1

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

35523

National Library
of CanadaBibliothèque nationale
du CanadaCANADIAN THESES
ON MICROFICHETHÈSES CANADIENNES
SUR MICROFICHENAME OF AUTHOR/NOM DE L'AUTEUR Dipak RoyTITLE OF THESIS/TITRE DE LA THÈSE "Electro-Optical Processing of Signals and Images Using CCDs"UNIVERSITY/UNIVERSITÉ Carleton UniversityDEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE DoctoralYEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ 1977NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE Professor M.A. Copeland

Permission is hereby granted to the NATIONAL LIBRARY OF
CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and neither the
thesis nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈ-
QUE NATIONALE DU CANADA de microfilmer cette thèse et
de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la
thèse ni de longs extraits de celle-ci ne doivent être imprimés
ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE _____ SIGNED/SIGNÉ Dipak RoyPERMANENT ADDRESS/RÉSIDENCE FIXE 100/C, BALLYGUNGE GARDENSCALCUTTA - 700029INDIA

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

ELECTRO-OPTICAL PROCESSING OF
SIGNALS AND IMAGES USING
CCDs

by

Dipak Roy, M.Eng.

A thesis submitted to the Faculty of
Graduate Studies and Research in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy.

Faculty of Engineering
Carleton University
Ottawa, Canada
1977

©

Dipak Roy

1977

The undersigned recommend to the Faculty of Graduate
Studies and Research acceptance of the thesis:

"Electro-Optical Processing of Signals and Images
Using CCDs",

submitted by Dipak Roy, B.E. & Tel. E, P.G. (Dip.Comp. Sc.), M.Eng.,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

MA Copeland

Thesis Supervisor

OR Booth

Chairman, Department of
Electronics

Carleton University

ABSTRACT

This thesis is based on the concept of functional convolution inside a CCD using an optical input method. The concept has been utilized to demonstrate a mask programmable transversal filter for electrical signals and an electrically programmable image processor. The important limitations of this approach have been studied in terms of accuracy linearity and noise. The major application potentials are explored, particularly for implementation of two-dimensional signal processing algorithms.

ACKNOWLEDGEMENTS

The author wishes to express his indebtedness to his supervisor, Professor M.A. Copeland, for his support, excellent guidance and encouragement during the course of this work.

The author also wishes to thank Dr. C.H. Chan for many interesting comments on the work. Many useful discussions with Mr. L. Berndt are gratefully acknowledged.

Special thanks are due to the fellow graduate students, namely, C.F. Rahim, Y.A. Haque, B.K. Ahuja and D.A. Harper for many stimulating discussions.

The author also wishes to express his deep appreciation to Mr. H. Kovalcik for many useful discussions on the CCD camera, and to Mr. P. Menard on the INTEL MC-48 microprocessor system.

The author thanks his girlfriend Tara for her support and encouragement. The speed and very efficient typing of Miss Catherine Maguire is gratefully acknowledged.

The financial assistance of the National Research Council of Canada and material support of the Bell Northern Research Laboratories, Ottawa are deeply appreciated.

TABLE OF CONTENTS

	<u>PAGE</u>
Chapter 1. INTRODUCTION	1
1.1 Introduction	1
1.2 Historical Background and Related Work	1
1.3 Thesis Organization	6
Chapter 2. SOME PRELIMINARY CONSIDERATIONS	9
2.1 Introduction	9
2.2 Description of the Processor	9
2.3 Experimental Arrangement	13
2.4 Experimental Demonstrations	22
2.5 Effect of Charge Transfer Inefficiency and Finite Width Pulse Sampling	28
Chapter 3. STUDY OF LIMITATIONS	33
3.1 Introduction	33
3.2 Frequency Response Error	34
3.3 Measurement of Linearity	46
3.4 Noise and Speed Considerations	52
3.5 Analysis of Limitations Arising from Optical Charge Injection	62
3.5(a) Transmittance of Light Through The CCD Electrode Structure	62
3.5(b) Relationship Between Storage Charge and Light Intensity	76
3.5(c) Effect of Lateral Diffusion	89
3.5(d) Clock Jitter Noise	96
3.6 Linearity of the Output Stage	97

	<u>PAGE</u>
Chapter 4. REALIZATION OF GENERAL FILTER FUNCTIONS.	118
4.1 Introduction	118
4.2 Negative Taps	118
4.2(a) Temporal Multiplexing	120
4.2(b) Spatial Multiplexing	129
4.3 Discussion	142
Chapter 5. PROCESSING OF SPACE VARYING SIGNALS	148
5.1 Introduction	148
5.2 Experimental Arrangements	149
5.3 Experimental Results	153
5.4 Performance Considerations	172
5.4(a) The Effect of Transfer Inefficiency	172
5.4(b) The Effect of Coefficient Quantization	177
5.5 Two-Dimensional Processing	179
5.6 Experimental Configuration	181
Chapter 6. CONCLUSIONS AND FURTHER WORK	185
References	188
Appendix A. Fabrication of On-Chip Masks	192

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

This thesis discusses a novel electro-optical processor that can be used for processing of electrical signals and optical images. A general description of the processor includes a charge coupled device (CCD) delay structure, illuminated either by a modulated light source or an object, and some means to control the charge gathering capability of the CCD delay elements. Justification for this work lies in the interest in obtaining one-dimensional linear transformations and filters for electronic signals in communication systems, and in the recent interest in obtaining transformations of optical images for coding, pattern recognition and bandwidth reduction.

1.2 HISTORICAL BACKGROUND AND RELATED WORK

Earlier than two decades ago, signal processing was typically carried out using analog components, and processing was mostly confined to frequency shaping, using passive and active filters. In the early 1950's the flexibility of a digital computer appeared extremely inviting for sophisticated signal processing applications. However, the early work on digital signal processing was mainly concerned with approximating or simulating an analog signal processing system using discrete-time mathematics. The continued development of digital computers, and the increasing use of digital signal processing have led to the development of extremely sophisticated signal processing algorithms.

Using these algorithms a variety of signal processing operations that have no analog equivalent can be efficiently performed. Further, special purpose digital hardware have emerged that can implement these algorithms in near real time. While digital signal processing is a rapidly growing field, there are many real time applications, such as speech communication, radar processing, etc, where speed, cost, and size are important factors in favour of the use of analog components.

With the advent of charge transfer devices, (charge-coupled devices and bucket brigade devices) in the period 1969-70, clock-controlled time delays of analog signals became possible. Previously, no simple and compact way of providing an accurate time delay of an analog signal was available. This opened up the possibility of implementing discrete-time algorithms in purely analog form. Although elimination of expensive A to D and D to A converters are a major attraction of this approach, other advantages include low cost, low power and monolithic implementation of a discrete-time analog signal processor.

Fundamental to the linear processing of discrete-time signals is the action of convolution which is based on a series of shifts and multiplication by weighted constants as below:

$$y(t) = [h_0 \times (t) + h_1 \times (t-T) + h_2 \times (t-2T) + \dots] \quad (1.1)$$

Since a charge coupled device (CCD) can move sampled analog signals across the semiconductor substrate, the shifting action necessary in a convolution is provided inside the CCD chip. The weighting of the time shifted signal can be achieved by capacitively sensing the signal charge packets at each delay stage, as in the "split-electrode" CCD

transversal filter [1.1]. In order to non-destructively weight the charge packets, area ratioing of a phase electrode in each delay element is done, in order to vary the sensing node capacitances according to an impulse response pattern ($h_0, h_1 \dots h_N$ in eqn. 1.1). This pattern once chosen, cannot be altered. The fixed weighting coefficient CCD convolvers or filters of this type are factory programmable, in the sense that the impulse response pattern is determined by a photo mask, which is used to define the metal on the filter chip. The advantage of fixed weighting is that it results in dramatic cost saving, because the CCD and weighting network can be integrated on a chip and mass produced. However, each filtering application requires a custom design, and this becomes cost effective only when the volume is sufficiently high.

Although real time programmable transversal filters have been reported in the literature [1.2, 1.3], much of the simplicity of fixed coefficient CCD filters are lost. Moreover, a number of chips are required to implement a modest size filter function. Consequently, non-identical device characteristics limit the available dynamic range [1.4]. Clearly, there are certain advantages of having a discrete time analog signal processor that does not sacrifice the simplicity and performance of the fixed coefficient CCD filters, yet is more readily programmable at the user's level. One such approach is investigated in this thesis.

The action of convolution discussed earlier can be obtained by moving a CCD potential well relative to a space varying illumination as shown in Fig. 1.1. Since a semiconductor is sensitive to light, the

Fig. 1.1. Convolution resulting from relative motion of a CCD potential well.

amount of charge collected in a potential well is dependent on the light intensity as well as the time the well spends at any particular position. A progressive summation takes place as the well moves past the spatially varying signal. The result of a convolution is available directly at the output of the CCD. A real advantage in terms of simplicity results from the elimination of the need to detect charge at every delay stage of the CCD register.

The electro-optical processor just discussed can be used to process either varying electrical signal or space varying optical images. In the former case, the coefficients of a transformation are encoded into an optical transparency making the arrangement mask programmable. In the latter case, the coefficients are applied as time varying electrical signal making the arrangement electrically programmable.

The work reported* in this thesis is not devoted to any specific application. Rather, the general potential of the novel electro-optical processor has been investigated by a number of different experiments both in the context of electrical signal processing and image processing applications. The important performance limitations have also been studied in terms of accuracy, linearity and noise.

It should be mentioned that the concept of performing correlation using an incoherent light source was conceived before the CCD became available. The advantages of high speed analog-multiply and parallel processing capability inherent in an optical system have

* A part of the work has also been reported elsewhere [1.5, 1.6, 1.7] by the author.

led to the development of a number of optical correlators [1.8, 1.9, 1.10]. Most of these correlators employ an integrating detector, for example, a vidicon or a photographic plate, and some means to scan a mask image across the face of the detector. The scanning arrangement provides the relative motion necessary to perform a convolution or a correlation operation. Movement of the detector instead of the optical mask provides identical results.

It is obvious from the above discussion, that a real advantage in terms of system simplicity is obtained, if the scanning operation can be performed within the detector itself. The charge coupled device structure provides such an arrangement and is therefore very suitable for electro-optical signal processing. The use of CCD as a non-scanning correlator was also proposed by the researchers of the Naval Electronics Research Laboratories [1.11, 1.12]. However, up to this date no detailed results are available. The objective of this thesis is to investigate the advantages and limitations of various concepts involved in electro-optical processing of signals and images using the charge coupled device. The concept of linear image processing, discussed in this thesis, is electronically controllable and can be extended to actual two-dimensional processing of images. Such processing capability of a charge coupled device is demonstrated for the first time here.

1.3 THESIS ORGANIZATION

The contents of this thesis are divided into two parts: the first part deals with processing of time varying signals (i.e., electrical

signals; the last part deals with processing of space varying signals i.e., images. The contrast between these two types of processing is highlighted in Chapter 5 prior to the discussion of image processing.

Chapter 2 discusses the principle of operation of the electro-optical signal processor, the experimental considerations and some preliminary results to demonstrate the action of convolution.

Chapter 3 is devoted to the study of limitations arising from such use of a charge coupled device with an optical method of injecting charge. Some preliminary experimental results on accuracy, linearity and noise have been presented for an electrical filter function. The possible sources of errors, non-linearity and noise have also been theoretically investigated. The discussions will generally apply to image processing as well, since both types of processing are functionally similar.

Chapter 4 is devoted to the techniques of realizing any general filter function. The problem of achieving negative taps has been solved using two approaches. The limitations of each approach are discussed in terms of practical realizations. Also discussed in this chapter are the advantages of the parallel processing ability of the electro-optical processor for processing large quantities of data in near real time.

Chapter 5 deals with image processing. The action of correlation of a linear image pattern with a set of predetermined coefficients has been experimentally demonstrated. The components of the Discrete Fourier Transform have been extracted from the results of the correlation. Discussions on two-dimensional processing of an

image have also been presented along with some initial results.

Conclusions appear in Chapter 6. Appendix A discusses the fabrication considerations involved in making on-chip metal masks for realizing very accurate filter functions.

PROCESSING OF TIME-VARYING SIGNALS

CHAPTER 2

SOME PRELIMINARY CONSIDERATIONS

2.1 INTRODUCTION

In this chapter, the electro-optical signal processor is introduced in a specific context, as an alternate approach to the split electrode CCD filter. The experimental arrangements necessary to perform linear filtering or transformation of an electrical signal are discussed in the context of a compact mask programmable structure. Finally, the concept is validated by experimental results, that compare favourably with the theoretical predictions.

2.2 DESCRIPTION OF THE PROCESSOR

Most signal processing applications of charge transfer devices (CTD's) rely on the fact that these devices can provide accurate, clock-controlled time delays of an analog signal. Of particular interest is the implementation of transversal filters [2.1] which, as a general class, use time delayed waveforms of a given input signal, weighted and summed with the proper signs to generate a wide range of filter characteristics with a finite impulse response (FIR). The schematic diagram of a tapped delay line transversal filter is shown in Fig. 2.1(a). Two sets of tap weights and a differential summing amplifier are used to provide tap weights of both polarities. The method indicated in the figure is functionally similar to that used in a split electrode structure [1.1]. The analog signal is

(a)

(b)

Fig. 2.1. Schematic diagram of (a) a tapped delay line transversal filter; (b) a split-electrode transversal filter.

introduced at the input of the CCD register and, as it is clocked through the system, the magnitudes of the charge packets beneath the CCD electrodes are sensed by measuring either the total current to the CCD electrodes or the potentials of the "floating" CCD electrodes [2.2]. Tap coefficients are obtained by area manipulation of the electrodes as indicated by the position of the split in a schematic plan of the electrodes shown in Fig. 2.1(b). It should be appreciated that in this mode of operation N tap weighted samples of the analog signal are simultaneously summed, thereby requiring the need for charge detection at every delay stage of the register.

An alternative technique for transversal filter realization that obviates the need for charge detection at every stage of the register is to use progressive summation approach shown in Fig. 2.2(b) as against the simultaneous summation approach shown in Fig. 2.2(a). Tap weighted charge can be generated inside the various potential wells of a charge coupled device by making use of an intensity modulated light source and a photographic mask which controls the amount of light reaching the various electrodes. A progressive summation is done inside a given potential well as the well is created and then progressively clocked to the output end of the device. Thus, the result of a convolution is directly obtained at the output of the device. Since the various multiplications necessary for a convolution are governed by the transmission ability of an optical mask, then by manipulating the ratio of the opaque to clear areas above each electrode according to an arbitrary pattern, a mask-programmable transversal filter can be obtained.

(a)

$D_1 = \text{UNIT DELAY}$

(b)

Fig. 2.2. Two approaches for transversal filter realization:
(a) simultaneous summation; (b) progressive summation.

A major advantage of this approach is the simplicity of the arrangement. A CCD line imager, an LED and the capability of producing photoemulsion masks can provide one with a powerful signal processing tool. Since charge is sensed only once at the final output, as in a linear CCD imager, comparable performances can be expected. Further, a commercial area array imager can be used for multiple processing of a signal, e.g., spectrum analysis.

2.3 EXPERIMENTAL ARRANGEMENT

The basic components of the electro-optical signal processor consist of an LED with associated drive circuits, a CCD imager and a coefficient mask designed according to a desired filter function. A schematic of the experimental set up is shown in Fig. 2.3. The analog signal to be processed is used to modulate the output of a light emitting diode. The radiation from the LED uniformly illuminates a photographic mask which controls the amount of light reaching the charge coupled device. The device used was a CCD delay line 112-elements long and was fabricated using 2-phase polysilicon gate N-channel process [2.3]. The devices were not designed specifically for imaging purposes. The element length in the direction of charge flow was $24\mu\text{m}$ and the width was $660\mu\text{m}$. Fig. 2.4(a) shows the device structure including the off chip output stage used for signal reconstruction. The timing diagram of the necessary clocks are shown in Fig. 2.4(b). The LED was turned on when there was a valid potential well under any one of the phase electrodes. Since the amount of photogenerated carriers collected at the silicon surface is proportional to both the

Fig. 2.3. A block diagram of the experimental setup.

(a)

(b)

Fig. 2.4 (a) The device structure; (b) timing diagram of the clocks. (The devices were obtained from the Bell Northern Research Laboratories, Ottawa)

intensity of light reaching the silicon surface and the on-time of LED (see section 3.5(b)), the LED can be either intensity modulated or duty cycle modulated as indicated in Fig. 2.4(b). In either case, the maximum on time of the LED was kept within the on-time of ϕ_1 and the off-time of the ϕ_2 clock pulse to avoid carrier generation during transfer. The mechanism of photogeneration of carriers and its limitations on the performance of the processor are discussed in detail in the following chapter. In most of the experiments discussed in this thesis the LED was intensity modulated by placing it in the collector circuit of a transistor amplifier (see Fig. 3.7). A series resistor in the transistor base circuit caused the sampled signal voltage to be converted to a linear base current, which was amplified to provide the collector current drive for the LED. Duty cycle modulation of the LED can be achieved by a linear ramp generator (a constant current source charging a capacitor) and a comparator which compares the output of the ramp generator to the sampled signal voltage. The circuit for intensity modulation is considerably simpler than that for duty cycle modulation and may be realized with higher accuracy at lesser expense. Both modulation schemes can be utilized in order to achieve multiplication of the input signal with another signal as is required in the chirp-Z-transform [2.4].

In most of the experiments discussed in this thesis, the LED used was the Monsanto MV5024 which has a luminous intensity in the range of 2 - 3 mcd. Since it is desired to have uniform illumination over all parts of the optical mask (to avoid weighting the tap coefficients

due to irregularities in the beam) knowledge of the spatial intensity distribution of the LED light is necessary. Fig. 2.4 shows the spatial intensity distribution of the MV5020 series. The figure indicates intensity relative to the peak intensity of over 98% at $\pm 5^\circ$, for the LED used. Thus a spacing of 1cm between the LED and the optical mask ensures almost uniform illumination over a square area of about 0.015cm^2 . By increasing the distance of separation between the LED and the coefficient mask even better uniformity of illumination can be obtained. However, this may require higher drive requirement for the LED in order to ensure adequate light intensity at the semiconductor surface. Using an LED with higher luminous intensity (e.g., MV4) and lensing arrangements, the image of a 35mm negative can be directly focussed onto the device. This configuration can be employed for picture processing experiments discussed in Chapter 5. No lenses were used in the experiments here for processing electrical signals. The LED was mounted on a micrometer arrangement for accurate positioning of the LED over the photomask. Small adjustments in the positions of the LED can be made in all three directions by observing the electrical output from the charge-coupled device.

The coefficient mask was cut in rubylith according to a desired impulse response pattern from which was subsequently generated a suitably reduced glass positive in a two step reduction process. The original artwork was designed 250 times the actual size to minimize error in the preparation of the coefficient mask. The glass mask was placed on top of the device header and aligned using a microscope. Small alignment errors can be corrected by observing the electrical

Fig. 2.5. Manufacturer's data showing the spatial intensity distribution of the Monsanto MV5020 series. The LED used in the experiment was MV5024.

impulse response and repeating the alignment procedure.

Since the mask is placed on the device reader, there is a non-zero distance of separation between the mask and actual chip. This along with any lack of collimation of the light beam results in a small blurring of the mask image. The situation is illustrated in Fig. 2.6(a). Assuming that no lenses are used other than the one embodied in the LED, blurring occurs around the area under the edge of an opening in the glass mask. The blurring therefore has an effect of widening the mask pattern which is primarily dependent on the distance of separation between the LED and the glass mask, the distance of separation between the glass mask and the CCD, device size and the length of the mask pattern. The approximate dimensions are shown in the figure. The blurring effect is visible in the impulse response. Fig. 2.6(b) shows the impulse response of a rectangular opening that covers 10 CCD delay elements. The effect of blurring on the frequency response can roughly be estimated by considering this effect as a convolution of the ideal impulse response with a rectangle of appropriate width. For the case of Fig. 2.6(b) this width is approximately 3 CCD delay elements for a distance of separation less than a centimeter between the LED and the mask. Since convolution in the time domain is equivalent to multiplication in the frequency domain, the idealized frequency response will be multiplied by the frequency response of the rectangle. This will produce a $\sin x/x$ roll-off* of the

* A three tap rectangle has a frequency response given by $\sin(3\pi f/f_c) / (3\pi f/f_c)$; where f_c is the clock frequency.

(a)

(b)

Fig. 2.6(a) Geometrical arrangement to illustrate blurring of the mask image; (b) the effect of blurring on a rectangular window with the LED placed less than 1 cm away from the mask.

ideal frequency response. The width of the convolving rectangle can further be reduced by using a larger distance of separation between the LED and the glass mask (adequate light intensity was available by placing the LED 1 - 2cm away from the photomask). Thus if a filter passband is restricted to only a small fraction of the Nyquist frequency limit (slowly changing tap weight patterns), the blurring error in the passband is small. Without using any collimating lens, filters with a passband extending up to 20% of the Nyquist frequency limit have been realized. The blurring error, however, may be significant for filters with fast changing tap patterns.

The problem of lack of collimation of the incident light beam may be overcome by using a collimating lens between the LED and the glass mask. An alternate solution would be to use an on-chip mask. This also reduces any misalignment error. However, the ease of programmability is lost with this approach.

In an attempt to realize more accurate filters with fast changing tap patterns, an on-chip metal mask was fabricated by evaporating aluminum over a pyrox-coated device and subsequently etching away the undesired aluminum according to the tap pattern. The results were not very satisfactory. The major problems in fabricating on-chip masks are defects in pyrox and random etching of aluminum. Since we had a small number of devices to work with, this approach had to be dropped. Some discussions of the fabrication steps involved in making on-chip masks and some initial results are presented in Appendix A. The experimental results reported in the main body of this thesis were all obtained using glass masks. The filter passbands were restricted to

well below the Nyquist frequency limit.

2.4 EXPERIMENTAL DEMONSTRATIONS

An initial illustration of the action of convolution is shown in Fig. 2.7 (top trace). The optical mask used for this experiment was a rectangular window that exposed 80 CCD delay elements. The input to the LED shown in the bottom trace of the figure was a rectangular pulse of width equivalent to 35 time samples. The resultant convolution sum has the shape of a trapezoid of total length equivalent to 115 time samples. The experimental results agree with the theoretical prediction, governed by the classical convolution equation:

$$y(k) = \sum_{n=0}^{\infty} x(n)h(k-n) \quad (2.1)$$

where,

$\{x(n)\}$ is the input sequence

and

$\{h(n)\}$ is the tap weight sequence as given by the coefficient mask.

Note that in order to obtain a stable scope display the input signal was repeated after every 128 time samples. The CCD clock frequency was 20 KHz.

The convolution of two time sequences as discussed above corresponds to multiplication of the frequency spectra of the sequences. The action of filtering is thus obtained if the tap sequence $\{h_n\}$ approximates the impulse response of a given filter. The most straight-

Fig. 2.7. The result of convolution of two rectangles. The bottom trace is the input signal and the top trace is the output from the CCD. Horizontal sensitivity is 1ms/div.

forward design approach is to obtain a finite length impulse response by truncating an infinite duration impulse response sequence of the ideal filter. Such simple truncation or "rectangular windowing" of the actual impulse response however, gives rise to "smearing" of the desired spectrum and rather high peak amplitude of the side lobes. [2.5]. By tapering the window smoothly at each end towards zero the height of the side lobes are reduced at the expense of a wider main lobe. Two important window functions of this type are the Hamming and the Blackman window functions. These window functions by themselves provide adequate information on the performance of the processor as they provide fairly large side lobe attenuations. Further, they all have only positive coefficients which can be easily implemented using a single CCD channel.

The initial experiment was performed on a 91-tap Hamming window function. The impulse response of the window function is governed by the relationship:

$$h_n = 0.54 - 0.46 \cos \left(\frac{2\pi n}{N-1} \right); 0 \leq n \leq N-1 \quad (2.2)$$

where,

N is the total number of taps.

The coefficient mask, according to eqn. (2.2), was first cut in ruby-lith, from which a suitable reduced glass positive was made. The largest tap, which has a value of one (occurs at $n = \frac{N-1}{2}$) was made to correspond to a maximum permissible channel width of $600\mu\text{m}$. Each tap window was allowed to expose an entire delay element i.e., $24\mu\text{m}$ in the direction of charge flow. Fig. 2.8(a) shows a photograph of the actual photomask used in the experiment. The observed impulse response,

(a)

(b)

Fig. 2.8. (a) The mask pattern; (b) The impulse response of the Hamming window function.

i.e., the response obtained at the output of the CCD corresponding to a single light flash, is shown in Fig. 2.8(b). The light flash was of duration less than the on-time of the ϕ_1 clock pulse and was repeated every 128 CCD clock pulse in order to obtain a stable scope display. To a first approximation, there is a good agreement between the theoretical and observed impulse response.

The frequency response of the Hamming window function was measured by using a wave analyzer. The experimental and the ideal response are shown in Fig. 2.9. The ideal theoretical response was computed on the basis of the system transfer function (see Fig. 2.2) given by:

$$H(z) = \sum_{n=1}^N h_n z^{-n} \quad (2.3)$$

where,

$$z = e^{j\omega T}$$

ω = signal frequency in rad/sec.

and T = sampling clock period (i.e. CCD clock period)

In Fig. 2.9 the response is plotted against normalised signal frequency, which is the ratio of signal frequency to clock frequency.

Because of the sampling action, the spectrum repeats at every harmonic of the clock frequency i.e., when the normalised signal frequency is an integral number. Meaningful information is contained up to Nyquist frequency limit given by $\omega T = \pi$ i.e., when the normalised signal frequency is equal to one-half.

Fig. 2.9 Frequency response of the Hamming window function.

The results shown in Fig. 2.9 indicate fairly good agreement with the theory. About 40dB stop band attenuation has been obtained with fairly predictable passband characteristics. The small errors in the stop band region are due to the tap weight errors, which are discussed in detail in the following chapter.

2.5 EFFECT OF CHARGE TRANSFER INEFFICIENCY AND FINITE WIDTH PULSE SAMPLING

Eqn. (2.3), which was used to calculate the frequency response, assumes that the delay elements of the charge coupled device are ideal. However, in a charge coupled device the propagation of signal samples is accomplished by transferring charge between delay stages. Thus inefficient transfer process can cause charge corresponding to a signal sample to be left behind. This particular behaviour of a CCD delay element has been modelled [2.6] as a non-ideal delay element, represented by Z^{-1} , by considering an averaged transfer inefficiency ϵ per delay element as below:

$$Z^{-1} = \frac{1-\epsilon}{1-\epsilon Z^{-1}} Z^{-1} \quad (2.4)$$

By replacing Z^{-n} in eqn. (2.3) by Z^{-n} and assuming $\epsilon \ll 1$, the transfer function of a CCD transversal filter can be obtained as:

$$H_T(Z) = \sum_{n=1}^N h_n \exp [-n\epsilon(1-Z^{-1})] Z^{-n} \quad (2.5)$$

The consequence of eqn. (2.5) is a reduction in bandwidth of the filter. The effect of transfer inefficiency is more pronounced on filters having passbands close to the Nyquist frequency limit.

In determining the effect of transfer inefficiency the product $N\epsilon$ is important. Filters with larger number of taps are more susceptible to this effect. Fig. 2.10(a) shows a plot of the theoretical frequency response of the 91-tap Hamming window function for two rather extreme values of ϵ . Fig. 2.10(b) shows a similar plot for a 32-tap Blackman window function discussed in the following chapter. The effect of transfer inefficiency in the stopband region is more pronounced in the former case because of the longer line length used.

Negligible variation from the ideal response is expected in the experimental case for the measured ϵ around 0.0015. ϵ was computed from the frequency response curve by operating the device as an electrical delay line [2.7]. (ϵ ranges from 10^{-4} to 10^{-5} in modern devices.) It is in principle possible to circumvent the effect of transfer inefficiency by properly weighting the tap coefficients [2.8]. However this requires an accurate knowledge of ϵ which may vary from device to device and is clock frequency dependent.

In addition to modification of the filter transfer function for charge transfer inefficiency, it is also necessary to incorporate the effect of finite width pulse sampling at the input and at the output in order to accurately predict the filter response. This can be easily achieved [2.9] by multiplying the transfer function $|H_T(Z)|$ by the frequency (ω) dependent function

$$\sin\left(\frac{\omega\tau_i}{2}\right) \sin\left(\frac{\omega\tau_o}{2}\right) / \left(\frac{\omega\tau_i}{2}\right) \left(\frac{\omega\tau_o}{2}\right)$$

Fig. 2.10(a) The effect of transfer inefficiency on the Hamming window response.

Fig. 2.10(b) The effect of transfer inefficiency on a 32-tap Blackman window response.

where τ_i and τ_o represent the width of the input and output sampling pulse. Clearly, the effect of finite width pulse sampling is to introduce additional roll-off of the frequency response, significantly near the Nyquist frequency limit.

CHAPTER 3

STUDY OF LIMITATIONS

3.1 INTRODUCTION

In the previous chapter we have demonstrated the use of an LED/CCD structure for processing electrical signals. In this chapter we shall discuss the performance of the signal processor in terms of accuracy, linearity and noise. This chapter is essentially divided into two parts: sections 3.2, 3.3 and 3.4 respectively deal with the experimental results on frequency response errors, harmonic distortions and noise of the output signal. The important sources of limitation are mentioned along with the discussions of the experimental arrangements and results. Theoretical investigations into these sources of limitations appear from section 3.5 onwards. Most of these limiting effects arise from the use of a charge coupled device as central to the processor, and the optical method of inputting charge.

The experimental vehicle used was a 32-tap Blackman window filter function which was primarily chosen for two reasons: (a) the window function has large side lobe attenuations, which will cause non-ideal effects to show up readily (b) the window function can be easily implemented using a single CCD channel since all taps are positive. The last point is important, as the results will apply also to linear image processing (discussed in Chapter 5) which utilizes a single CCD channel. The additional sources of limitations for image processing and for filters with bipolar tap weights are discussed in their specific contexts in the following chapters. Only 32 taps

were chosen in order to minimize the errors due to transfer inefficiency and blurring.

3.2 FREQUENCY RESPONSE ERROR

The basic action of a transversal filter involves weighting and summing of time delayed samples of a given input signal waveform. Thus errors in the weighting coefficients give rise to non-ideal frequency characteristics. An important consideration in filter design is the knowledge of the sensitivity of the transfer function to errors in the coefficient value. In general, the filter transfer function $|H(Z)|$ is most sensitive to errors in the largest coefficients; and the region of highest sensitivity is the region where $|H(Z)|$ is small, i.e., the stop band region. In the processor under discussion the important sources of tap weight errors are: tap weight quantization error i.e., the errors introduced in the tap weights due to inherent quantization in the mask making procedure; errors arising from random variation of the processing parameters of the device, such as, variation of transmittance of light through the multilayer electrode structure of the device; lateral diffusion of carriers generated in the neutral bulk; and errors associated with the optics, such as non-uniformity of the light source and blurring of the mask image. The errors associated with the optics, as discussed in the previous chapter, can be greatly improved upon using proper lensing arrangements or on-chip masks. The errors due to variation of light transmittance along the CCD channel can be potentially important and are discussed in section 3.5(a). The ultimate limit on accuracy with which a desired

filter function can be implemented is dependent on the mask resolution.

The mask making procedure involves scaling the tap weight vectors so that the maximum tap weight equals the largest value permitted by the device size (the channel width) and then rounding off all remaining tap weights to the nearest integral multiple of the minimum co-ordinatograph spacing. The original artwork was drawn at 250 times the actual size of the device. Thus the quantization step size Q is given by 250 times the ratio of the channel width to the co-ordinatograph resolution. The round off error can be assumed to be uniformly distributed between $-Q/2$ to $Q/2$ with zero mean and variance $Q^2/12$. The mask quantization places an ultimate limit on the accuracy of a filter and therefore it is necessary to investigate the frequency response error due to this particular source of coefficient errors.

If $\{h'(n)\}$ is the sequence that results due to mask quantization of an original sequence $\{h(n)\}$, then

$$h'(n) = h(n) + e(n) \quad (3.1)$$

where $\{e(n)\}$ is the error sequence. For each value of n , $e(n)$ is a number that satisfies $|e(n)| \leq Q/2$. The error response of an arbitrary phase filter is thus

$$E_A(e^{j\omega T}) = \sum_{n=0}^{N-1} e(n) \cdot \exp[-j\omega nT] \quad (3.2)$$

where ω is the frequency in rad/sec.

and T is the clock period.

In a linear phase filter the tap weights are symmetrical with the property that

$$h(n) = h(N-1-n) \quad (3.3)$$

The quantized sequence $\{h'(n)\}$ also retains the above property, producing an error sequence that is symmetrical around $n = \frac{N-1}{2}$. The error response of a linear phase filter is [3.1],

$$E_L(e^{j\omega T}) = \begin{cases} \left[\sum_{n=0}^{(N-2)/2} 2e(n) \cos\{\omega(n - \frac{N-1}{2})\} \right] \exp\{-j\omega(N-1)T/2\} & \text{for } N \text{ even} \\ \left[\sum_{n=0}^{(N-3)/2} 2e(n) \cos\{\omega(n - \frac{N-1}{2})\} + e(\frac{N-1}{2}) \right] \exp\{-j\omega(N-1)T/2\} & \text{for } N \text{ odd} \end{cases} \quad (3.4)$$

The variance of the response error [3.2] for a linear phase filter is

$$\sigma_L^2(\omega) = \begin{cases} \frac{Q^2}{12} \cdot \left[1 + 4 \times \sum_{n=1}^{(N-1)/2} \cos^2(\omega n) \right] & \text{for } N \text{ odd} \\ \frac{Q^2}{12} \cdot \left[4 \times \sum_{n=1}^{N/2} \cos^2\{\omega(n-1/2)\} \right] & \text{for } N \text{ even} \end{cases} \quad (3.5)$$

For an arbitrary phase filter the error variance is .

$$\sigma_A^2 = N Q^2 / 12 \quad (3.6)$$

Thus it is possible to define a signal to noise ratio due to mask quantization as below,

$$\text{SNR} \left\{ \begin{array}{ll} \leq 10 \log_{10} \left[\sum_{n=0}^{N-1} h^2(n) / \frac{Q^2}{12} \cdot (2N-1) \right] & \text{for linear phase}^* \\ = 10 \log_{10} \left[\sum_{n=0}^{N-1} h^2(n) / \frac{Q^2}{12} \cdot N \right] & \text{for arbitrary phase} \end{array} \right. \quad (3.7)$$

Eqn. (3.7) provides a measure of the response error due to mask quantization. This error normally affects the filter stop band region and places an ultimate limit on the minimum stop band attenuation achievable, although the practical limit is often set by other sources of error such as those arising from the tolerances of the device parameters. Since the achievable minimum stop band attenuation is a measure of inaccuracies of the tap weights, in the following experimental study this fact has been exploited to determine the accuracy with which a filter function is being implemented.

An appropriate tool to study the tap weight errors is the Blackman window function which provides a peak side lobe amplitude of about -58dB. This window function has all positive tap weights and thus is easily implementable using a single CCD channel. The impulse response of the Blackman window function is given by [3.1],

* Strictly speaking,

$$\sigma_L^2 \leq \begin{cases} (2N-1) \frac{Q^2}{12} & \text{for } N \text{ odd} \\ 2N \frac{Q^2}{12} & \text{for } N \text{ even} \end{cases}$$

$$h(n) = 0.42 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) + 0.08 \cos\left(\frac{4\pi n}{N-1}\right) \quad (3.8)$$

A sketch of the impulse response profile and a photograph of the actual photomask used in the experiment are shown in Fig. 3.1(a) and (b) respectively. The number of taps were restricted to only 32 in order to minimize the errors associated with the optics (see section 2.3). The original artwork was 250 times the actual size of the device. A device with a channel width of 600 μ m was used and a co-ordinatograph resolution of 1 mil provided a quantization step that is equivalent to 12 $\frac{1}{2}$ bits of accuracy in a digital system. For the 32-point linear phase Blackman window function this would provide theoretically a worst case signal to noise ratio of about 78dB. If the channel width is reduced by one-half, the above signal to noise ratio decreases by approximately 6dB. Fig. 3.2 shows a plot that compares the response due to quantized and non-quantized (ideal) tap weights. The ideal response is plotted with 16 bits of accuracy. The figure shows some departure of the quantized response from the ideal in the region where the attenuation is close to -80dB. Thus with the device dimensions indicated here the tap weight errors introduced by the mask making procedure are very small and for normal filtering applications with stop band attenuation ranging from -40 to -50dB the quantization error can be virtually ignored. However, for the Blackman window function the observed response has been compared with the quantized response in order to evaluate the tap weight inaccuracies.

(a)

(b)

Fig. 3.1 (a) Theoretical impulse response profile;
(b) The actual tap weight mask used in the experiment.

Fig. 3.2 A comparison between the ideal and quantized response.

For experimental evaluation of response error the window function was implemented on three different devices. The clock frequency was 20 KHz. Correlated double sampling was used at the output stage (see section 3.4) in order to improve the signal to noise ratio of the charge detection stage. The observed impulse response and the frequency response of one of the devices is shown in Fig. 3.3(a) and (b) respectively. The bottom figure indicates a peak first side lobe attenuation of about -56dB. The other two devices indicates peak first side lobe attenuations around -50dB. The experimental results of all three devices along with the ideal (quantized) are shown in Fig. 3.4. Deviation from the theory is mainly observed in the stop band region.

In view of these results it is of interest to relate the stop band response errors to the tap weight errors. This was done following the approach of Beartsch et al [3.3]. Assuming a tap tolerance, an independent error term was added to each of the quantized tap values. The error term is given by,

$$e(n) = \text{Tolerance} \times \text{Max}^m\{h(n)\} \times \text{Random No.} \quad (3.9)$$

where the random number is uniformly distributed between +1 and -1. Based on eqn. (3.8) an ensemble of frequency responses was calculated for different tap weight tolerances. Fig. 3.5 shows a typical ensemble of responses for 1% tap weight tolerances. For clarity only 8 responses are shown. As expected the figure shows that only the stop band characteristics are significantly altered due to coefficient errors. The response ensemble thus generated was scanned to determine minimum stop band attenuation for each response. The minimum stop band

(a)

(b)

Fig. 3.3 (a) The observed impulse response, and
(b) the frequency response of one of the devices;
scale: vertical 10dB/div, horizontal 1KHz/div.

Fig. 3.4 The ideal and observed frequency responses of three different devices.

Fig. 3.5 Eight response ensemble for 1% tap weight tolerance.

TABLE 3.1

RESULTS OF COMPUTER SIMULATION
ON MINIMUM STOP BAND ATTENUATION

TAP TOLERANCE	AV. MIN. ^m STOP BAND ATTN.
0.5%	-52.34dB
1.0%	-46.80dB
3.0%	-37.06dB
5.0%	-32.45dB

attenuation averaged over 50 runs for different tap tolerances are listed in Table 3.1. The number of runs was determined by comparing the average after every ten runs. For example, the absolute difference in the average minimum stop band attenuation between forty and fifty runs was less than 0.5dB. Comparing the experimental results of Fig. 3.4 with the Monte Carlo simulation results of Table 3.1, it can easily be appreciated that all three devices exhibit tap weight tolerances better than 1% (close to 0.5%). In spite of various potential sources of tap weight errors, the results are truly encouraging. In this context it should be pointed out that tap weight tolerances better than 1% are hard to achieve even with carefully designed split electrode CCD filters [3.3, 3.4].

3.3 MEASUREMENT OF LINEARITY

Measurement of harmonic distortions at the output of the processor is important to establish the linearity of the processor. There are two functionally different stages in the processor: at the input, light power is converted to charge, and at the output each charge packet is converted to an equivalent voltage. The overall linearity of the processor is thus dependent on both the input and the output stages. These stages have been separately analysed in sections 3.5(b) and 3.6. The experimental results are briefly discussed in this section.

The relationship between the stored charge and light intensity is very linear. The amount of photocharge collected in a CCD potential well over a given time is weakly dependent on the

depletion layer width. The non-linearity of the input stage is, thus, expected to be small provided the light intensity modulator circuit exhibits good linearity. This circuit was designed to provide a current drive for the LED in accordance with the sampled amplitude of a given input signal. The total light flux available from the LED is directly related to the forward current. For the LEDs chosen (MV5024 & HP 5082-4650), the radiated power output increases linearly with the forward current (Manufacturer's data). The linear current drive for the LED was obtained by using bipolar transistors and CMOS sampling gates. Fig. 3.6 shows the harmonic contents of the LED drive current corresponding to an input signal amplitude of 1.5 volt (r.m.s.) which was about half the maximum allowable signal. The detailed circuit arrangement along with a plot of the total harmonic distortions against the input signal amplitude is shown in Fig. 3.7. The allowable range of the input signal was limited by the dynamic range of the output stage. Typically a maximum input signal of about 1.8v r.m.s. was used (the allowable range of input signal is dependent on various factors: position of LED w.r.t. the device, on-time of the LED and the filter function). Within this range the harmonic distortions are small, as can be seen from the figure.

The major source of non-linearity is expected to arise from the output stage, where the successive signal charge packets are sensed on a floating diffusion and brought to the outside world through a source follower buffer stage. The non-linearity at the output is 1) due to the dependence of the sense diffusion junction capacitance on the sense diffusion potential, and 2) due to the non-linear

Fig. 3.6 Harmonic content of the LED drive current.

Fig. 3.7 Circuit diagram for intensity modulation and total harmonic distortion of the LED drive current.

relationship between the gate voltage and drain current of a source follower stage. The harmonic distortions at the output depend on the bias charge input to the sense diffusion. The results of measurements for two bias conditions are shown in Fig. 3.8. Curves A and B correspond to two different positions of the LED. The bias charge for curves B is about twice as much as for curves A. The second and the third harmonic distortions were measured by feeding in a low frequency signal (about 40Hz) to the input of the Blackman window function. The increase of the total harmonic distortions with the increase of bias charge is evident from the figure. The theoretically predicated second harmonic distortions (see section 3.6) assuming that the input stage is perfectly linear, agree to within 4dB over most of the signal range. This fact tends to confirm that overall non-linearity is mostly from the output stage, consequently, the linearity of the electro-optical processor can be improved by using an on-chip operational amplifier that will constrain the voltage of the sense diffusion.

Fig. 3.8. Second and third harmonic distortions of the output signal at two different bias charge conditions.

3.4 NOISE AND SPEED CONSIDERATIONS

The maximum speed of operation of the electro-optical processor is limited by the desired signal-to-noise ratio. This is because increasing the speed of operation reduces the integration time, thereby reducing the available signal amplitude at the output. Thus noise measurement is important to the maximum speed of operation. An attempt was made to measure the noise at the output of the processor. The results are discussed here. It should be mentioned, however, that the results may not be indicative of what could be achieved using optimized up-to-date devices.

The major noise sources that are encountered in a charge coupled device [3.5, 3.6, 3.7] can be separated into three categories: those associated with the input, those associated with transfer, and those associated with the output. Apart from random fluctuation of the luminous intensity of the LED, the optical injection of charge is accompanied by shot noise. The output noise due to this source is dependent on the filter transfer function. During the whole time of existence of a potential well, thermal generation of electron-hole pairs (a random process) contributes charge to the potential well and gives rise to dark current noise. The interface state trapping noise is introduced when charge packets are shifted down the CCD line. If an excess of noise charge is transferred into a particular storage location during one transfer, an identical deficiency of charge remains behind, thereby introducing correlation between adjacent charge packets. At the end of the CCD, reset noise is introduced in the charge read-out process. This noise is due to thermal noise in

the reset circuit which involves charging of a capacitor through the reset MOSFET. The MOS amplifier following the sense diffusion introduces thermal noise and $1/f$ noise. All the noise sources discussed above have also been discussed in the context of the split electrode CCD filter [3.4]. Contrary to the split electrode filter, the dark current noise and the surface state noise are not directly dependent on the filter transfer function, since these noises are introduced following the coefficient multiplications.

For measurement of noise the Blackman window filter function was used. The charge coupled device and the output circuits were mounted on a PC board with metal plating on either side. One side of the plate was used as the ground plane and wide bus lines were isolated on the other side for use as positive and negative power supplies. The signal ground and the digital ground were separated and connected together at only one point. Metal film resistors were used to reduce $1/f$ noise of the resistors. At the output signal recovery circuit, correlated double sampling [3.8] was used, to eliminate the reset noise. A schematic representation of the circuit is shown in Fig. 3.9(a). Fig. 3.9(b) shows a sketch of the sense diffusion potential. When ϕ_R^- is turned on, the sense diffusion is reset to a voltage V_R . Because of thermal noise in the channel of the reset MOSFET there is a noise due to the reset operation. The correlated double sampling technique shown in Fig. 3.9(a) provides an approach for removal of this reset noise, by effectively taking the difference in voltage of the sense diffusion before and after signal charge transfer to the sense diffusion.

(a)

(b)

Fig. 3.9 (a) Circuit arrangements for correlated double sampling;
 (b) a sketch of the sense diffusion potential at node A.

In other words, the circuit makes use of the high correlation of the reset noise between two samples when the reset MOSFET is off. When ϕ_C is turned on after ϕ_R has gone off, the node B is clamped to ground. The signal charge is transferred to the sense diffusion after ϕ_2 goes off. The change in the sense diffusion potential reflected at the output of the source follower is amplified and held on capacitor C_H after ϕ_S is turned on. The noise voltages introduced by charging of capacitors C_1 , C_2 and C_H is much smaller than the reset noise voltage at the sense diffusion because of larger capacitance values used for these capacitors. White et al [3.8] have shown that when T_{CS} (see fig. 3.9(b)) is 1/2 the CCD clock period, any time varying component (noise signal) between ϕ_C and ϕ_S clock pulse interval encounters a transfer function that introduces a zero at zero frequency and at every even harmonic of the clock frequency. Thus an additional effect of correlated double sampling is to suppress the 1/f noise (provided the 1/f corner is low) of the source follower MOSFET and harmonics of the clock feed through signal. Fig. 3.10 shows the source follower output and the output of the amplifier at node C (see Fig. 3.9(a)) for bias input only. The gain of the amplifier stage was 3.85. The LED on time was 24 μ sec.

After careful inspection of the clock voltage waveforms, solution of grounding problems and heavy filtering on power supplies, the noise spectrum was observed using the HP-3580A spectrum analyzer. Fig. 3.11(a) shows the noise power spectrum obtained by only applying a DC bias at the input. The fundamental and the second harmonic of

Fig. 3.10 Top trace is the source follower output (vertical scale: 5v/div) and the bottom trace is the output at node C (see Fig. 3.9(a)) before sample and hold (vertical scale: 2v/div).

Fig. 3.11 (a) Noise spectra of the output signal. The fundamental and the second harmonic of the clock feed through signal is also seen in the figure;
 (b) Noise spectra upto the Nyquist frequency limit.
 Scale: 10dB/div.

the clock feedthrough signal are considerably small, as can be seen from the photograph. The zero dB point corresponds to the peak of the transfer function, and the fundamental clock noise is seen to be down by 70dB without using any low pass filter. Fig. 3.11(b) shows the noise spectra in expanded scale up to the Nyquist frequency limit (10KHz). The analyzer bandwidth was 100Hz. The result indicates strong evidence of $1/f$ noise. Most of this noise was generated at the CCD output stage since the noise spectrum was seen to be almost independent of the integration time and the LED position with respect to the device. The estimated noise voltage against clock frequency is shown in Fig. 3.12. The integrated noise up to 10KHz is about 0.7mv. Using a low pass filter with cut-off at the Nyquist frequency limit and an rms voltmeter the total noise was measured to be about 0.81mv, which is reasonably close to the value obtained by integrating the curve of Fig. 3.12 up to the Nyquist frequency limit. The noise referred to the output of the CCD is thus 0.21mv. A signal to noise ratio (peak to rms) can be estimated by measuring the available signal swing at the output. However, the output signal swing from the CCD is dependent on the filter function, integration time and power of the incident light beam. The devices used were capable of providing a peak to peak signal swing of about 2 volts, and thus a signal to noise ratio of about 73.5dB. With 24 μ sec integration time (i.e., the on time of the LED with suggested luminous intensity of 3mcd, typically) the available peak to peak output swing was about 1.4 volts providing a signal to noise ratio of about 70dB.

Fig. 3.12 A plot of output noise voltage against frequency.

Reducing the integration time by one-half will reduce the signal to noise ratio by approximately 6dB. However, output linearity improves through reduction of bias charge input to the sense diffusion. Table 3.2 lists the measured signal to noise ratio and harmonic distortions for different integration times. The results are typical of experimental conditions and devices used. Although the clock frequency for all the measurements was only 20KHz, with 4 μ sec on time of the LED (the on time of LED is slightly less than the on time of ϕ_1 clock pulse), the clock frequency can be increased to about 100KHz. Since most of the noise is confined to low frequency (within 3KHz), identical signal to noise ratio can be expected at 100KHz clocking rate. The signal to noise ratio can be increased by increasing the radiated power output of the LED. The ultimate speed of a CCD register, however, is limited by transfer inefficiency.

TABLE 3.2

	INT. TIME = 24 μ sec	INT. TIME = 10 μ sec	INT. TIME = 5 μ sec
S/N RATIO	70dB	63dB	57dB
2nd HARMONIC DISTORTION	-38dB	-45dB	-50dB

NOTE: The 2nd harmonic distortions were measured at approximately one-half the peak output signal levels.

3.5 ANALYSIS OF LIMITATIONS ARISING FROM OPTICAL CHARGE INJECTION

The operation of the electro-optical processor relies on optical injection of charge into the CCD potential wells. Before reaching the silicon surface the incident light travels through a fairly complicated CCD electrode structure and consequently is attenuated. Inside the silicon photocarriers are generated both within and outside the depletion layer edge. The ones generated outside are free to diffuse in any direction and may not be collected by the correct potential well. Additional limitations may arise from any jitter in the LED drive clock from sample to sample. To investigate such limitations it is necessary to first examine the dependence of light transmittance and photogenerated carriers on the device parameters and operating conditions, and then to use the basic relationships to study the limiting effects. The discussions made in the following sections are mainly theoretical in nature. Some of the important conclusions derived from the study of a CCD imager have been used to meet the present goal.

3.5 (a) TRANSMITTANCE OF LIGHT THROUGH THE CCD ELECTRODE STRUCTURE

In all the experiments discussed in this thesis, the charge-coupled device was front side illuminated by an LED. Because of a fairly complicated electrode structure a considerable fraction of the incident photons is lost by reflection and absorption before reaching the silicon substrate. The devices used in the experiments were composed of a layer of pyrox, polysilicon and oxide. The interfaces between these layers, having non-identical indices of refractions,

give rise to multiple reflections and absorption of light within the polysilicon gate layer. The light sensitivity of the device is thus limited by reflection and absorption and is a function of the wavelength of the incident radiation. In the electro-optical signal processor, since only one colour light is used, wavelength dependent transmittance does not affect the accuracy of the system. However, variation of transmittance along the CCD channel will give rise to tap weight errors. In this section an attempt has been made to evaluate the light transmittance of the experimental devices and to determine its sensitivity to changes in the layer thicknesses. The following analytical model on light transmittance is based on reported works of Brown and Chamberlain [3.9], Anagnostopoulos and Sadasiv [3.10] and Van de Wiele [3.11].

For simplicity of mathematical treatment Fig. 3.13 shows the cross-section of light path for normal incidence [3.9]. The electric field light vector incident at the interface $(j+1)$ is denoted by E_j^+ , while the reflected component is denoted by E_j^- . The substrate is assumed to be semi-infinite and the amplitude of the reflected component is taken to be zero. As indicated in the figure, each absorbing layer (polysilicon gate and silicon substrate) is characterized by a complex index of refraction:

$$N_j = n - jK \quad (3.10)$$

The imaginary part is called the extinction coefficient which is related to the wavelength dependent absorption coefficient $\alpha(\lambda)$ through the relation

$$K = \frac{\alpha \cdot \lambda}{4\pi} \quad (3.11)$$

Fig. 3.13 The light path through various medium.

The reflected and transmitted electric field amplitudes are related to the incident electric field amplitude by Fresnel coefficients of reflection (γ) and transmission (t) respectively. At an interface ($j+1$) between materials with refractive indices N_j and N_{j+1} , the Fresnel coefficients of reflection γ_{j+1} and transmission t_{j+1} are given by,

$$\gamma_{j+1} = \frac{N_j - N_{j+1}}{N_j + N_{j+1}} \quad (3.12)$$

and

$$t_{j+1} = \frac{2N_j}{N_j + N_{j+1}} = (1 + \gamma_{j+1}) \quad (3.13)$$

For the complete structure composed of Air/Pyrox/Polysilicon/Silicon dioxide/Silicon substrate the resulting equation [3.9, 3.11] relating the incident electric field to the transmitted electric field into the silicon is given by,

$$\begin{bmatrix} E_0^+ \\ E_0^- \end{bmatrix} = (t_1 t_2 t_3 t_4)^{-1} \begin{bmatrix} e^{j\delta_1} & \gamma_1 e^{-j\delta_1} \\ \gamma_1 e^{j\delta_1} & e^{-j\delta_1} \end{bmatrix} \begin{bmatrix} e^{j\delta_2} & \gamma_2 e^{-j\delta_2} \\ \gamma_2 e^{j\delta_2} & e^{-j\delta_2} \end{bmatrix} \\ \times \begin{bmatrix} e^{j\delta_3} & \gamma_3 e^{-j\delta_3} \\ \gamma_3 e^{j\delta_3} & e^{-j\delta_3} \end{bmatrix} \begin{bmatrix} 1 & \gamma_4 \\ \gamma_4 & 1 \end{bmatrix} \begin{bmatrix} E_4^+ \\ 0 \end{bmatrix} \quad (3.14)$$

where,

$$\delta_1 = \frac{2\pi}{\lambda} N_1 d_1$$

$$\delta_2 = \frac{2\pi}{\lambda} N_2 d_2 = \frac{2\pi}{\lambda} (n_2 - jk_2) d_2$$

$$\delta_3 = \frac{2\pi}{\lambda} N_3 d_3$$

and d_1 , d_2 and d_3 are thickness of pyrox, polysilicon and silicon dioxide layer respectively.

The ratio of the transmitted power in the silicon substrate to the incident power is given by the transmittance,

$$T_4 = \left| \frac{N_4}{N_0} \right| \left| \frac{E_4^+}{E_0^+} \right|^2 \quad (3.15)$$

The above equations are particularly suited for implementation in a digital computer. Eqn. (3.15) coupled with eqn. (3.14) exhibits a number of maxima and minima within the visible spectrum [3.11] as shown in Fig. 3.14. The transmittance curve shown is particular to the values of layer thicknesses indicated in the figure. The magnitude and the shape of the curve will change for changes in the thickness of any of the layers. The light transmittance of the experimental device for three different colours of LEDs normally available (red, yellow and green) are shown in Table 3.3. The layer thicknesses and the light wavelengths are indicated in the table. The values of refractive indices of different layers were taken from [3.11].

Fig. 3.14 A plot of transmittance against wavelength for $d_1=0.617\mu\text{m}$, $d_2=0.303\mu\text{m}$ and $d_3=0.125\mu\text{m}$ [3.11].

TABLE 3.3

Thickness of Pyrox, $d_1 = 0.825\mu\text{m}$

Thickness of Polysilicon, $d_2 = 0.29\mu\text{m}$

Thickness of Oxide, $d_3 = 0.11\mu\text{m}$

Colour of Light	Wavelength	% Transmittance
Red	$0.66\mu\text{m}$	24.11
Yellow	$0.589\mu\text{m}$	53.49
Green	$0.565\mu\text{m}$	33.41

NOTE: Wavelengths were obtained from manufacturer's data
(Monsanto Catalog)

The results of table 3.3 show that for the devices used, yellow light has the highest transmittance. This is because a peak of the transmittance spectra occurs at a wavelength close to that of the yellow light. In most of the experiments however, red LEDs were used mainly because they are easily available with relatively high radiated power outputs and uniform illuminations. The devices used in the experiments were not designed specifically for imaging applications and hence exhibit rather low transmittance. It is possible to design devices with an optimum transmittance around $\lambda=0.6\mu\text{m}$ to provide an overall enhancement of the transmittance spectra [3.10]. Because of considerable absorption within the polysilicon layer attempts have been made to use transparent metal oxide electrodes to improve the transmittance spectra. Transmittances as high as 80%, nearly uniform in the range of $\lambda=0.4$ to $0.8\mu\text{m}$ have been reported [3.12].

In the context of the electro-optical signal processor it is important to discuss the sensitivity of the transmittance spectra to changes in the layer thicknesses. This is to recognize the fact that variation of transmittance along a CCD channel introduces errors in the coefficient values of a filter function. The mathematical equations governing the relationship and layer thicknesses are very complex, and thus an analytical study and consequently a general conclusion is difficult. However once the transmittance equations have been programmed into a computer it is relatively straight forward to perturb the values of the layer thicknesses from their nominal values and record the transmittance variations.

To study the sensitivity of the transmittance function eqns. (3.14) and (3.15) have been solved separately for different values of the thickness of each layer, i.e. oxide, polysilicon and pyrox layers. The nominal values of thicknesses used are the same as shown in table 3.3. A $\pm 2\%$ change in the thickness has been considered for each layer, although in a practical device the variations of thicknesses along a CCD channel are expected to be much less. Fig. 3.15 shows the variation of light transmittance for changes in the gate oxide thickness alone. For the particular device parameters and within the tolerance range considered red and green light have almost identical variation. The figure shows that small changes in the oxide thickness do not drastically change the transmittance spectra, for example, for 1% variation of oxide thickness the transmittance for red light changes by less than 0.5%. Fig. 3.16 shows the effect of change in the pyrox layer thickness on the transmittance. Changes in the pyrox layer thickness produce wider variation of transmittance than that produced by similar changes in the gate oxide thickness. The largest variation of transmittance occurs with changes in the polysilicon layer thickness as shown in Fig. 3.17. This is because of the light absorbing property of the polysilicon layer. Yellow and green light are more sensitive to this change than red light. Comparing the results of Figs. 3.15, 3.16 and 3.17 it is seen that the transmittance function is most sensitive to small changes in the polysilicon layer thickness. Thus uniformity of the thickness of this layer is important.

Fig. 3.15 Variation of transmittance as function of variation of oxide thickness.

Fig. 3.16 Variation of transmittance as a function of variation of pyrox layer thickness.

Fig. 3.17 Variation of transmittance as a function of variation of polysilicon layer thickness.

The above discussion reveals that the parameters that play an insignificant role in the electrical operation of a charge coupled device may become important for optical injection of charge. The results shown in Figs. 3.15, 3.16 and 3.17 provide a rather discouraging picture on the achievable accuracy of the processor. The important question that now arises is how much variation of the layer thicknesses is expected along a CCD channel. All the filters reported in this thesis (implemented on different devices and using red LEDs) have been realized with reasonably good accuracies. The results on a Blackman window function implemented on three different devices (see section 3.2) indicate tap weight inaccuracies better than 1%. However the number of taps i.e., the number of CCD delay bits exposed to light, was only 32, consequently only a small portion of the active area of the device was used. The light transmittance model discussed here, however, does not take into account the effect of the geometry of the device. The variation of layer thicknesses within a CCD cell can be accounted for by considering an average transmittance. The resulting transmittance error will be small if the variation of this average transmittance from cell to cell is very small.

The consideration used for optimum design of a CCD imager may help reduce the transmittance sensitivity to layer thickness variations. Using 0.1 μ m pyrox film on top of the polysilicon gate, with the other layer thicknesses as before, the transmittance sensitivity is drastically reduced. Fig. 3.18 shows the effect of changes in the

Fig. 3.18 Variation of transmittance as a function of variation of poly layer thickness.

polysilicon layer thickness on the variation of light transmittance for 0.1 μm pyrox film. For 1% change in the poly layer thickness the change in transmittance for red light is less than 4%. The effect is much less for changes in the gate oxide or pyrox layer thickness. Since the light absorbing polysilicon layer has the most disturbing effect on transmittance, better results can be expected using transparent metal oxide electrodes [3.12]. However, the limitations arising from strong dependence of light transmittance on the thicknesses of the electrode layers must be recognized.

3.5 (b) RELATIONSHIP BETWEEN STORED CHARGE AND LIGHT INTENSITY

Fundamental to the operation of the electro-optical signal processor is photogeneration of carriers inside the semiconductor.

It is therefore important to examine the basic relationship between the photogenerated charge and light intensity. The mechanisms of the photogeneration process is briefly discussed in this section, two important limitations arising from this process are discussed in the following two sections.

The light absorption mechanisms of a semiconductor follow Lambert's Law of absorption [3.13]:

$$I(x) = I_0 e^{-\alpha x} \quad (3.16)$$

where,

$I(x)$ is the radiant flux density of wavelength λ as a function of depth x ,

I_0 is the radiant flux density at the silicon surface
and,

α is the wavelength (λ) dependent absorption coefficient. The penetration depth is defined as the distance from the surface at which the light intensity has fallen to $1/e$ of the surface intensity. In other words, the penetration depth ' x_p ' is equal to $1/\alpha(\lambda)$. The longer wavelengths of light have lower absorption coefficients consequently larger penetration depth. The penetration depths of red, yellow and green light are shown in Table 3.4. The devices used in the experiments produced depletion layer widths around $1\mu\text{m}$ for 10 volts at the gate with substrate grounded. Increasing the substrate bias to -5 volts (N-channel device) increased the depletion layer width to about $1.3\mu\text{m}$. It is easily seen from table 3.4 that a part of the incident photons is absorbed within the depletion layer width and the rest inside the neutral bulk. The optical generation rate of carriers at a distance x from the surface is given by [3.14],

$$g(x) = \frac{\alpha I_0 \lambda}{hc} e^{-\alpha x} \quad \text{carriers/cm}^3 \quad (3.17)$$

where,

h is the Planck's constant,

and

c is the velocity of light.

TABLE 3.4

WAVE LENGTHS	PENETRATION DEPTH
RED (0.66 μ m)	3.33 μ m
YELLOW (0.589 μ m)	2.17 μ m
GREEN (0.565 μ m)	1.75 μ m

In order to evaluate the photogenerated charge, it is important to first examine the geometrical structure of a CCD cell. Fig. 3.19(a) shows a 2-phase CCD cell along with a sketch of the surface potential profile when the ϕ_2 gate is biased in the deep depletion mode. Because of the presence of a fringing field towards the direction of the storage potential well under the "on" gate, most of the carriers generated within the cell will end up in the storage potential well. For a simple first order one-dimensional analysis, the equivalent CCD imagery cell of Fig. 3.19(b) is assumed.

With the gate electrode biased in the deep depletion mode, the semiconductor is divided into two regions: depletion region and neutral region. In the depletion region, all generated carriers will be collected at the surface by the normal electric field. In the neutral region between $X=W$ to $X=L$, the photogenerated carriers will tend to diffuse in all directions and those reaching the depletion layer edge will be collected at the surface. To calculate the total photogenerated charge over a given integration time, the convenient approach [3.10] is to solve the steady state continuity equation separately for the depletion and the neutral region.

In the depletion region, the 1-dimensional steady state continuity equation for electrons (an N-channel device is considered throughout the course of this discussion) is,

$$\frac{dJ_d}{dx} = -q[g(x) - u(x)] \quad (3.18)$$

Fig. 3.19 (a) Cross-sectional view of a two-phase CCD cell;
 (b) simple structure considered for analytical study.

where,

J_d is the electron current density in the depletion region,

q is the electronic charge,

and

g and u are generation and recombination rate respectively.

The recombination term constitutes a dark current component and can be considered to be independent of the external generation source. Thus the photo current density J_{dp} flowing towards the surface (ie., the negative x direction) can be easily obtained by integrating eqn. (3.18) without the recombination term from the depletion layer boundary to the surface.

$$J_{dp} = \frac{q I_0 \lambda}{hc} (1 - e^{-\alpha W}) \quad (3.19)$$

Clearly, as the depletion layer width W is increased more carriers are generated inside the depletion region, as a result J_{dp} increases.

In the neutral region a part of the photogenerated carrier is lost by recombination before reaching the depletion layer edge. Invoking the condition of low level injection* and consequently neglecting the drift field, the current density in the neutral region is,

$$J_n = q D_n \frac{dn'(x)}{dx} \quad (3.20)$$

* The condition is violated at large light intensities. For high level injection condition, an effective diffusion constant $D_n' (= 2D_n)$ can be assumed in place of D_n in eqn. (3.20).

and the continuity equation is,

$$\frac{dJ_n}{dx} = -q \left[g(x) - \frac{n'(x)}{\tau_n} \right] \quad (3.21)$$

where

D_n is the diffusion constant for electrons

$n'(x)$ is the carrier concentration in excess of the equilibrium concentration n_{po} at a distance x beneath the silicon surface

τ_n is the lifetime for electrons in the neutral region and is related to the diffusion length L_n by the relation $L_n^2 = D_n \tau_n$

The coupled equations (3.20) and (3.21) can be combined to,

$$\frac{d^2 n'(x)}{dx^2} - \frac{n'(x)}{L_n^2} = - \frac{\alpha I_0 \lambda}{h C D_n} e^{-\alpha x}$$

The solution of the above equation is of the form

$$n'(x) = C_1 e^{x/L_n} + C_2 e^{-x/L_n} - \frac{\alpha I_0 \lambda L_n^2}{h C D_n (\alpha^2 L_n^2 - 1)} e^{-\alpha x} \quad (3.22)$$

normally $\alpha L_n \gg 1$.

Assuming a perfect ohmic contact at the back of the substrate at $x=L$ and a zero total carrier concentration at the depletion layer edge at $x=W$, the boundary conditions for eqn (3.22) are,

$$\left. \begin{aligned} n'(x) \Big|_{x=L} &= 0 \\ n'(x) \Big|_{x=W} &= -n_{po} \end{aligned} \right\} \quad (3.23)$$

Which leads to the final solution:

$$n'(x) = \frac{n_{p0} \sinh[(X-L)/L_n]}{\sinh[(L-W)/L_n]} - \frac{\alpha I_0 L_n^2 \lambda}{h C D_n (\alpha^2 L_n^2 - 1)} \left\{ e^{-\alpha x} - \frac{e^{-\alpha L} \sinh[(X-W)/L_n] + e^{-\alpha W} \sinh[(L-X)/L_n]}{\sinh[(L-W)/L_n]} \right\} \quad (3.24)$$

The first term constitutes a component of dark current. The second term represents photogenerated carriers. The photo current density flowing by diffusion into the depletion region can be obtained by substituting eqn. (3.24) into eqn. (3.20) and replacing X by W . The resultant photo current density is,

$$J_{np} = \frac{q I_0 \alpha L_n^2}{h C (\alpha^2 L_n^2 - 1)} \left\{ \alpha e^{-\alpha W} + \frac{e^{-\alpha L} - e^{-\alpha W} \cosh[(L-W)/L_n]}{L_n \sinh[(L-W)/L_n]} \right\} \quad (3.25)$$

The total photocurrent density J_p , flowing into the surface is,

$$J_p = J_{dp} + J_{np} \quad (3.26)$$

If the penetration depth is much larger than the depletion layer width, the contribution of J_{np} is greater than J_{dp} . When the light penetration depth $1/\alpha$ is much smaller than $(L-W)$, which is normally the case in the visible spectrum, an increase in L does not significantly affect J_{np} .

For an integration time T the surface charge density due to photogenerated carriers is,

$$P(T) = \int_0^T J dt$$

$$= \frac{qI_0\lambda}{hC} \left[(1-e^{-\alpha W}) + \frac{\alpha L_n^2}{(\alpha^2 L_n^2 - 1)} \right] \left\{ \alpha e^{-\alpha W} + \frac{e^{-\alpha L} - e^{-\alpha W} \cosh[(L-W)/L_n]}{L_n \sinh[(L-W)/L_n]} \right\} T \quad (3.27)$$

The surface photocharge density is thus dependent on the depletion layer width which is again related to the surface charge through the surface potential:

$$W = \sqrt{\frac{2\epsilon_s}{qN_A} (\phi_s + |V_{SUB}|)} \quad (3.28)$$

and the surface potential ϕ_s is given by [3.15],

$$\phi_s = \left(V_G - V_{FB} - \frac{|P|}{C_{ox}} + \frac{\epsilon_s q N_A}{C_{ox}^2} \right) - \sqrt{\frac{\epsilon_s q N_A}{C_{ox}^2} \left\{ \frac{\epsilon_s q N_A}{C_{ox}^2} + 2(V_G - V_{FB} - \frac{|P|}{C_{ox}} + |V_{SUB}|) \right\}} \quad (3.29)$$

where,

V_G , V_{FB} and V_{SUB} are the gate, flat band and substrate bias voltage respectively

ϵ_s is the dielectric constant of silicon

C_{ox} is the oxide capacitance per unit area

N_A is the substrate doping density

and

P is the surface charge density.

The nonlinearity arising in the relationship between the photocharge density P and the incident light intensity I_0 is through the dependence of the charge density on the depletion layer width. Therefore it is necessary to examine the nature of this dependence. Fig. 3.20 shows a plot of the normalized current density ($J_{hc}/qI_0\lambda$) against the depletion layer width for different diffusion lengths. Higher diffusion lengths represent lower recombination in a given volume and consequently higher current densities. The nonlinearities indicated in the curves of figure 3.20 are only significant over a wide variation of the depletion layer width. In a practical device the depletion layer width changes by a very small amount from no charge to full charge condition. The maximum charge handling capability of a two-phase charge coupled device is governed by the difference in surface potential between the surface under the storage and the transfer electrode [3.15]. In the devices used the variation of the depletion width from no charge to full charge condition is not more than $0.4\mu\text{m}$, consequently non-linearity arising from the dependence of the photocharge density on the depletion layer width is small. Fig. 3.21 shows a plot of photocharge density against light intensity for varying integration time (on time of the LED). The results were obtained by simultaneously solving equations (3.27), (3.28) and (3.29). The parameters used in the plot are typical of the devices used, with the important ones listed

Fig. 3.20 A plot of normalised photocurrent density against depletion layer width for various diffusion lengths for 0.66 μm incident light wavelength.

2

in the figure. Also shown in the figure is the locus of the maximum charge density that can be handled by the device. As discussed before the saturation charge density is dependent on the surface potentials under the thick and thin oxide regions through the flat band voltages, oxide capacitances, doping density and substrate bias voltage. The curves shown in the figure can be used to establish an operating point of the processor by taking care of the fact that summing action takes place inside the potential wells of the CCD. Because of this summing action, the depletion layer widths towards the end of the CCD line during charge injection are less than the depletion layer widths at the beginning of the CCD delay line. However, this is not expected to introduce any serious tap weight error because the dependence of surface charge on the depletion layer width is weak and also the change in the depletion layer width is small. High degree of linearity of the photogeneration process is evident from Fig. 3.21. Direct dependence of the surface charge on light intensity and integration time is illustrated in the figure. Eqn. (3.27) can be considerably simplified by assuming that the substrate thickness is large and is greater than the diffusion length. This assumption is usually good because normally $L > 200\mu\text{m}$ and L_n is in the range of 60 - 120 μm . The simplified expression is,

$$P = \frac{qI_0\lambda}{hc} [1 - e^{-\alpha W} + F e^{-\alpha W}] T \quad (3.30)$$

where

$$F = \frac{\alpha L_n}{\alpha L_n + 1}$$

Eqn. (3.30) has been referred to in later discussions.

3.5(c) EFFECT OF LATERAL DIFFUSION

The relationship between the light intensity and the stored charge density has been discussed in section 3.5(b). The model presented there was based on the assumption that all the minority carriers generated in the neutral bulk travelled towards the depletion layer boundary so that 1-dimensional continuity equation was valid. In a real situation however, the minority carriers generated in the neutral region are free to diffuse in all directions. In a uniformly illuminated semiconductor because of the absence of a gradient of carrier concentration (assuming uniform transmittance and a single wavelength of light) parallel to the silicon surface, lateral diffusive tendency is non-existent. However, any variation or modulation of intensity along the device surface will cause lateral diffusion of carriers. In the electro-optical processor, the tap weight mask in front of the device provides an effective modulation of light intensity parallel to the silicon surface, for example, a tap weight of "one" followed by a "zero" correspond to two adjacent CCD cells one of which is fully illuminated and the other completely dark. Lateral diffusion due to uneven distribution of photogenerated carriers will tend to fill areas of low minority carrier concentration, ultimately causing resolution loss. This resolution loss will be reflected in the impulse response of a filter in terms of tap weight errors. In the above example, the impulse response will indicate decrease in the output corresponding to the tap weight of "one" and increase in the output corresponding to the tap weight of "zero". The resolution degradation due to lateral diffusion is spatial-frequency dependent. Thus the amount of error

introduced is dependent on the nature of the tap weight pattern (mask). For fast changing tap patterns, i.e., for filters with passband close to the electrical Nyquist frequency limit the effect of lateral diffusion will be more pronounced because of high spatial frequency content of the mask pattern.

In order to determine the effect of lateral diffusion, it is necessary to solve the continuity equation in two-dimensions. Following the approach of Crowell [3.16] and Seib [3.17] a two-dimensional generating function is considered as a monochromatic illumination of maximum intensity I_0 that penetrates the silicon surface in the X-direction and is spatially modulated in the Y-direction (parallel to the silicon surface).

$$g(x,y) = \frac{I_0 \lambda \alpha}{2hc} (1 + \cos ky) e^{-\alpha x} \quad (3.31)$$

where,

k is the angular spatial frequency

$k = 2\pi/\lambda_s$, λ_s is the spatial wavelength

The two-dimensional continuity equation for excess electrons is,

$$\frac{d^2 n'}{dx^2} + \frac{d^2 n'}{dy^2} - \frac{n'}{L_n} = -\frac{I_0 \lambda \alpha}{2hcD_n} [1 + \cos ky] e^{-\alpha x} \quad (3.32)$$

Since the forcing function on the right hand side of the linear equation is the sum of a term constant in Y and another cosinusoidally varying in Y , a general solution is of the form [3.18],

$$n'(x,y) = n_0'(x) + n_k'(x) \cos ky \quad (3.33)$$

With the boundary conditions (similar to 1-D case)

$$\left. \begin{aligned} n'(W,y) &= -n_{p0} \\ n'(L,y) &= 0 \end{aligned} \right\} \quad (3.34)$$

the resultant solution for the excess photocarrier distribution n'_p is,

$$\begin{aligned} n'_p(x,y) = & -\frac{I_0 \lambda \alpha L_n^2}{2hCD_n(\alpha^2 L_n^2 - 1)} \left[\left\{ e^{-\alpha x} - \frac{e^{-\alpha L} \sinh[(X-W)/L_n] + e^{-\alpha W} \sinh[(L-X)/L_n]}{\sinh[(L-W)/L_n]} \right. \right. \\ & \left. \left. + \left\{ e^{-\alpha x} - \frac{e^{-\alpha L} \sinh[(X-W)/L_k] + e^{-\alpha W} \sinh[(L-X)/L_k]}{\sinh[(L-W)/L_k]} \right\} \cos ky \right] \end{aligned} \quad (3.35)$$

where

$$L_k = \sqrt{\frac{L_n^2}{k^2 L_n^2 + 1}}$$

Clearly when $k=0$ i.e., for uniform illumination L_k becomes equal to L_n and eqn. (3.35) becomes the solution of one-dimensional continuity equation. Using a similar approach and assumptions as for the one-dimensional case (see eqn. 3.30), the stored charge density at the surface for an integration time T is,

$$\begin{aligned} P(T) = & \frac{qI_0 \lambda}{2hC} \left[(1 - e^{-\alpha W} + \frac{\alpha L_n}{(\alpha L_n + 1)} e^{-\alpha W}) \right. \\ & \left. + (1 - e^{-\alpha W} + \frac{\alpha L_k}{\alpha L_k + 1} e^{-\alpha W}) \cos ky \right] T \end{aligned} \quad (3.36)$$

A convenient representation is,

$$P(T) = P_0 + P_k \cos ky \quad (3.37)$$

where,

$$P_k = \frac{qI_0\lambda}{2hc} \left(1 - e^{-\alpha W} + \frac{\alpha L_k}{\alpha L_k + 1} e^{-\alpha W} \right) T$$

$$P_0 = P_k \Big|_{k=0}$$

It is clear from eqn. (3.37) that for penetration depth $\frac{1}{\alpha} < L_k$

(i.e. $\alpha L_k > 1$) the modulation term P_k is greater than that for $\frac{1}{\alpha} > L_k$

(i.e. $\alpha L_k < 1$). In other words, the carriers generated at a distance greater than L_k have more opportunity to diffuse laterally. Further, as the angular spatial frequency k increases L_k decreases, and consequently more carriers are subjected to lateral diffusion. The ratio P_k/P_0

is the response of the imager for spatially modulated input signal.

This ratio, called the "modulation transfer function" (MTF), is plotted in Fig. 3.22 for red, yellow and green light using the device parameters indicated in section 3.5(b). The MTF roll-off for red light, having higher penetration depth, is faster than that for yellow and green. The depletion layer width was only $1\mu\text{m}$ because of rather high substrate doping of the experimental devices. The figure shows that at the spatial Nyquist frequency of 208.3 cycles/cm, which is the reciprocal of twice the intercell spacing ($24\mu\text{m}$), the MTF is about 80% for red light.

Fig. 3.22 The effect of lateral diffusion on imager response.

In the context of the electro-optical signal processor it is appreciated that filters with fast changing impulse response pattern are more susceptible to the effect of lateral diffusion. This is because fast changing filter functions have high spatial frequency contents. In terms of electrical frequency these filters have passbands extending close to the electrical Nyquist frequency limit. The roll-off indicated in Fig. 3.22 in the spatial frequency domain is equivalent to convolution in the space domain. The image of a bar object will therefore tend to broaden due to lateral diffusion as shown in Fig. 3.23 for an isolated bar object. The electrical frequency response of a filter will roll-off in a manner identical to the MTF roll-off. Using red light therefore a roll-off of about 20% is expected at the Nyquist frequency limit. This roll-off is smaller than the roll-off produced by 100% hold at the output of a sampled data system. However, in order to minimize passband error, the passband of a filter can be restricted to a fraction of the Nyquist frequency limit. If a filter passband is restricted to less than one-fifth the clock frequency, the effect of lateral diffusion (even using red light) is small for the experimental devices. If the intercell spacing is reduced in order to increase geometric resolution, the effect of lateral diffusion increases.

The results discussed here are for a rather small depletion layer width. Normally, CCD imagers are designed to produce a depletion layer width of about $3\text{-}4\mu\text{m}$ [3.17] under normal operating conditions. The effect of lateral diffusion for these devices are only significant for incident wavelengths above $0.8\mu\text{m}$.

Fig. 3.23 The effect of lateral diffusion on an isolated rectangular bar. Smearing of the input profile results due to lateral diffusion.

3.5(d) CLOCK JITTER NOISE

Since the amount of stored charge in a CCD potential well is proportional to both the light intensity and integration time, random fluctuation of light intensity or the width of the LED drive clock from sample to sample will introduce noise at the output. It is possible to estimate this noise at the output of a filter (characterized by an impulse response sequence $\{h(n)\}$) by considering a modified input sequence.

$$x'(n) = x(n) [1 + e(n)] \quad (3.38)$$

where, $e(n)$ is the uncorrelated error sequence. The non-ideal filter output is given by,

$$\begin{aligned} Y'(n) &= \sum_{k=0}^{N-1} h(k) [1+e(n-k)]x(n-k) \\ &= Y(n) + Y_e(n) \end{aligned} \quad (3.39)$$

where $Y(n)$ is the ideal output and $Y_e(n)$ is the error output. Thus the signal to noise ratio is independent of the filter function and is given by the variance of the error sequence. If only the LED drive clock pulse width varies from sample to sample which can be specified by a tolerance (given by the ratio of the deviation of the clock pulse width from its nominal value to the nominal value), the output noise to signal ratio is given by,

$$NSR = 10 \log_{10} \frac{(\text{Tolerance})^2}{3} \quad (3.40)$$

With 0.1% tolerance the signal to noise ratio is 65dB. The example emphasizes the importance of proper clock circuit design. Similar arguments can be made for random fluctuation of the light intensity.

3.6 LINEARITY OF THE OUTPUT STAGE

The linearity of the charge input stage is discussed in section 3.5(b). The next logical consideration is the linearity of the output stage. In this section a simple model is developed to determine the harmonic distortion introduced by the output stage.

The method used to detect charge at the end of a CCD line is similar to one first described by Kosonocky and Carnes [3.19]. The charge detection stage consisted of an N^+ "reset" diffusion and an N^+ "sense" or "floating" diffusion as shown in Fig. 3.24(a).

A fixed reverse bias is applied to the reset diffusion and the sense diffusion is periodically reset to a desired potential by applying a pulse voltage to the reset electrode. The sense diffusion is left floating before each time the charge sensing operation is initiated.

As shown in Fig. 3.24(a) the last electrode adjoining the sense diffusion is a transfer electrode (placed on thick oxide) which provides a path for charge to flow into the sense diffusion from the last ϕ_2 storage well. A DC voltage is placed at this transfer electrode to eliminate any clock feedthrough from this electrode to the sense diffusion. The various capacitances associated with the sense diffusion are indicated in the figure. The time arrangement of the clocks and a sketch of the sense diffusion potential is shown in Fig. 3.24(b). The MOSFET formed

C_L = Capacitance of the line connecting the sense diffusion to the MOSFET gate.

C_G = Gate capacitance of the MOSFET gate

C_{ϕ_R} = Overlap capacitance between the sense diffusion and the last phase electrode, the reset electrode.

C_{ds} = Voltage dependent junction capacitance of the sense diffusion.

Fig. 3.24(a) The output stage of the charge-coupled device. Various capacitances associated with the sense diffusion are shown in the figure.

Fig. 3.24(b) Time arrangements of the CCD clocks and a sketch of the sense diffusion potential.

by the sense and reset diffusion and the reset electrode is normally operated in the triode region so that the steady state potential attained by the sense diffusion during the on time of ϕ_R is threshold independent. The charge dumped into the sense diffusion during the off period of ϕ_2 clock appears as a voltage input to the source follower MOSFET. The non-linearity arising at the charge sensing stage is due to the dependence of the sense diffusion junction capacitance C_{ds} on the sense diffusion voltage. In addition to this, non-linearity is introduced by the source follower stage because of a non-linear relationship between the gate voltage and drain current. These two stages are analyzed separately to develop a simple analytical model to calculate the harmonic distortion of the output stage.

A simple model representing the operation of the charge sensing stage is shown in Fig. 3.25(a). The constant capacitances associated with the sense diffusion are lumped together and represented as a fixed capacitance C . Considering an abrupt junction, the junction capacitance of the sense diffusion is given by [3.20],

$$C_{ds}(V) = \frac{C_{dso}}{\left(1 + \frac{V_{SD} + |V_{SUB}| + V}{\phi_B}\right)^{1/2}} \quad (3.41)$$

where,

V_{SD} is the reset potential of the sense diffusion

$|V_{SUB}|$ is the magnitude of the substrate bias voltage

V is the voltage on the sense diffusion

Fig. 3.25(a) A model representing the charge sensing stage;
 (b) a sketch showing the nature of variation of C_{ds} with voltage.

ϕ_B is the contact potential of the junction:

$$\phi_B = \frac{KT}{q} \ln \frac{N_A}{N_i}$$

C_{dso} is the zero bias junction capacitance of the sense diffusion:

$$C_{dso} = \left(\frac{\epsilon_s q N_A}{2\phi_B} \right)^{1/2} \times \text{Area}$$

A sketch of variation of C_d with the sense diffusion voltage is shown in Fig. 3.25(b). For an electronic charge of magnitude Q_2 on C_{ds} , the relation governing the charge and voltage V_0 is given by,

$$\int_0^{-Q_2} dQ = \int_0^{V_0} C_{ds}(v) dv$$

Using the expression for C_{ds} , one obtains on integration and some algebraic manipulation,

$$V_0 = \frac{Q_2^2}{C_{dso}^2 4\phi_B} - \frac{Q_2}{C_{dsr}} \quad (3.42)$$

where,

C_{dsr} is the junction capacitance when the sense diffusion is at reset potential:

$$C_{dsr} = \frac{C_{dso}}{\left(1 + \frac{V_{SD} + |V_{SUB}|}{\phi_B} \right)^{1/2}}$$

It is clear from (3.42) that to calculate the available voltage V_0 from the charge sensing stage a relationship between Q_2 and Q , the total charge input to the charge sensing stage, is necessary. The following approximation,

$$Q_2 = Q \cdot \frac{C_{d_{av}}}{(C + C_{d_{av}})} \quad (3.43)$$

assumes that the charge sensing stage is perfectly linear, i.e., C_{d_s} is voltage independent and can be represented by an average capacitance $C_{d_{av}}$. The overall non-linear transfer characteristics can then be computed from eqn. (3.42). The average capacitance is considered as,

$$C_{d_{av}} = \frac{C_{d_{so}}}{\left(\frac{1 + V_{SD} + |V_{SUB}| + V_{av}}{\phi_B} \right)^{1/2}} \quad (3.44)$$

A reasonably good estimate for V_{av} is,

$$V_{av} = \frac{1}{2} \left[\frac{-Q}{(C + C_{d_{sr}})} \right] \quad (3.45)$$

The approximation (3.43) introduces only small error at higher voltage swing. A plot of V_0 against Q is shown in Fig. 32.6 for a hypothetical device for different values of the fixed capacitance C . The actual numerical solution and the solution of eqn. (3.42) based on approximation (3.43) are compared in the figure. It is seen that for values of C higher than $C_{d_{so}}$, the linearity of the charge sensing stage improves.

Fig. 3.26 Transfer characteristics of the charge sensing stage.

as well as the agreement between the approximate and the actual solution.

Normally, C is higher than $C_{d_{so}}$ because of high intended aspect ratio of the source follower MOSFET. The maximum detectable charge is limited by the condition $(V_{SD} + V_0) \geq V_T$, the threshold voltage of the source follower MOSFET.

If we now consider that the input stage of the processor is linear (input non-linearity is small as discussed in section 3.5(b)) so that the charge Q_2 delivered to the sense diffusion junction capacitance comprises of a bias charge $\bar{Q}_2$ and a small signal charge $\hat{Q}_2 \cos \omega t$:

$$Q_2 = \bar{Q}_2 + \hat{Q}_2 \cos \omega t \quad (3.46)$$

substituting for Q_2 into eqn. (3.42) results

$$V_0 = V_b + v_1 \cos \omega t + v_2 \cos 2\omega t \quad (3.47)$$

where,

$$V_b = \frac{\bar{Q}_2^2}{C_{d_{so}}^2 \cdot 4\phi_B} - \frac{\bar{Q}_2}{C_{d_{sr}}} + v_2$$

$$v_1 = \frac{\hat{Q}_2}{C_f}$$

$$C_f = \left(\frac{\bar{Q}_2^2}{C_{d_{so}}^2 \cdot 2\phi_B} - \frac{1}{C_{d_{sr}}} \right)$$

and

$$v_2 = \frac{q_2^2}{C_{dso}^2 \cdot 8\phi_B} = v_1^2 \cdot \frac{C_f^2}{C_{dso}^2 \cdot 8\phi_B}$$

The small signal model of the charge sensing stage thus consists of a bias charge dependent capacitor C_f (at any given bias C_f is a fixed capacitor) in parallel with a second harmonic generator. Knowledge of the bias charge is thus important in order to determine the non-linearity introduced by the charge sensing stage. Further the bias charge input to the charge sensing stage is dependent on the filter pattern due to charge summation taking place inside the CCD potential wells. An estimation of bias charge input to the sensing stage can be obtained from Fig. 3.21 in section 3.5(b) provided the incident light intensity is known. If Q_{in} is the total charge introduced in a uniformly illuminated CCD cell, the charge input to the sensing stage is $\sum_n Q_{in}$, which therefore is dependent on the filter pattern. For predicting the experimental results this bias charge has been estimated from the output source follower voltage swing. Fig. 3.27 shows a plot of variation of bias voltage V_b and capacitor C_f with bias charge Q . The relevant device parameters are indicated in the plot. V_b determines the bias voltage applied to the source follower stage. The variation of C_f with the bias charge is important as it governs the non-linearity of the charge sensing stage. The ratio

$$\frac{v_2}{v_1} = \frac{C_f^2}{C_{dso}^2 \cdot 8\phi_B} v_1 \quad (3.48)$$

Fig. 3.27 Variation of V_b and C_f with input bias charge Q (the parameters are representative of the experimental device)

Fig. 3.28 A plot of the ratio of the 2nd harmonic to fundamental against the fundamental for various values of bias charge input to the sense diffusion. The device parameters are shown in Fig. 3.27.

increases with higher input bias charge because of increasing C_f .

Fig. 3.28 shows a plot of eqn. (3.48) for increasing bias charge.

The slope of the curves can be used to determine the distortion at any given amplitude of the fundamental component. For example, at $\bar{Q} = 3\text{pc}$, $v_1 = 0.3\text{v}$, the 2nd harmonic distortion is about -45dB; at $v_1 = 0.6\text{v}$, the 2nd harmonic distortion is -39dB. The results shown in Fig. 3.28 indicate that harmonic distortion introduced by the charge sensing stage is not negligible and should not be ignored in computing the overall harmonic distortion.

Considering now the source follower stage as shown in Fig. 3.29(a) the voltage V_G appearing at the gate of the source follower MOSFET is

$$V_G = V_{SD} + V_0 \quad (3.49)$$

Substituting V_0 from eqn. (3.47) the gate voltage can be expressed as

$$V_G = V_{gb} + v_1 \cos \omega t + v_2 \cos 2\omega t \quad (3.50)$$

where V_{gb} is the bias voltage given by

$$V_{gb} = V_{SD} + V_b$$

Since the source follower MOSFET is operated in the saturation region [3.21]

the output voltage V_{OUT} for an input V_G is given by,

$$V_{OUT} = \beta R [V_G - V_{OUT} - V_T]^2 \quad (3.51)$$

where,

$$\beta = \frac{U_n^* \text{cox}}{2} \frac{W}{L}$$

U_n^* is the effective surface mobility of electrons

C_{ox} is the oxide capacitance per unit area

W/L is the aspect ratio of the MOSFET

V_T is the threshold voltage, an average fixed value

is considered for the threshold voltage for simplicity

R is the load resistor

Solution of eqn. (3.51) for an assumed constant V_T (for a given bias and small signal condition the assumption is justified) is,

$$V_{OUT} = V_G - V_T + \frac{1}{2BR} - \frac{1}{2BR} [1 + 4BR(V_G - V_T)]^{1/2} \quad (3.52)$$

To calculate the total harmonic distortion at the output of the source follower, V_G is replaced by eqn. (3.50),

$$\begin{aligned} V_{OUT} = & V_{gb} - V_T + \frac{1}{2BR} + v_1 \cos \omega t + v_2 \cos 2\omega t \\ & - \frac{1}{2BR} [1 + 4BR(V_{gb} - V_T)]^{1/2} \\ & \times \left\{ 1 + \frac{4BR}{1 + 4BR(V_{gb} - V_T)} (v_1 \cos \omega t + v_2 \cos 2\omega t) \right\}^{1/2} \end{aligned} \quad (3.53)$$

Under the normal operating condition the term $\frac{4BR}{1 + 4BR(V_{gb} - V_T)} (v_1 + v_2) < 1$.

Consequently, the binomial expansion of the terms inside the second bracket is possible. Ignoring terms with powers higher than two, and rearranging, leads to the various harmonic components as below:

Fundamental component;

$$D_1 = v_1 - \frac{1}{2} \left(\frac{k}{\beta R} \right)^{1/2} v_1$$

Second harmonic component;

$$D_2 = v_2 - \frac{1}{2} \left(\frac{k}{\beta R} \right)^{1/2} v_2 + \frac{1}{16} \left(\frac{k^3}{\beta R} \right)^{1/2} v_1^2$$

3rd harmonic component;

$$D_3 = \frac{1}{8} \left(\frac{k^3}{\beta R} \right)^{1/2} v_1 v_2$$

where,

$$k = \frac{4\beta R}{1 + 4\beta R (V_{gb} - V_T)}$$

The output bias component is

$$D_0 = V_{gb} - V_T + \frac{1}{2\beta R} - \left(\frac{1}{k\beta R} \right)^{1/2}$$

A plot of the total harmonic distortion as a function of the amplitude of the input fundamental component is shown in Fig. 3.29(b). The relevant device parameters are indicated in the plot. At lower bias charge, i.e., at higher gate bias voltage, the harmonic distortion is less than that at higher bias charge. At $\bar{Q} = 3\text{pC}$ and $v_1 = .3\text{v}$, the total distortion is about -41dB. Comparing the results of Fig. 3.28 an increase in distortion of about 4db is seen under the same bias charge and voltage condition. The effect of substrate bias and the output bias voltage was considered in computing the average threshold voltage. A lower value of the load resistor increases the distortion. The curves shown in Fig. 3.29(b) are for the $10\text{k}\Omega$ resistor which was used in the experiment.

Fig. 3.29(a) The output source follower stage;
 (b) total harmonic distortion at the output for different bias charge input to the sense diffusion.

Fig. 3.30. The source follower output voltage waveform.
Scale: vertical 0.5v/div.

Experimental measurements of harmonic distortions were performed on the 32-tap Blackman window function. The voltage swing available at the output of the source follower is dependent on the bias charge $\bar{Q}$ which in turn is dependent on the light intensity, integration time and also the filter function. This bias charge is also the key parameter for estimation of output non-linearity. Fig. 3.30 shows the output of the source follower for an integration time of about 24 μ sec and for a DC input only. Noting the output voltages corresponding to $V_G = V_{SD}$ and $V_G = V_{SD} + V_b$ an estimation of V_b can be obtained from the following equation and assuming a fixed average threshold voltage,

$$V_G - V_T \approx V_{OUT} + \sqrt{V_{OUT}/\beta R} \quad (3.54)$$

for $\beta R > 1$

Once V_b is known, $C_{d_{av}}$, $\bar{Q}$ and C_f can be determined based on the relevant equations discussed earlier. The measured harmonic distortions for 10 μ sec LED on time are shown in Fig. 3.31 (a) and (b) for two different positions (1cm and 1/2cm away from the device, respectively) of the light source. The predicted 2nd harmonic distortions in both cases agree to within 4dB of the experimental results. This tends to support our previous conclusion that the non-linearity of the photogeneration process is small. The predicted third harmonic distortions are well below the measured results and are excluded from the figures. However, the contribution of the third harmonic component to the total harmonic distortion, in each case, is small. Hence only the prediction of the dominant 2nd harmonic component is important. Fig. 3.31(b) exhibits

Fig. 3.31 (a) Measured and predicted harmonic distortions.

Fig. 3.31 (b) Measured and predicted harmonic distortions at a bias charge higher than that of figure (a).

higher harmonic distortions because of higher bias charge input to the charge sensing stage.

The simple model developed in this section is useful from the point of view of device design and selecting the optimum operating range.

CHAPTER 4

REALIZATION OF GENERAL FILTER FUNCTIONS

4.1 INTRODUCTION

Discussed in the previous chapters were the basic principle of operation, and the limiting effects of an LED/CCD electrical signal processor. The experimental results presented in this context established the performance of the processor in terms of accuracy, linearity and noise. This section shall discuss the various practical alternatives of realizing a general filter function and the relative merits of each approach. A general filter function has tap weights of both polarities (+ve and -ve). The discussion, so far, was restricted to window functions having only positive taps. The initial discussions in this chapter are devoted to methods of achieving tap weights of both polarities. At the end of this chapter, the various application potentials of electro-optical signal processing are explored.

4.2 NEGATIVE TAPS

A charge coupled device handles only one polarity of charge. Consequently, circuit solutions are necessary to provide tap weights of both polarities. In order to achieve this, separate storage sites have to be assigned either in the same or in a separate CCD register, as shown in Fig. 4.1. In the former method, the outputs at two adjacent time slots are combined and hence this method is termed Temporal multiplexing. In the latter method, the outputs from two separate CCD registers are combined at the same time and is termed

(a) TIME MULTIPLEXING

(b) SPACE MULTIPLEXING

Fig. 4.1 Two approaches for negative tap realization.

Spatial multiplexing. Both these approaches are discussed in detail in the following sections.

(a) Temporal multiplexing

In this scheme, two adjacent storage sites are used to represent each tap weight. In other words, every alternate storage site represents a negative charge storage site. Thus, by using only one CCD channel and alternately switching the output of the CCD between the positive and the negative input terminals of a differential amplifier, tap weights of both polarities are obtained. As shown in Fig. 4.1(a), a sample and hold circuit is necessary because of 1-bit delay in the availability of the negative (or positive, depending upon the configuration used) tap weighted signal. Since two adjacent storage sites separated by a delay of one clock period represent one tap weight, the LED modulated by the signal should be turned on only once every two CCD clock periods.

In order to appreciate this method of any general filter realization, the operation of the processor is diagrammatically illustrated in Fig. 4.2. For ease of understanding a device with only four storage sites is considered. a_0 , a_1 , a_2 and a_3 correspond to the openings (transparent areas) in the mask pattern and the x 's correspond to the signal strength. The charge packets gathered at the respective storage sites during a flash of the LED light are represented by the products of corresponding a 's and x 's. The direction of transfer of charge from one storage site to another is indicated by an arrow head. The output at any clocking instant is the sum of the

The output at $(2n+2)$ CCD clock pulse.

$$\begin{aligned}
 Y_n &= (a_0 - a_1)x_n + (a_2 - a_3)x_{n-1} \\
 &= h_0x_n + h_1x_{n-1}
 \end{aligned}$$

Fig. 4.2 A diagram showing the operation of the processor in time multiplexed mode.

diagonal elements along the arrows and is indicated in the figure. The LED is flashed during every alternate clock cycle, and the output from the differential amplifier is also sampled in synchronism with the light flash.

It is clear from the figure that to realize an N-tap filter, a CCD line with 2N delay bits is required. The output at the (n+1)th input-output sampling clock pulse is:

$$\begin{aligned} Y_{n+1} &= (a_0 - a_1)x_n + (a_2 - a_3)x_{n-1} + \dots (a_{2N-2} - a_{2N-1})x_{n-N+1} \\ &= h_0x_n + h_1x_{n-1} + \dots h_{N-1}x_{n-N+1} \end{aligned} \quad (4.1)$$

Thus, by proper choice of a_0 , a_1 , etc., tap weights of both polarities can be achieved. There are various possible ways a_0, a_1, a_2 , etc. can be chosen. In order to achieve a maximum dynamic range (i.e., tap weighted signal), one of the two a 's making up a tap weight can be assigned a value of zero. In other words, two adjacent tap weights of +1 and -1 can be obtained by assigning a value of 1 to a_0 and a_3 and a value of zero to a_1 and a_2 . This design scheme, however, leads to a fast changing mask pattern. As discussed before, such fast changing mask patterns are more susceptible to blurring and lateral diffusion effects. Blurring occurs because of finite distance of separation between the CCD and the glass mask and any lack of collimation of the incident light beam. Since two successive outputs are subtracted small blurring can introduce serious tap weight errors. This error however, can be reduced by using on chip metal mask as discussed in appendix A. Since an act of differentiation takes place at the output, a reasonably smooth mask pattern can be obtained for commonly used filter functions, by using an impulse response

integration approach. The basic design rule consists of making a_2 equal to the sum of a_0 and a_1 , a_4 equal to the sum of a_2 and a_3 and so on. An example of this approach is discussed in the context of experimental demonstration. In general, this approach may lead to out of bound growth of the values of a 's unless the original tap weight vectors are properly scaled and a suitable bias applied to the tap weights. Thus, in general, a smoothly varying mask pattern is obtained at the cost of available dynamic range.

Fig. 4.3(a) shows the necessary details in the circuit arrangement at the output of the CCD to achieve time multiplexing. The clocking sequence is shown in Fig. 4.3(b). As shown in the diagram, the output from the CCD is routed through two sample and hold circuits to the two input ends of a differential amplifier. Though, in principle, only one sample and hold circuit is needed to hold the negative (or positive) tap weighted signal for one delay bit, two of them were used for balancing the two channels. The final output from the differential amplifier is made available only during the on-time of ϕ_{S_2} . Since the input-output sampling clock (ϕ_{S_2}) is operated at one-half the speed of the CCD drive clocks, there is a consequent reduction in the bandwidth capability of the system by one-half.

As a demonstration of this approach, a 21-tap, 2-cycle sine wave filter was realized using the photoemulsion mask. The impulse response of the filter was governed by:

$$h_n = -\sin\left(\frac{4n\pi}{N-1}\right); 0 \leq n \leq N-1 \quad (4.2)$$

(a)

(b)

Fig. 4.3 : (a) Circuit arrangements (at the output of the CCD) to achieve time multiplexing; (b) clocking sequence.

Fig. 4.4 (a) The mask pattern for a 21-tap filter designed using the time multiplexed approach.
 (b) A sketch of the desired impulse response.

A sketch of the mask pattern is shown in Fig. 4.4(a). The mask covers 42 CCD delay elements. The desired impulse response is shown in Fig. 4.4(b). The width of each CCD delay element is shown in Fig. 4.4(a) by dashed lines. In designing this particular filter, an impulse response integration approach was used to achieve a smooth mask pattern. The tap weights given by eqn. (4.2) were scaled and encoded into the optical mask according to the following recursive relation

$$a_0 = k \quad ; \quad k \text{ is a suitably chosen number, } k=0 \text{ in this particular case.}$$

$$a_1 = k + h'_0$$

$$a_2 = a_1$$

$$a_3 = a_2 + h'_1$$

$$\vdots$$

$$a_{2m} = a_{2m-1} \quad ; \quad m \text{ is any integer.}$$

$$a_{2m+1} = a_{2m} + h'_m$$

The sequence $\{h'_n\}$ denotes the scaled tap weight vectors.

The observed impulse response is shown in Fig. 4.5(a) after sample and hold and low pass filtering at the output of the differential amplifier. The CCD clock frequency was 20 KHz and the cut off frequency of the low pass filter was 5 KHz. The measured and the theoretical frequency responses are shown in Fig. 4.5(b). Good agreement is evident from the figure. The agreement of the observed side lobes with the theory is clearly encouraging, and justifies design of filters with higher stop band attenuations. The ideal theoretical results were based on transfer function derived from the block diagram representation of the time multiplexed scheme shown in Fig. 4.6.

(a)

(b)

Fig. 4.5 (a) The observed impulse response of the filter shown in Fig. 3.7.
 (b) The frequency response.

Fig. 4.6. Block diagram representation of the time multiplexed scheme.

One of the major limitations of the time multiplexed scheme is the requirement of $2N$ -CCD delay elements for an N -tap filter. The effect of transfer inefficiency will be more pronounced compared to a filter realized with N -delay elements. The input-output sampling rate is one-half the CCD clock rate consequently, there is a reduction in the bandwidth capability of the system by one-half.

The concept of alternate CCD potential well carrying negative information can have interesting applications in image processing, e.g., edge detection of an image [4.1]. Since two consecutive output samples are subtracted from each other, any difference in the brightness of an image will produce an output signal. To illustrate this situation, the output of the system for a trapezoidal image pattern is shown in the bottom trace of Fig. 4.7. As seen from the figure, the output is only present when there is a change in the profile of image pattern.

(b) Spatial Multiplexing

In this scheme, the signal is divided in space between two CCD line imagers as shown in Fig. 4.1(b). The modulated light source uniformly illuminates two parallel CCD channels with separate mask patterns, so designed that subtractions between the outputs from the two CCD channels give rise to the desired filter response. Thus, for an N -tap filter, two parallel CCD registers, each with N delay elements, are required. This method has the advantage over the previous one, in that the input-output sampling clocks are operated at the same speed as the CCD drive clocks; consequently, there is no reduction in the

Fig. 4.7 An illustration of the use of the principle of the time multiplexed approach for edge detection of an image. The bottom trace shows the output corresponding to a trapezoidal image pattern shown in the top trace.

bandwidth capability of the system. The obvious disadvantage is the need for two CCD line imagers.

There are various possible ways to design two mask patterns that lead to the desired impulse response pattern on subtraction. To attain a maximum dynamic range, one of the two CCD channels can be assigned to carry only the negative tap weighted signal. An example of this approach is shown in Fig. 4.8, for a band pass filter. This approach may be useful for on-chip metal masks. An alternate approach is to apply a bias to the normalised coefficients, such that, all the coefficients are of same polarity. Thus, the impulse response of one of the CCD channels is the desired impulse response superimposed on a bias level. By having a separate bias channel, the applied bias is cancelled out on subtraction. Although there is a reduction of dynamic range, there are some practical advantages of this approach:

- (a) the design procedure is straightforward
- (b) a smoothly varying mask pattern is obtained for commonly used filters (low pass, band pass, etc, with passband well below the Nyquist frequency limit)
- (c) a number of filter functions can be implemented in an area-array imager using one common bias channel.

This is the most significant advantage of the "bias" approach.

The results reported here are based on the above design principle. A 91-tap low pass filter was made according to the equation:

Fig. 4.8 Impulse response of a band pass filter with centre frequency at $1/2$ the Nyquist frequency. One of the CCD channels can be used for positive taps and the other for negative taps. The impulse response can be multiplied by the Hamming window function to obtain higher stop band attenuation. The Q of a filter depends on the number of taps used.

$$h_n = \frac{\sin[0.2\pi(n-45)]}{0.2\pi(n-45)} ; 0 \leq n \leq 90 \quad (4.3)$$

The sequence $\{h_n\}$ obtained from the above equation was scaled down by a factor of two and a bias of 0.5 was added to all the coefficients prior to mask making. Fig. 4.9(a) shows the actual photomask employed in the experiment. The rectangular bias channel can be seen at the top of the photograph. The impulse response of the two CCD channels prior to subtraction are shown in Fig. 4.9(b). Small blurring of the mask image can be seen at the ends of the impulse response. The final impulse response is shown in Fig. 4.9(c). The observed impulse response clearly shows the $\sin x/x$ behaviour as expected. The measured frequency response is shown in Fig. 4.10 along with the ideal. The ideal response takes into account the effect of transfer inefficiency (0.00075 per transfer) and the effect of finite width pulse sampling at the input and output as discussed in section 2.5. The general agreement between theory and experiment is good. A small passband error (within 4dB) is observed at the passband edge. The measured response appears to have some undesired roll-off. The filter was implemented using two separate CCD chips and light sources, and therefore may have been subjected to response errors (apart from blurring) arising from mismatch between the CCD channels of gain and transfer inefficiency.

To account for the effect of non-uniform channel gains and transfer inefficiencies, the frequency response equation can be written as below (see eqn. 2.5):

(a)

(b)

(c)

Fig. 4.9

- (a) A photograph showing the designed mask pattern for a 91-tap low pass filter.
 (b) The outputs from the two CCD channels prior to subtraction.
 (c) The final impulse response.

Fig. 4.10 The observed and theoretical frequency response of a 91-tap low pass filter.

$$\begin{aligned}
 H(e^{j\omega t}) = & \alpha_1 \sum_{n=0}^{N-1} h'_n \exp[-n\epsilon_1(1-\cos\omega t)] \exp[-jn(\epsilon_1 \sin\omega t + \omega t)] \\
 & - \alpha_2 \sum_{n=0}^{N-1} h''_n \exp[-n\epsilon_2(1-\cos\omega t)] \exp[-jn(\epsilon_2 \sin\omega t + \omega t)] \quad (4.4)
 \end{aligned}$$

where $\{h'_n\}$ is the tap weight sequence which, when subtracted from the sequence $\{h''_n\}$ yields the desired set of tap weights. In the above example h''_n is the bias applied to the actual tap weights to obtain h'_n . ϵ_1 is the transfer inefficiency of the CCD channel corresponding to the tap weight sequence $\{h'_n\}$. ϵ_2 is the transfer inefficiency of the other channel. The above equation is derived on the basis that the responses of the two channels are subtracted to obtain the desired filter response. The effect of non-uniform channel gains have been incorporated into eqn. (4.4) by post-multiplying each of the channel transfer function by a different gain parameter (α_1 and α_2). Under normal operating frequencies the assumption of frequency independence of the gain parameter is justified [2.7]. Since the gain parameter appears as a direct multiplication factor in eqn. (4.4), the effect of small variations of gain between the channels is more serious than similar variations of transfer inefficiencies. The effect of gain variation on the frequency response of the low pass filter is shown in Fig. 4.11 for the case when α_2 is 98% of α_1 . Also shown in the figure is the ideal response for uniform gain and zero transfer inefficiency. The passband ripple is increased because of inexact subtraction between the responses of the two channels. There will be reversal of the shape of the response curve if the values of α_1 and α_2

Fig. 4.11 The effect of gain variation on the frequency response of the low pass filter.

are reversed. The transfer inefficiency affects both the magnitude and phase response of each channel as is evident from eqn. (4.4).

The subtraction between the two channel responses can be represented vectorially by two phasors, which in absence of transfer inefficiency are exactly 180 degrees apart. Both amplitudes and angular displacements of the phasors are affected by transfer inefficiency. Non uniform transfer inefficiency, thus, have a tendency to increase the peak amplitudes of the side lobes. The situation is illustrated in Fig. 4.12 for 25 and 50% variation of transfer inefficiency between the two channels.

The above discussion is mainly to recognize the effect of non-uniform channel characteristics on the frequency response of a filter. Any practical implementation of the space multiplexed scheme should incorporate two CCD channels on the same chip. The variation of the parameters can thus be kept small enough to produce any appreciable effect on the frequency response.

Using the bias approach a band pass filter was implemented by Hamming windowing a Cosine wave. This is to obtain a large stop band attenuation. The impulse response of the 91-tap filter was given by

$$h_n = \left[0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \right] \cos\left(\frac{4\pi n}{N-1}\right); \quad 0 \leq n \leq 90 \quad (4.5)$$

The observed impulse response is shown in Fig. 4.13. The measured frequency response along with the ideal is shown in Fig. 4.14. The figure indicates very good agreement in the pass band region, but some deviation from the theory in the stop band region. Considering the fact that two separate CCD chips were used, the observed response appears promising. The design method used to implement this particular

Fig. 4.12 The effect of transfer inefficiency variation on the frequency response of the low pass filter.

Fig. 4.13 The impulse response of a Hamming weighted bandpass filter.

NORMALISED SIGNAL FREQ.

Fig. 4.14 The frequency response of the bandpass filter.

filter can be used to achieve a high Q bandpass filter by using a longer CCD delay line. The centre frequency can be tuned by varying the CCD clock frequency.

The total harmonic distortion at input (LED drive current) and the 2nd harmonic distortion at the output were measured and the results are shown in Fig. 4.15. The measurement procedure involved applying a signal of frequency lying at the centre frequency of the bandpass filter and then changing the input signal frequency to half the centre frequency so that the second harmonic component lies in the pass band of the filter. The results are consistent with the discussions made in chapter 3. No rigorous noise measurements were done, some initial measurements indicate a signal to noise ratio better than 60dB for an integration time of about 20 μ sec.

4.3 DISCUSSION

In the previous sections we have discussed two approaches for realizing any general filter functions. Comparison between these two approaches indicates superiority of the space multiplexed scheme over the time multiplexed scheme from the point of view of ease of design of the coefficient mask, bandwidth capability and line length requirement. The experiments discussed in the context of the space multiplexed scheme were performed by using two separate CCD chips and consequently were subjected to errors arising from non-uniform channel characteristics. Using two CCD channels on the same CCD chip will greatly reduce these errors. The commercially available

Fig. 4.15 Harmonic distortions at the input and the output of the Hamming weighted bandpass filter.

linear imager (Fairchild CCD 110) is particularly suitable for realizing a general filter function using the time multiplexing approach. The imager has 256 interleaved photoelements [4.2] as shown in Fig. 4.16. The organization of the imager permits storing of the positive and the negative lagged products into two separate CCD registers which are shielded from light. Transfer of charge from the integrating sites into the CCD registers takes place when the transfer gate is turned off. The two CCD registers are time multiplexed into a common output. Thus the restriction on the bandwidth of the time multiplexed scheme discussed in section 4.2(a) is removed. The necessary clocking arrangement for operating the device in the signal processing mode is also shown in Fig. 4.16. A maximum of 128 taps can be implemented using the device.

The discussions in the previous sections have demonstrated the use of an LED/CCD structure for linear filtering applications. The use of an optical system and a charge-coupled device can offer significant advantages in terms of high-speed analog multiply (multiplications take place at the velocity of light) and parallel processing ability of a charge-coupled device. Consequently, the major application potential of the electro-optical signal processor will be in the areas requiring large quantities of data to be processed in near real time. One such application is obtaining the Discrete Fourier Transform of an electrical signal. An N-point Discrete Fourier Transform is defined by

$$Y_K = \sum_{n=0}^{N-1} X_n [\cos(2\pi nk/N) - j \sin(2\pi nk/N)]; k=0, 1, \dots, N-1 \quad (4.6)$$

Fig. 4.16 (a) Organisation of the Fairchild's CCD 110 linear imager;
 (b) the clocking sequence to operate the device as a time multiplexed convolver.

Fig. 4.17 Circuit arrangement for obtaining 100 point DFT. The output switch closes once every 100 CCD horizontal clocks for bias cancellation.

Where Y_k represents the output frequency component corresponding to the input samples $\{X_n\}$. By using an area-array imager, the coefficients can be encoded onto an optical transparency such that each column of the imager corresponds to a set of cosine or sine coefficients for a given value of k . As a specific example, Fig. 4.17 shows the circuit arrangement for obtaining a 100 point DFT (for real input signal only $N/2$ components are unique) using a 100×100 CCD imager. Since the sine coefficients corresponding to $k=0$ form the bias channel, application of a fixed bias to the coefficients does not require any additional bias channel. The output corresponding to this channel is held for 100 horizontal register clock periods and is subtracted from the output of every other channel. The resultant cosine and the sine components of the 100-point DFT is obtained at the output of the differential amplifier. A 100 point transform requires 100 vertical register clock pulses, at a clock rate of 20KHz. This time corresponds to 5msec.

The arrangement shown in Fig. 4.17 can in principle be employed for simultaneously correlating an unknown input signal with a library of reference signals encoded onto an optical mask. Chirp-Z-transform can also be implemented electro-optically, similar to the split-electrode technique [4.3]. The arrangement however, necessitates separate CCD chips and light sources for realizing the real and imaginary transform. Premultiplication of the input signal with the chirp function can be obtained by duty cycle modulation (in addition to intensity modulation) of the LED, thereby eliminating the need for an analog multiplier.

PROCESSING OF SPACE-VARYING SIGNALS

8

CHAPTER 5

PROCESSING OF SPACE VARYING SIGNALS

5.1 INTRODUCTION

In the previous chapter we have discussed an LED/CCD structure for processing of time varying electrical signals. The input function, which varied in time, was correlated with a known function encoded onto an optical mask. The same general principle can be effectively utilized for processing of signals that vary in a spatial manner. The coefficients of a desired transformation can now be applied as a time varying signal to the LED. The space varying function being processed is available in the form of a focussed image. That is, the role of the "known" and the "unknown" function has been reversed from the previous discussion. Through the reversal of roles between the "known" and the "unknown" function, the arrangement becomes electrically programmable. The coefficients can be generated using an algorithm and applied as electrical signal input to the system.

The concept discussed above is directly applicable to processing of linear images, where the entire processing is done inside a CCD line imager. The initial discussions in this chapter refer to linear images. The experiments discussed in that context demonstrate correlation of an image pattern with a set of predetermined coefficients. A linear transformation (e.g., Fourier or Hadamard) can be obtained from the results of such correlation, as has been demonstrated in the following experiments. The applications of processing one-dimensional

variables involve monitoring of fluid level or diameter of a wire, etc. for industrial control.

The concept of linear image processing presented here can be extended to two-dimensional processing by a combination of analog and digital techniques for applications like bandwidth reduction, coding and pattern recognition. A discussion to this effect is incorporated at the end of this chapter.

5.2 EXPERIMENTAL ARRANGEMENTS

A block diagram of the experimental set-up for processing linear images is shown in Fig. 5.1. The coefficient memory shown in the figure holds the coefficients of a desired transformation in digital form. In an actual application a microprocessor unit can be used to generate the coefficients. These coefficients are sequentially clocked out to modulate either the light intensity or the integration time as indicated by the dashed lines in the figure. The coefficients of a desired transformation in general, are both positive and negative numbers.

In the experiments discussed here all the coefficients were stored as positive numbers by suitably scaling and applying a bias to the coefficients. Consequently, it is necessary to examine the effect of this bias at the output of the CCD. In the context of the CCD image processor, a convenient representation of the correlation operation is given by

$$Y(n+m) = \sum_{n=1}^{\infty} a_n x_{n+m} \quad (5.1)$$

Fig. 5.1 A block diagram of the image processor.

Where the sequences $\{a_n\}$ and $\{x_n\}$ respectively represent (i) the image samples* and (ii) the coefficients superimposed on the bias level, $m=0$ corresponds to an output $Y(N)$ at the N th CCD clock pulse period, i.e., when a set of coefficients moving relative to the image completely overlaps the image pattern. Hence for $m \geq 0$ i.e., after $(N-1)$ CCD clock pulses, any fixed bias applied to the coefficients will be multiplied by all the image samples and will be reflected at the output as a fixed offset which can be subtracted out in a differential amplifier.

The two types of modulation discussed at the beginning of this section (LED intensity or CCD period) can be meaningful in different applications. In situations where the LED can shine through a photographic negative, light intensity modulation may be convenient. On the other hand, integration time modulation would allow direct processing of an image of a scene which the CCD is looking at. For LED intensity modulation the clocking arrangements can be exactly similar to that of electrical signal processing as shown in Fig. 5.2(a). Since the LED drive clock is shut off before charge transfer, carriers can be integrated in the transfer channel itself. In other words, a simple CCD delay line with imaging capability (transparent gates) can be used.

In the case of integration time modulation, if integration is done in the transfer channel, the CCD clocks have to be modulated as shown in Fig. 5.2(b). The on time of the integrating clock, i.e. ϕ_1

* The sampling action of the image is provided by the CCD with a pitch equivalent to the length of each delay element.

Fig. 5.2 Time arrangements of the clocks for (a) intensity modulation; (b) clock period modulation.

in the figure (b) is modulated keeping the off time of ϕ_1 and on time of ϕ_2 fixed. The photocharge collected at the ϕ_2 storage sites are reflected at the output as an additional bias component after (N-1) CCD clock pulses. This additional bias component is proportional to

$$\sum_{i=1}^N a_i T_2 \text{ where } T_2 \text{ represents the on time of the}$$

ϕ_2 clock. This situation however does not arise if devices with separate integrating sites (e.g., Fairchild's CCD 110/110F linear imager) are used. The photon-electrons can be integrated under the photogates and shifted laterally to the CCD storage registers, which are protected from the incident illumination. The availability of two separate CCD registers in the CCD 110/110F device, with independent transfer gates will allow the accumulation of positive lagged products in one register and negative lagged products in the other register (see Fig. 4.16(a)). However, if negative coefficients are handled in the above manner, for two dimensional processing, a specially designed area array imager (with two vertical CCD registers for each column of sensor elements) would be required. Alternately, the use of a bias to handle negative coefficients reduces the dynamic range. The method is however straightforward to use and the concept of linear image processing can be directly translated to an area-array device. This is the method discussed here.

5.3 EXPERIMENTAL RESULTS

This section is devoted to discussion of experimental results that validate the concept of linear processing of images. The

initial results presented here demonstrate the action of correlation between a linear image pattern and a set of known coefficients. Different image patterns have been correlated with the cosine and sine coefficients of varying frequencies. Since the Discrete Fourier Transform (DFT) can be viewed as a similar correlation, the Fourier components can be extracted from the results. Obtaining other transforms (Hadamard, Haar, etc.), image contour extraction, matching, etc. are also examples of correlation, and therefore, are within the scope of the present work.

In the demonstration experiment the correlation coefficients chosen were the cosine and the sine coefficients of a 32-point DFT given by,

$$Y_K = \sum_{n=0}^{N-1} a_n \cos(2\pi nk/N) - j \sum_{n=0}^{N-1} a_n \sin(2\pi nk/N);$$

$$K = 0, 1, \dots, N-1 \quad (5.2)$$

where the Y 's represent the output Fourier components and the a 's represent the samples of the image. The above equation clearly indicates that a real (or imaginary) Fourier component given by a certain value of K is obtained after a set of N real (or imaginary) coefficients corresponding to the same value of K have been correlated with the N image samples. Thus an N -point DFT requires N cosine and N sine coefficients for each value of K . This calls for a total of $2N^2$ coefficients. For real $\{a_n\}$ however, there are only $N/2$ unique (non-aliased) frequency components of $\{a_n\}$. Thus a total of N^2 coefficients are required. If only cosine or sine transform are

desired the requirement is further reduced by one-half. In a real application such a large number of coefficients if required in obtaining a transform of an image have to be generated using algorithms.

The speed with which a component of a transform is obtained is dependent on the CCD clock rate. Each component is obtained after N clock pulses. Thus a complete DFT would require N^2 clock pulses. Considering the volume of computations involved the speed is comparable to digital techniques using efficient and dedicated hardware.

In the actual experiment the real and imaginary coefficients were stored separately in a bank of PROM's. The coefficients were scaled down by a factor of 2 and a bias of 0.5 was applied before quantization into 8-bit words. These coefficients were then clocked sequentially along the rows of the coefficient matrix as shown in Fig. 5.3. The LED light intensity was modulated by the coefficients following D/A conversion. An optical mask pattern was used to define a linear image. A detailed block diagram of the circuit arrangement is shown in Fig. 5.4. The output of the D/A converter is sampled and converted to LED drive current in a way similar to that discussed in the context of electrical signal processing. A part of the cosine and the sine coefficients at the output of the D/A converter is shown in Fig. 5.5(a) and (b) respectively. The swept frequency nature of the coefficients is evident from the figure. The straight line portions in both figures correspond to coefficients for $K=0$, followed by $K=1$ and so on.

$$c[k,n] = \frac{1}{2} \cos\left(\frac{2\pi nk}{N}\right) + .5 \quad \text{for cos trnsfm.}$$

Fig. 5.3 Arrangement of coefficient memory.

Fig. 5.4 A block diagram of the circuit arrangement used to perform a 32-point DFT of a linear image.

(a)

(b)

Fig. 5.5 (a) The cosine and (b) the sine coefficients used to modulate the LED light intensity.

Fig. 5.6(a) shows a square wave mask pattern used to define an image. The equivalent temporal representation of the image is shown in Fig. 5.8(b). The theoretical results of correlation of the image pattern with the cosine and the sine coefficients are shown in part in Figs. 5.7(a) and (b) respectively. The results are plotted on the basis of eqn. (5.3), and do not take into account the effect of transfer inefficiency and coefficient quantization. In both figures, the horizontal axis represents the time samples and the vertical axis represents amplitude. The figures are drawn inverted because of the inverting action of the CCD. The components of the cosine and sine transform are obtained every 32nd sample at the output, i.e., at the end of correlation of the image pattern with each set of 32 coefficients. These sample points are marked in the figures as $K=0, 1, 2, 3$, etc. The effect of bias and coefficient quantization on the components of the cosine transform (computer) of the image pattern of Fig. 5.6(a) are shown in Table 5.1. With the application of a bias to the coefficients an output of 8 corresponds to an output of 0 in actual case. Thus a reduction of dynamic range is evident. Eight-bit quantization does not produce significant error. If the errors are compared to the peak output of 16, a signal to noise ratio of about 72dB is obtained. This is consistent with the theoretical discussions to follow.

The experimental results of correlation with the cosine and sine coefficients are shown in Fig. 5.8(a) and (c) respectively. For the sake of comparison, respective theoretical results are reproduced in Figs 5.8(b) and (d). The agreement appears to be good.

(a)

(b)

Fig. 5.6 (a) The bar pattern used in the experiment.
(b) Temporal representation of the bar pattern.

Fig. 5.7 (a) Theoretical plot of correlation of the bar pattern with the cosine coefficients.

Fig. 3.7 (b) Theoretical plot of correlation of the bar pattern with the sine coefficients.

(a)

(b)

Fig. 5.8 (a) Experimental results of correlation of the bar pattern with the cosine coefficients.
(b) Theory.

(c)

(d)

Fig. 5.8 (c) Experimental results of correlation of the bar pattern with the sine coefficients.
(d) Theory.

TABLE 5.1

The effect of bias and coefficient quantization
on the components of the cosine transform of Fig. 5.8(a)

COMPONENT	IDEAL COEFS.		QUANTIZED COEFS. WITH $\epsilon = 0$
	Actual	With Bias Applied	
K=0	<u>16</u>	<u>16</u>	<u>16</u>
K=1	0	8	8.000
K=2	<u>2</u>	<u>9</u>	<u>9.004</u>
K=3	0	8	8.000
K=4	0	8	8.000
K=5	0	8	8.000
K=6	<u>2</u>	<u>9</u>	<u>8.996</u>
K=7	0	8	8.000
K=8	0	8	8.000
K=9	0	8	8.000
K=10	<u>2</u>	<u>9</u>	<u>9.004</u>
K=11	0	8	8.000
K=12	0	8	8.000
K=13	0	8	8.000
K=14	<u>2</u>	<u>9</u>	<u>8.996</u>
K=15	0	8	8.000

NOTE: The non-zero components are underlined.

The theoretical results take into account the effect of transfer inefficiency (see section 5.4), quantization of the coefficients and small blurring of the mask image (which can be seen at the output corresponding to a single light flash). Also taken into account is the effect of the last segment ($K=15$) of the sine coefficients on the results of the cosine correlation and vice-versa. This occurs because the sine coefficients are stored following the cosine coefficients and because of recycling of the coefficients in order to get a stable scope display. This effect is reflected only on the first 31 samples of the results of cosine and sine correlation and does not affect the components of the transform.

For an additional demonstration, a 32-element rectangular image pattern was chosen. Fig. 5.9(a) shows the in contact aluminum mask (which avoids blurring) chosen for the experiment. The temporal representation of the image is shown in Fig. 5.9(b). The experimental and theoretical results of correlation with the cosine coefficients are shown in Fig. 5.10(a) and (b) respectively. The zero line indicated in figure (b) corresponds to applied bias at the input. Fig. 5.11(b) shows some of the cosine and the sine components of the 32-point transform. The component shown at the centre of the photograph is the component of the cosine transform for $K=0$ (DC). The components to the left of the DC component are components of the transform corresponding to $K=12, 13, 14$ and 15 respectively from the left end of the photograph. The components to the right of the DC component are components of the cosine transform corresponding to $K=1, 2, 3, 4$ respectively from the right of the centre. In this

(a)

(b)

Fig. 5.9 (a) Rectangular window mask pattern.
(b) Temporal representation of the mask image.

(a)

TIME

(b)

Fig. 5.10 Results of correlation of the rectangular mask image with the cosine coefficients.
 (a) Experiment; (b) theory.

(b)

Fig. 5.1] The components of the cosine and sine transforms.
(a) a sketch to indicate the coordinates
(b) observed components.

particular case only the $DC(K=0)$ component of the cosine transform has a non-zero value. The other components have equal heights corresponding to the applied bias at the input. This result is intuitively justifiable since the particular image pattern is a DC signal as far as the DFT is concerned.

We have thus experimentally demonstrated correlation of a linear stationary image with a set of known coefficients in the specific context of a DFT. Any other transformation for image representation on an orthogonal basis, e.g., Hadamard, Haar, etc. can be similarly implemented. The concept of correlation can be applied for matched filtering, edge detection, etc. An example of contour extraction using the second difference scheme [5.1] is shown in Fig. 5.12. The second derivative is approximated by: $a(n) - 2a(n-1) + a(n-2)$. A possible set of coefficients to achieve this operation is shown in the figure along with the results of correlation with a linear image pattern. The output bias level is obtained by first operating with an input corresponding to the bias applied to the coefficients. The actual coefficients are applied next. The output bias level is established after N clock pulses. Subtracting the following time samples from this bias level, bias cancellation can be achieved. The technique of Fig. 5.12 can be used to achieve any general correlation. One such application involves matching an image against a template [4.1] for the purpose of pattern recognition.

Fig. 5.12 A possible set of coefficients for image contour extraction is shown in the middle. The results of correlation is shown in the bottom.

5.4 PERFORMANCE CONSIDERATIONS

In chapter 3 we have discussed in detail the performance limitations of electro-optical processing of signals. Since the basic operation of both types of processing (signals and images) is similar, the limitations discussed in the context of electrical signal processing are equally applicable to linear image processing. The sources of noise and non-linearity discussed in chapter 3 will similarly affect the performance of a linear image processor. Variation of transmittance of light along a CCD channel, and lateral diffusion of carriers generated in the neutral bulk, formed potential sources of tap weight errors in the electro-optical transversal filter. For the case discussed here, alternately, these sources cause degradation of the image reproduced by a CCD imager. That is, an image different from the one focussed onto the device will be processed. The above limitations arise from technology limitations of the charge coupled device and are discussed in detail in chapter 3. In this section we shall consider the effect of digital quantization of coefficients and the charge transfer inefficiency inherent in a charge coupled device. Since the concept of image processing leads to an electronically programmable arrangement the coefficients of a desired transformation can be predistorted with relative ease in order to correct for the effect of transfer inefficiency.

5.4 (a) THE EFFECT OF TRANSFER INEFFICIENCY

The operation of the image processor involving correlation of a set of coefficients with a linear image pattern is diagrammatically

illustrated in Fig. 5.13. For simplicity only 4 CCD delay elements are considered. a_1, a_2, a_3 and a_4 denote the image samples at the four CCD storage locations. The X 's correspond to the coefficients that either modulate the intensity of the light or the period of the CCD clocks. The amount of charge gathered at any storage site during the m th CCD clock period is represented by the product of X_m and the corresponding a . The charge transfer loss per delay stage is denoted by ϵ and the transfer inefficiency by η (where $\eta=1-\epsilon$). Once in every CCD clock cycle, η times the stored charge is transferred from one storage site to another in the direction indicated by the arrowhead in the figure. The output at any clock period is the sum of the diagonal elements along the arrows. For an N -element CCD imager the output at $(N+1)$ th clock period represents the result after a set of N coefficients have been correlated with an image pattern

$$Y(N+1) = \eta^N a_1 x_1 + \eta^{N-1} [a_2 x_2 + (N-1)\epsilon a_2 x_1] + \eta^{N-2} [a_3 x_3 + (N-2)\epsilon a_3 x_2] \\ + \dots + \eta [a_N x_N + \epsilon a_N x_{N-1}] \quad (5.3)$$

In the absence of charge transfer loss, i.e., $\epsilon=0$, the ideal output at $(N+1)$ th clock period is,

$$Y_I(N+1) = \sum_{i=1}^N a_i x_i \quad (5.4)$$

(Terms with powers of ϵ greater than 2 have been neglected.)

Output at the 5th clock pulse.

$$Y(5) = n^4 a_1 x_1 + n^3 (a_2 x_2 + 3\epsilon a_2 x_1) + n^2 (a_3 x_3 + 2\epsilon a_3 x_2) + n (a_4 x_4 + \epsilon a_4 x_3)$$

Fig. 5.13 Operation of the processor with non-zero transfer inefficiency.

The limitations imposed by charge transfer loss can theoretically be circumvented by selecting coefficients to invert the effect introduced by the charge transfer inefficiency. If $\hat{x}$'s represent the modified coefficients, comparison of equations (5.2) and (5.3) leads to the modified Jth coefficient as,

$$\hat{x}_J = \frac{\sum_{i=0}^{J-1} (-1)^i \eta^{(J-1-i)} \epsilon^i (N-J-1-i)! x_{(J+i)}}{\eta^N} \quad (5.5)$$

In order to explain the implications of eqn (5.5) the result of a correlation between a 32-element rectangular image (all a 's=1) and a set of 32 coefficients of all unity values (all x 's=1) is considered in Fig. 5.14. The theoretical results before and after correcting the coefficients for a transfer inefficiency of 0.002 per delay stage have been compared with the ideal (eqn. 5.4) in the figure. The figure demonstrates that the desired output ($Y(33)$ in the figure) has been corrected by modifying the coefficients according to eqn. (5.5).

Correction for the effect of transfer inefficiency may present two practical difficulties. First, an accurate knowledge of ϵ is necessary. Second, the modified coefficients have to be correctly quantized in a limited number of digital bits. The first problem is more serious since the transfer inefficiency parameter may vary widely from device to device and is clock frequency dependent. However, since the experimental arrangement is electrically programmable,

Fig. 5.14 Results of a correlation between a 32-element rectangular image with a set of 32 coefficients of all unity value.

it may be possible to correct for the effect of transfer inefficiency by trial and error with relative ease.

5.4 (b) THE EFFECT OF COEFFICIENT QUANTIZATION

The errors in the coefficients of the desired transformation are introduced by quantization of the coefficients in a limited number of digital bits. As discussed before the negative coefficients are handled by suitable scaling and application of a bias, so that prior to quantization the coefficients lie between +1 and 0. The coefficients of well known transformations e.g., Fourier, Hadamard, Sine, Cosine, etc., however, lie between +1 and -1. Thus conversion according to a bias causes a reduction of the output dynamic range by one-half. However the bias b_0 applied to the coefficients is reflected at the output as a fixed offset proportional to

$$\sum_{i=1}^N a_i b_0 \quad \text{for stationary images.}$$

Consequently, bias cancellation followed by signal amplification is possible.

The quantization of the coefficients in 'b' number of digital bits provides a quantization step size of 2^{-b} . The coefficients are all rounded off to the integral multiple of the above step size. The errors introduced by this process lie within $\pm$ one-half of the least significant bit. The probability distribution of the error has a variance [3.1]

$$\sigma_e^2 = \frac{2^{-2b}}{12}$$

(5.6)

The output noise variance is thus

$$\sigma_f^2 = \sigma_e^2 \sum_{n=1}^N (a_n)^2 \quad (5.7)$$

where $\{a_n\}$ represents the image sample sequence. Since a_n is real and positive for all values of n , the peak output signal is given by

$$Y_p = \sum_{n=1}^N a_n \quad (5.8)$$

A signal to noise ratio can thus be defined as,

$$\text{SNR} = 10 \log_{10} \{R/\sigma_e^2\} \quad (5.9)$$

where,

$$R = \frac{\left(\sum_{n=1}^N a_n \right)^2}{\sum_{n=1}^N (a_n)^2}$$

For more than one non-zero image sample in $\{a_n\}$, $R > 1$ consequently the signal to noise ratio is better than $1/\sigma_e^2$. In general,

$$\text{SNR} \geq (6.02b + 10.79) \text{dB} \quad (5.10)$$

The major implication of eqn. (5.10) is that the signal to noise ratio increases approximately by 6dB for each bit added. The upper bound on R is N and the lower bound is 1. Thus the worst case signal to noise ratio ($R=1$) with 8-bit quantization is approximately 59dB. Thus 8-bit quantization will provide adequate signal to noise ratio for most applications. This requirement is compatible with easy commercial availability of 8-bit word memory systems or microprocessors.

5.5 TWO-DIMENSIONAL PROCESSING

The concept of processing a linear image functionally inside a CCD line imager can be extended to two-dimensional image processing by using an area-array CCD imager. In view of the particular interest in obtaining a two-dimensional transformation of an image, discussions in this section are devoted to a possible circuit configuration starting from the basic definition of a two-dimensional transformation.

If a matrix A represents an image that is sampled in space, a general separable linear transformation on the image matrix A is defined as [5.2]

$$[\alpha] = [U]^t [A] [V] \quad (5.11)$$

where $[\alpha]$ is called the transform domain of the image, $[U]$ and $[V]$ are unitary operators, and the superscript t denotes the matrix transpose. The possible set of orthogonal systems for decomposition of images is infinite; the common ones are Fourier, Hadamard, Haar, cosine, sine, etc. In all these cases $[U]=[V]$. The consequence of separability is that column transformation followed by row transformation yields the complete two-dimensional transformation. To elaborate, let us consider the specific case of the DFT of a two-dimensional finite area sequence $a(m,n)$ [3.1] which is zero outside the interval $0 \leq m \leq M-1$ and $0 \leq n \leq N-1$;

$$Y(k,l) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} a(m,n) e^{-j(2\pi mk/M)} e^{-j(2\pi nl/N)} \quad (5.12)$$

(for square matrix $M=N$)

Equation (5.12) can be expressed as:

$$Y(k,l) = \sum_{n=0}^{N-1} G(k,n) e^{-j(2\pi nl/N)} \quad (5.13)$$

where

$$G(k,n) = \sum_{m=0}^{M-1} a(m,n) e^{-j(2\pi mk/M)}$$

The function $G(k,n)$ corresponds to an M -point, 1-dimensional DFT, one for each value of n . The complete 2-dimensional DFT is then obtained by implementing N -point 1-dimensional DFT's, one for each row of the sequence $G(k,n)$. In other words, the 2-dimensional DFT is obtained by first transforming the columns and then transforming the rows of the result. This is also true for any other linear separable transformation discussed before.

The method of obtaining a 1-dimensional transform of a linear image can thus be applied to obtain the transformation of the columns by using an area-array imager. Transforms of all the columns are obtained at the same speed as that for a single column. The output from the CCD can be quantized and processed digitally to obtain the transforms of the rows.

The speed of the system is limited by the clocking rate of the vertical registers. Assuming that the digital system can put out a component of a transformation at N times the horizontal clock rate for an $N \times N$ array, the complete transformation is obtained in (N^2+N) vertical clock pulses. The additional N clock pulses are due to the need to establish the output bias level due to the bias applied

to the coefficients. For a 100×100 array, at 20KHz vertical clock rate, this time is approximately 0.5 sec. Thus real time (30 frames/sec.) 2-dimensional transformation is not quite achievable at the clock rates mentioned. Consequently, applications would be envisaged perhaps in the areas where TV frame rates are not necessary, such as in processing photographic negatives, where this approach would reduce cost compared to digital computations. If the entire processing is done digitally, the number of arithmetic operations involved, in general, is $2N^2 \log_2 N$ [5.2]. The combination of analog-digital technique discussed would reduce this number by one-half. A programmable filter [1.4], instead of a digital system, can be used to obtain the transform of the rows.

5.6 EXPERIMENTAL CONFIGURATION

In order to experimentally demonstrate the concept of two-dimensional image processing, an attempt was made to build a processor using the BNR 100×100 line transfer CCD imager*. However, at the time of writing this thesis the work is still continuing.

A detailed block diagram of the set up to perform a correlation operation in the vertical direction is shown in Fig. 5.15. Figure (a) shows a block diagram of the digital network at the output of the imager. The output from the CCD horizontal register is quantized (at 1MHz data rate) to subtract the bias voltage associated with each channel because of an applied bias to the input coefficients. The bias levels are established after the first one-hundred vertical

* This work is being supported by a Communications Research Centre contract entitled: "Research Study to Build a Demonstration Camera With Charge-Coupled Device Imagers for Performing Optical Image Processing Within the Imager", CRC Contract OSU77-00097

Fig. 5.15. (a). Output and (b) input circuit arrangement for the area-array image processor.

clocks and are held in RAM-1 in 100 eight-bit words. The subsequent outputs are held in RAM-2 and subtracted using 2's complement arithmetic. To obtain a 2-D transform, the CCD outputs every one hundred vertical clock pulses represent given components of a transform and are subtracted from the bias levels held in RAM-1. A programmable filter can be incorporated at the output of a D/A converter to compute the transform of the rows.

A block diagram of the input circuit is shown in Fig. 5.15(b). A microprocessor unit (INTEL MC-48) is used to put out a coefficient once every vertical clock period, (10KHz rate) to modulate the light intensity of the LED. The microprocessor is supplied with two pulses: one requesting a byte of information (at 10KHz rate) and the other signalling the beginning of a new frame (0.01 sec).

To operate the device in the camera mode (in pulse response), the LED is flashed once every 100 vertical clock pulses and the output from the CCD is directly fed to the monitor. The LED light is passed through a collimating lens to uniformly illuminate a photographic negative. A condensing lens placed between the negative and the device focusses the image onto the device.

A very initial demonstration of the processor working in the camera mode is shown in Fig. 5.16(a). Fig. 5.16(b) shows the convolution of the image with a rectangular pulse of width covering 11 vertical clocks. Smearing of the letters due to the convolution can be seen from the figure. For displaying the output from the CCD imager, Z-axis modulation of an oscilloscope was used. Although the problem of optics and certain circuit problems are yet to be solved, the preliminary results indicate the capability of linear image processing.

OF/DE

(a)

(b)

Fig. 5.16 (a) A very initial demonstration of the 100 x 100 CCD camera;
(b) convolution of the image with a rectangle results in smearing of the image.

CHAPTER 6

CONCLUSIONS AND FURTHER WORK

In this thesis, we have demonstrated the use of an LED/CCD structure to perform transversal filtering of electrical signals. The multiplications necessary in a convolution operation are obtained by means of an optical mask placed between the LED and the CCD, and the time-shifting of the weighted analog signal is provided by the CCD. The simplicity of this arrangement should be compared with the split electrode CCD filter, where programmability, i.e., changing the filter characteristic, can only be achieved by fabricating a new device.

The problem of achieving negative taps has been studied by using two approaches. In the time multiplexed scheme, the outputs at two successive CCD clock periods are combined and in the space multiplexed scheme, the outputs from two separate CCD channels are combined to obtain both positive and negative taps. Both these approaches have been experimentally demonstrated. Comparison between them indicates the superiority of the space multiplexed scheme from the point of view of ease of design of the coefficient mask, bandwidth capability and line length requirement.

The limitations of electro-optical signal processing was studied using the Blackman window function, which provides large side lobe attenuation. This filter function was realized with better than 1% tap weight inaccuracies. The major source of tap weight errors arises from variation of transmittance of light (along the channel) through the multilayer electrode structure of a charge coupled device.

The light transmittance was found to be particularly sensitive to the changes in the thickness of the polysilicon layer. The frequency response errors arising from lateral diffusion of carriers, generated beyond the depletion layer boundary, can be made negligible by using lower wavelength light or by increasing the depletion layer width.

The study of linearity of the processor indicates that the optical charge injection process is fairly linear. Non-linearity of the output signal mainly arises from the non-linearity of the charge detection output stage of the charge coupled device. Thus, overall non-linearity can be reduced by using an on-chip operational amplifier. Initial measurement of noise indicates strong evidence of $1/f$ noise at the source follower output buffer stage used. The signal to noise ratio is dependent on the speed of operation of the processor. Increasing the speed of operation reduces the level of the output signal, thereby lowering the signal to noise ratio.

Most of the limitations mentioned above arise from technology limitations of charge coupled devices. Thus better performances can be expected with improved design of CCD imagers.

The concept of processing a time-varying electrical signal has also been applied to process space varying images. The coefficients, in this case, are applied as a time varying signal, which results in a programmable image processor. The negative coefficients are handled by applying a bias to the coefficients. The Discrete Fourier Transform of a linear image pattern has been demonstrated experimentally. The concept can be extended to actual two-dimensional processing of images. A detailed scheme to this effect has been presented.

This thesis has demonstrated the basic capabilities of electro-optical processing of signals and images using functional convolution in CCD's. The future effort should be directed towards applications where the parallel processing capabilities of such a system can be exploited for processing a large volume of data e.g., the DFT of an electrical signal and the linear processing of images.

References:

- [1.1] D.D. Buss, D.R. Collins, W.H. Bailey & C.R. Reeves, "Transversal filtering using charge-transfer devices", IEEE J. Solid State Circuits, vol. 8, no. 2, April 1973, pp. 138-146.
- [1.2] J.J. Tiemann, W.E. Engeler & R.D. Baertsch, "A surface charge correlator", IEEE J. Solid State Ckts., vol. 9, Dec. 1974, pp. 403-410.
- [1.3] D.R. Lampe, M.H. White, J.L. Fagan & J.H. Mims, "An electrically reprogrammable analog transversal filter", ISSCC Conference Digest, Philadelphia, Feb. 1974, pp. 156-157.
- [1.4] Y.A. Haque, M.A. Copeland, "An electronically programmable transversal filter", Proc. IEDM 1976, pp. 27-30.
- [1.5] M.A. Copeland, D. Roy, J.D.E. Beynon & F.Y.K. Dea, "An optical CCD convolver", IEEE J. Solid State Ckts., Vol. 11, no. 1, Feb. 1976. (special issue on charge coupled devices), pp. 84-87.
- [1.6] D. Roy & M.A. Copeland, "Electro-optical processing of signals and images using CCDs", Proc. IEDM 1976, pp. 239-242.
- [1.7] D. Roy, M.A. Copeland & C.H. Chan, "Electro-optical processing of signals and images", to be presented at the NATO/AGARD Conf. on CCD/SAW Devices, Ottawa, October 1977.
- [1.8] J.K. Parks, "Optical-correlation detector for the audio frequency range", J. Acoust. Soc. Amer., vol. 37, Feb. 1965, pp. 268-277.
- [1.9] K. Bromley, "An optical incoherent correlator", Optica acta, vol. 21, no. 1, 1974, pp. 35-41.
- [1.10] R.P. Bocker, "Matrix multiplication using incoherent optical techniques", Applied Optics, vol. 13, no. 7, July 1974, pp. 1670-1676.
- [1.11] I. Lagnado & H.J. Whitehouse, "Signal processing image sensors using charge coupled devices", Proc. of International Conf. Tech. and Applications of CCDs, Sept. 1974.
- [1.12] M.A. Monahan, R.P. Bocker, K. Bromley, A.C.H. Louie, R.D. Martin & R.G. Shepard, "The use of CCDs in electro-optical processing", Proc. of International Conf. Tech. and Applications of CCDs, September 1975..

- [2.1] C.H. Seguin & M.F. Tompsett, "Charge transfer devices", Academic Press Inc., 1975., pp. 216-218.
- [2.2] A.A. Ibrahim, L. Sellers, T. Foxall & W. Steenart, "CCDs for transversal filtering applications", Proc. IEDM, Dec. 1974, pp. 240-243.
- [2.3] M.A. Copeland, D. Roy & C.H. Chan, "A multiplexed video bandwidth CCD delay line", ISSCC Conference Digest, Philadelphia, Feb. 1974, pp. 146-147.
- [2.4] L.R. Rabiner, R.W. Schafer & C.M. Rader, "The Chirp-Z transform algorithm", IEEE Trans. Audio and Electroacoustics, vol. 17, June 1969, pp. 86-92.
- [2.5] L.R. Rabiner & B. Gold, "Theory and application of digital signal processing", Prentice-Hall, Inc., 1975.
- [2.6] D.D. Buss & W.H. Bailey, "Applications of charge transfer devices to analog signal processing", IEEE intercon. Technical papers, 26-29 March 1974, Article 9/1.
- [2.7] D. Roy, C.F. Rahim, & M.A. Copeland, "Distortion effects due to multiplexing of charge transfer structures", IEEE Electron Devices, vol. 24, no. 6, June 1977, pp. 679-684.
- [2.8] D.D. Buss, W.H. Bailey & D.R. Collins, "Analysis and applications of analog CCD circuits", Proc. of 1973 Int. Symposium on Circuit Theory, Toronto, April 1973.
- [2.9] B.P. Lathi, "Signals, Systems and Communications", John Wiley & Sons, Inc., 1965.
- [3.1] A.V. Oppenheim & R.W. Schafer, "Digital signal processing", Prentice-Hall Inc., 1975.
- [3.2] D.S.K. Chan & L.R. Rabiner, "Analysis of quantization errors in the direct form for finite impulse response digital filters", IEEE Trans. Audio and Electroacoustics, vol. 21, no. 4, August 1973, pp. 354-366.
- [3.3] R.D. Baertsch, W.E. Engeler, H.S. Goldberg, C.M. Puckette & J.J. Tieman, "The design and operation of practical charge-transfer transversal filters", IEEE J. Solid State Ckts., vol. 11, no. 1, Feb. 1976, pp. 65-74.
- [3.4] R.W. Brodersen, C.R. Hewes & D.D. Buss, "A 500 stage CCD transversal filter for spectral analysis", IEEE J. Solid State Ckts., vol. 11, no. 1, February 1976, pp. 75-84.

- [3.5] J.E. Carnes & W.F. Kosonocky, "Noise Sources in a Charge Coupled Device", RCA Review, vol. 33, June 1972, pp. 327-343.
- [3.6] K.K. Thornber & M.F. Thomsett, "Spectral density of noise generated in CCD's", IEEE Trans Electron Devices, vol. 20, no. 1, 1973.
- [3.7] J.E. Carnes, W.F. Kosonocky & P.A. Levins, "Measurements of noise in charge coupled devices", RCA Review, vol. 34, Dec. 1973, pp. 553-565.
- [3.8] M.H. White, D.R. Lampe, F.C. Blaha & J.A. Mack, "Characterization of surface channel CCD image arrays at low light levels", IEEE J. Solid State Ckts., vol. 9, no. 1, Feb. 1974, pp. 1-13.
- [3.9] R.W. Brown & S.G. Chamberlain, "Quantum efficiency of a silicon gate charge-coupled optical imaging array", Phys. Stat. Sol., vol. 20, 1973, pp. 675-685.
- [3.10] C. Anagnostopoulos & G. Sadasiv, "Transmittance of air/SiO₂/Polysilicon/SiO₂/Si structures", IEEE J. Solid State Ckts. (Corresp.), vol. 10, no. 3, June 1975, pp. 177-179.
- [3.11] F. Van der Wiele, "Anti-reflection films and multilayer structures", NATO advanced study institute on solid state imaging, Universite Catholique de Louvain, Sept. 3-12, 1975.
- [3.12] D.M. Brown, M. Ghezzi & M. Garfinkel, "Transparent metal oxide CID imager", IEEE J. Solid State Ckts., vol. 11, no. 1, Feb. 1976, pp. 128-132.
- [3.13] S.M. Sze, "Physics of Semiconductor devices", Wiley-Interscience, 1968, Chapter 12.
- [3.14] G.C. Salter, "Investigation of silicon solar cells with field induced junctions", Master's Thesis, Carleton University, January 1975.
- [3.15] D. Roy, "A multiplexed CCD analog delay line", Master's Thesis, Carleton University, April 1974.
- [3.16] M.H. Crowell & E.F. Labuda, "The silicon diode array camera tube", BSTJ, May-June 1969.
- [3.17] D. Seib, "Carrier diffusion degradation of modulation transfer function in charge coupled devices", IEEE Trans. Electron Devices, vol. 21, no. 3, March 1974, pp. 210-217.

- [3.18] D.H. Harper, "The modulation transfer function of front illuminated charge coupled imaging systems", Master's Thesis, University of Waterloo, March 1976.
- [3.19] W.F. Kosonocky & J.E. Carnes, "Charge coupled digital circuits", IEEE J. Solid State Ckts., vol. 6, no. 5, October 1971, pp. 311-319.
- [3.20] W.M. Penny & L. Lau [ed.], "MOS integrated circuits", Van Nostrand, 1972.
- [3.21] W.N. Carr & J.P. Mize, "MOS/LSI design and application", McGraw-Hill, 1972.
- [4.1] A. Rosenfeld & A.C. Kak, "Digital picture processing", Academic Press, 1976.
- [4.2] "MOS/CCD data book", Fairchild Semiconductor Co.
- [4.3] D.D. Buss, R.L. Veenkant, R.W. Brodersen and C.R. Hewes, "Comparison between the CCD CZT and the digital FFT", CCD technology and applications Conf., September 1975.
- [5.1] P.I. Suciú & D.A. Hodges, "Image contour extraction with analog MOS circuit techniques", IEEE J. Solid State Ckts., vol. 12, no. 1, February 1977, pp. 65-72.
- [5.2] H.C. Andrews & B.R. Hunt, "Digital Image Restoration", Prentice-Hall Inc., 1977.

APPENDIX A

FABRICATION OF ON-CHIP MASKS

As discussed before, in-contact metal mask reduces coefficient errors due to lack of collimation of the incident light and virtually eliminates any misalignment error. Such metal masks may be useful in applications which demand simultaneous but fixed processing of an electrical signal using an area-array device, e.g., the DFT of a signal.

An on-chip metal mask was fabricated for a fast changing filter function. Aluminum was first evaporated onto a Pyrox-coated device. Photolithographic techniques were then used to define the window to be etched away. For the chip size of 151 x 135 sq.mils, the mask making facility available was inadequate. Extra step and repeat processes were therefore necessary to step four different quarters of the designed mask into a single complete mask pattern.

Fig. A-1(a) shows a fast changing filter pattern (passband extending beyond $0.4f_c$) on a glass mask. The observed impulse response shown in Fig. A-1(b) clearly indicates that the blurring error is unacceptable in this case. Fig. A-2(a) shows a microphotograph of a similar filter pattern etched on deposited metal. The observed impulse response is shown in Fig. A-2(b). Comparing this result with that of Fig. A-1(b), a drastic improvement can be seen. However, the measured frequency response indicated tap weight errors believed to be due to random etching of aluminum.

(a)

(b)

Fig. A-1 (a) The optical mask for a fast changing tap pattern.
(b) The corresponding impulse response.

(a)

(b)

Fig. A-2 (a) Micro-photograph of a fast changing filter function etched on deposited aluminum.
(b) The impulse response.
Comparison with Fig. 3.14 reveals improvement in the filter response.

The major problems encountered with this approach were:

- (i) defects (pinholes) in the Pyrox, and
- (ii) random etching of the pattern.

The first problem caused total failure of devices and uncontrollable random etching caused coefficient errors. One major cause for random etching was the contoured nature of the deposited metal because of the two-level poly layer structure of the device. Another cause was the non-uniform deposition of Pyrox. It is therefore believed that by using spin-on oxide to flatten the surface before metallization and using improved etching methods, the problem of random etching can be greatly solved. Because of a limited number of available devices, experiments with on-chip masks could not be carried any farther. The fabrication details of an aluminum mask is listed below.

Photolithography

Dry bake:	30 min. at $160^{\circ}\text{C} \pm 5^{\circ}\text{C}$
Photoresist application:	flood with Waycoat IC photoresist
Photoresist spin:	30 sec. at 5000 RPM
Pre-bake:	30 min. at $105^{\circ}\text{C} \pm 5^{\circ}\text{C}$
Exposure:	1.25 sec in strong U.V.
Develop:	45 sec. in Waycoat IC developer
	45 sec. in clean Waycoat IC developer
	Dip in Waycoat IC thinner.
	15 sec. in clean Waycoat IC thinner
	Dip in methanol.
	15 sec. in clean methanol
	Blow dry with N_2 .

Aluminum Etch

4 min. in phosphoric acid (60°C)

15 min. DI water rinse.

Photoresist Removal

10 min. in J100 at 100°C

5 min. in boiling tric

2 min. in clean boiling tric

1 min. in boiling acetone

15 min. DI rinse.

Aluminum Annealing

Temp. 450°C furnace

1 min. preheat at mouth

5 sec. push to centre

10 min. at centre

5 sec. pull to mouth

1 min. cool at mouth

END

5

08

78

FIN