\phi(x)$. Zernike noted that light due to the constant is focused near $\alpha = 0$ in the Fourier domain, whereas light due to $\phi(x)$ is distributed throughout plane P_2 in Fig. 5. A phase shift of $\pi/2$ at $\alpha = 0$ in plane P_2 can be introduced by a phase plate whose transmission is represented by $H(\alpha)$:

$$H(\alpha) = \begin{cases} e^{j\pi/2}; & |\alpha| \leq \frac{1}{2} \\ 1; & \text{else.} \end{cases} \tag{51}$$

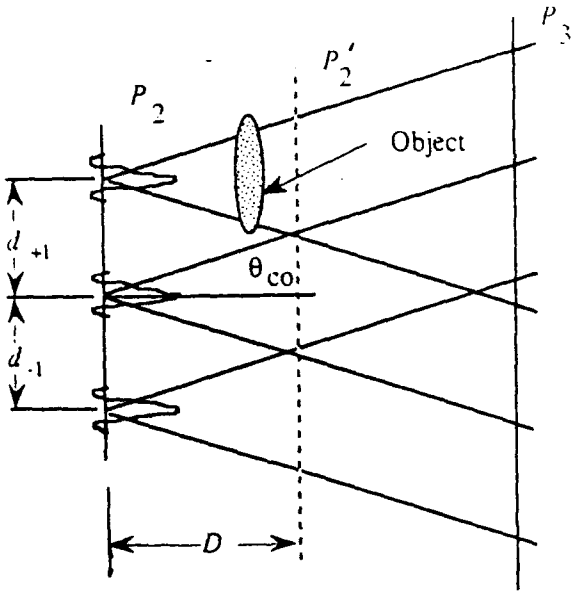


Fig. 6. Fresnel diffraction from three $\mathrm{sinc}(x/d_0)$ sources located in plane P_2.

This phase shift causes the amplitude in plane P_3 to become $j[1+\phi(x)]$ so that the intensity is approximately $1+2\phi(x)$; the phase object is, therefore, rendered visible. To generalize Zernike's procedure, we can decompose $g(x)$ into a sequence of $N/2$ sine waves. The Fourier transform is then a sequence of sinc functions as given by (42). In the Fourier plane, there is a strong $\mathrm{sinc}(\alpha L)$ function located at $\alpha = 0$ due to the constant in $g(x)$. As in Zernike's original method, a $\pi/2$ phase shift at $\alpha = 0$ produces bright background against which the phase information appears dark whereas a $-\pi/2$ phase shift produces dark background against which the phase information appears bright.

F. White-Light Holography

We introduce students to white-light holography with the aid of Figs. 5 and 6. A sinusoidal grating in plane P_1 of Fig. 5 is illuminated by white light so that three polychromatic sources are formed at plane P_2 according to (43). As we see from Fig. 6, light from the separate sources do not overlap in the region between planes P_2 and P_2'. It is in this region that one or more beams can be spatially modulated in magnitude or phase or both by an object. The white-light fringe pattern formed at plane P_3 of Fig. 6 is now modulated by light diffracted from the object to produce a white light hologram.

V. Conclusions

The mathematics associated with diffraction theory can be simplified and the physical aspects of diffraction emphasized through an analysis of interference from three sources. The examples presented are designed to stimulate the reader to incorporate a three-source discussion in the teaching of physical optics.

References

[1] E. Hecht, *Optics*, Second Ed. New York: Addison-Wesley, 1987.

[2] G. Fowles, *Introduction to Modern Optics*. New York: Holt, Rinehart, and Winston, 1968.

[3] B. D. Guenther, *Modern Optics*. New York: Wiley, 1990.

[4] F. Zernike, "A precision method for measuring small phase differences," *J. Opt. Soc. Am.*, vol. 40, pp. 326–328, 1950.

[5] R. E. Kinzly, "A new interferometer capable of measuring small optical path differences," *Appl. Opt.*, vol. 6, pp. 137–140, 1967; D. Wang and T.-g. Chen, "Holographic real-time three-slit interferometer," *Appl. Opt.*, vol. 27, pp. 1298–1301, 1988.

[6] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw-Hill, 1968, p. 111.

[7] M. Born and E. Wolf, *Principles of Optics*. New York: Macmillan, 1964, p. 415.

[8] E. L. O'Neill, "Spatial filtering in optics," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 56–65, 1956.

[9] L. J. Cutrona, E. N. Leith, C. J. Palermo, and L. J. Porcello, "Optical data processing and filtering systems," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 386–400, 1960.

[10] M. Born and E. Wolf, *Principles of Optics*. New York: Macmillan, 1964, p. 419.

[11] M. Born and E. Wolf, *Principles of Optics*. New York: Macmillan, 1964, p. 424.

[12] B. J. Chang, R. Alferness, and E. N. Leith, "Space-invariant achromatic grating interferometers: Theory," *Appl. Opt.*, vol. 14, pp. 1592–1600, 1975.

B. D. Guenther received the B.S. degree in physics from Baylor University, Waco, TX, in 1960, and the M.S. and Ph.D. degrees in physics from the University of Missouri, Columbia, in 1963 and 1968, respectively.

He was employed by the U.S. Army Missile Command from 1968 to 1979. From 1968 through 1972, he was involved in research in solid-state physics. From 1973 to 1975, he was the Army Representation to the Air Force High Energy Laser program and was technically involved in the investigation of new laser concepts. Beginning in 1975, he was engaged in the field of optics, specializing in the areas of holography, optical processing, and image processing. He is currently the Director of Physics at the U.S. Army Research Office, Research Triasngle Park, NC, and Adjunct Professor at Duke University, Durham, where he conducts research in femtosecond technology.

Anthony VanderLugt received the B.S.E.E. and M.S.E.E. degrees from the University of Michigan, Ann Arbor, and the Ph.D. and DS.c. degrees from the University of Reading, Reading, England.

He was a member of the research staff at the Institute of Science and Technology at the University of Michigan from 1959 to 1969, where he worked on optical spatial filtering techniques, Fourier optics, and applications of holography. In 1969, he joined the Harris Corporation as Director of Research at the Electro-Optics Center in Ann Arbor. From 1973 to 1979, he managed research and development groups at the Government Systems Division of Harris in Melbourne, FL. From 1979 to 1986, he was a Senior Scientist in the Advanced Technology Department, where his major area of interest was optical signal processing. In 1986, he joined the faculty of North Carolina State University, Raleigh, as Professor of Electrical and Computer Engineering; he teaches courses on optical signal processing and continues his research work in this field. He was the American Editor for *Optica* from 1970 to 1974 and a member of the Advisory Board from 1974 to 1979. He is currently a Topical Editor for *Applied Optics*.

Dr. VanderLugt is a Fellow of the Optical Society of America, a Fellow of the Society of Photographic Instrumentation Engineers, and a member of Sigma Xi, Tau Beta Pi, and Eta Kappa Nu.

APPENDIX N

EFFICIENT COMPUTATION OF NEAR-FIELD DIFFRACTION PATTERNS USING SUBSAMPLED CONVOLUTION

Accepted for Publication in Optical Engineering

Efficient Computation of Near-Field Diffraction Patterns using Subsampled Convolution

Dan Owen Harris

Siecor Corporation

489 Siecor Park - HE

Hickory, NC 28603-0489

Phone: (704) 323-6568

Fax: (704) 323-6580

Abstract

We introduce "subsampled convolution" and convolution with an "aggregate kernel" as computationally efficient techniques for discrete convolution of the free-space propagation kernel with generalized and rectangular aperture functions, respectively, in computing near-field diffraction patterns. We develop FFT-based algorithms for implementing subsampled and aggregate kernel convolution, then demonstrate the memory and execution time economy of these techniques through computational examples.

Keywords: diffraction patterns, near-field diffraction, discrete convolution, subsampled convolution, aggregate kernel.

1. Introduction

Analysis and simulation of optical systems often requires calculation of diffraction patterns. Since a closed form solution for diffraction from an arbitrary aperture is not always possible, numerical evaluation of diffraction patterns is sometimes necessary. In the far-field, the diffracted light is accurately described by the Fourier transform of the input aperture;[1] efficient numerical calculation of the Fourier transform is easily achieved through the use of Fast Fourier Transform (FFT) algorithms.[2] Methodologies for calculating near-field diffraction patterns for arbitrary apertures are more complicated, however. Various techniques for calculating near-field diffraction patterns are available,[1,3] with numerous approaches for reducing numerical computational complexity (e.g., refs. 4 and 5).

In this paper, we investigate algorithms based on the well-known concept of subsampling to implement efficient numerical convolution for calculating near-field diffraction patterns. We begin with background in Section 2, discussing numerical convolution based on the Fresnel-Kirchhoff equation for determining near-field diffraction. In section 3, we propose *subsampled convolution* and convolution with an *aggregate kernel* to efficiently calculate diffraction pattern sequences for generalized and rectangular apertures, respectively. Section 4 describes FFT-based implementations of the efficient convolution algorithms, while performance of the algorithms is verified in Section 5 through two computational examples.

2. Background

The two-dimensional physical model for near-field diffraction is shown in Figure 1. A plane wave of monochromatic light with wavelength λ propagates along the z-axis and encounters an aperture with amplitude weighting a(x). The amplitude distribution of the light along the ξ-axis in the output plane may be written as

$$f(\xi) = \int_{-\infty}^{\infty} a(x)h(\xi - x)dx\,, \tag{1}$$

where h(x) is the free-space propagation kernel.[1]

We use the free-space propagation kernel as derived from the Fresnel-Kirchhoff equation[1,3], which is given by

$$h(x) = \frac{1}{\sqrt{j\lambda r}}\left[\frac{r+z}{2r}\right]\exp\left[-j\frac{2\pi r}{\lambda}\right],$$

where

$$r = \sqrt{z^2 + x^2}\,. \tag{2}$$

To solve equation (1) numerically, we sample the continuous functions a(x) and h(x), then convolve the resulting discrete sequences. With conventional discrete convolution techniques, we sample both the aperture and kernel functions at the same rate; according to the Nyquist criterion, the function with the largest spatial bandwidth dictates the minimum sampling rate,[6] which in turn affects array size and computation time.

In typical applications, aperture weighting is described by Gaussian[7] or Hermite-Gaussian[8] functions; these weighting functions are generally truncated in practical applications due to the finite extent of elements in optical systems.[9] We have plotted the envelope of the spatial frequency spectrum of a truncated Gaussian aperture in Figure 2, along with the spectrum of a Fresnel-Kirchhoff kernel; for this particular example, the aperture $a(x)=\exp[-4x^2/(L/2)^2]\mathrm{rect}(x/L)$, while $z=L$ and $\lambda=L/40$ in the kernel. The high frequency content of the kernel is much more significant than that of the aperture weighting function, implying that the Nyquist sampling rate is greater for the kernel. Therefore, the Nyquist rate of the kernel function determines the array sizes and computation time for the discrete convolution of these two functions.

One distressing outcome of sampling both the aperture and kernel at the Nyquist rate of the kernel is that the resulting output will also be sampled at the Nyquist rate of the kernel, even though it can be sufficiently represented at a lower sampling rate closer to that required for the aperture. This occurs because the convolution of two functions in the space domain produces a

spectrum which is the product of the two input spectrums; as is illustrated in Figure 2, the output spectrum closely resembles that of the aperture, with the high frequency components significantly reduced from those of the kernel. Ultimately, by sampling both the aperture and kernel at the kernel Nyquist rate and convolving, we calculate more values of the output sequence than are necessary, wasting computer memory and needlessly increasing computation time.

3. Efficient Discrete Convolution by Subsampling

In this section, we develop algorithms which circumvent calculation of extraneous output samples for convolution of aperture and kernel sequences with mismatched bandwidths. If we sample both the kernel h(x) and aperture a(x) at the Nyquist rate for the kernel $1/x_0$, then according to arguments in the previous section, we lose little information if we subsample the convolution of the two functions at a lower sampling rate $1/\xi_0$ which is close to the required aperture rate. If the input aperture is encoded in N_{ap} sample points and ξ_0/x_0=m, where m is a positive integer, we write the subsampled convolution as

$$f(l\xi_0 + px_0) = \sum_{k=0}^{N_{ap}-1} a(kx_0)h(l\xi_0 + px_0 - kx_0)\,. \tag{3}$$

Here, p is an integer between 0 and m-1 which represents the location of the first subsampled point in the output sequence. If N_{ap} is an integer multiple of m, equation (3) is equivalent to

$$f(l\xi_0 + px_0) = \sum_{i=0}^{m-1} \sum_{k'=0}^{\frac{N_{ap}}{m}-1} a(k'\xi_0 + ix_0)h\left[(l-k')\xi_0 + (p-i)x_0\right]. \tag{4}$$

At this point, we introduce a new notation for subsampled arrays such that the argument corresponds to the ξ_0 index of the sequence while a subscript identifies the x_0 index [e.g., $a_i(k)\equiv a(k\xi_0+ix_0)$]. Using this notation, equation (4) can be rewritten as

$$f_p(l) = \sum_{i=0}^{m-1} \sum_{k'=0}^{\frac{N_{ap}}{m}-1} a_i(k') h_{p-i}(l-k') \tag{5}$$

or, alternatively,

$$f_p(l) = \sum_{i=0}^{m-1} a_i(l) * h_{p-i}(l)\,, \tag{6}$$

where * denotes convolution. According to equation (6), we may represent the subsampled output function as the sum of m discrete convolutions between aperture and kernel functions which are subsampled at a rate of $1/\xi_0$; we call this technique *subsampled convolution*. Note that $f_p(l)$ is exactly the sequence we would obtain by subsampling the discrete convolution between an aperture and kernel sampled at x_0; therefore, the results of subsampled calculations will always agree completely with those computed by standard convolution techniques.

For the special case of a rectangular aperture, we may simplify subsampled convolution further. If the sample spacing ξ_0 is chosen such that $a_0(j)=a_1(j)=\ldots=a_{m-1}(j)$ for all j, then equation (6) becomes

$$f_p(l) = a_0(l) * \left[\sum_{i=0}^{m-1} h_{p-i}(l) \right]. \tag{7}$$

Here, we have collapsed m subsampled kernel functions into a single sequence which we call the *aggregate kernel*, and we require only a single convolution between functions sampled a $1/\xi_0$.

4. FFT methods

To evaluate discrete convolutions, we invoke the convolution theorem and apply FFT algorithms; this approach, which is commonly used in digital signal processing applications, implements an efficient circular convolution.[10] To guarantee that the results of circular convolution are identical to those of the direct convolutions of equations (6) and (7), we must "pad" our aperture and kernel arrays with zeros so that the number of sampled points per array in the computation (N_c) equals the sum of nonzero points in the discrete aperture (N_a) and kernel (N_k)

functions.[2] We write this constraint as

$$N_c = N_a + N_k \,. \tag{8}$$

The finite size of N_k causes an abrupt cutoff in the kernel function which is physically inaccurate. As is shown in Figure 3, this is equivalent to hard-limiting the cone of rays that emanate from each point in the aperture function; hence, the output calculation is accurate only for points on the ξ-axis that have a contribution from all points in the input aperture. Using this fact with equation (8), we find that the number of usable points in the output sequence (N_o) is given by

$$N_o = N_k - N_a = N_c - 2N_a \,. \tag{9}$$

The FFT-based implementation of equation (6) is written as

$$f_p(l) = FFT^{-1}\left\{\sum_{i=0}^{m-1} FFT\left[a_i(l)\right] \times FFT\left[h_{p-i}(l)\right]\right\} . \tag{10}$$

With equation (10), we can use arrays that are N_c/m points in length to perform the calculations; this smaller array size reduces both memory requirements and computational complexity.

To determine the reduction in computational complexity afforded by equation (10), we first note that a single FFT requires $N\log_2 N$ complex multiplications.[10] For conventional FFT-based convolution we require two N_c-point FFT operations, an N_c-point complex multiplication, then an N_c-point inverse FFT, resulting in a computational complexity of

$$C = 3N_c \log_2(N_c) + N_c \,. \tag{11}$$

In the algorithm of equation (10), we are required to perform 2m (N_c/m)-point FFT's, m (N_c/m)-point multiplies, then an (N_c/m)-point inverse FFT. For this calculation, our computational complexity is reduced to

$$C_{sc} = \frac{2m+1}{m} N_c \log_2\left[\frac{N_c}{m}\right] + N_c \,. \tag{12}$$

For the typical case where $1 << m <<< N_c$, we find that ratio of the two computational complexities is approximately

$$\frac{C_{sc}}{C} \approx \frac{2}{3} \quad . \tag{13}$$

When the aggregate kernel method of equation (7) is implemented with FFT-based convolution, we can calculate the diffraction pattern as

$$f_p(l) = FFT^{-1}\left\{ FFT\left[a_0(l)\right] \times \sum_{i=0}^{m-1} FFT\left[h_{p-i}(l)\right] \right\} . \tag{14}$$

Since we now represent the diffracted light distribution with a single convolution of two arrays which are N_c/m points long, our computational complexity is only

$$C_{ak} = 3\frac{N_c}{m}\log_2\left[\frac{N_c}{m}\right] + \frac{N_c}{m} \quad , \tag{15}$$

and the resulting ratio of complexities for $1 << m <<< N_c$ is

$$\frac{C_{ak}}{C} \approx \frac{1}{m} \quad . \tag{16}$$

We can extend the methods described above to calculate the diffraction pattern resulting from a two-dimensional input aperture. For the conventional convolution, we would use the same approach except we would need to calculate two-dimensional FFT's and multiply two-dimensional arrays together. Given that a two dimensional FFT requires $2N^2\log_2 N$ multiplications[11], we find that for conventional convolution, a two-dimensional diffraction pattern calculation requires

$$C' = 6N_c^2 \log_2(N_c) + N_c^2 \tag{17}$$

multiplications. Using the subsampled convolution technique, the computational complexity is reduced to

$$C'_{sc} = \frac{4m+2}{m^2} N_c^2 \log_2\left[\frac{N_c}{m}\right] + \frac{N_c^2}{m} \quad . \tag{18}$$

For the case where $1 << m << N_c$, we find that the ratio of the two computational complexities becomes

$$\frac{C'_{sc}}{C'} \approx \frac{2}{3m} \quad , \tag{19}$$

which is equivalent to the one-dimensional aperture ratio scaled by 1/m. For the aggregate kernel method, the two-dimensional computational complexity is given by

$$C'_{ak} = 6\frac{N_c^2}{m^2}\log_2\left[\frac{N_c}{m}\right] + \frac{N_c^2}{m^2} \quad , \tag{20}$$

and the ratio with C' for $1 << m << N_c$ reduces to

$$\frac{C'_{ak}}{C'} \approx \frac{1}{m^2} \quad , \tag{21}$$

which is again scaled by 1/m as compared to a one-dimensional calculation.

5. Computational examples

To illustrate the power of subsampled convolution and the aggregate kernel, we offer two computational examples. In Figure 4, we plot a portion of half of the symmetric normalized intensity pattern due to plane wave diffraction from an aperture with a truncated Gaussian amplitude weighting function; for this particular calculation, we use parameters identical to those used to compute the frequency domain functions of Figure 2. The solid line in Figure 4 represents the diffraction pattern calculated from the conventional FFT-based convolution technique while the circles are points calculated using our subsampled convolution method. For the conventional calculation, $m=\xi_0/x_0=1$, $N_c=1024$, $N_a=80$, $N_k=944$, and all of the 360 plotted points are drawn from the $N_0=864$ usable points described in Section 4; in the subsampled calculation m=4, resulting in values of $N_c=256$, $N_a=20$, $N_k=236$, with all 80 plotted points drawn from the $N_0=216$ usable points.

Here, subsampled convolution sufficiently tracks the mainlobe of the diffraction pattern

($0 \le x \le L/2$) and the envelope of the sidelobes in the region $L/2 \le x \le L$. As we increase x, the sidelobe width increases, allowing subsampled convolution to track the sidelobe signature reasonably well when $x \ge L$. For some purposes, we might consider the portion of the pattern between L/2 and L to be undersampled (even though it appears to meet the Nyquist criterion for the conventional convolution output) because it does not track sidelobe structure exactly; hence, the subsampled calculation might not not be sufficient for applications where one is attempting to predict the correct number of maxima and minima in the shadow region of $L/2 \le x \le L$. The high frequency oscillations in this region are relatively small in magnitude, however, so most of the diffraction pattern information is retained in the subsampled output. Therefore, subsampled convolution is appropriate for many applications, in particular those which require calculating average intensity levels over finite apertures.

Even though little information is lost when we use subsampled convolution, we have a significant economy of computational resources. The array size (and in turn, memory usage) is only one-fourth of that for conventional FFT-based convolution, while the execution time of the computer program used to calculate the subsampled convolution was 63% of that required for a similar program using the conventional technique.

The pattern shown in Figure 5 represents plane wave diffraction from a rectangular aperture $a(x) = \text{rect}(x/L)$ and is calculated using an aggregate kernel; all other conditions are identical to those used in the calculations of Figure 4. Comparing Figures 4 and 5, we see that the quality of the diffraction pattern represented by the aggregate kernel is identical to that of subsampled convolution. The required memory usage for subsampled and aggregate kernel convolution is also the same, but the computational efficiency of the aggregate kernel method is much better: the execution time of the aggregate kernel program is just 29% of that required for conventional FFT-based convolution.

6. Conclusions

We have introduced the method of subsampled convolution for calculating the near-field diffraction patterns that result from plane wave illumination of arbitrary aperture functions. For typical applications, this method provides adequate representation of the diffraction pattern while requiring significantly less memory and computation time than conventional discrete convolution techniques. Convolution with an aggregate kernel may be used for rectangular aperture functions to reduce computation time even further.

The author wishes to thank P. Melsa for helpful discussions concerning FFT implementations, Z.-J. He for computing diffraction patterns to validate algorithm performance, and reviewers for constructive comments regarding manuscript revision.

This work was performed while the author was with North Carolina State University and The University of Alabama in Huntsville, and was supported in part by the U. S. Army Research Office and BellSouth Enterprises, Inc.

References

1. J. W. Goodman, *Introduction to Fourier Optics*, San Francisco: McGraw-Hill Book Company, 1968, ch. 3.
2. D. F. Elliot and K. R. Rao, *Fast Transforms: Algorithms, Analyses, Applications*, New York: Academic Press, Inc., 1982.
3. A. Papoulis, *Systems and Transforms with Applications in Optics*, New York: McGraw-Hill Book Company, 1968, ch. 3.
4. J. J. Stamnes, *Waves in Focal Regions*, Bristol: Adam Higher, 1986, chs. 7 and 14.
5. E. A. Sziklas and A. E. Siegman, "Mode calculations in unstable resonators with flowing suturable gain. 2: fast Fourier transform method," *Appl. Opt.*, vol. 14, pp. 1874-1879, 1975.
6. A. V. Oppenheim and A. S. Willsky, *Signals and Systems*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1983, p. 519.
7. M. V. Klein and T. E. Furtak, *Optics*, New York: John Wiley & Sons, Inc., 1986, ch. 7.
8. D. Marcuse, *Light Transmission Optics*, Princeton, NJ: Van Nostrand, 1972.
9. A. VanderLugt, "Bragg cell diffraction patterns," *Appl. Opt.*, vol. 21, pp. 1092-1100, 1982.
10. R. A. Roberts and C. T. Mullis, *Digital Signal Processing*, Reading, MA: Addison-Wesley Publishing Co., 1987, ch. 5.
11. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1975, ch. 6.

Figure Captions

Figure 1. Diagrammatic representation of the physical model used for near-field diffraction.

Figure 2. Spectrums of input aperture, Fresnel-Kirchhoff kernel, and resulting output pattern. Here, the aperture is Gaussian and truncated at the e^{-4} amplitude level, the propagation distance along z is equal to the aperture width, and the wavelength is 1/40 of the aperture width. Note that for clarity, we plot only the envelope of the oscillatory aperture and output spectrums in the high frequency region.

Figure 3. The finite size of the kernel yields the same effect as limiting the cone of rays which emanate from each point in the sampled aperture function. To arrive at a correct output distribution, there must be a contribution from each point in the input aperture. Therefore, accurate calculations exist only on the shaded portion of the output axis, a distance equal to the kernel length less the aperture length.

Figure 4. Calculated diffraction pattern resulting from convolution of the aperture and kernel used to generate the spatial frequency plots in Figure 2. Solid line represents conventional FFT-based convolution while circles denote subsampled convolution with m=4.

Figure 5. Calculated diffraction pattern resulting from convolution of the kernel used for frequency plots in Figure 2 with a rectangular aperture. Solid line represents conventional FFT-based convolution while circles denote aggregate kernel convolution with m=4.

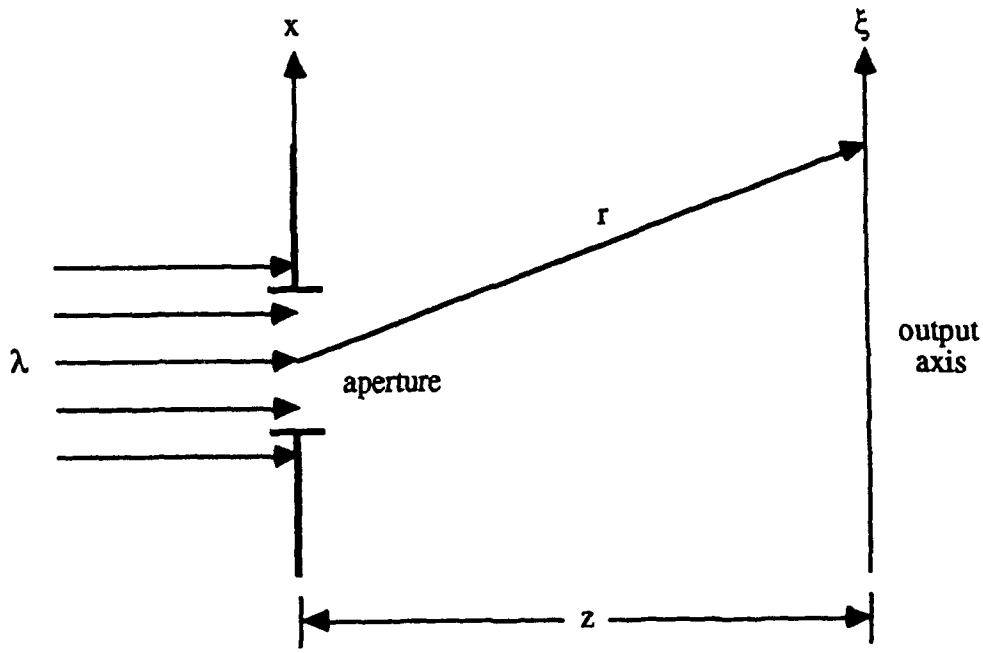


Figure 1.

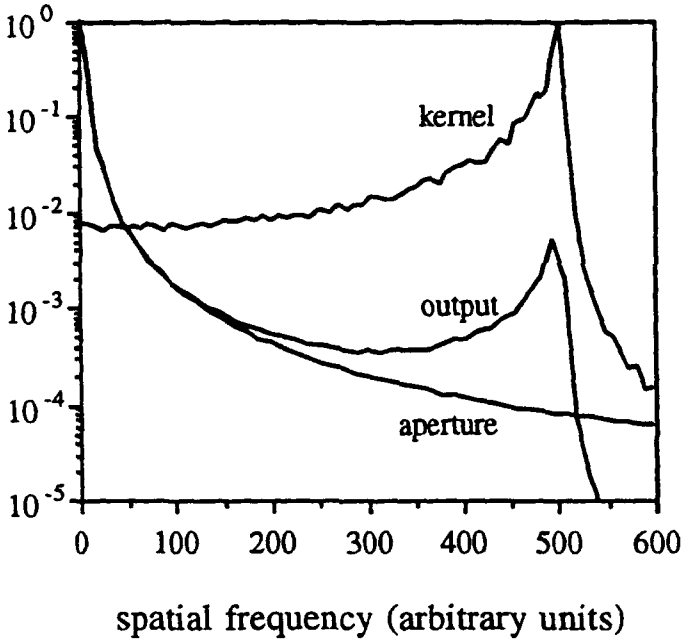

10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
kernel
output
aperture
0
100
200
300
400
500
600
spatial frequency (arbitrary units)

Figure 2.

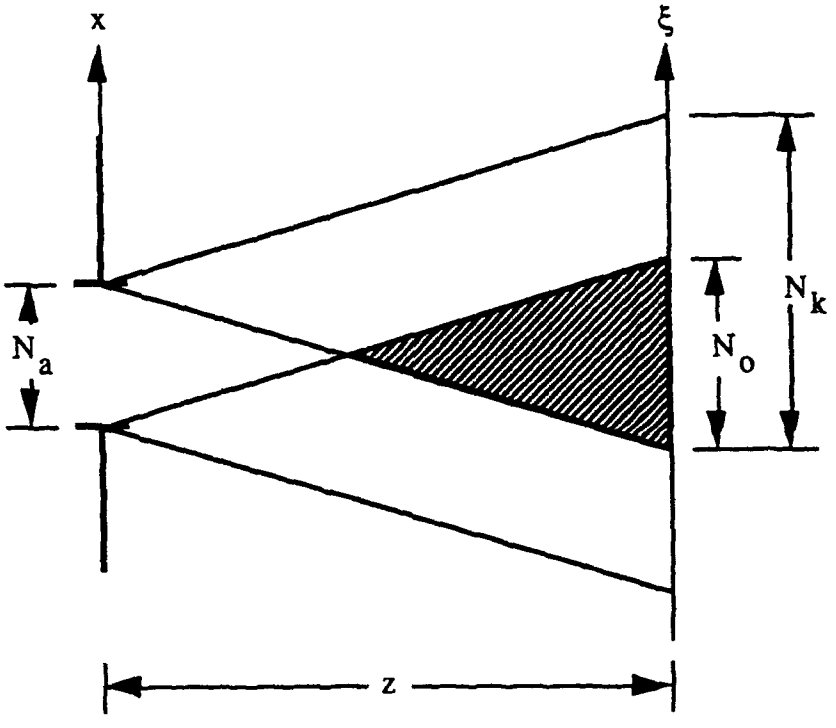


Figure 3.

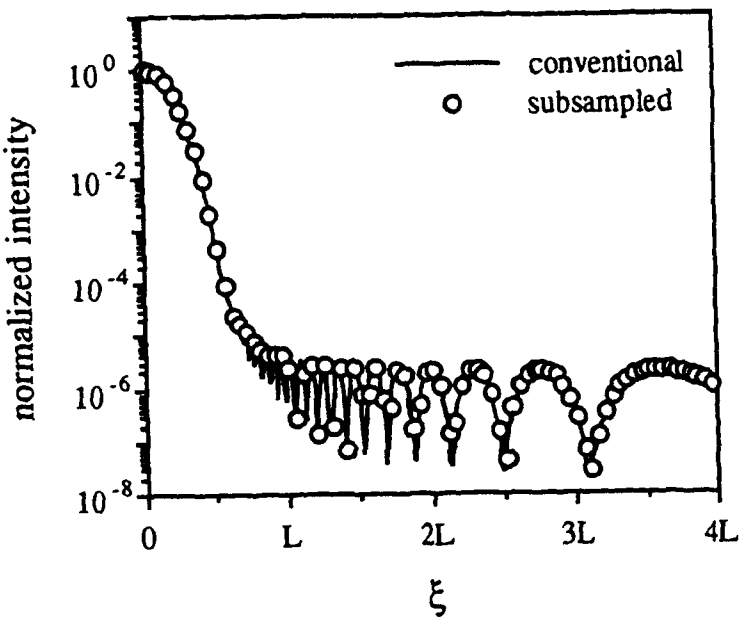

conventional
subsampled
normalized intensity
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
0
L
2L
3L
4L
ξ

Figure 4.

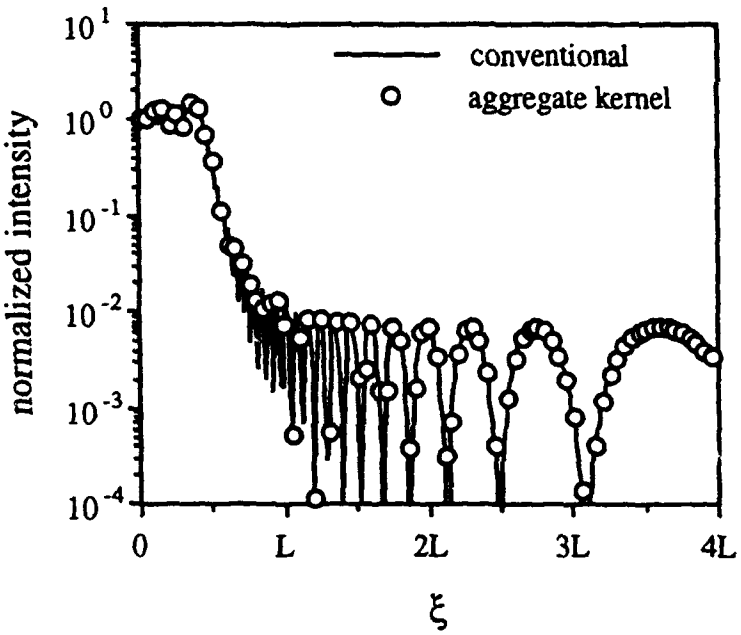

conventional
aggregate kernel
normalized intensity
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
L
2L
3L
4L
ξ

Figure 5.