NPS-012-78-002PR

NAVAL POSTGRADUATE SCHOOL

Monterey, California

A SUMMARY OF THE

FOUNDATION DEVELOPMENT PROGRAM

March 1978

Report for the Period

1 July 1976 to 30 September 1977

Approved for public release; distribution unlimited.

Prepared for:
Chief of Naval Development
Washington, D. C. 20360

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Tyler F. Dedman, USN
Superintendent

Jack R. Borsting
Provost

The work reported herein was supported by the Chief of Naval Development, Washington, D. C. 20360

This report was prepared by:

TABLE OF CONTENTS

FY 1977 FOUNDATION EXPLORATORY DEVELOPMENT PROGRAM

I. Introduction

The principal thrust of the exploratory development program at the Naval Postgraduate School (NPS) stems from the School's mission:

> To conduct and direct advanced education of commissioned officers, and to provide such other technical and professional instruction as may be prescribed to meet the needs of the Naval Service; and in support of the foregoing to foster and encourage a program of research in order to sustain academic excellence.

In fulfillment of the exploratory development program objectives and within the above constraints, the Naval Postgraduate School

> Conducts exploratory development (6.2) deriving from scientific program areas or in other areas specifically requested by the Navy.

In addition, NPS performs exploratory development, where uniquely qualified, for other agencies of the Department of Defense and, in defense related efforts, for other Federal Government agencies. NPS also furnishes consulting services for the Navy and, where specifically qualified, for other agencies of the Department of Defense and in defense related efforts for other Government agencies.

II. Background

The Navy has developed the Naval Postgraduate School as an academic institution which uses university educational methods to address the special graduate education requirements of the Navy. The Superintendent of the Naval Postgraduate School is a Rear Admiral of the line of the Navy. He is supported by a dual management structure, part military and part civilian. The faculty, mostly civilians, is responsible for the academic programs and, in support of these, conducts an active research effort. The military staff specifies the educational needs of the Navy, in terms of curricula, and provides administrative logistic support.

At the Naval Postgraduate School, as in other academic institutions, the faculty is organized into departments. The department represents a resource center of faculty members with allied disciplinary specialization. Currently, the departments at the Naval Postgraduate School include: Computer Science, Mathematics, Administrative Science, Operations Research, Physics and Chemistry, Electrical Engineering, Meteorology, Aeronautics, Oceanography, Mechanical Engineering, and National Security Affairs. Inter-disciplinary groups which have effective departmental status include Electronic Warfare, Anti-submarine Warfare, and Command, Control and Communications (C3).

Development projects are largely conducted by the individual faculty members on a project basis. Projects typically originate from proposals prepared either by individual faculty members or by groups of faculty members from the same or different departments.

The development program is divided administratively into two parts. First is the Sponsored Development Program. This program includes projects awarded by sponsoring agencies to an individual principal investigator. The principal source of funds is the various commands and laboratories of the Naval Material Command. The sponsored program constitutes about 80% of the total NPS development program. The Foundation Development Program is based on a grant from the Chief of Naval Development. This program is administered internally by a Research Council, made up of faculty members appointed by the Provost. The Dean of Research is the Chairman of the Research Council. The Council meets periodically to review faculty proposals, allocate funding, and review results of completed projects.

Appendix I identifies each project by title and category or type funding (Exploratory Development is 6.2 research). The 6.2 category was funded by the Chief of Naval Development, Washington, DC, 20360.

This report summarizes the FY 1977 and FY 197T Foundation Exploratory Development Program.

Title:	Investigation of a Sequential Test Median
Investigator:	D. R. Barr, Professor of Operations Research
Sponsor:	Foundation Research Program (6.2)
Objective:	To assess properties of a sequential test based on sample medians, including expected sample size and operating characteristics.
Summary:	Empirical evidence was obtained through simulation which demonstrates potential usefulness of the proposed test. This evidence suggested the sample number has asymptotic geometric distribution; efforts were undertaken to demonstrate this analytically, but to date have not been successful.
Publications Forthcoming:	It is anticipated that a journal article will be published on this material as soon as one or two open questions about the procedure are settled.
Theses Directed:	C. Y. Ching, "An Investigation of the Sequential Median Test," M.S. Thesis, NPS (1978). S. H. Wie, "A Comparison of Two Sequential CEP Tests," M.S. Thesis, NPS (1978).

FURTHER REMARKS ON THE SEQUENTIAL MEDIAN TEST

1. INTRODUCTION

Testing hardware systems to determine whether they satisfy operational specifications is frequently an expensive and time-consuming task. In some circumstances, it is possible to decrease the consumption of resources in testing by using sequential test procedures. A sequential test requires that limits be placed upon the desired precision of the test (probabilities of errors), and that the alternatives under test (hypotheses) be clearly specified. Such a test with two hypotheses is conducted by taking one observation (test value) at a time and, after each observation, making one of the decisions

a. Terminate and reject the null hypothesis;

b. Terminate and reject the alternate hypothesis;

c. Take another observation.

Since a sequential procedure requires that a decision (a - c above) be made after each observation, it is necessary to be able to obtain the observation value before another test is run (assuming it does not terminate at that point). In the case of testing a system where several test runs may be scheduled in a relatively short span of time, this means that the measure being used in the test must essentially be obtainable in real time, without appreciable processing delays. Similarly, some "field computation," again without long delays, may be required in order to reach one of the decisions a - c.

2. Description of the Sequential Median Test

Let the components of an IID sequence $\{X_1, X_2, \ldots\}$ be observed sequentially, and let M_k denote the sample median after $2k - 1$ observations have been made. We consider a sequential test of $H_0: X \sim F_0$ vs. $H_1: X \sim F_1 \leq F_0$ based on

the statistics M_k; $k = 1, 2, 3, \ldots$ and sequences $\{(a_k, r_k)\}$ of continuation intervals whose end points are "acceptance numbers" a_k and "rejection numbers" r_k. We assume F_0 and F_1 are absolutely continuous. The basic idea is to determine $\{(a_k, r_k)\}$ such that, if the test terminates when M_k is not between a_k and r_k (with appropriate terminal action), the procedure will have prespecified size α and power $1 - \beta$.

The test we consider is based on medians of odd sized samples because the distributions involved are simpler, from a computational standpoint, than those for arbitrary sized samples. Even so, the situation is challenging, primarily because the sequentially observed sample medians are not independent. In previous work, we have developed a numerical method of determining the acceptance and rejection numbers, and have applied the method to a CEP problem like that introduced above.

3. An Application

An example where such a procedure is useful is in testing air-to-ground weapon systems. A standard measure of performance for such systems is the "circular error probable" (CEP), which is defined to be the median radial miss distance. Frequently, it has been assumed that the weapon (say, bomb) impacts under fixed drop conditions follow a bivariate normal distribution, and that by rotating from the "target plane" to a "normal plane" (perpendicular to the bomb trajectory at impact), the distribution is approximately circular normal (bivariate normal with mean $\underline{0}$ and covariance matrix $\sigma^2 I$). Under these conditions, the radial miss distances are Rayleigh distributed, and the familiar formula $\text{CEP} = \sqrt{2 \ell n 2}\ \sigma \approx 1.1774\ \sigma$ holds. There has been increasing citicism in recent years of using radial miss distributions based on bivariate normal impact distributions in the underlying plane. Common criticisms cite poor fit of the resulting radial model in the tails, especially in the upper tail due to a relatively high frequency of large miss distances in many data sets. Since there is apparently a serious question about the validity of such models in a substantial number of cases, tests of CEP by corresponding traditional tests of σ^2 from an underlying normal model may be of doubtful validity. On the other hand, wide use continues to be made of models based on bivariate normal impact distributions, with corresponding faith in the

implications of these models. The sample median, therefore, seems to be an attractive alternative in such cases, because it is generally more robust with respect to poor fit in the tails than are statistics sufficient for σ^2. In addition to its robustness, the proposed procedure has the feature that it can be used with truncated samples with little or no modification, unlike some of its competitors. The procedure is flexible, and can be adapted to meet various constraints. For example, one can accommodate taking observations in batches, also the procedure can be modified so as to preclude termination before some specified number of observations have been taken.

The procedure is potentially of great usefulness for application to problems within DOD. For example, it should prove useful in connection with test and evaluation of a variety of weapons systems, as well as equipment components testing and quality control applications.

4. Results

Progress was made on the problems of determining asymptoti properties of this procedure; even so, many challenging problems remain. The author continues to devote effort to this problem area as time to do so becomes available.

Title: Aerodynamic Stabilization of Gaseous Discharges

Investigator: Oscar Biblarz, Associate Professor of Aeronautics

Sponsor: Foundation Research Program (6.2)

Objective: The main objective is to define practical aerodynamic means for stabilizing discharges for electrical laser. Other applications in cold plasma chemistry are also envisaged. A particular objective was to become acquainted with supersonic flow and vortex flows as stabilizing agents and, furthermore, to examine the theoretical aspects of aerodynamic stabilization. This is one part of a continuing project.

Summary: The work undertaken complements the activities in turbulence stabilization which have been underway at the Naval Postgraduate School. First-hand experience with supersonic flow and vortex-flow equipment was obtained. The problem of analyzing the effects of turbulence in a discharge was given some perspective, and the classical or Schottky solution for the positive column in a flow discharge was re-examined in the light of the stability problem. The program will continue with more studies with turbulence.

Conference Presentations: O. Biblarz, E. Wassertrom and Y. Crispin, "Stabilization of Flow Discharges by Supersonic Flows", presentation at the 30th Annual Gaseous Electronics Conference, 18-21 Oct. 1977, Palo Alto, CA, (abstract in conference record and in bulletin of the American Physical Society.

Publications: O. Biblarz, J. L. Barto, and H. A. Post, "Gas Dynamic Effects on Diffuse Electrical Discharges in Air", Israel Journal of Technology, Vol. 15, No. 1-2.

O. Biblarz, J. L. Barto, and H. A. Post, "Gas Dynamic Effects of Diffuse Electrical Discharges in Air", 19th Israel Annual Conference on Aviation and Astronautics, Tel Aviv-Haifa, March 2-3, 1977, (full paper on

conference record).

Y. Khait and O. Biblarz, "Influence of Turbulence on diffuse Electrical Gas Discharges Under Moderate Pressures". Paper submitted to the Journal of Applied Physics .

O. Biblarz, "Aerodynamic Stabilization of Gaseous Discharges", Technical Report, NPS-67Bi77111, 1 November 1977.

Conference presentation at 30th Annual Gaseous Electronic Conference, 18-21 October 1977, Palo Alto, CA

Stabilization of Glow Discharges by Supersonic Flows*. E. WASSERSTROM, Y. CRISPIN and O. BIBLARZ, Technion, Haifa, Israel. -- It is known that there is a strong coupling between electrical discharges and gas flows. In our work, this interaction is studied for supersonic flow[1], where the gas temperature is strongly coupled to the kinetic energy of the flow. Experimental results of supersonic flow discharges in cylindrical and conical tubes are described. It is found that the flow stabilizes the discharge and reduces the gas temperature. The results of some numerical calculations of a simplified model indicate that the computed exit temerature agrees rather well with our experiments. Our work is relevant to the stabilization of electric discharges in lasers.

---

1. E. Wasserstrom, Y. Crispin, J. Rom,and J. Shwartz: "The Interaction between Electrical Discharges and Gas Flow," to be published in the Journal of Applied Physics.

---

*Submitted by Oscar Biblarz

INFLUENCE OF TURBULENCE ON DIFFUSE ELECTRICAL GAS DISCHARGES UNDER MODERATE PRESSURES

Y. KHAIT and O. BIBLARZ*
Technion - Israel Institute of Technology
Haifa, Israel

ABSTRACT

The problem of a corona discharge under the influence of a turbulent flow is considered. This problem represents an extension of the already complicated situation of discharges in very non-uniform fields in non-moving gases, and our approach is restricted to providing a qualitative description of important phenomena. The geometry studied is that of pin anodes and a wire mesh (plate) cathode immersed in a flow where controlled turbulence is generated by the presence of grids and screens. The action of these is discussed as to their desirable effects on the discharge. The applicability of the theory of locally developed turbulence is mentioned along with semi-quantitative effects of turbulence on the transport coefficients. The fact that stationary distributions cannot be expected to occur under our conditions points out the usefulness of a pulsed-discharge analogy. Estimates of the relative power consumption, turbulent-to-laminar, can be made with the information given herein.

*Permanent Address: Naval Postgraduate School, Monterey, California.

PROOF OF ARTICLE TO BE PUBLISHED IN THE
ISRAEL JOURNAL OF TECH., VOL 15, No. 1-2.

Gas Dynamic Effects on Diffuse Electrical Discharges in Air

O. Biblarz, J. L. Barto and H. A. Post
Naval Postgraduate School,
Monterey, CA 93940 USA

ABSTRACT

Effects of velocity from flows with and without grid-generated turbulence on the power input to a diffuse discharge are presented. Density dependences and some electrodes are also reported. Air has been used because, convenience aside, it presents some of the chemical complexities of molecular laser mixtures. Improvement of the diffuse discharge power is noted as a result of turbulent flows at subatmospheric density conditions — as much as 500 times the maximum no-flow power. Density changes were obtained with a diffuser; the discharge geometry ordinary-mode consists of an upstream pin-rack electrode and a downstream wire-mesh electrode, typically 2.9 cm apart. A model for the interaction of the ionization/recombination region at the anode pin tips is presented which describes some of the observations of the behavior of breakdown voltage with the flow. The voltage just prior to the glow collapse decreases somewhat in the ordinary (or anode-to-cathode) flow mode and increases in the reverse and cross-flow modes. The current just prior to the glow collapse increases markedly under the action of a turbulent flow.

NOMENCLATURE

$D_{i,e}$	— diffusion coefficient for ions, electrons
D_a^*	— ambipolar diffusion coefficient (laminar or turbulent)
d	— physical gap
E	— electrical field
E_b	— electric field at breakdown (just prior to the glow collapse)
e	— electric charge
I_b	— current at breakdown (just prior to the glow collapse)
$j_{i,e}$	— current density for ions, electrons
L	— characteristic length for axial changes
L_r	— characteristic length for radial changes
$n_{i,e}$	— number density for ions, electrons
n^*	— charge density
$\dot{n}$	— net rate of charge production
R	— "hyperboloidal" pin radius
r	— radial coordinate
U	— convective field velocity
V	— discharge voltage
V_b, V_{b0}	— breakdown voltage (just prior to the glow collapse), reference value
V_i	— inception voltage
x	— axial coordinate
Z	— ionization coefficient
α	— recombination coefficient (two or three body)
ρ, ρ_0	— density, atmospheric value
δ_0	— anode extension for no flow
$\mu_{i,e}$	— mobility for ions, electrons
τ	— characteristic time

INTRODUCTION

as dynamic stabilization has been demonstrated to an important means of increasing the power into ffuse discharges in molecular gases (Shwartz and asserstrom, 1975; Brown and Davies, 1972; Bibrz and Nelson, 1974; Wasserstrom et al., 1976). These gases are found in molecular lasers, MHD devices, atmospheric discharges, and others. Turbulence, vorticity, and supersonic flow with shocks have been demonstrated to accomplish more than just bulk convective cooling of the discharge; an interaction between the flow field and the ions takes place

which somewhat affects the fundamental nature of the glow discharge. For example, the so-called corona inception voltage and the breakdown voltage of the glow are modified by the flow, while at the same time the current becomes considerably more homogeneous and greatly enhanced.

Our work has focused on the effects of turbulence, an effort which involves several related questions. We have been searching for an eddy size distribution or spectrum of turbulence which most favorably acts on the discharge (*Biblarz and Nelson*, 1974; ~~*Wasserstrom et al.*, 1976~~; *Aunchman*, 1974), i.e., for some optimum which represents the most effective coupling of flow power to discharge power. We have found eddy frequencies between 10^2 and 10^3 Hz very favorable. The convective velocity itself has a stabilizing effect on the discharge, the understanding of which turns out to be important in the sorting out of the action of turbulence. The observed increase in current and power with velocity is nearly proportional to the square of the convective velocity (*Nygaard*, 1965; *Biblarz*, 1974). Colateral to these studies has been our interest in the effects of density, interelectrode spacing, and electrode orientation. Since our experiments are carried out at pressures higher than optimum for discharges desirable for laser applications, it is significant to know the effects of reducing the density together with changes of interelectrode spacing. The orientation (and to some extent the polarity) of the pin-to-plate discharge geometry with respect to the flow field has a bearing on discharge activity, and such knowledge turns out to be important in the understanding of gas dynamic effects, as will be shown in this paper. Finally, our experiments have been conducted with atmospherical air where chemical complexities in the discharge are inherent; since we are observing relative changes in discharge performance and since air itself is a difficult discharge medium, we anticipe our results to be relevant to molecular lasers. Oxygen, for example, is used to increase laser power in CO–N_2 laser mixtures (*Keren et al.*, 1975); also, we expect our work to show qualitative trends in CO_2-mixtures.

The plane in the positive point-to-point geometry is represented in our experiments by a wire mesh which minimizes blockage of the flow. Even though an array of pins is used to fill the interaction volume, the anode is analyzed on the basis of the activity at a single pin with some statistical allowance for the behavior of the set (*Barto*, 1976). This type of discharge is "anode controlled" since the cathode plays only a secondary role (*Nasser*, 1971). When the electric field has reached a sufficiently high value, an ionization region develops in the neighborhood of the pin tips and it is this region that has the most profound effect on the discharge characteristics. The flow interacts with this region and a geometrical model which describes the anode extension, and which is discussed in this paper, seems to reflect some of the observed features of the discharge.

The flow can interact with discharge phenomena that occur in comparable times. Table I indicates a few relevant characteristic times. Some values are obtained using a characteristic length of one mm, which is the order of magnitude of the observed anode region.

TABLE I
CHARACTERISTIC TIMES

	Characteristic Time (s)	*Reference*
Diffusion (laminar)	10^{-1}	
Diffusion (turbulent)	10^{-4}	
Flow time (10–100 m/s)	10^{-4} –10^{-5}	
Turbulent eddies	10^{-1} –10^{-3}	*Biblarz and Nelson*, 1974
Current pulsing	5×10^{-4} – 2×10^{-5}	*Nasser*, 1971
Energy transfer, electrons to atoms	10^{-5} –10^{-6}	*Meek and Craggs*, 1953
Recombination	10^{-3} –10^{-6}	*Meek and Craggs*, 1953
Maximum positive ion drift ($E_r = 10^6$ V/m)	10^{-5} –10^{-6}	
Negative ion formation (O_2)	10^{-8}	*Brown*, 1966
Streamer propagation	10^{-8}	*Nasser*, 1971
Maxwellian distribution	10^{-8} –10^{-10}	*Meek and Craggs*, 1953
Excitation	10^{-12}	*Meek and Craggs*, 1953
Ionization	10^{-12}	*Meek and Craggs*, 1953

As may be evident from Table I, gas dynamic phenomena have times comparable to the recombination time, current pulsing, and ion drift among others. The effect of flow on recombination itself may be overshadowed by the effect of flow on the current, as discussed in a later section of this paper. It is interesting to note here that current pulsing

(*Nasser*, 1971; *Loeb*, 1965), a statistical fluctuation of the current which manifests itself only at the onset of the corona and just prior to breakdown, has characteristic times close to those of the high end of the turbulence spectrum. Since the maximum positive ion drift time is not much faster than flow times, it may be surmised that within the anode extension, and particularly in regions of low electric field charge, convection should be significant. Diffusion times are substantially lower than flow times and it is this difference which signals the possibility of using different characteristic lengths for diffusion and for convection in the equations (see Model and Additional Results section). An important feature of discharges in air results from the dominant electron-attaching property of oxygen which changes the electron energy distribution and electron density (*Keren et al.*, 1975) and which cannot be affected by turbulence (see Table I).

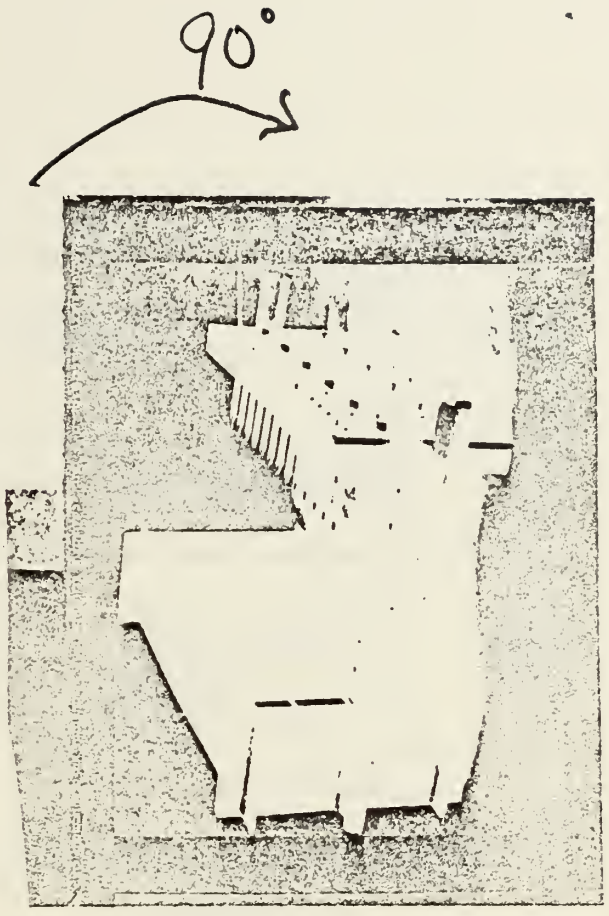


Fig. 1. Electrode system.

EXPERIMENTAL RESULTS

Facility

The experimental facility (*Post*, 1976) utilized consists of a plexiglass tunnel which is fed from an air compressor. Speeds in excess of 180 m/s can be obtained at relatively large flow rates with this facility. For subatmospheric operation, a diffuser is mounted after the test section. This diffuser was also made of plexiglass with an overall length of 1.04 m and wall divergence angle of 2.5 deg. No flow separation was observed in this diffuser. The test section houses the electrode system shown in Fig. 1, and the turbulence screen which is also evident in the photograph (Plate III attached to the pin rack). The interelectrode spacing is adjustable but most of the results reported here correspond to a spacing of 2.9 cm. The anode consists of 39 pins forming the corners of rectangles 1.27 cm by 0.64 cm; the pin length is 1.57 cm. The test section cross-sectional area is 64 cm^2.

Most of the experiments were done with the electrode orientation indicated in Fig. 2a (ordinary mode). Some reverse flow and cross flow experiments were also carried out with the configuration shown in Figs. 2b and 2c. Although data of Loeb (1965) indicate that under no-flow conditions in air an optimum of the power into the discharge just prior to the collapse of the flow is obtained with a pressure of 500 torr (for a 2.9 cm gap), our systems was only capable of operating down to about 640 torr (or ρ/ρ_0 or 0.85) because of the high blockage of the plate/electrode unit and because of chocking upstream of the electrodes.

Electrical and fluid dynamic measurements were taken with conventional equipment (*Biblarz and Nelson*, 1974; *Post*, 1976). All measuring equipment was calibrated before use and corrections for humidity changes were made on the measured data (*Post*, 1976). The screen used (Plate III) provided the best low frequency, high intensity eddies (*Aunchman*, 1974; *Post*, 1976). Plate III consists of an aluminum plate laminated on phenolic with mixed arrays of holes of 1.0 cm and 0.3 cm diameter. The plate is mounted directly upstream of the pin rack and its area is smaller than the test section area to allow for a boundary layer "bleed". In general we found that the pressure drop across the plate was a good indicator of its ability to stabilize the discharge (*Aunchman*, 1974). Turbulence intensities of the order of 10% are generated by Plate III.

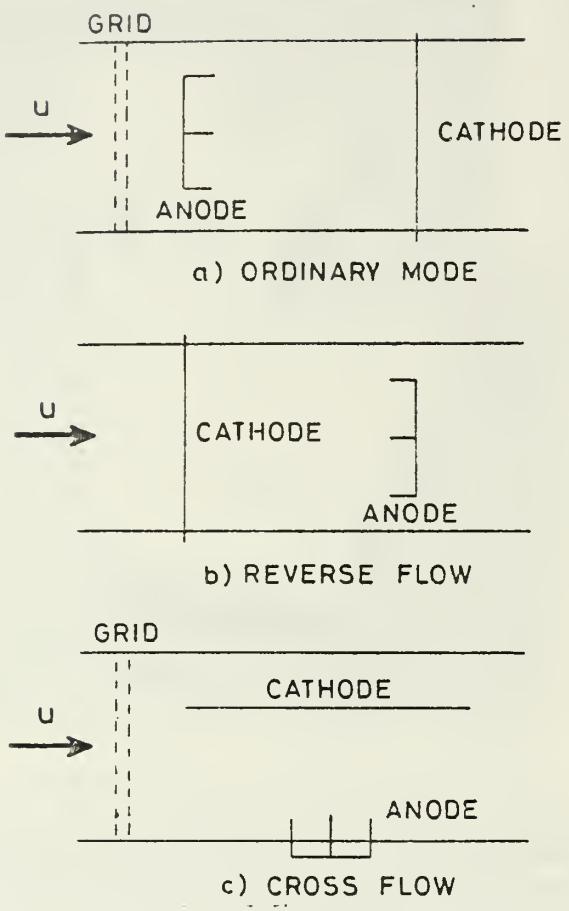


Fig. 2. Various orientations of electrodes and turbulence grids tested.

Results

Curves of power divided by the mass flow rate vs. the flow rate are indicative of flow effects (*Shwartz and Lavie*, 1975); any departure from the appearance of a rectangular hyperbola by the curves is to be attributed to flow effects. Figure 3 shows our experimental results for the ordinary mode in these coordinates; the power shown is the maximum diffuse or pre-arcing power. For the curves where the flow was turbulent the power was eventually limited by the power supply, whereas for the laminar flow curves the results were mass flow rate limited. The combined effect of reduced pressure and turbulent flow at maximum observed performance (*Post*, 1976) resulted in an increase of discharge power of 75% over ambient laminar flow at the same flow rate. The initial downward trend of the curves in Fig. 3 indicates a constant power input; the slope then becomes positive as the flow interaction becomes significant (beyond 60 m/s). Pure convection seems to saturate in its ability to affect the discharge whereas turbulence does not. Error bars are shown in Fig. 3 indicating the reliability of the measurements.

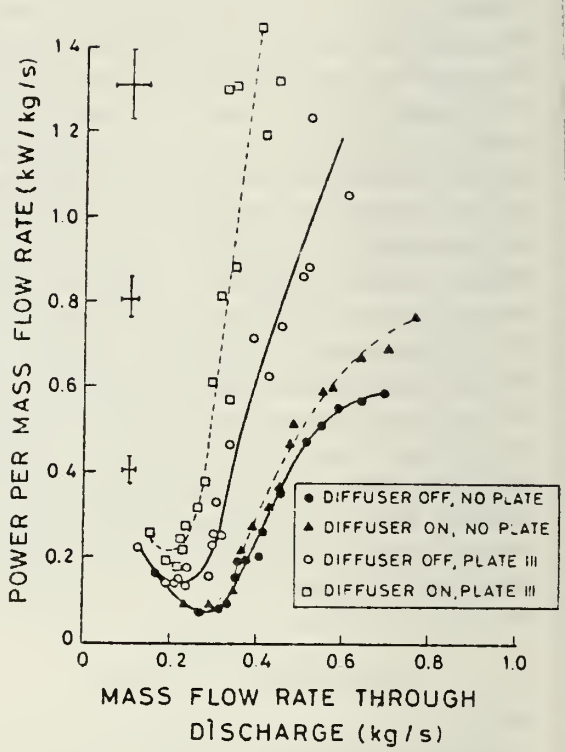
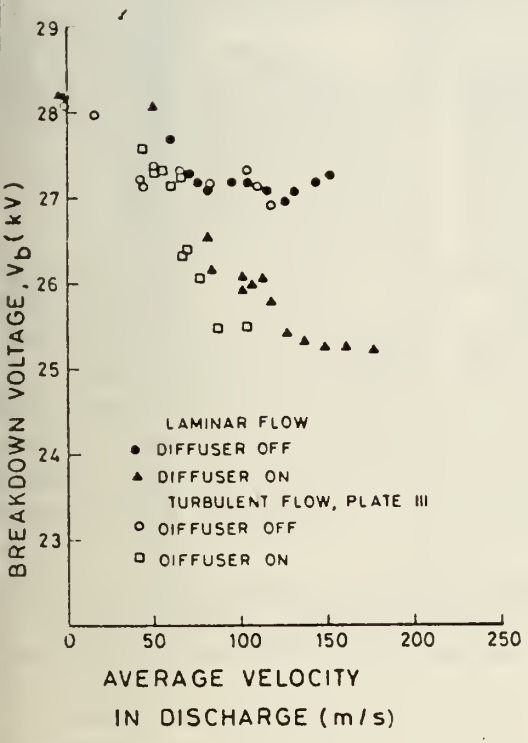
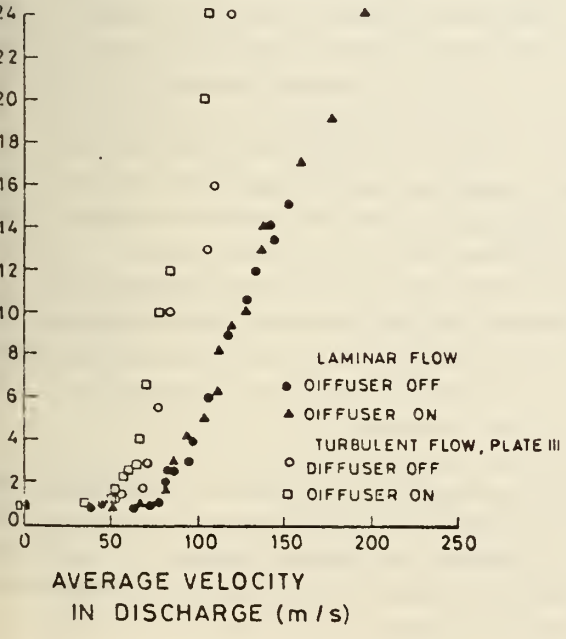


Fig. 3. Power per mass flow rate vs. mass flow rate through the discharge.

Figure 4 displays breakdown voltage data as a function of velocity. Clearly, the major parameters affecting the breakdown voltage are the density and the velocity (fixed spacing); turbulence appears to have only a secondary effect. The decrease of V_b with velocity is discussed in the next section in terms of the anode extension. Figure 5 is the corresponding plot for the breakdown or maximum diffuse current vs. average flow velocity; here, turbulence and velocity appear to be the controlling variables with the density playing only a secondary role. The behavior of the maximum power follows closely that of the current, since the magnitude of the change in voltage is relatively very small. In addition to spreading the discharge at each pin (see Fig. 6), turbulence causes all of the pins to "light-up" more or less equally; under non-turbulent conditions the pins at the outer periphery appear to be more active for the 2.9 cm



4. Breakdown voltage vs. average velocity through the arge



5. Breakdown current vs. average velocity through the dis-ge.

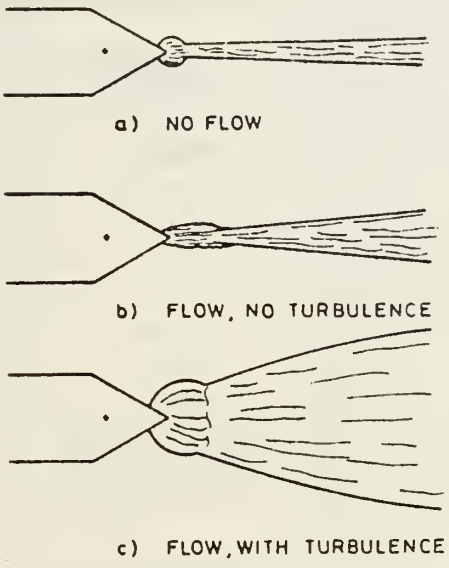


Fig. 6. Visual character of various glow modes without diffuser

spacing; for spacings under 2 cm all the pins seem to be active under all conditions. Figure 7 shows photographs of the various discharges. The no-flow photograph was a ten-second exposure. Here, the general appearance of the glow is that of a low intensity discharge extending between each pin and the cathode. The bright region at the bottom in the no-flow photograph was caused by a breakdown spark which took place during the time of exposure. In contrast, the laminar and turbulent flow photographs were one-second exposures. In the laminar flow photograph, the individual pin discharges appear to be more diffuse than in the no-flow case and an intense short narrow glow extends from the anode tips to the cathode. As evident from the turbulent flow photograph, the discharge here is even more diffuse with a wider bright region emerging from the anode region; turbulent mixing seems to spread out the current and homogenize the current density throughout the interaction volume. The current-voltage characteristics of the diffuse discharge display the effects of convection and density reduction as typified in the information plotted in Fig. 8.

Figures 9–12 will be discussed as part of the next section in which a model for the anode extension is presented.

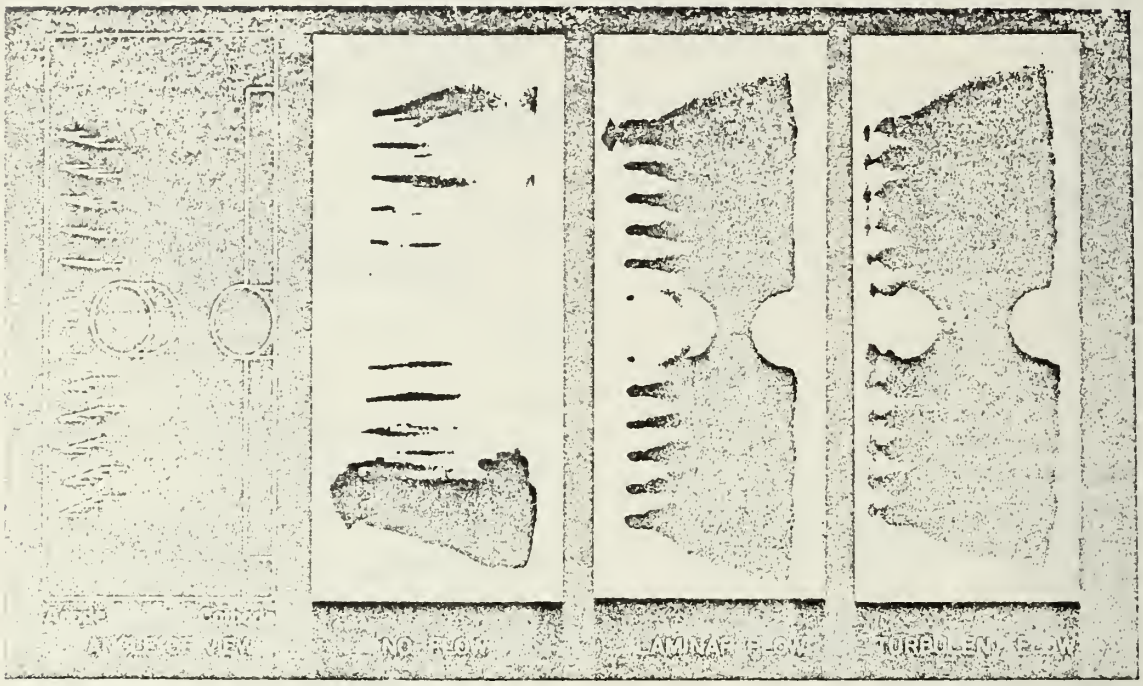


Fig. 7. Discharge appearance.

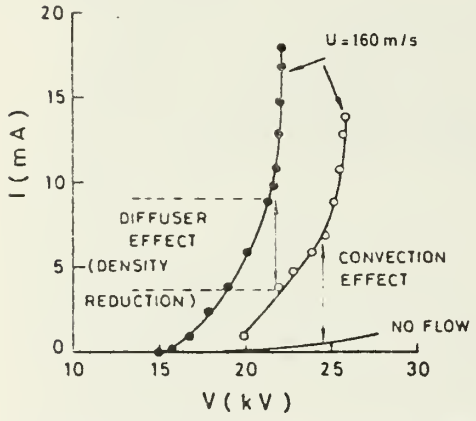


Fig. 8. Typical observed $I-V$ shift due to diffuser induced density reduction.

MODEL AND ADDITIONAL RESULTS

Model

The observations depicted in Fig. 6 indicate definite affects of the flow on the radiative recombination pattern at the anode tips. We propose to discuss here an elementary model which is based on a geometrical interpretation of changes attributable to diffusion and convection within the anode region. In this region, which exists already without any flow, intense ionization and recombination take place. Even though the anode region includes the sheath, we shall assume the region to be quasineutral for the simple reason that the sheath is considerably smaller (*Biblarz et al.*, 1975) and least affected by convection due to its proximity to the wall (*Dolson and Biblarz*, 1976). If there are changes in the anode region attributable to the flow, then we can think of them as changes in the effective physical gap which in turn will affect breakdown-gap related parameters such as V_c and V_b. In particular, it has been observed that the flow extends the anode region in the ordinary mode of operation and that this results in a decrease in the breakdown voltage.

From Figs. 6 and 7 it is seen that the diffusion field for the no-flow case is more or less spherical about the pin tip. Convection stretches out this diffusion region into an oblate spheroid and turbulence fattens the region considerably.

Since we are disregarding effects of convection on the sheath, the anode extension can be described with the quasineutral equations (*Shwartz and Wasserstrom*, 1975) where diffusion is ambipolar and where the effect of convection is included (*Biblarz and Dolson*, 1974; *Sutton and Sherman*, 1965). The species current equations are

$$\bar{j}_i = en_i\mu_i\bar{E} + en_i\bar{U} - eD_i\nabla n_i, \quad (1)$$

$$\bar{j}_e = -en_e\mu_e\bar{E} + en_e\bar{U} - eD_e\nabla n_e, \quad (2)$$

here the symbols are defined in the Nomenclature. he total current is $\bar{J} = \bar{j}_i - \bar{j}_e$. In the ambipolar egion, $n_i = n_e = n$ and the basic equations that escribe the system become

$$\nabla \cdot \bar{E} = 0, \quad (3)$$

$$\nabla \cdot \bar{j}_{i,e} = e\dot{n}. \quad (4)$$

olutions to Eq. (3) would have to be "patched" to xisting sheaths and subsequently to the external rcuit. Equation (4) assumes steady flow; we shall reat turbulent flows as pseudosteady where the ffects of fluctuations appear in the form of greatly nhanced transport coefficients. The form of the net roduction term, $\dot{n}$, that is ususal in positive column nalyses (*Cobine*, 1958; *Jaeger et al.*, 1976) includes onization by electron collision but two-body recom- ination (i.e., $Zn - \alpha n^2$); this is not appropriate for ear-atmospheric operation where three-body pro- esses dominate. We shall use, therefore, the three- ody form of Hinnov and Hirschberg (1961)

$$\dot{n} = Zn - \alpha n^3. \quad (5)$$

he above set of equations is incomplete without a uitable energy equation (*Sutton and Sherman*, 1965) hich yields the electron temperature and the gas emperature, if Joule heating is appreciable, and ome information regarding the flow field. If we liminate the electric field from Eqs. (1) and (2) in he usual way (*Brown*, 1966) we obtain

$$n\nabla \cdot \bar{U} + \bar{U} \cdot \nabla n - \nabla \cdot (D_a \nabla n) = Zn - \alpha n^3, \quad (6)$$

ere D_a is the ambipolar diffusion coefficient. The bove equation can be specialized to a simplified ersion of the conditions for the ordinary mode (see ig. 2a). Assuming that U is only in the axial irection and constant and that diffusion is radial ainly and symmetric, we can write Eq. (6) as

$$U\frac{\partial n}{\partial x} - \frac{1}{r}\frac{\partial}{\partial r}\left(rD_a\frac{\partial n}{\partial r}\right) = Zn - \alpha n^3. \quad (7)$$

he above is a non-linear, second order partial ifferential equation which we shall not attempt to olve here. It is useful, however, to suitably non- imensionalize (*Kline*, 1965) the equation so that characteristic parameters may be estimated. First of all, let us define n^* such that

$$Zn^* - \alpha n^{*3} = 0. \quad (8)$$

This corresponds to an equilibrium condition in which ionization and recombination locally balance each other. Note that n^* is a function of the electron temperature through the Saha equation (α is also a function of electron temperature (see *Hinnov and Hirschberg*, 1961 or *Mitchner and Kruger*, 1973)). The parameter n^* will be used for the characteristic density since $n/n^* \leqq 1.0$. As stated in the Introduction, there are really two characteristic lengths in the problem, one for the x-variation which is primarily convective and another for the r-variation which is primarily diffusive. With this in mind, and introducing D_a^* as the ambipolar diffusion coefficient for the condition n^*, we can rewrite Eq. (7) as follows:

$$\left(\frac{U}{ZL_x}\right)\frac{\partial n}{\partial x} - \left(\frac{D_a^*}{ZL_r^2}\right)\frac{1}{r}\frac{\partial}{\partial r}\left(rD_a\frac{\partial n}{\partial r}\right) = n - n^3, \quad (9)$$

where

$$n \leftrightarrow n/n^*,$$

$$x \leftrightarrow x/L_x,$$

$$r \leftrightarrow r/L_r,$$

$$D_a \leftrightarrow D_a/D_a^*.$$

Since we have made the variables of order one, it is reasonable to assume that each group of coefficients that appears in Eq. (9) is also of order one. Hence, we can arrive at an estimate for the characteristic lengths:

$$L_x \cong \frac{U}{\alpha n^{*2}}, \quad (10)$$

$$L_r \cong \sqrt{D_a^*/\alpha n^{*2}}. \quad (11)$$

Within a cylinder characterized by a length L_x and a radius L_r we anticipate to find the anode region. The accuracy of the above estimates will depend on the extent to which the electron temperature is constant in the ambipolar region because we have inserted n^* and D_a^* through the differential. Note also that we have assumed a thermally stable system.

Equations (10) and (11) are the desired interim r- ults. Equation (10) agrees with a solution of Eq.

(9) in which diffusion is completely neglected. Equation (11) is consistent with work reported in Mitchner and Kruger (1973) where a solution of Eq. (9) without the convective terms is shown to be of the "boundary-layer type". We can state now that the condition for which diffusion can be neglected in the x-direction is one for which $L_x \gg L_r$. Note that L_r will be greater for turbulent flows than for laminar ones.

It is clear that, other things being equal, L_x will increase with increasing convection velocity. If this corresponds to a lengthening of the ionization/recombination region then our geometric model appears to be appropriate. In particular, from Fig. 4, we have found that the breakdown voltage decreases with increasing velocity; Fig. 9 shows data additional to Fig. 3 for V_b/V_{bo} where V_{bo} is the no-flow value. In this figure, an eventual increase in V_b is also apparent (a more appropriate variable for the ordinate is U/d because it collapses the data for various gaps into one curve). The eventual increase in breakdown voltage, if it really exists, is not easily explained. It is known that turbulence causes the electric field to rise because of greater charge losses (*Shwartz and Wasserstrom*, 1975). This turbulence (either from the wake of the pins or grid generated) increases in intensity with increasing velocity, but it is not clear that results for a fixed column discharge apply to our case. Another possibility lies in the effects of convection on recombination; it is reasonable to expect that L_r is shrinking because n^* is increasing very rapidly; since the current in the discharge is roughly proportional to U^2 and since appreciable electron heating may take place at the higher currents, we can perhaps attribute the rise in voltage to changes in recombination length arising from an increase in the denominator in Eq. (10).

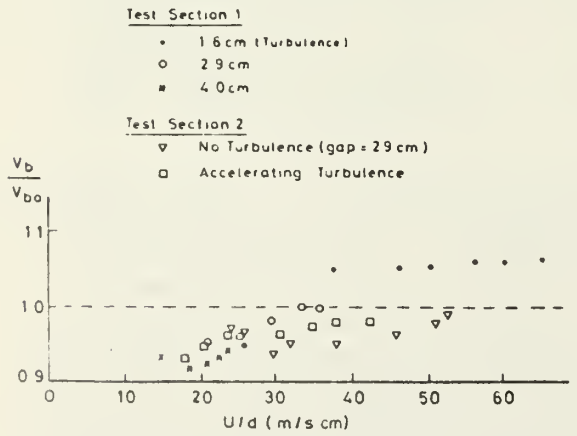


Fig. 9 Voltage ratio vs. U/d in the ordinary flow mode.

Additional Results

Figure 10 shows V_b/V_{bo} data for the reverse and cross-flow modes. Here, according to our model, the anode extension which exists for no flow is eventually, at high enough speeds, completely blown out of the interelectrode region leaving the physical gap as the smallest (or breakdown) gap. It is evident from the data in Fig. 10 that the voltage increase levels out at the high velocity end. These results are consistent with those of Biblarz and Nelson (1974) and Werner (1950). If d is the physical gap and δ_0 the anode extension for no flow, then the maximum increase in the breakdown voltage is simply

$$\left(\frac{V_b}{V_{bo}}\right)_{max} = \frac{d}{d-\delta_0}. \tag{12}$$

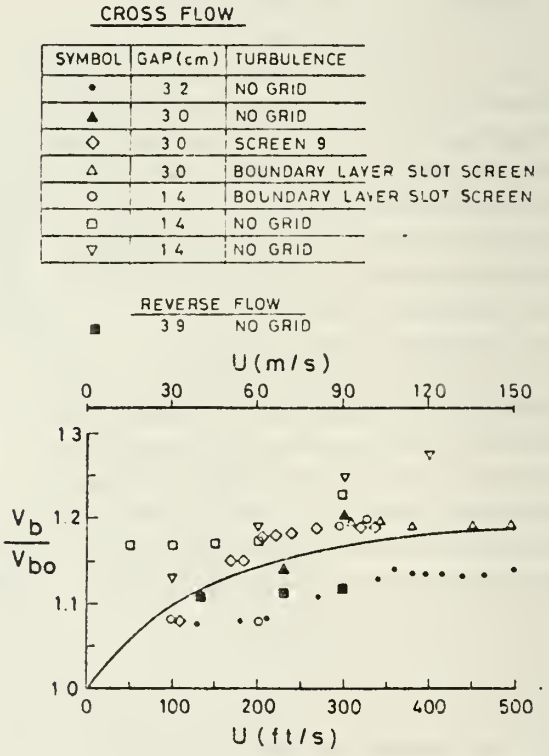


Fig. 10. Voltage ratio vs. velocity in reverse and cross-flow modes.

We are implying, of course, that the breakdown voltage is proportional to the distance between the anode or anode extension and cathode (*Barto*, 1976; *Nasser*, 1971). The sheaths at these pressures are very thin. In order to arrive at an estimate of δ_0 we must note that it should be a function of the electric field distribution outside of the anode region. Assuming a hyperboloidal point-to-plane distribution (*Barto*, 1976) for the electric field in the neutral region we have arbitrarily chosen the anode region to lie where $E \geqq 20$ kV/cm; this is a value of the electric field which is greater than the 15kV/cm needed for ionization by electron collision in air. Therefore

$$E(x) = \frac{R}{R + x} E_p, \tag{13}$$

where

$$E_p = \frac{2V}{R \log(4d/R)} \tag{14}$$

and $\delta_0 = x$ when $E(x) = 20$kV/cm.

Using values calculated with Eq. (13) and taking the data for the leveling off value of V_b/V_{b0} from cross-flow and reverse flow measurements, we arrive at the information shown in Table II and in Fig. 11. It must be kept in mind that the data come from various velocities and levels of turbulence. In Eq. (14) we use $R = 0.05$ cm and $V_b = 7.8d + 4$ (kV) as suggested by our experimental conditions (*Barto*, 1976). The trend is well indicated by the calculations, and the agreement between theory and experiment appears to be reasonably good. This then is supporting evidence for the anode extension model.

TABLE II

REVERSE AND CROSS FLOW MAXIMUM VOLTAGE RATIOS

Gap Length d(cm)	*Anode Ext.* δ_0(cm)	*Effective Gap* $d - \delta_0$(cm)	*Voltage Ratio Eq.* (12)	(*Max*). *Experiment*
1.0	0.22	0.78	1.28	—
1.4	0.27	1.13	1.24	1.27
1.45	0.27	1.18	1.23	1.26
2.0	0.34	1.66	1.20	—
3.0	0.45	2.55	1.18	1.20
3.2	0.47	2.73	1.17	1.15
4.0	0.56	3.44	1.16	1.13

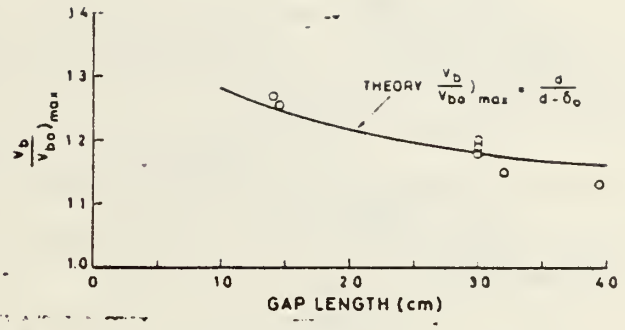


Fig. 11. Summary of maxium breakdown voltage ratio vs. gap length, reverse and cross flow only.

Our observations during the cross-flow experiments revealed a host of interesting effects, and since these may have a bearing to E-beam lasers they will be briefly discussed here. Figure 12 is a summary schematic of our observations. The ions within the anode extension, δ, are affected by two orthogonal forces, namely, convective drag in the flow direction and the electric force in the cathode direction. The net effect is shown in Fig. 12 as a displacement to the right which acts to lengthen the interelectrode gap and to distort the local field at the anode tip region. The displacement of the anode region was such that apparently it could affect the downstream pins (1.27 cm away!). Instabilities were observed in the second and third rows of pins (b and c in Fig. 12) in the form of minute but numerous transient arcs. These minute arcs were not strong enough to trip the power supply; they, however, spread into the first row of pins just prior to breakdown (when the power supply is actually tripped). Since breakdown streamers are energetic enough to be uninfluenced by the flow, the displacement of the anode region shows up as an increase in the breakdown voltage, as discussed previously. Even while these numerous arcs were evident, a blue-violet glow region (region 1 in Fig. 12) marked an upstream boundary for these arcs. This glow region was visible only from the first and second row of pins and it always "leaned into the wind". Downstream from the last set of pins, a blue-green glow was evident the extent of which decreased with increasing velocity; the length, x_{rec}, in Fig. 12 was of the order of 5 cm. In ~~Brown~~ Barto (1976) it is shown to follow a recombination controlled pattern which is proportional to the inverse of the convective velocity.

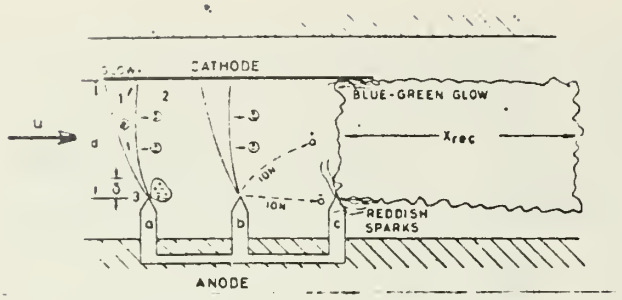


Fig. 12. Schematic representation of cross-flow current distribution observations.

CONCLUSIONS

The power into a diffuse discharge in air has been enhanced by 500 times by the combined action of flow, turbulence and reduced density. To some extent, effects of velocity, turbulence, and density have been separated allowing individual scrutiny of each. The experiments have been undertaken in a relatively high flow facility with a "difficult" molecular gas mixture.

By far the most pronounced effect is the enhancement in the magnitude of the current. Turbulence, for example, makes all the anode pins active, which is a difficult task to accomplish in practice with gaps over 2 cm (no individual ballasting of the pins was used); the increase of current, however, was more than the number of pins that were "activated". Turbulence considerably enhances and homogenizes the anode region apparently allowing each pin to flow more substantial amounts of current. Of course, velocity also shows up as an important parameter, but there is no pure laminar flow at the pin tip because it is in a wake where vorticity is being shed; still, the current has been observed to possess a square or higher (*Barto*, 1976) dependence on the convective velocity. It is interesting to note that density changes by themselves did not seem to affect the results shown in Fig. 5; this is somewhat surprising in the light of the dependence of the Townsend coefficient on the density (*Nasser*, 1971; *Cobine*, 1958). Figure 6 shows some density enhancement but not a substantial one. Turbulent flows have been seen to stabilize our discharges in which Joule heating and thus thermal instabilities are not as substantial as in the laser discharges (*Brown and Davies*, 1972; *Wasserstrom et al.*, 1976); it may therefore be concluded that the effect is a fundamental one.

With the anode region model we have shown that it is possible to interpret the action of convection in terms of a physical extension or contraction of the anode region. This incidentally can also be argue to apply to results with parallel plate electrodes (*Barto*, 1976). In the ordinary flow mode, a minimum of V_b/V_{b0} is obtained of about 0.90–0.95 corresponding to a velocity of about 75 m/s; the subsequent increase in breakdown voltage is at the present not a conclusive result. The reverse and cross-flow modes saturate at a V_b/V_{b0} of about 1.25, which is postulated to correspond to the limit imposed by the physical gap. The results for the breakdown voltage indicate that a relation exists of the form

$$\frac{V_b}{V_{b0}} = f\left(\frac{U\tau}{d}, \cdots\right), \tag{15}$$

where τ is some relevant characteristic time.

REFERENCES

Aunchman, L. J., 1974. Controlled tubulence as a design criterion for electrical discharge convention lasers, MS Thesis, Naval Postgraduate School, Monterey, Calif.

Barto, J. L., 1976. Gasdynamic effects on an electric discharge in air, MS Thesis, Naval Postgraduate School, Monterey, Calif.

Biblarz, O., 1974. Stabilization of a medium pressure discharge by turbulent flows, 27th Annual Gaseous Electr. Conf., Houston, Texas.

Biblarz, O. and R. C. Dolson, 1974. Voltage profiles in the vicinity of non-emitting electrodes, 14 Engr. MHD Symp., UTSI, Tenn.

Biblarz, O. and R.E. Nelson, 1974. Turbulence effects on an ambient pressure discharge, *J. Appl. Phys.*, **45**, 633–637.

Biblarz, O., R. C. Dolson and A. M. Shorb, 1975. Anode phenomena in a collision-dominated plasma, *J. Appl. Phys.*, **46**, 3342–3346.

Brown, S. C., 1966. *Introduction to Electrical Discharges in Gases*, John Wiley and Sons, New York.

Brown, C. O. and J. W. Davies, 1972. Closed cycle performance of high power electric discharge laser, *Appl. Phys. Lett.*, **21**, 480–481.

Cobine, J. D., 1958. *Gaseous Conductors*, Dover Edition

Dolson, R. C. and O. Biblarz, 1976. Analysis of the voltage drop arising from a collision dominated sheath, *J. Appl. Phys.*, **47**, 5280–5287.

Hinnov, E. and J. G. Hirschberg, 1961. Electron-ion recombination in dense plasmas, *Phys. Rev.*, **125**, 795.

Jaegle, E. F., L. Oster and A. V. Phelps, 1976. Growth of thermal constrictions in a weakly ionized gas discharge in helium, *Phys. Fluids*, **19**, 819.

Keren, H., P. Avivi and F. Dothan, 1975. The influence of oxygen on CO-laser performance, IEEE J. Quantum Electronics, **QE-11**, 590.

Kline, J. S., 1965. *Similitude and Approximation Theory*, McGraw-Hill, New York.

Loeb, L. B., 1965. *Electrical Coronas*, University of California Press, Berkeley, Calif.

Meek, J. M. and J. D. Craggs, 1953. *Electrical Breakdown of Gases*, Oxford University Press, pp. 149–150.

Mitchner, M. and C. H. Kruger, 1973. *Partially Ionized Gases*, John Wiley and Sons, New York.

Nasser, E., 1971. *Fundamentals of Gaseous Ionization and Plasma Electronics*, Chap. 11, John Wiley and Sons, New York.

Nygaard, K. J., 1965. "Electric wind" gas discharge anemometer, *Rev. Sci. Instrum.*, 21, 67.

Post, H. A., 1976. Sub-ambient controlled turbulence effects on discharge stabilization for laser applications, MS Thesis, Naval Postgraduate School, Monterey, Calif.

Shwartz, J. and V. Lavie, 1975. Effects of turbulence on a weakly ionized plasma column, *AIAA J.*, 13, 647-652.

Shwartz, J. and E. Wasserstrom, 1975. The role of gas flow and turbulence in electric discharge lasers, *Isr. J. Technol.*, 13, 122–133.

Sutton, G. W. and A. Sherman, 1965. *Engineering Magnetohydrodynamics*, McGraw-Hill, New York.

Wasserstrom, E., Y. Crispin, J. Rom and J. Shwartz, 1976 The interaction between electrical discharges and gas flow, AIAA Paper No. 76–315, San Diego, Calif.

Werner, F. D., 1950. An investigation of the possible use of the glow discharge as a means for measuring air flow characteristics, *Rev. Sci. Instrum.*, 21, 67.

NPS-67Bi77111

NAVAL POSTGRADUATE SCHOOL

Monterey, California

AERODYNAMIC STABILIZATION OF GASEOUS DISCHARGES

O. Biblarz

Final Report for Period 1 Oct 1976 - 30 Aug 1977

NAVAL POSTGRADUATE SCHOOL

Monterey, California

Rear Admiral I. W. Linder
Superintendent

Jack. R. Borsting
Provost

This report reflects work performed during AY 1976-77 while the author was on temporary duty at the Department of Aeronautics at the Technion in Haifa, Israel.

The author would like to thank his colleagues and students at the Technion for helping make his stay a memorable experience.

This report was prepared by:

Oscar Biblarz

O. BIBLARZ
Associate Professor of Aeronautics

Reviewed by:

R. W. BELL, Chairman
Department of Aeronautics

Released by:

for W. M. Tolles

R. R. FOSSUM
Dean of Research

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS67Bi77111	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AERODYNAMIC STABILIZATION OF GASEOUS DISCHARGES		5. TYPE OF REPORT & PERIOD COVERED 1 Oct 1976 - 30 Aug 1977 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) O. Biblarz		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE 1 November 1977 13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Glow Discharges
Stability
Aerodynamics
Lasers

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

This report covers some aspects of the formulation of the stabilization problem by aerodynamic means. Results of experimental work are not included but are referred to in appropriate places. The classical solution for ambipolar diffusion in tne positive column of the glow discharge must be extended to reflect the influence of current and pressure. The influence of current is given by the inclusion of appropriate energy equations and the pressure influence by tne addition of two-body recombination. It was found that, from dimensional analysis, the characteristic lengths for ambipolar and heat diffusion for

DD FORM 1 JAN 73 1473 (Page 1) EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

helium do not match. A re-formulation of the problem is thus necessary. The recombination term does not change the eigenvalue nature of the diffussion equation except at the stationary or local equilibrium limit.

TABLE OF CONTENTS

I. INTRODUCTION

In the high energy laser field, the electric discharge convection laser (EDCL) offers good efficiency as well as simplicity of handling relative to gasdynamic and chemical lasers. Progress towards higher laser outputs from EDCLs, however, has been impeded by the collapse of the glow discharge at the increased currents and pressures desired. This glow collapse (usually referred to as arcing or constriction) represents the practical upper limit of the electrical power that may be pumped into the necessary non-equilibrium processes. It has been shown[1-4] that aerodynamic stabilization can be used very successfully in overcoming the glow-collapse limits. It should be pointed out here that the laser application is but one application of the relatively new interdisciplinary field of cold plasma chemistry.

Aerodynamic stabilization is based on two separate effects usually referred to as convection (laminar) and mixing. The purely convective effect is said to decrease the residence time of the particles through the discharge and thereby to counteract unstable mechanism that lead to the glow collapse[5]. The mixing effect, which may be due to vortex flows or to turbulence from boundary layers or grids, has been observed to have powerful stabilizing characteristics. Convection scales up with flow velocity whereas mixing depends on the intensity, spectral distribution, and location of the eddies or vortices.

II. SPECIFICS

At the Technion, the author had the opportunity to participate in experimental work dealing with vortex-flow stabilization[6,7] and with supersonic flow stabilization[6,8]. These techniques complement the work that has been underway at the Postgraduate School on turbulence stabilization of discharges. Some results of the Technion work have already appeared (see Refs. 6-11) and more will be forthcoming.

A significant portion of the author's time was dedicated to a study of theoretical aspects of discharge stabilization. The author was privileged to colaborate with Dr. Y. Khait in establishing a perspective regarding some of the author's previous results with turbulence stabilization. This effort resulted in the writing of a paper titled "Influence of Turbulence on Diffuse Electrical Gas Discharges Under Moderate Pressures," which is being submitted for publication. Other theoretical work involved a close look at the Schottky solution for the positive column in a glow discharge with the aim at establishing its use for a stability formulation. This effort was fruitful on two intermediate aspects of the problem and these are included here as Appendixes A and B.

A proposal was written to the United States-Israel Binational Science Foundation[10] which extensively outlines the progress of the entire effort. If awarded, the cooperation of the Technion and the Postgraduate School would be formalized (however, the budget would provide neither salaries for principal investigators nor equipment for the USA experiments).

III. DISCUSSION

The classical solution to the ambipolar diffusion of charges in the positive column[12,13] of a glow discharge offers no hints as to how striations and constriction, among other important observations, manifest themselves in the discharge. As pressure and/or current are increased, instabilities grow in the discharge and it is important to have a sufficiently complete model of the physics which reflects these effects. In other words, before flow effects such as (laminar) convection and turbulence are introduced it is necessary to re-examine some assumptions inherent in the Schottky solution.

A partial list of phenomena not reflected in the Schottky solution may be given as follows:

1) effect of sheaths at the walls
2) effect of electrodes and intermediate regions
3) effect of recombination in the gas (pressure and composition)
4) effects of the energy balance, including radiation and the elevation of the electron temperature

The list can be rather long. We closely examined the energy balance because the instabilities are believed to be thermal disturbances (i.e., the temperature becomes non-homogeneous) and recombination in the gas because it closely reflects the effect of pressure and composition. For simplicity, a two-dimensional (infinite parallel plate) geometry was considered, see Fig. 1.

When E_L is constant, $\nabla \cdot \bar{J} = 0$ implies that $J_x = J(y)$ or that $n = n(y)$ because $J_y = 0$; moreover, $\nabla \times \bar{E} = 0$ implies that $E_y = E(y)$. The quasineutral equations for steady state may be written as[10,14]

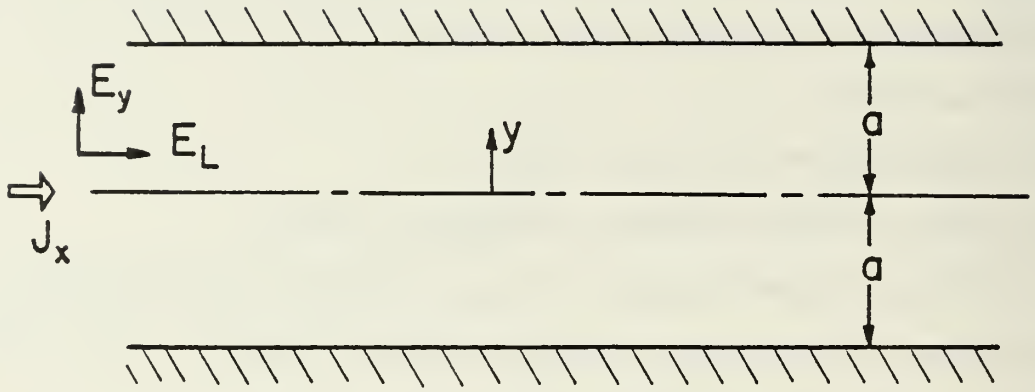


Figure 1. Geometry of Positive Column

$$- \partial/\partial y \left[\frac{D_a}{2T} \frac{\partial}{\partial y} (nT_e)\right] = \nu_i n - \alpha n^2 \quad (1)$$

$$- \partial/\partial y \left[K \, \partial T/\partial y\right] = \bar{J}\cdot\bar{E} = en\mu_e E_L^2 \quad (2)$$

$$2 \frac{me}{m_N} \nu_{eN} \frac{3}{2} k(T_e - T) = e\mu_e E_L^2 \quad (3)$$

$$p = NkT \quad (4)$$

with boundary conditions

$$\text{at} \quad y = 0 \qquad \partial n/\partial y = \partial T/\partial y = 0$$

$$\text{at} \quad y = \pm a \qquad n \cong 0 \quad \text{and} \quad T = T_o$$

In the above formulation x is the axial coordinate and y the interplate coordinate, n the charge concentration in the ambipolar region, T_e the electron temperature and T the gas/ion temperature, N is the concentration of the neutrals, D_a the coefficient of ambipolar diffusion, K the thermal conductivity, m_e the electron mobility, ν_i and α the ionization and recombination coefficients respectively, E_L the axial electric field, ν_{eN} the collision frequency between electrons and neutrals, e the electronic charge, p the pressure, and m_e and m_N the mass of the electrons and neutral respectively. There are several assumptions inherent in this formulation, namely that $N>>n$, $T_e>T$, and that p = constant . In addition, we are using a symmetry condition in the boundary conditions and we are neglecting of wall space charges. Glow discharge data[12,13], in addition, indicate that E_L as well as T_e appear to be constant in the discharge at least before constricting. The electron energy equation, Eqn. 3, should have to reflect the energy dependance of the collision cross sections and, in addition,

non-elastic collisions -- since we are interested in laser operation. However, we shall not deal with Eqns. (3) and (4).

Using simple kinetic theory to describe the gas temperature dependance of the coefficients (D_a , μ_e , K , etc.) and defining a stationary state where

$$\nu_i n^* - \alpha n^{*2} = 0 \tag{5}$$

$$\text{or} \qquad n^* = \nu_i/\alpha$$

we can proceed to nondimensionalize Eqns. (1) and (2). We shall, however, not use the plate spacing to render y dimensionless, but introduce two characteristic lengths, L_n and L_T[15]. The gas temperature is made non-dimensional by the use of a wall temperature T_o . The resulting equations are:

$$-\, d/dy \left(T^{1/2} \frac{dn}{dy}\right) = n - n^2 \tag{6}$$

$$L_n \equiv \sqrt{\frac{T_e D_{ao}}{2\nu_{io} T_o}} \tag{7}$$

$$-\, d/dy \left(T^{3/2} \frac{dT}{dy}\right) = nT \tag{8}$$

$$L_T \equiv \sqrt{\frac{T_o K_o}{en^* \mu_{eo} E_L^{\,2}}} \tag{9}$$

It seems reasonable to expect that in a conventional (unconstricted) glow, the two characteristic lengths be about equal to each other and that

$$L_n/a \approx 0(1) \quad \text{and} \quad L_T/a \approx 0(1) \tag{10}$$

This expectation is also required by the eigenvalue nature of Eqns. (6) and (8). It may be shown (see Appendix B), that except for the equilibrium limit, Eqn. (6) is an eigenvalue equation, (the classical solution being a cosine when no recombination is present and when $T \approx$ constant). The nature of Eqn. (6) is discussed further at the end of this section.

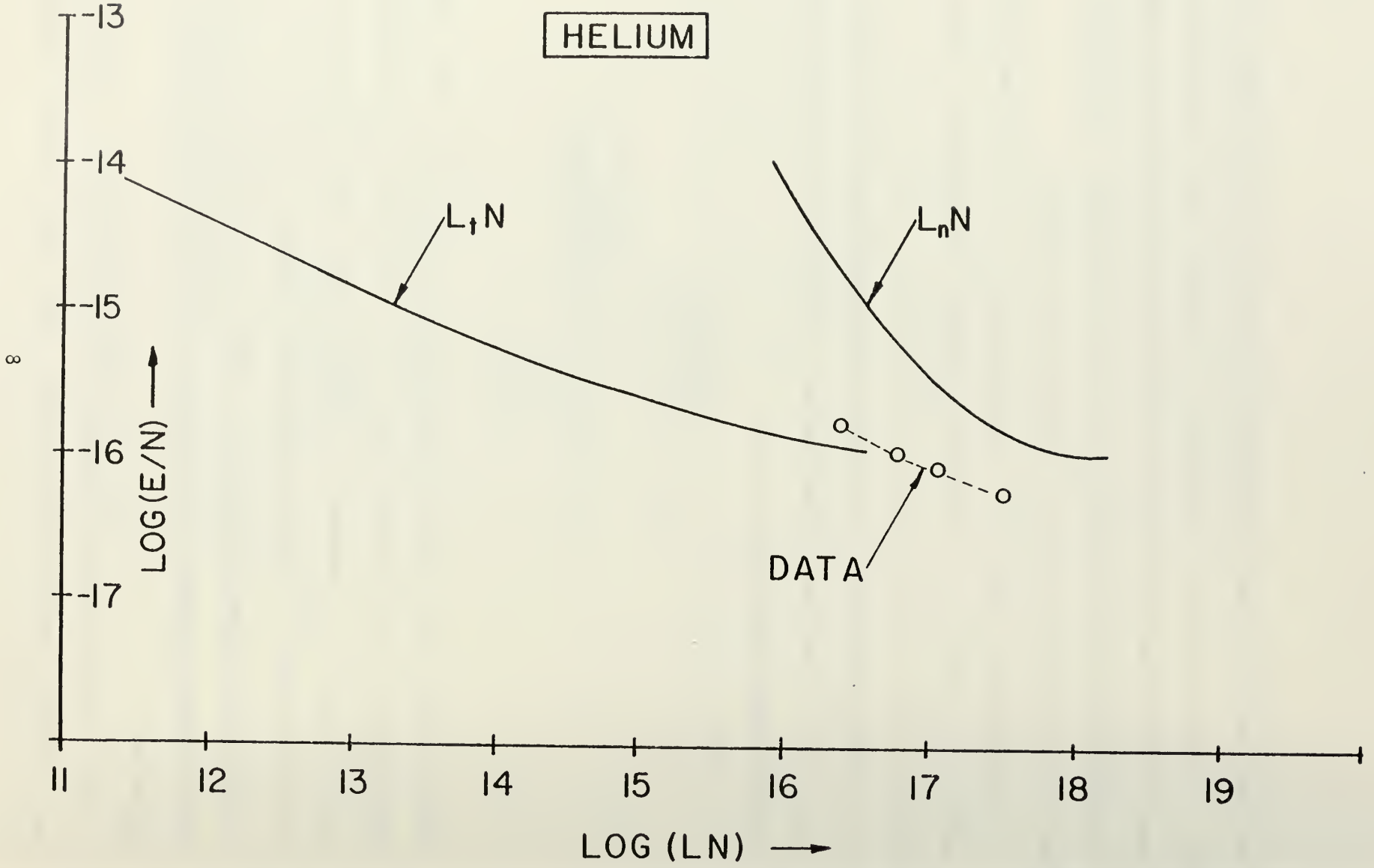
In the light of Eqns. (10), it is interesting to look at some actual values of L_n and L_T. Particularly extensive information for helium is reported in Ref. 16, and this is used in estimating the values shown in Fig. 2, which are limited to the range of applicability of the equations in Ref. 16. An appropriate form for the coefficients is obtained by introducing the gas density N_o,

$$L_n N_o = \sqrt{\frac{T_e (D_a N_o)}{2 T_o (\nu_i / N_o)}} \tag{11}$$

$$L_T N_o = (E/N_o)^{-1} \sqrt{\frac{T_o K_o \alpha}{e(\mu_{eo} N_o)(\nu_i / N_o)}} \tag{12}$$

The above forms show "similarity" since LN is a function of E/N only. Examining Fig. 2, it is evident that L_nN is not equal to L_TN, certainly not for large E/N, whereas L_TN follows closely data reported for helium in the literature[12,13] (the fact that we are looking at parallel plates and the data is for tubes is immaterial). Since in a discharge tube n and T share a common diameter, it is necessary to reconcile the discrepancy of the two curves.

It is possible, that the assumption[14] which brings T_e into Eqn. (1) and, therefore, into Eqn (11) is invalid. But, even if we were to eliminate


HELIUM
-13
-14
-15
-16
-17
LOG(E/N) ⟶
$L_t N$
$L_n N$
DATA
11
12
13
14
15
16
17
18
19
LOG (LN) ⟶

the factor T_e/T_o from the equation, the curves would diverge at the higher values of E/N. Constriction of the positive column is another possibility but in this case L_T would be expected to be greater than L_n for any given E/N and just the opposite seems to be true. Uncertainties in the estimate of α, the recombination coefficient, would affect the value of L_T but this estimate is the one that follows nicely the experimental values.

The method of reconciliation of the two lengths proposed here is based on the alternative that the equations previously written are incomplete and that, in particular, ν_i in Eqns. (1) and (7) should be replaced by a parameter (C_2) which makes up for the difference in the two curves. This matter is discussed in Appendix A of this report.

Consider now Eqn. (1) with constant coefficients. We proceed here to nondimensionalize it in a conventional way

$$- D_a\, n'' = \nu_i\, n - \alpha n^2 \tag{13}$$

$$n'(o) = 0 \quad \text{and} \quad n(\pm a) = 0$$

Let $n(o) = n_o$ and $\varepsilon = \alpha n_o/\nu_i$

Hence we get

$$- \hat{n}'' = \hat{n} - \varepsilon \hat{n}^2 \tag{14}$$

$$\hat{n}'(o) = 0 \quad \text{and} \quad \hat{n}(o) = 1$$

$$\hat{n}(\pm a/\sqrt{D_a/\nu_i}) = 0$$

Now the Shottky solution is one for which $\varepsilon = 0$, namely,

$$- \hat{n}'' = \hat{n} \tag{15}$$

$$\text{and} \qquad n = n_o \cos\left(\tfrac{\pi}{2}\, y/a\right) \tag{16}$$

It can be seen therefore that the characteristic length L should be

$$L = \sqrt{\frac{D_a}{\nu_i}} \equiv \frac{\bar{a}}{\pi/2} \tag{17}$$

On the other hand, for the equilibrium limit $\varepsilon = 1$ and the equation

$$-\hat{n}'' = \hat{n} - \hat{n}^2 \tag{18}$$

is of the "Boundary Layer" type, i.e., it is not an eigenvalue problem, see Ref. 14 pp 149-51 for the equivalent problem with three-body recombination. The solution to Eqn. 18 in an implicit form is

$$y = \log \left\{ \frac{\sqrt{\frac{3}{2}} + (\hat{n} + \frac{1}{2})^{1/2}}{\sqrt{\frac{3}{2}} - (\hat{n} + \frac{1}{2})^{1/2}} \right\} - 1.317 \tag{19}$$

or explicitly

$$n = \frac{3}{2} n_o \left[\frac{1 - e^{-(y + 1.317)}}{1 + e^{-(y + 1.317)}} \right]^2 - \frac{n_o}{2} \tag{20}$$

which is indeed of the boundary-layer type. The intermediate cases, $0 < \varepsilon < 1$, are represented by elliptic functions, and this matter is discussed in Appendix B of this report.

IV. CONCLUSIONS

Some deficiencies of the Schottky solution have been re-examined with the intent of establishing a more useful solution for the description of the transition to the glow collapse. At least two major, recent attempts in the literature (Refs. 16 and 17) base the steady solution for the charge profile on the Schottky result and these attempts have not succeeded in presenting a complete formulation of the problem. The Technion approach[10] is to base the perturbation on a physically complete solution of the ambipolar charge profile. This profile must include the ingredients of constriction and of the elevation of the electron temperature.

As shown in Appendix A, the characteristic length scales for ambipolar diffusion and heat diffusion (conduction only) may be reconciled if ν_{io} in Eqn. (7) is replaced by C_2 where $C_2 > \nu_{io}$ and C_2 is a function of E/N (i.e., the dependance given by the difference of the two curves in Fig. 2). What appeared to be a bonus, namely, an insight into the mathematical description of striations, needs further study in order for it to make more physical sense. The eigenvalue nature of the diffusion equation is not changed by the recombination term except at the limit when $\varepsilon = 1.0$.

APPENDIX A - PROBLEM REFORMULATION

Equation (1) in the text presupposes steady flow along with $n = n(y)$. Let us relax these suppositions and go back to the species flux equations which may be written as

$$\frac{\partial n}{\partial t} + \nabla \cdot \left[n(-\mu_e \bar{E} - \frac{D_e}{n} \nabla n) \right] = \dot{n} \tag{A1}$$

$$\frac{\partial n}{\partial t} + \nabla \cdot \left[n(\mu_i \bar{E} - \frac{D_i}{n} \nabla n) \right] = \dot{n} \tag{A2}$$

Combining and eliminating the electric field we get the ambipolar diffusion equation in the usual manner which for constant coefficients is

$$\partial n/\partial t - D_a \nabla^2 n = \dot{n} \tag{A3}$$

$$\text{or} \qquad \partial n/\partial t - D_a \left(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2} \right) = \nu_i n - \alpha n^2 \tag{A4}$$

The above assumes two-body recombination as well. Now, using Eqn. (A3) to eliminate $\nabla^2 n$ from Eqn. (A2) yields

$$\frac{\partial n}{\partial t} + \frac{\mu_i \nabla \cdot n\bar{E}}{1 - D_i/D_a} = \dot{n} \tag{A5}$$

But since $\nabla \cdot \bar{E} = 0$ in the ambipolar region

$$\frac{\partial n}{\partial t} + \frac{\mu_i}{1 - D_i/D_a} \bar{E} \cdot \nabla n = \dot{n} \tag{A6}$$

$$\text{or} \qquad \partial n/\partial t + \left(\frac{\mu_i}{1 - D_i/D_a} \right) \left(E_L \frac{\partial n}{\partial x} + E_y \frac{\partial n}{\partial y} \right) = \nu_i n - \alpha n^2 \tag{A7}$$

Let us work with Eqns. (A4) and (A7) in linear form, i.e., neglecting recombination. If

$$n(x,y,t) = f(y)\ g(x,t) \tag{A8}$$

Then, substituting into Eqn. (A4)

$$f\ g_t - D_a\ (f g_{xx} + g f'') = \nu_i f g - \ldots \tag{A9}$$

where $f'' = \dfrac{d^2 f}{dy^2}$ $\quad g_{xx} = \dfrac{\partial^2 g}{\partial x^2}$ and $g_t = \partial g/\partial t$

Dividing by fg both sides of (A9)

$$1/g\ (g_t - D_a\ g_{xx}) - \frac{D_a}{f} f'' = \nu_i \tag{A10}$$

At a given E/N, ν_i is constant so we may presume that

$$1/g\ (g_t - D_a\ g_{xx}) = \text{constant (or } - C_1) \tag{A11}$$

and

$$-\frac{D_a}{f} f'' = \text{constant (or } C_2) \tag{A12}$$

with

$$-C_1 + C_2 = \nu_i \tag{A13}$$

The constants C_1 and C_2 are positive and defined so that $C_2 \geq \nu_i$ as required by Eqn. (A12), which now replaces Eqn. (6) since

$$-D_a\ f'' = f\ C_2 \tag{A14}$$

and

$$L_f \equiv \sqrt{D_a/C_2}$$

Equation (A14) is the new Schottky equation. The constant C_1 appears in the companion partial differential equation

$$g_t - D_a\, g_{xx} + C_1\, g = 0 \tag{A15}$$

Now look at Eq. (A7). At the centerline $\partial n/\partial y = 0$ and $f(o) = f_o$ so that

$$\text{at } y = 0 \qquad \frac{\partial n}{\partial t} + \left(\frac{\mu_i}{1 - D_i/D_a}\right) E_L \frac{\partial n}{\partial x} = \nu_i\, n - \ldots \tag{A16}$$

Introducing $n = fg$ we obtain

$$g_t\, f_o + \left(\frac{\mu_i}{1 - D_i/D_a}\right) E_L\, g_x\, f_o = \nu_i\, f_o g \tag{A17}$$

We can cancel f_o in the linear equation above and obtain an equation in which no function of y appears

$$g_t + \left(\frac{\mu_i}{1 - D_i/D_a}\right) E_L\, g_x = \nu_i g \tag{A18}$$

At this point it is worthwhile to eliminate g_t between Eqns. (A18) and (A15) and obtain

$$D_a\, g_{xx} - C_1\, g = \nu_i g - \left(\frac{\mu_i}{1 - D_i/D_a}\right) E_L g_x$$

$$D_a\, g_{xx} + \frac{\mu_i E_L}{(1 - D_i/D_a)}\, g_x = (\nu_i + C_1) g$$

or

$$g_{xx} + \left(\frac{\mu_i E_L}{D_a - D_i}\right) g_x = \frac{C_2}{D_a}\, g \tag{A19}$$

Equation (A19) can be solved with

$$g = T(t)\ e^{\alpha x} \tag{A20}$$

where α is found from

$$\alpha = \frac{-\xi \pm \sqrt{\xi^2 + 4C_2/D_a}}{2} \qquad \xi = \frac{\mu_i E_L}{D_a - D_i} \tag{A21}$$

Note that $D_i < D_a$ and that ξ is anyway a positive parameter. Equation (A21) indicates that the roots are real and there is no periodic behavior in x -- that striations are not to be found with this description.

In order for n to oscillate in x, we must have a situation in which new constants C_1^* and C_2^* are introduced such that

$$C_1^* - C_2^* = \nu_i \tag{A22}$$

$$g_t - D_a\, g_{xx} = C_1^*\, g \tag{A23}$$

and

$$-D_a\, f'' = -C_2^*\, f \tag{A24}$$

Here, Eqn. (A24) will result in (non-periodic) hyperbolic functions in y and Eqn. (A23) in oscillatory functions in x provided that

$$\frac{4C_2^*}{D_a} > \left(\frac{\mu_i E_L}{D_a - D_i}\right)^2 \tag{A25}$$

As before we would like to have

$$C_1^* > \nu_i \quad (C_2^* = C_1^* - \nu_i) \tag{A26}$$

so as to match L_n to L_T.

The solution to Eqn. (A24) does not represent the usual diffusion controlled situation because in order to meet the symmetry condition at the centerline we must give up the physically sensible boundary conditions at the wall. Perhaps variable coefficients and the inclusion of recombination can remedy this problem with the boundary conditions. Or perhaps we should be willing to give up axial symmetry and consider some sort of contraction asymmetry which would likely have to include some reconbination because of the higher pressures. This, however shall not be pursued in this report.

APPENDIX B - INFLUENCE OF TWO-BODY RECOMBINATION OF SCHOTTKY SOLUTION

The diffusion equation given in the main text (Eqn. 14) is not suitable for elliptic funciton representation, therefore, a different form is developed below.

The diffusion equation with constant coefficients is (without consideration of whether C_2 shall be used)

$$- D_a \frac{d^2 n}{dy^2} = \nu_i n - \alpha n^2 \tag{B1}$$

Defining a stationary state $\nu_i n^* - \alpha n^{*2} = 0$, or

$$n^* = \nu_i / \alpha \tag{b2}$$

We can nondimensionalize the equation with

$$\hat{n} = n/n^* \quad \text{and} \quad \hat{y} = y/L_n$$

Obtaining

$$- \frac{d^2 \hat{n}}{d\hat{y}^2} = \hat{n} - \hat{n}^2 \tag{B3}$$

$$n'(o) = 0 \quad \text{and} \quad n = \beta \ , \quad 0 \leq \beta \leq 1$$

$$n(a/L_n) = 0$$

It is clear that the boundary condition at the centerline makes up for the disappearance of ε in Eqn. (B3). Now Eqn. (B3) can be integrated once to yield

$$\frac{d\hat{n}}{d\hat{y}} = \sqrt{\left(\frac{2}{3}\hat{n}^3 - \hat{n}^2\right) - \left(\frac{2}{3}\beta^3 - \beta^2\right)} \tag{B4}$$

or
$$\sqrt{\frac{3}{2}} \int_{\hat{n}}^{\beta} \frac{d\hat{n}}{\sqrt{n^3 - \frac{3}{2} n^2 - \beta^3 + \frac{3}{2} \beta^2}} = \int_{\hat{y}}^{\infty} d\hat{y} \qquad \text{(B5)}$$

The left-hand-side of the above equation can be represented in terms of elliptic functions. Let

$$(\hat{n}^3 - \frac{3}{2} \hat{n}^2 - \beta^2 + \frac{3}{2} \beta^2) = (n - a)(n - b)(n - c) \qquad \text{(B6)}$$

therefore

$$\left.\begin{aligned} a &= \frac{\left(\frac{3}{2} - \beta\right) + \sqrt{\left(\frac{3}{2} - \beta\right)\left(\frac{3}{2} + 3\beta\right)}}{2} \\ b &= \beta \qquad (0 \leq \beta \leq 1) \\ c &= \frac{\left(\frac{3}{2} - \beta\right) - \sqrt{\left(\frac{3}{2} - \beta\right)\left(\frac{3}{2} + 3\beta\right)}}{2} \end{aligned}\right\} \qquad \text{(B7)}$$

where $a \geq b \geq \hat{n} \geq c$

With information on a, b, and c we can enter any table for elliptic functions (see for example Ref. 18 or 19). Using Byrd and Friedman notation, formula 234.00, we have

$$\int_{\hat{n}}^{b} \frac{d\hat{n}}{\sqrt{(a - \hat{n})(b - \hat{n})(\hat{n} - c)}} = gF(\varphi, k) \qquad \text{(B8)}$$

where

$$g = \frac{2}{\sqrt{a - c}}, \qquad k^2 = \frac{b - c}{a - c}$$

and $$\varphi = \sin^{-1} \sqrt{\frac{(a - c)(b - \hat{n})}{(b - c)(a - \hat{n})}}$$

$F(\varphi,k)$ is tabulated. Note that for $\beta = 1$, we get a double root $(a = b)$ and the equation is not an elliptic integral (though its solution can still be found from the Tables). When $\beta = 1$ we get the boundary-layer behavior described by Eqn. (19) or (20) in the main text.

Except for $\beta = 1$, Eqn. B3 represents an eigenvalue problem. In order to check this, we solve for the value of the integral at the wall where $\hat{n} = 0$. Thus

$$\varphi = \sin^{-1} \sqrt{\frac{(a - c)b}{(b - c)a}}$$

For very small values of β, we obtain the cosine solution, namely,

$$\lambda = (\pi/2)^2 \qquad \text{where} \qquad L_n = a/\lambda^{1/2}$$

As β increases, we obtain progressively larger values until at $\beta = 1$ we get $\lambda = \infty$). Figure B1 shows this behavior for the eigenvalue. The results shown in Fig. B1 proved identical to results from a computer program at the Technion (for the combined diffusion and heat dissipation).

Increasingly higher degrees of recombination flatten out the density profile at the center and increase its slope at the walls. The radial electric field is thus enhanced. The variable coefficient solution is not anticipated to change the above results in any major way.

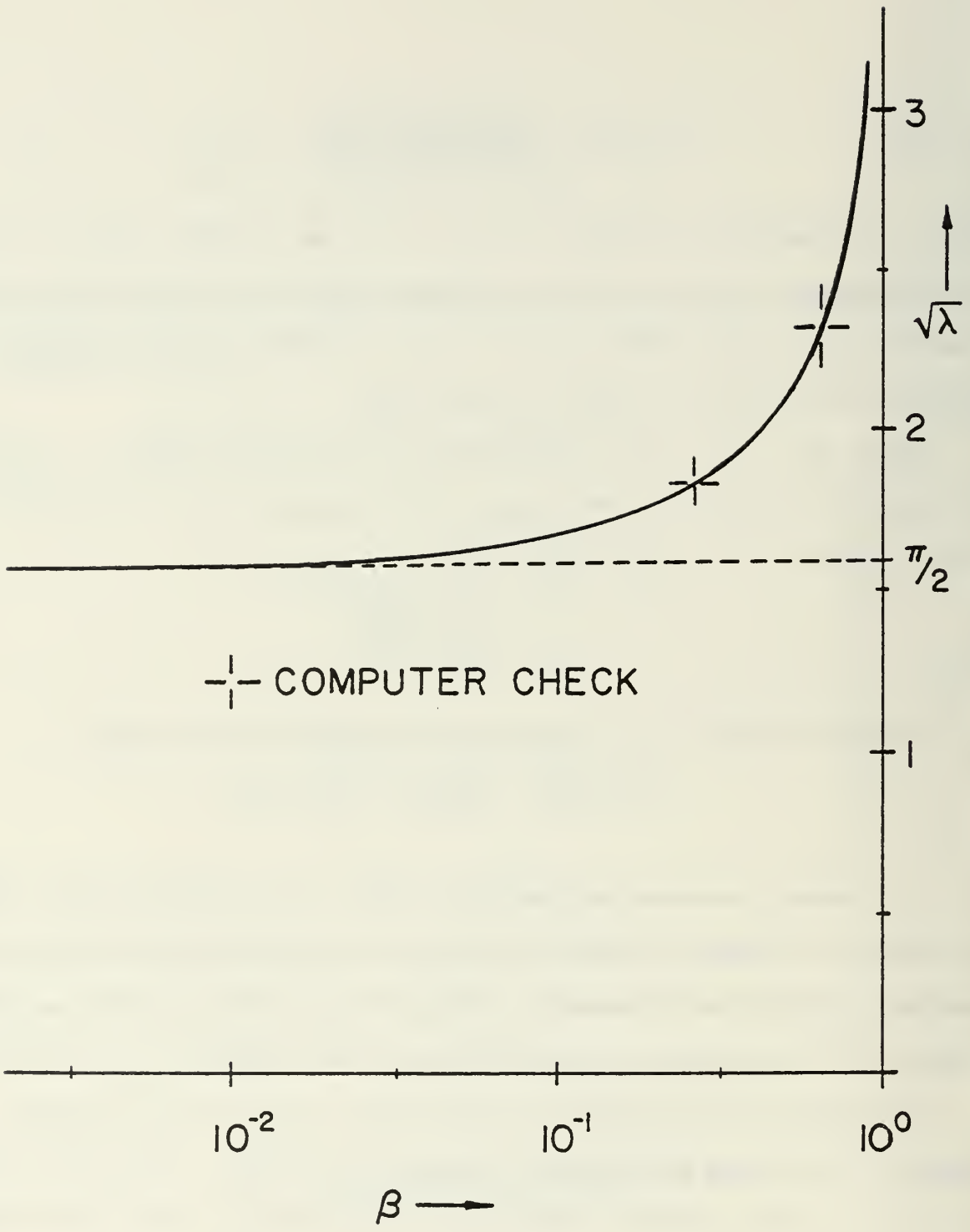


Figure B1. Eigenvalue for Increasing Degrees of Recombination

REFERENCES

1. Biblarz, O. and Nelson, R. J., "Turbulence Effects on an Ambient Pressure Discharge," JAP, 45, 633 (1974). Also Biblarz, O., Barto, J. L., and Post, H. A., "Gas Dynamic Effects on Diffuse Electrical Discharge in Air," Israel Journal of Technology, 15, No. 1-2 (1977).

2. Shwartz, J. and Wasserstrom, E., "The Role of Gas Flow and Turbulence in Electric Discharge Lasers," The Israel Journal of Technology, 13, 122 (1975).

3. Eckberth, A. C. and Owen, F. S., "Flow Conditioning in Electric Discharge Convection Lasers,""Rev. of Sci. Instruments, 43, 995 (1972).

4. McErlean, D. P. and Ortwerth, P. J., "Aerodynamically Stabilized Electric Discharges," AFAPL-TR-74-1 dated February 1974.

5. Wigham, W. C. and Wiegard, W. J.," Influence of Fluid Dynamic Phenomena on the Occurrence of Constriction in CW Convection Laser Discharges," Appl. Phys. Lett., 26, 554 (1975). Also Appl. Phys. Lett., 25, 633 (1974).

6. Wasserstrom, E., Crispin, Y., Rom, J., and Shwartz, J., "The Interaction Between Electrical Discharges and Gas Flow," AIAA paper No. 76-315, San Diego, 1976 (to appear in JAP).

7. Forte, N., "Experimental Investigation of the Effect of Swirling Flow on the Stability of the Discharge and on the Power of Electrical Lasers," M.Sc. Thesis Aero. Engr., Technion, Haifa (1977).

8. Crispin, Y., "The Interaction Between Electrical Discharges and Gas Flow," MSC Thesis Aero. Engr. Technion, Haifa (1976).

9. Shwartz, J. and Lavie, Y., "The Effects of Turbulence on a Weakly Ionized Plasma Column," AIAA J., 13. 647 (1975).

10. Wasserstrom, E., Shwartz, J., Striker, J., and Biblarz, O., "Aerodynamic Stabilization of Gaseous Discharges," Proposal to the United States - Israel Binational Science Foundation (BSF), dated 19 June 1977.

11. Wasserstrom, E., Crispin, Y., and Biblarz, O., "Stabilization of Glow Discharges by Supersonic Flow," Presented at 30th Gaseous Electronics Conference, Palo Alto, (Oct. 1977). Abs. in Bulletin of APS.

12. von Engle, A., Ionized Gases, second ed. pg. 245, The Clarendon Press, Oxford (1965).

13. Francis, G., "The Glow Discharge at Low Pressure," Encyclopedia of Physics, 22, Springer-Verlag, Berlin (1956).

14. Mitchner, M. and Kruger, C. H., Partially Ionized Gases, pp 146-155, Wiley Series on Plasma Physics, New York (1973).

15. Kline, S. J., Similitude and Approximation Theory, McGraw-Hill, New York (1965).

16. Jaeger, E. F., Oster, L., and Phelps, A. V., "Growth of Thermal Constrictions in a Weakly Ionized Gas Discharge in Helium," Phys. Fluids, 19, 819 (1976).

17. Ecker, G., Kroll, W., and Zoller, "Thermal Instability of the Plasma Column," Phys. Fluids, 7, 2001 (1964). Also, Phys. Fluids, 10 1037 (1967).

18. Byrd, P. F., and Friedman, M. D., Handbook of Elliptic Integrals for Engineers and Scientists, Second Ed., Springer-Verlag, New York (1971).

19. Milne-Thompson, L. M., "Elliptic Integrals," Handbook of Mathematical Functions, Abramowitz, M. and Stegun, I, Editors, NBS, (June 1964).

LIST OF FIGURES

DISTRIBUTION LIST

		No. of Copies
1.	Defense Documentation Center Cameron Station Alexandria, VA 22314	2
2.	Library Naval Postgraduate School Monterey, CA 93940	2
3.	Dean of Research Naval Postgraduate School Monterey, CA 93940	2
4.	Chairman Department of Aeronautics Naval Postgraduate School Monterey, CA 93940	1
5.	Professor A. E. Fuhs Department of Mechanical Engineering Naval Postgraduate School Monterey, CA 93940	1
6.	Professor A. W. Cooper Department of Physics and Chemistry Naval Postgraduate School Monterey, CA 93940	1
7.	Professor Oscar Biblarz Department of Aeronautics Naval Postgraduate School Monterey, CA 93940	5
8.	Professor M. Mitchner Mechanical Engineering Department Stanford University Stanford, CA 94305	1
9.	Dr. Alan Garscadden AFAPL/POD High Power Branch AF Aero Propulsion Laboratory Building 450/Room D101 Wright-Patterson AFB, OH 45433	1
10.	Dr. A. V. Phelps JILA Boulder, CO 80309	1

No. of Copies

11. Professor B. E. Cherrington — 1
Gaseous Electronics Laboratory
University of Illinois
Urbana, IL 61801

12. Professor J. Shwartz — 1
Department of Aeronauticsl Engineering
Israel Institute of Technology
Technion City, Haifa 32000 ISRAEL

13. Professor E. Wasserstrom — 1
Department of Aeronautical Engineering
Israel Institute of Technology
Technion City, Haifa 32000 ISRAEL

14. Mr. Y. Crispin — 1
Department of Aeronautical Engineering
Israel Institute of Technology
Technion City, Haifa 32000 ISRAEL

15. Commander — 1
Naval Air System Command
Department of the Navy
Attn: Dr. H. R. Rosenwasser, Code AIR 310C
Washington, DC 20360

16. Mr. John A. Satkowski — 1
Office of Naval Research
Power Program, Code 473
Washington, DC 20360

Title: Shaped Charges For Neutralization of Buried Mines

Investigator: A. E. Fuhs, Professor of Mechanical Engineering

Sponsor: Foundation Research Program (6.2)

Objective: To develop a technique to neutralize buried mines. The technique is to be used by EOD officers and crew.

Summary: Mines may be buried at depths from a few inches to 6-8 feet. The buried mines are detected and located by magnetic means. Error in location may be comparable to buried depth. Some means of placing an explosive charge near the mine is required. A shaped charge can open a large hole through which a charge can be rammed.

An assembly of charges was designed, built, and tested. The assembly of charges consisted of a 2-inch diameter shaped charge with ½ lb HE located on top of a cylinder containing liquid explosive. The cylinder had a coaxial tube which allowed passage of the metal jet from the shaped charge. It was found that the liquid explosive spray lined the hole; it was also found that the cylinder for the liquid explosive was rammed down the hole. As a result of these early tests, a series of tests is being planned using solid explosive located below the shaped charge. A time delay detonator will be used for the solid explosive.

Publication: (See supporting information that follows)

NAVAL POSTGRADUATE SCHOOL

FOUNDATION PROJECT ON UNDERWATER SHAPED CHARGES

Principal Investigator

Professor A. E. Fuhs

Students

LT Richard Funk
LT Larry Kryske
LT Dana Vugteveen

As the result of Foundation funding and subsequent additional funding from EODF, a series of experiments were conducted at EODF Indian Head, MD. The attached material describes the results. For FY 78 additional funding was obtained from EODF. It should be noted that EODF supports the tests with their own manpower.

The students who have worked or currently are working on the project are as follows:

LT Richard Funk, USN, LT Larry Kryske, USN and LT Dana Vugteveen, USN.

The attached material is:

1. Abstract from LT Funk's Thesis.
2. Invention disclosure on explosive injection by shaped charges.
3. List of contacts at other research and development organizations.
4. Sketch of shock attenuator for mine neutralization weapon.
5. Description of LT Funk's test at EODF.
6. Test plan for tests on 22-27 May 1978.

1.

EXPLOSIVE ORDNANCE DISPOSAL MINEHUNTING SYSTEM AND MINE NEUTRALIZATION USING SHAPED CHARGE FOLLOW-THROUGH TECHNIQUE(U)

A brief historical review of mine warfare is given as background. The NATO First Scenario serves as the threat scenario which provides the impetus for an improved mine countermeasures technique. The experience of clearing Haiphong harbor and the Suez Canal showed that minesweeping of any sort is a difficult, tedious, and lengthy process totally devoid of any glamour. The prerequisite of any mine countermeasures operation is training. The Navy Explosive Ordnance Disposal community, which is tasked with determining the extent of mine fields, has the equipment and expertise to initiate this new technique in mine countermeasures. This technique entails locating and cataloging all bottom objects in United States ports. To counter a mine threat these surveys would have to be conducted annually. Experiments were conducted which demonstrated the praciticality of using the shaped charge follow-through concept to neutralize buried ordnance.

2.

INVENTION DISCLOSURE ON EXPLOSIVE INJECTION BY SHAPED CHARGES

INVENTION DISCLOSURE ON

EXPLOSIVE INJECTION BY SHAPED CHARGE

by

Richard Funk and Allen Fuhs

There are several military tasks for which the ability to inject an explosiv desirable. The military applications are described later.

Recent tests have shown that either a liquid or solid explosive can be injec through sand or mud a considerable distance by use of a shaped charge. The metal jet from the charge opens the hole for injection. Figure 1 shows a cross section an arrangement for injecting liquid explosive. Figures 2 and 3 show the apparatu the shaped charge, in assembled and disassembled condition. Figures 1 and 3 can compared to identify components.

Figure 4 is the sketch of apparatus for injecting a solid explosive through sand, or mud. The detonator at the top initiates the detonation wave in the shap charge and starts the time sequence leading to detonation of the high explosive b the shaped charge. This will be discussed in more detail shortly in connection w Figure 6.

Consider now the sequence of events during the explosion of the shaped charg liquid explosive (LE). At the moment of initiation of detonation, the apparatus sitting on the surface of the sand; see Figure 5(a). The velocity of the metal j shown in Figure 5(b) is about 10000 m/sec. Figure 5(a) shows the condition when jet has bored a hole of 0.2 m or so. The metal jet causes a shock wave in the sa Notice the piston which forces the LE from the chamber. Motion of the piston relative to the chamber can be ascertained using the reference marks.

In Figure 5(c) the metal jet and slug from the shaped charge have bored a large hold in the sand. The chamber is moving into the hole due primarily to the hot, high-pressure gas formed by the explosion. A spray of liquid explosive develops due to piston motion. In Figure 5(d) the chamber continues to move into the hole bored by the slug and the metal jet. The spray of LE coats the sand.

About a millisecond later the sand begins to collapse as sketched in Figure 5(e). The high pressure gases have expanded reducing the pressure on the top of the chamber. As a result, the spray starts to go in both directions. Depending on the rate of collapse of the sand, the chamber may be trapped part way in the hole. The final condition is shown in Figure 5(f). The dark lines are sand soaked with liquid explosive (LE). Figure 5(e) is based on observations from tests conducted at EODF, Indian Head, MD, July 1977.

Figure 6 shows a sequence of events similar to Figure 5 but with a solid explosive being injected. The injected charge is pushed down the hole by high pressure gases as shown in Figures 6(c) and 6(d). By the time the sand collapses, the injected charge is at the bottom of the hole as can be seen in Figure 6(e). After a delay sufficient to allow complete collapse of the sand, the injected charge is exploded as shown in Figure 6(f).

There are numerous applications which will be summarized below.

Mine Neutralization. Buried mines can be effectively neutralized. In the Navy application, the mines can be buried in mud or sand in a harbor or coastal region. In the USMC application, the mines are of the land variety buried in soil or sand ashore.

Torpedo Warhead. When used in ASW role, the torpedo would strike a submarine. The ballast tank wall would be penetrated followed by the 2 to 3 meters of water inside the tank. The explosive would strike the pressure hull and explode; the pressure hull will have been weakened by the metal jet and slug.

When used against major combatant surface vessels, the antitorpedo water and fuel tanks along the waterline are penetrated. The solid explosive is injected de within the tanks and explodes at that deep location.

Antipersonnel/Antitank Mine. The apparatus shown in Figure 6(a) could be turned upside down and buried 2 to 3 meters in sand, soil, or mud. When detonated the metal jet and slug bore a hole upward. Slightly later, the solid explosive pops up and explodes. Similar mines exist; however, this mine could be buried ver deep (2 to 3 meters). Detection of a plastic, deeply-buried mine would be very difficult.

Anti Bunker. Bunkers frequently consist of a wood or concrete structure with sand overlay. The sand may be a few meters in thickness. The shaped charge would fired horizontally into the sand; the solid explosive would be injected to a locat in contact with wood or concrete. Obviously, the protection offerred by the sand negated.

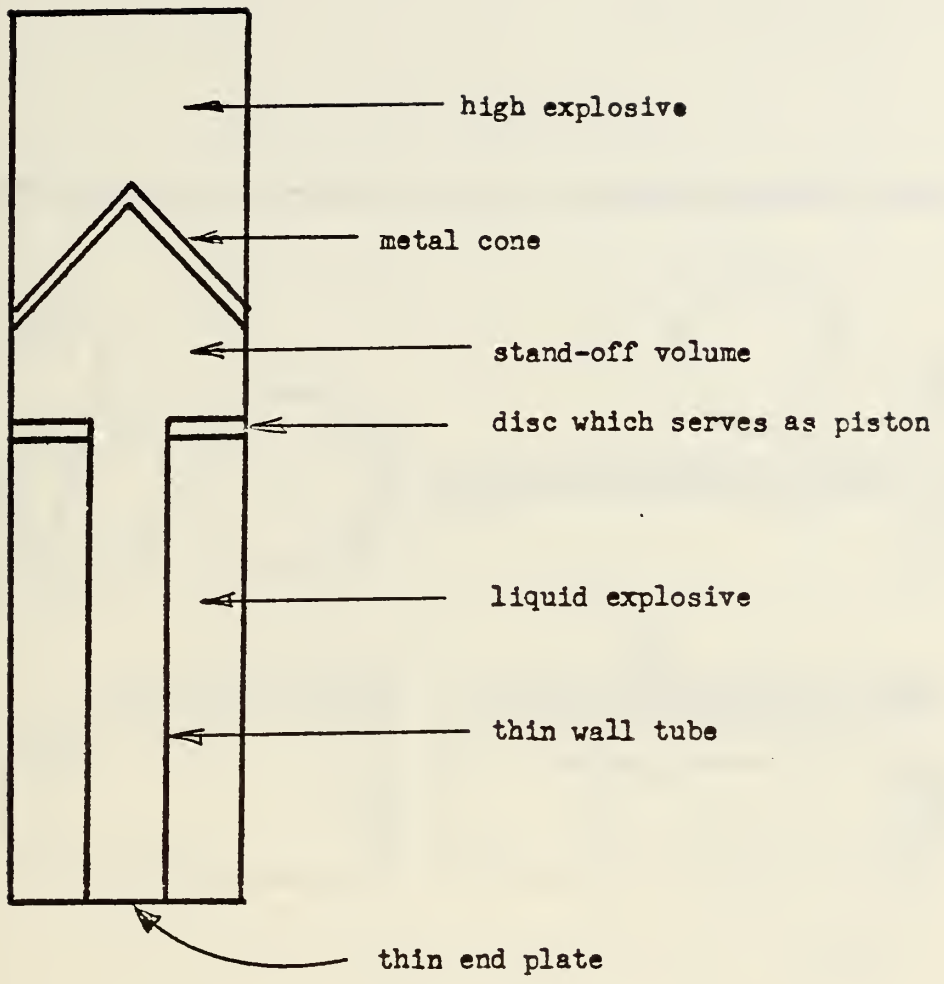


Figure 1. Arrangement for injection of liquid explosive.

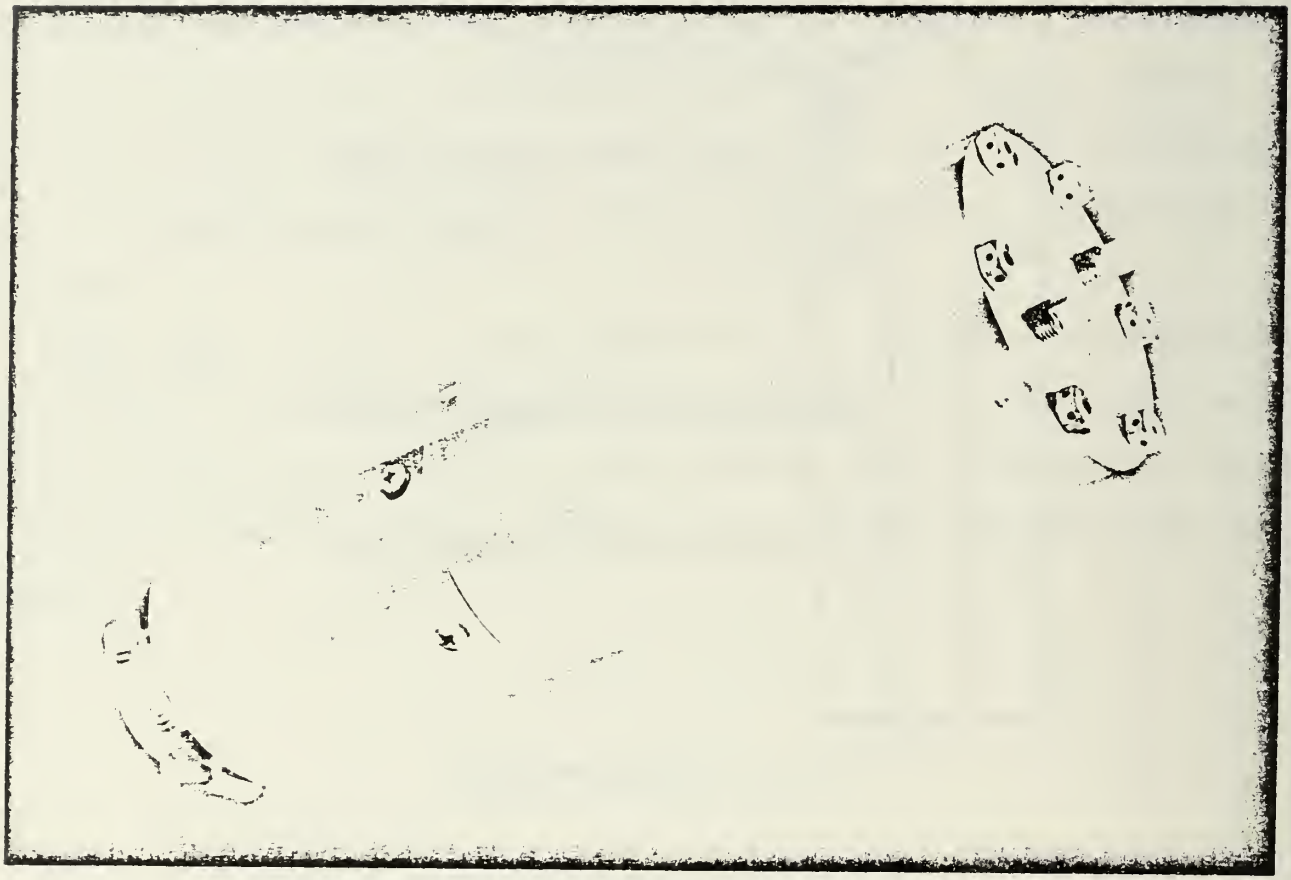


Figure 2. Assembled cylinder for injecting liquid explosive into sand, mud, or soil. Hose clamp on left clamps shaped charge. Two 3/8" pipe plugs are for filling chamber with liquid explosive.

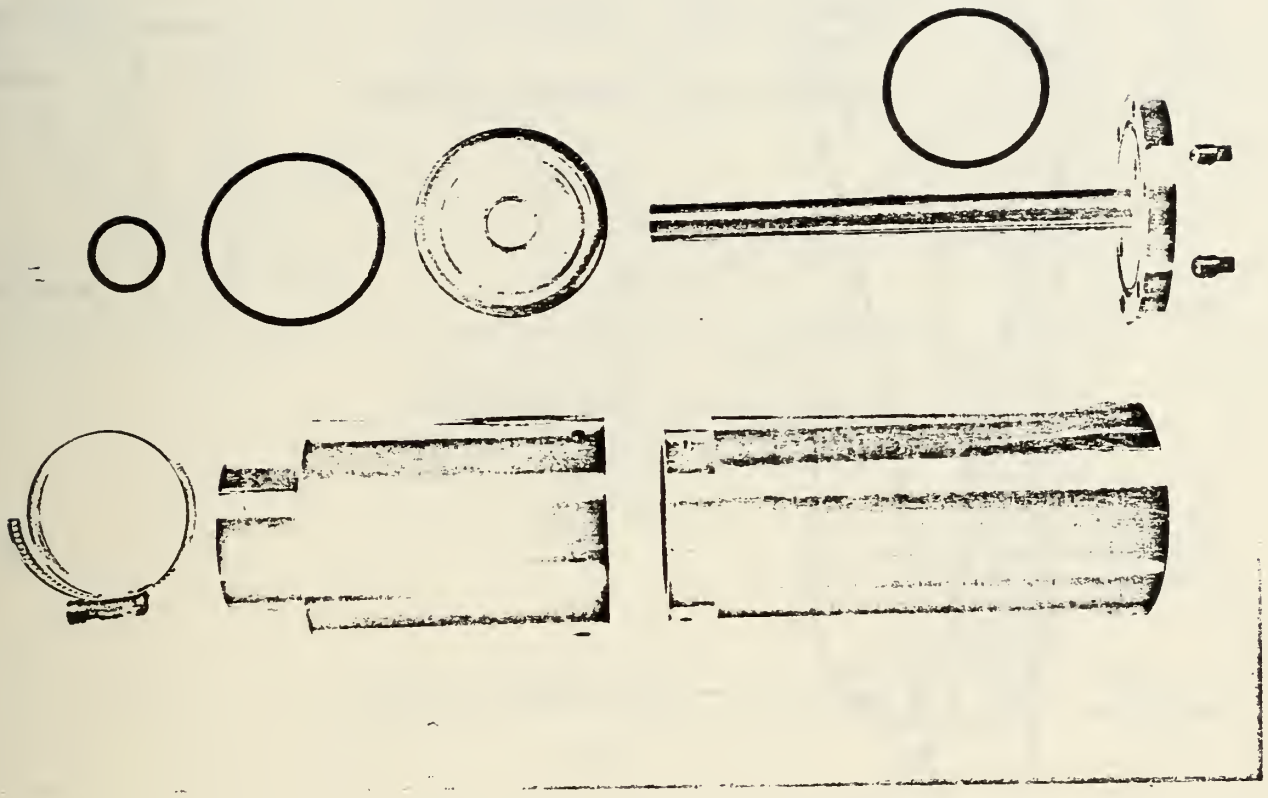


Figure 3. Disassembled cylinder showing components. From left to right across the top are the following: two O-rings, disc which shears and becomes a piston, tube inserted in end plate, an O-ring, and 3/8" pipe filler plugs. From left to right across the bottom are the following: hose clamp, hollow cylinder to provide stand-off space for shaped charge, and hollow cylinder to hold liquid explosive.

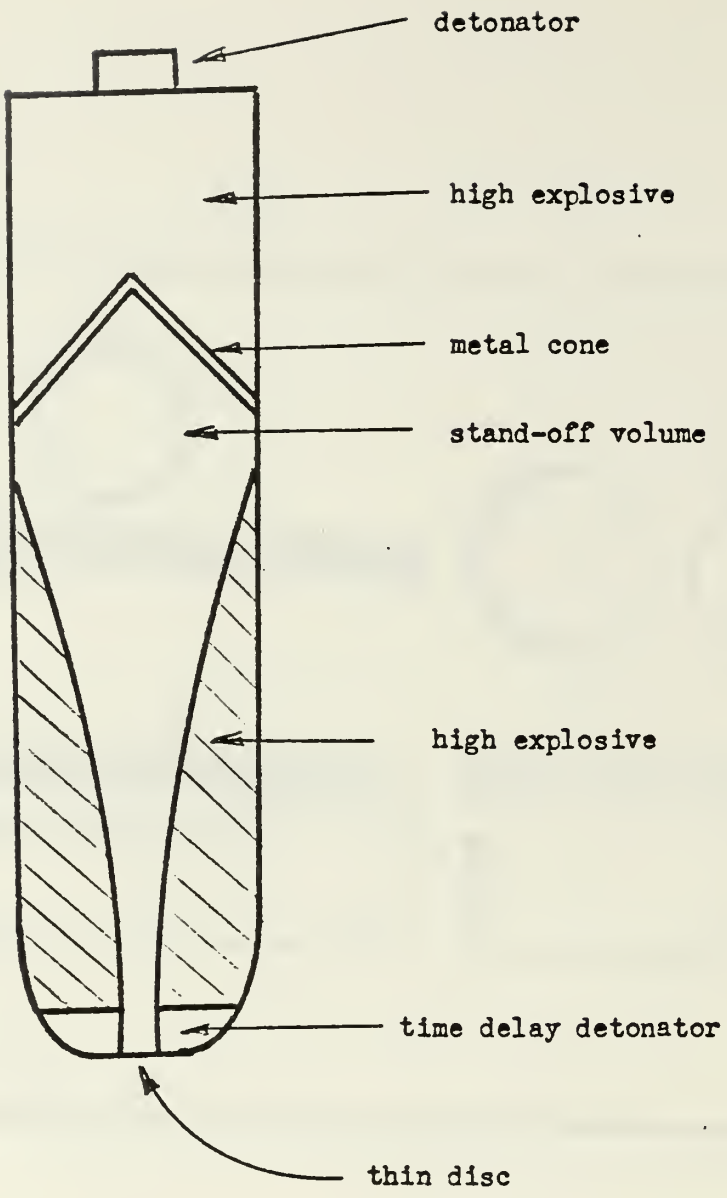
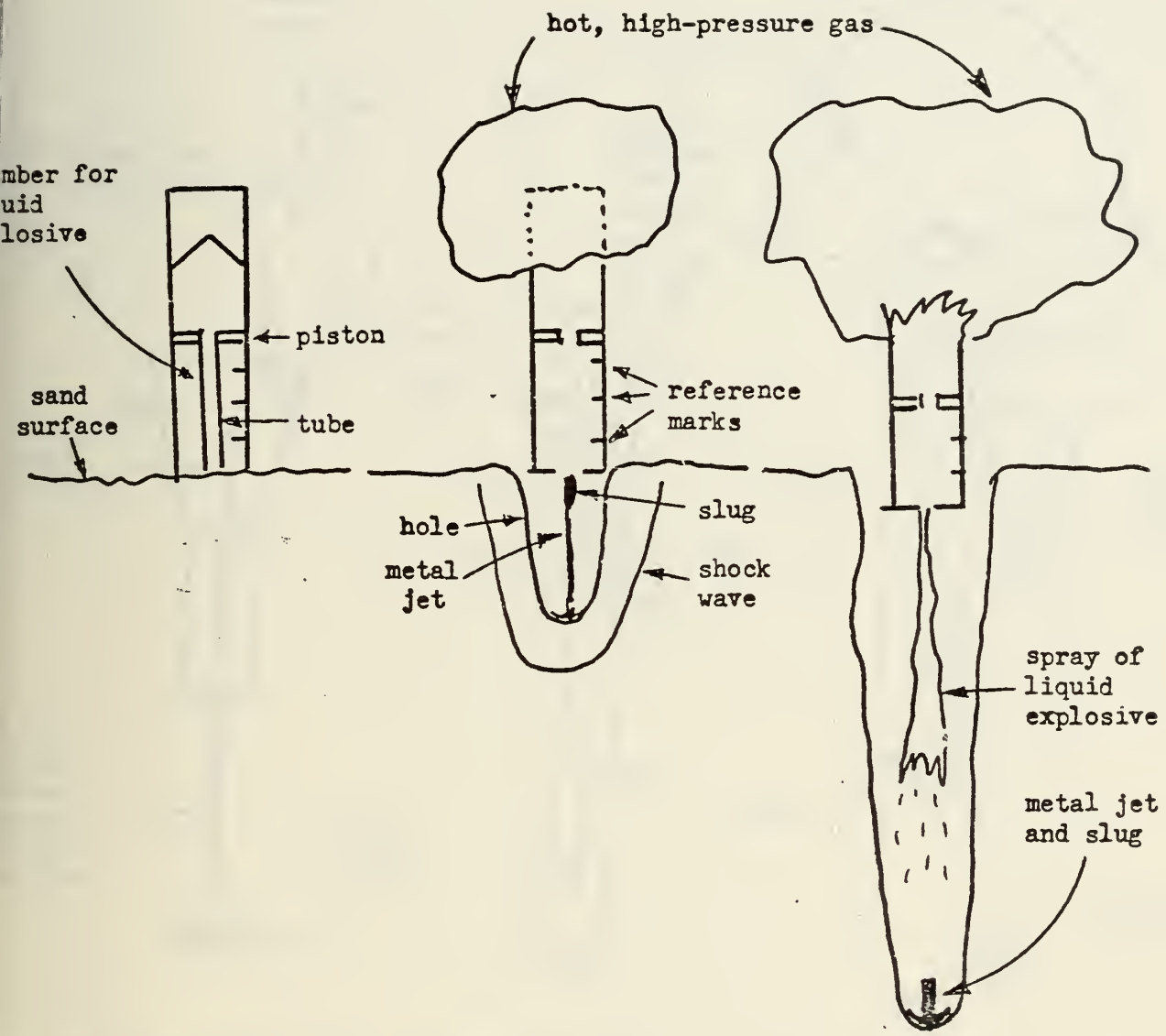


Figure 4. Arrangement for injection of solid explosive.

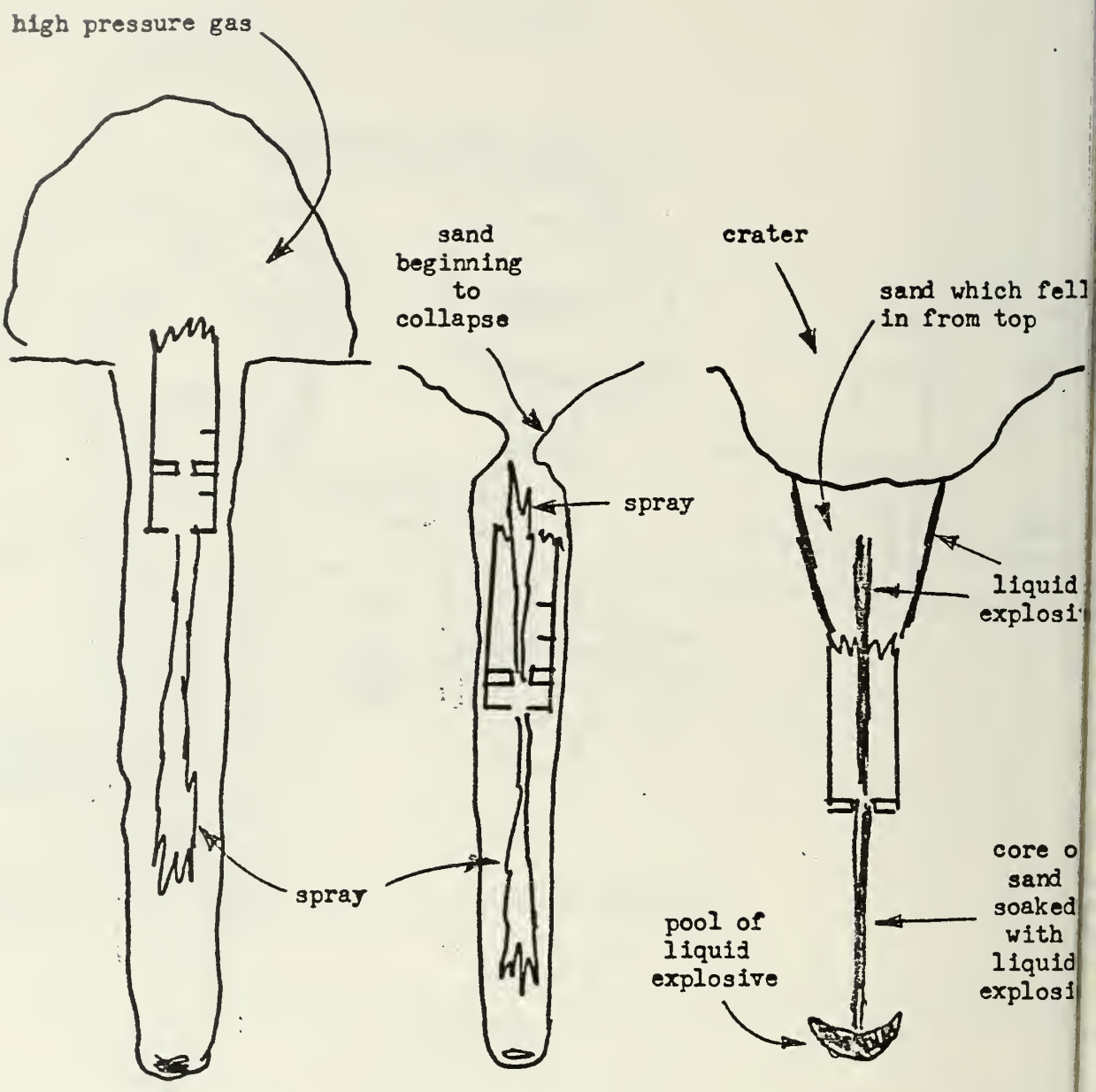


(a) At moment of initiation of detonation

(b) A few 10's of microseconds later

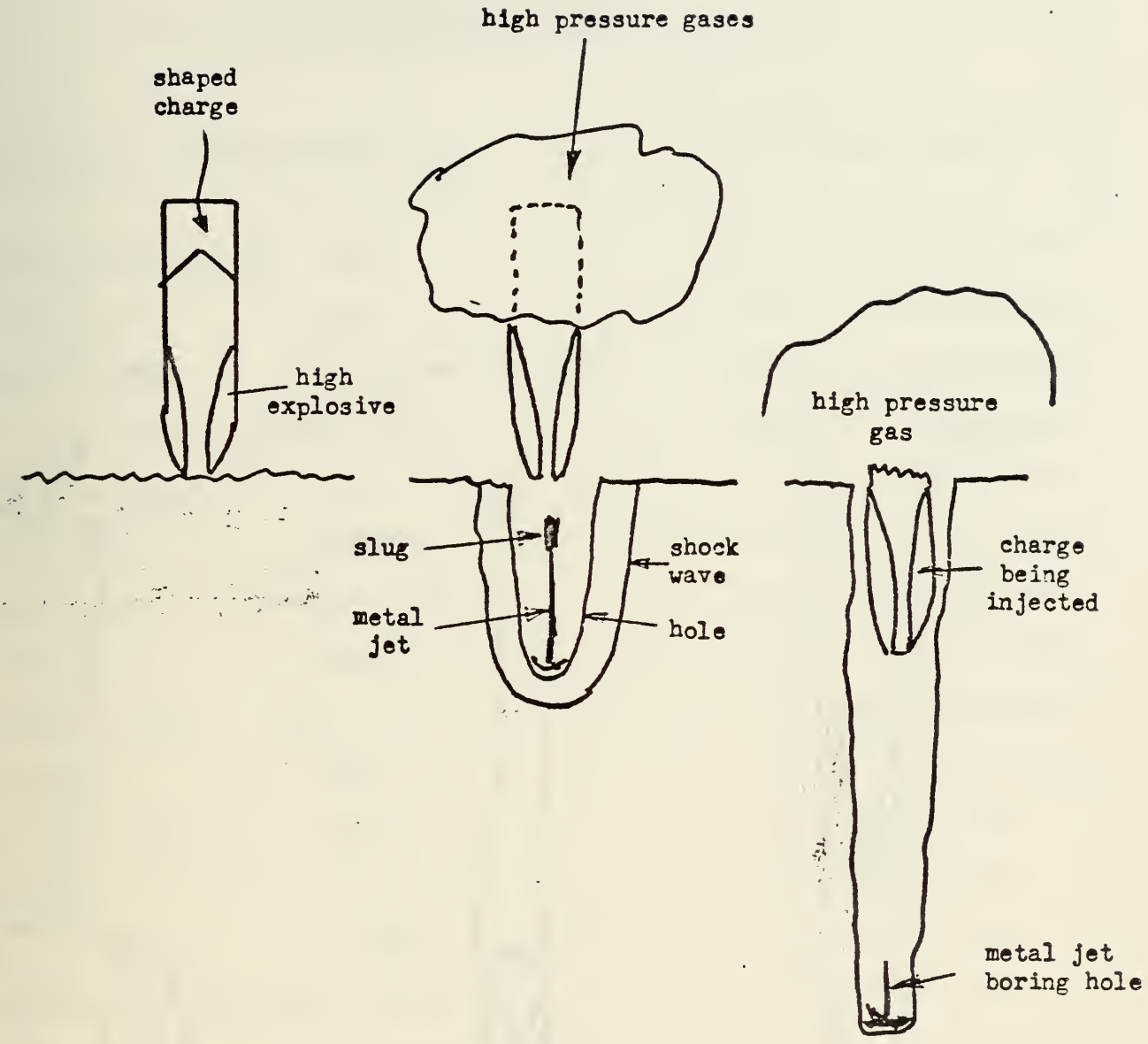
(c) A few 100's of microseconds later

Figure 5. Sequence of events.



(d) A few 100 microseconds (e) Millisecond (f) Final condition

Figure 5 (Continued). Sequence of events.

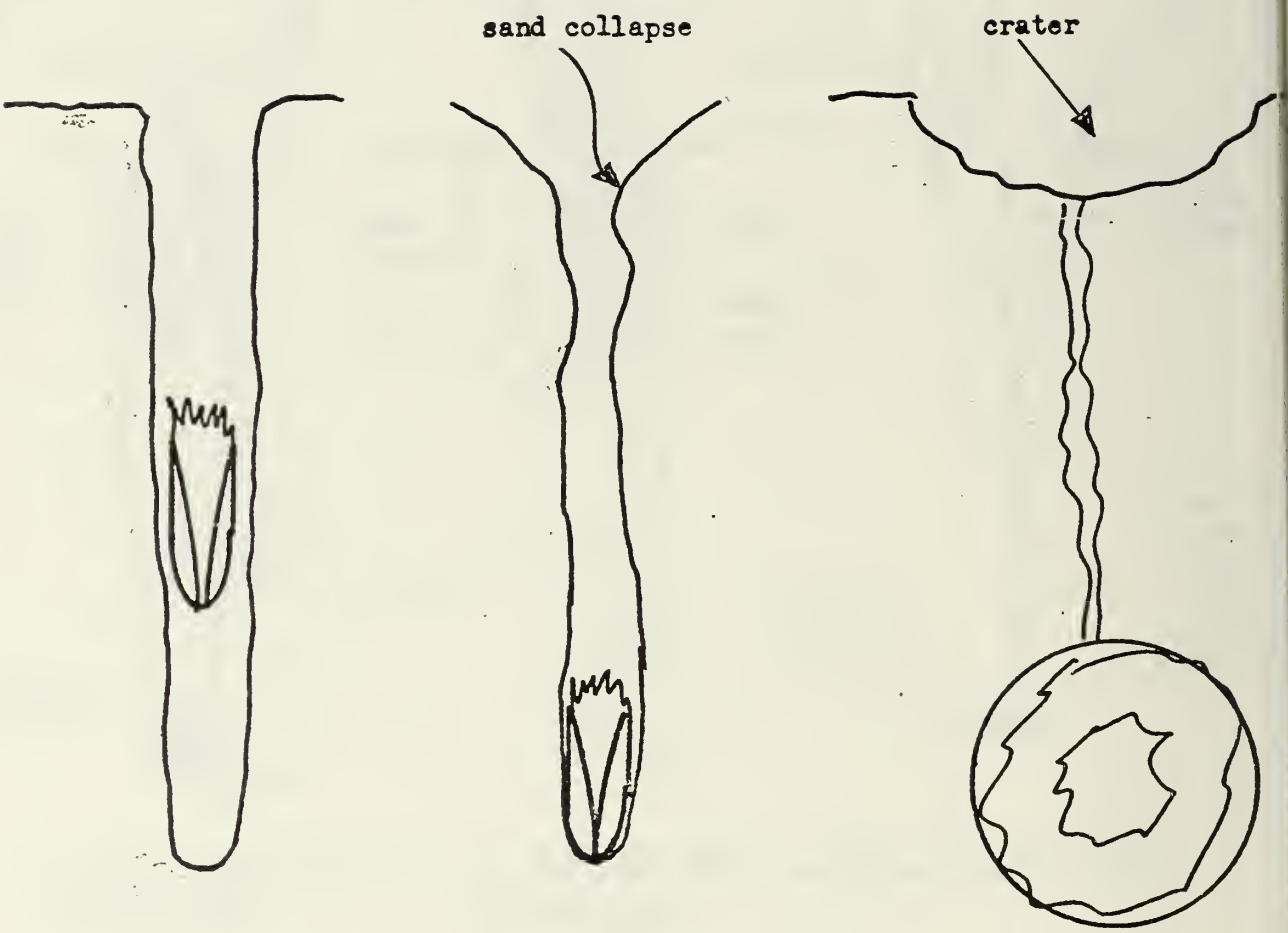


(a) Initial condition

(b) A few 10's of microseconds later

(c) Solid explosive entering hole

Figure 6. Sequence of events for injection of solid explosive.



(d) Solid explosive moving along hole

(e) Sand collapse

(f) Detonation of solid explos

Figure 6 (Continued). Sequence of events for injection of solid explosive.

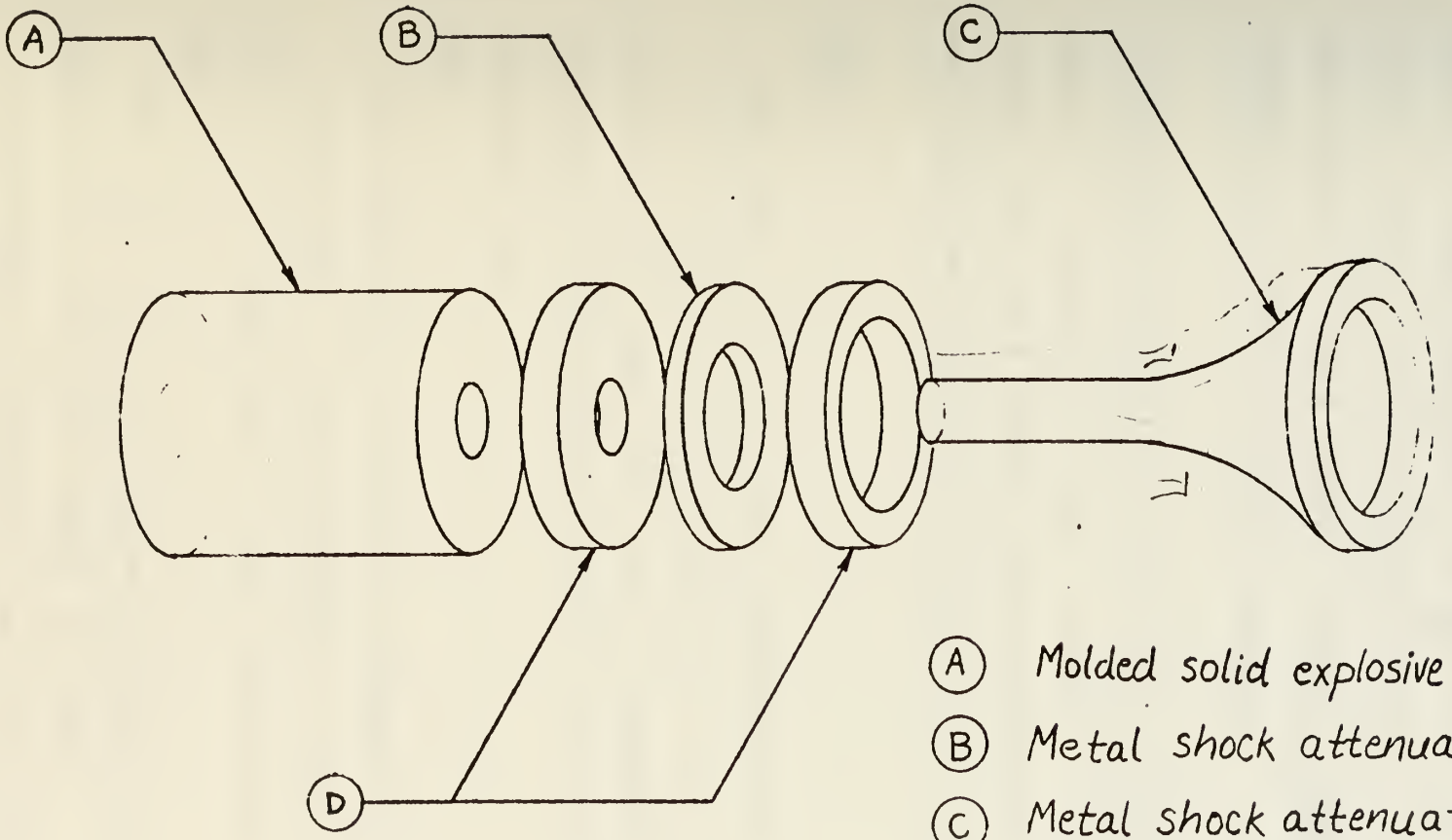
3.

LIST OF CONTACTS AT OTHER RESEARCH AND DEVELOPMENT ORGANIZATIONS

Individual and Organization		Extent of Contact
Colonel Robert Gomez	BRL	Visit to Laboratory
Dr. Dietrich	BRL	Visit to Laboratory
Donald Phillips	WOL/NSWC	Visit to Laboratory
CAPT Walton Boyer	WOL/NSWC	Visit to Laboratory
Julius Simmons	BRL	Visit to Laboratory
William Sykes	DTNSRDC	Visit to Laboratory
Paul W. Cooper	Sandia Corp.	Visit to Laboratory
Dr. Salive	DTNSRDC	Visit to Laboratory
Dr. Julius Enig	WOL/NSWC	Visit to Laboratory
Jim Hammond	NCSL	Visit to Laboratory
Robert Obrachta	ONR	Visit to Office
Jim Bailey	ONR	Visit to Office
Dr. V. C. Dawson	WOL/NSWC	Visit to Office
Frank Romano	Code 034 NAVSEA	Visit to Office
Colonel	AFAL Eglin AFB	Visit to

4.

SKETCH OF SHOCK ATTENUATOR FOR
MINE NEUTRALIZATION WEAPON



(A) Molded solid explosive
(B) Metal shock attenuator
(C) Metal shock attenuator funnel
(D) Polymer shock attenuators

Exploded view of shock attenuator assembly and solid explosive of Mine Neutralization Weapon

Naval Postgraduate School Lt. L.M. Kryske 7/II/78

5.

DESCRIPTION OF LT FUNK'S TEST AT EODF FY 77

DENSITY OF SAND: DRY: 67.2 lbs/cubic foot

WET: 112.25 lbs/cubic foot

Tests 1-6 were conducted with the built in stand-off of the shaped charge, one cone diameter (c.d.). In all tests other than numbers 17-20 a wooden box stand, figure 1, was used to support the shaped charge.

Al screening was used in tests 1-6 as teller plates the spacing of the screens was every one inch in the dry then the distance the sand settled was noted when the water was added and this gave a correction factor for the depth of the screens.

TEST 1

Dry sand, 27" in a plastic garbage can, 1 c.d. stand-off

Penetration: Jet/slug penetrated the bottom of the can, leaving a 3/8" diam. hole tnen it penetrated the 1st 3/4" oak board of the supporting pallet and cracked the board 4" below that one.

TEST 2

Dry sand, 27" in a plastic garbage can, 1 c.d. stand-off

Penetration: good penetration to the 23" level and then burned wood chips found on top of the screen at the 24" level.

TEST 3

Wet sand, depth of sand/water mixture in the can was 25" therefor the correction factor is 0.926.

Penetration: penetrated the entire can leaving an exit approx. 3/8" in diam., hole close to the center of the can

implying that there was little bending of the jet/slug on traveling through the sand/water mixture.

TEST 4

Wet sand, settled depth of wet sand 25" therefor the correction factor is 0.926.

Penetration: The jet/slug penetrated the entire can leaving an exit hole of approx. 1/4" and again it was close to the center of the can.

TEST 5

Wet sand in a 4'X4'X4' plywood box with a correction factor of 0.943. A stand-off of 1 c.d. was used.

Penetration: 28.25" corrected

TEST 6

Wet sand in a 4'X4'X4' plywood box with a correction factor of 0.943. A 1 c.d. stand-off was used.

Penetration: 29.25" corrected

In the tests that follow the number and placement of the teller plates was changed. One foot square Al screens were buried at depths of 26" and 25" in the dry sand then a one foot square 1/8" thick Al plate (7075-T6) was buried at 24" then the cans were filled with sand. Water was then added to the cans which were gently agitated until they stopped bubbling.The amount of settling was then noted to be used in establishing the correction factor.

TEST 7

2 c.d. stand-off, corrected plate depth 21.75"

Penetration: Slight dent in the plate about two handfulls of burned wood chips resting on the top of the plate.

TEST 8

2 c.d. stand-off, corrected plate depth 21.34"

Penetration: 3/8" hole in the plate and a pressure rip was induced in the bottom of the can.

TEST 9

3 c.d. stand-off, corrected plate depth 22.0"

Penetration: severe tearing of the plate resulting in a hole 3/4"X2", exit hole in the bottom of the can approx. 3/8", penetration continued through the 1st oak board, but there was no sign of any penetration of the 2nd board.

TEST 10

3 c.d. stand-off, after noting the depth of original settling the can was filled to the top with more sand and water this resulted in a corrected plate depth of 24.39".

Penetration: 1/2" hole in the plate with very little tearing. Bottom of the can ruptured by pressure.

TEST 11

4 c.d. stand-off, corrected plate depth 21.9"

Penetration: 1" hole in plate
1/2" hole in 1st oak board
3/8" hole in 2nd oak board
1/2" deep hole in the ground

TEST 12

4 c.d. stand-off, added an extension to the can so that the distance to the bottom was 30" instead of 27", corrected

plate depth 24.28".

Penetration: 1/2" hole in plate with some tearing jet/slug missed the pallet boards as it exited the can, but it did dent the ground.

TEST 13

5 c.d. stand-off, corrected depth to the plate 21.75"

Penetration: 1 3/4"X3/4" hole in plate
5/8" hole in 1st oak board
1/8" hole in 2nd oak board
1/2" hole in the ground

TEST 14

5 c.d. stand-off, refilled can after settling just as in test 13, corrected depth to the plate 24.35".

Penetration: 1/2" hole in the plate caused mostly by the tearing action, penetration was to the bottom of the can but not through it.

All remaining test were done in 4'X4'X4' plywood boxes. All stand-offs totaled 4 c.d. of air before the jet/slug started penetrating the wet sand. Teller plates for these tests consisted of two Al plates and seven (7) Al screen squares. These were buried at the original depths of; one plate each at 30" and 36" and one square of Al screening each at 12", 15", 18", 21", 24", 27", and 33". The Injector was designed at NPS and had a built in stand-off of 3 c.d. through the injection chamber. For an exploded veiw of the injector see figure 2.

TEST 15

TEST 15

Wet sand with a water head, corrected depth of plate 27.25" correction factor is 0.909, water depth was 37" to the surface of the sand. S.c. and holder placed in garbage bag.

Penetration: 3/8" hole in plate, overpressure split plate in two, pressure damage only to the screen at 33(30.0")

TEST 16

Wet sand, corrected depth of plate 27.8" correction factor is 0.927.

Penetration: 1/4" hole in plate with tearing, pressure tearing of the screen at 33(30.6").

In these last four test the stand-off refers to the distance from the base of the cone to the tube running the length of the injection chamber.

TEST 17

Dry sand, 1 c.d. stand-off (total 4 c.d.), poor technique in digging out the box lead to poor results but pictures when developed should be consulted.

Penetration: No red dye found on the surface. At 5.5" dye starts and had a circular pattern. Bottom cylinder of injector found at a depth of 24" with an O-ring found at 27" and filler plugs found at 30". There was no deformation of the plate.

TEST 18

Dry sand, 1.25 c.d. stand-off (total of 4.25 c.d)

Penetration: No red dye on the surface, ring of dye with a bullseye presistant to the 18" level then the bullseye

starts to move off center, at the 21" level the ring is still 6" in diam.. The bottom cylinder was found at the 24" level and sloped down to the 27" level, the bottom plate off the injector was found on the plate at 30" but it had completly turn itself around from what its original position had been. The plate was badly cracked, approx. 75% of its width, and the red dye was evedent to the 33" level.

TEST 19

Wet sand with a water head, 1 c.d. stand-off (total of 4 c.d.). Corrected depth of plate 28.15" and the correction factor 0.938. The water depth to the surface of the sand was 31".

Penetration: The box and tower were scattered all over the range so depth of penetration of the different parts was impossible to determine. The lower cylinder of the injector was recovered. The plate was bowed but other wise unmarked. Large amounts of the oil based red dye were found at what had been a depth of 42.2" corrected.

TEST 20

Wet sand with a water head, 1 c.d. stand-off (total of 4 c.d.). Corrected depth of plate 27.95" and correction factor of 0.932. The water head depth was 36" to the surface of the sand.

Penetration: Cylinder impacted with the plate and was recovered, impact only causing the plate to be dented. Dye and small pieces of screen were found on the bottom of the box.

THINGS TO INVESTIGATE

1. Use scale model steel cylinders to simulate the mine.

2. Check penetration of the shaped charge in just water using the 4 c.d. air stand-off.

3. Use different cone material like zinc.

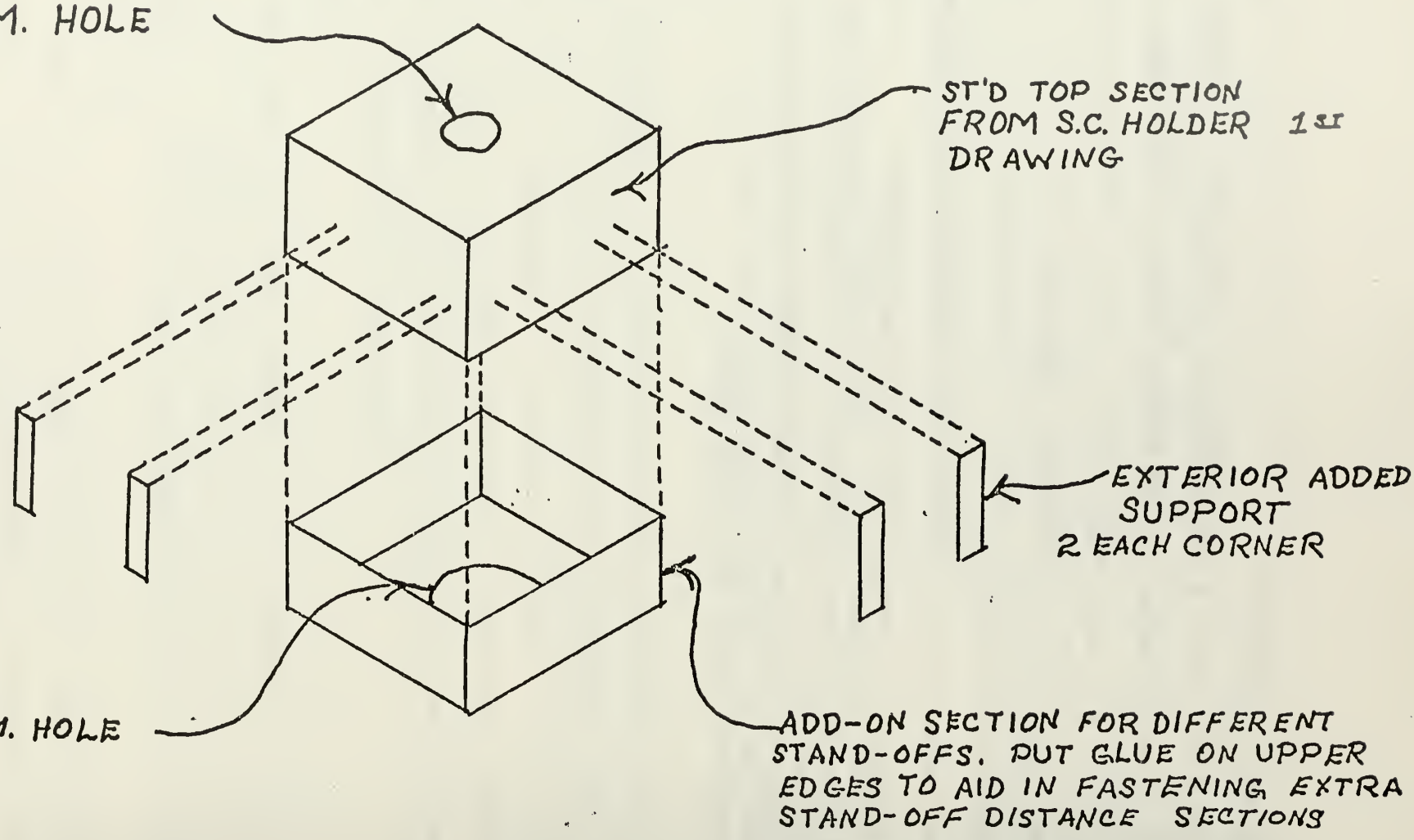
4. Redesign the injector

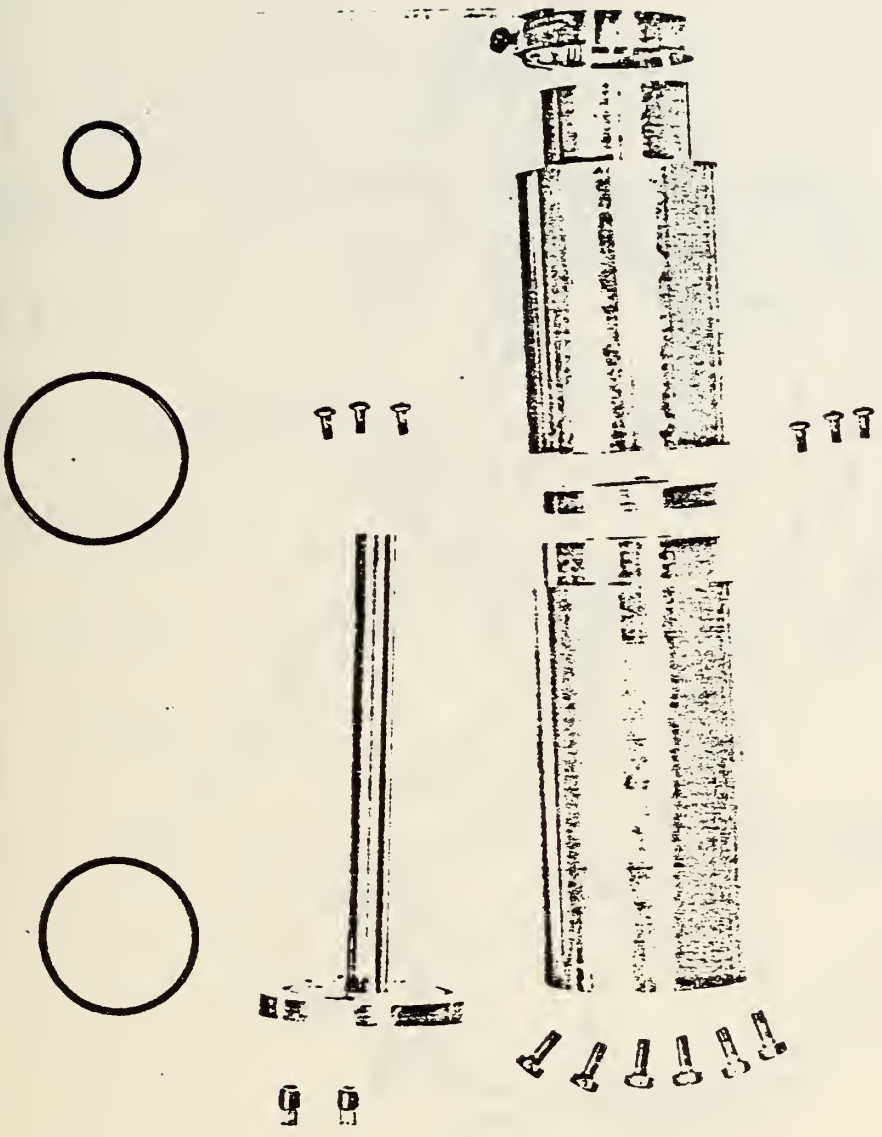
a. make the s.c. holder out of plastic so that the shock wave will not be coupled with the base cylinder.

b. make the base cylinder more aerodynamically stable.

4. Check for optitimum distance between the base of the cone and the top of the center tube.

S.C. HOLDER 2ND DRAWING





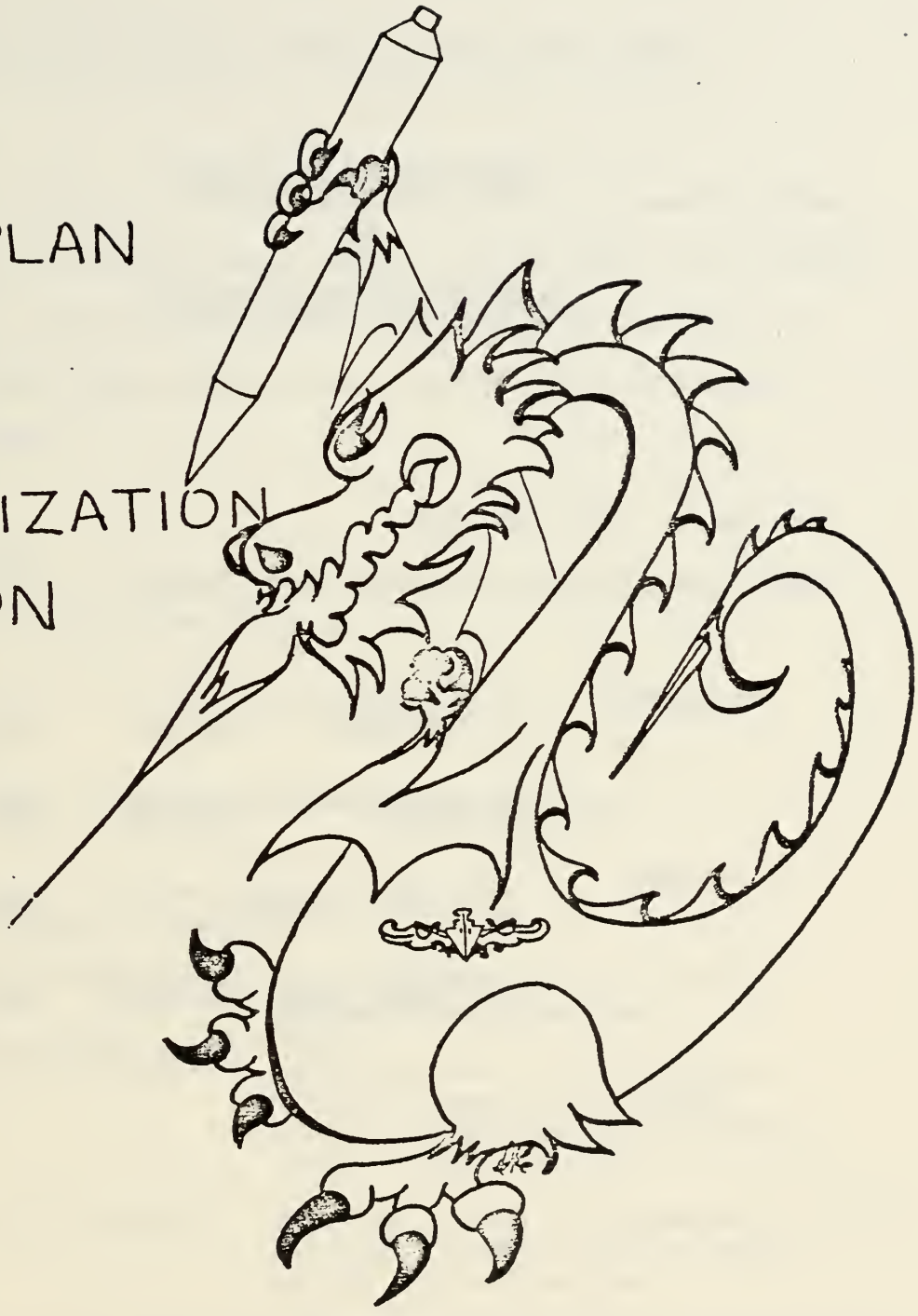
EXPLOPED VEIW OF INJECTOR

FIGURE 2

6.

TEST PLAN FOR
MINE NEUTRALIZATION WEAPON TESTS
ON 22-27 MAY 1978

TEST PLAN FOR MINE NEUTRALIZATION WEAPON



TO BE CONDUCTED

AT

EODF, INDIAN HEAD

22-27 MAY 1978

TEST PLAN FOR EVALUATION

OF

MINE NEUTRALIZATION WEAPON

Test Dates: 22-27 May 1978

Test Site: NAVFAC, EOD, Indian Head, Maryland

SUBMITTED:

D. VUGTEVEEN	LT, USN	Date
L. KRYSKE	LT, USN	Date

APPROVED:

A.E. FUHS, ADVISOR	Date
E.W. RICE, ~~TECH. DIR.~~ EODF Director of Research	Date

Shaped Charge Research Team

TEST PLAN FOR EVALUATION OF MINE NEUTRALIZATION WEAPON

1. Objectives:

   a. To determine optimum stand-off for the Jet Research Center, Inc. underwater shaped charge (Y-1076) against sand targets.

   b. To establish the technical feasibility of insertion of a solid stage explosive into a cavity formed by a shaped charge.

2. Period of Tests:

   21 May 1978 - 27 May 1978

3. Testing Site:

   U.S. Naval Facility, EOD, Indian Head, Maryland

4. Synopsis of Schedule of Events:

20 May 1978 (Saturday)	Depart Monterey. PM arrive Dulles; proceed to accomodations.
21 May 1978 (Sunday)	Establish contact with EODF Duty Officer; liaison with assigned contact. Survey test site, install/ check-out test equipment.
22 May 1978 (Monday)	Establish contact with ~~Technical Director~~ Dr Rice. Conduct stand-off tests against dry sand targets.
23 May 1978 (Tuesday)	Conduct evaluation of mine neutralization weapon (MNW) against dry sand targets.
24 May 1978 (Wednesday)	Conduct evaluation of MNW against wet sand targets.
25 May 1978 (Thursday)	Conduct follow-up evaluation of MNW (wet and/or dry sand targets)

26 May 1978 (Friday)	Backup day for testing. Prepare equipment and samples for return to Monterey.
27 May 1978 (Saturday)	Depart for Monterey.

5. Personnel:

Dr. A.E. FUHS, Chairman, Department of Mechanical Engineering, Naval Postgraduate School, Advisor

LT L. KRYSKE, USN Test project officer

LT D. VUGTEVEEN, USN Test project officer

6. Requirements:

a. Test Equipment -- see Table 1.

b. Material -- see Table 2.

c. Support Services:

1) Contact personnel for Sunday positioning of equipment.

2) Photographer for period 22-26 May 1978.

3) Ordnance personnel for explosive handling.

4) Hypervelocity range (including use of flash radiography equipment).

7. Safety Precautions:

Follow range SOP, no special safety precautions.

8. Transportation/Handling Hazards:

Follow all standard transportation and handling procedures.

9. Explosive Weights:

Each test will consist of detonating one one-half pound explosive shaped charge manufactured by Jet Research Center, Inc. (Y-1076).

10. Disposal Procedures:

Any excess explosives will be returned to storage. That which is in excess and cannot be returned will be destroyed by detonation on the final testing day at the discretion of the ordnance safety officer.

11. General Preparations:

a. Sun 21 May 1978

1) Locate instrumentation in blockhouse; run all necessary cables.

2) Prepare screens for testing; color code and label. Sizes and quantities to prepare:

3) Attach foil switches to mounting hardware.

4) Arrange short test rigs for testing: locate on pallets, attach eye-bolts and guy-wires; attach pressure transducer stand (see drwg); attach pressure transducer housing to first unit (see drwg); fill all units with sand and insert foil switches (see drwg); attach shaped charge positioning hardware (see drwg).

5) Time permitting, arrange test apparatus for 23 May testing.

b. Mon 22 May 1978 (10 shots)

1) Check out instrumentation. Position X-ray heads for testing (see drwg).

2) Place short test rig at hypervelocity site. Attach film cassettes to test rig (see drwg).

3) Attach instrumentation cables. Insert pressure transducers. Position sand bags (see drwg).

4) Insert shaped charge in positioning hardware for stand-off (SO) of 1 cone diameter (CD). Insure shaped charge (SC) is perpendicular to surface of test rig. Place detonator into SC; tape into place.

5) Steps 2-4 will be repeated 9 more times to determine SC penetration at the following SOs:

2 CD	5 CD	8 CD
3 CD	6 CD	9 CD
4 CD	7 CD	10 CD

Even CD stand-off shots will be conducted with flash radiography; odd shots may be fired with or without flash radiography depending upon time available for testing.

6) Prepare long test rigs for testing on 23 May 1978.as before.

c. Tues 23 May 1978 (3 shots)

1) If necessary continue tests from previous day.

2) Position X-ray heads for MNW tests. (see drwg).

3) Place long test rig at hypervelocity site. Attach film cassettes to test rig (see drwg).

4) Attach instrumentation cables. Insert pressure transducers. Position sand bags (see drwg).

5) Insert MNW in mounting hardware. Insure device is perpendicular to test rig surface. Place detonator into SC, tape into place.

6) Repeat steps 3-5 for two more shots.

7) Prepare 3 test rigs for following days' testing using similiar procedures as before. Use foil switches designated for underwater use. Caulk all openings, fill with water and allow

to stand overnight.

d. Wed 24 May 1978 (3 shots)

1) Repeat procedures in steps c - 1) - 6) for three shots into wet sand targets.

2) Prepare 3 test rigs for next day's testing (wet and/or dry set-ups as determined by test and range officers).

e. Thur 25 May 1978 (3 shots)

1) Repeat procedures in steps c - 1) - 6) for 3 shots.

2) Time permitting prepare garbage can rigs for additional stand-off data. Fill one can with sand, placing a coded teller screen at one inch intervals. Rap can to settle sand, place last screen on the top of the sand. Invert second can (with bottom cut out) and affix to first with ordnance tape. Continue sand/screen filling until achieve sand depth of $4\frac{1}{2}$ - 5 feet.

f. Fri 26 May 1978 (5 shots)

1) Place SC in positioning hardware as before. Insert detonator into SC and tape into place. Repeat for remaining SC.

12. Final Preparations and Test Procedures:

a. In paragraph 11 all set-ups ended with the placement of detonators in the SC. Follow standard hook-up procedures to detonating circuit of hypervelocity range and detonate. For last five tests follow standard hook-up procedures to blasting machine and detonate.

b. After range has been certified safe, the test officers will inspect all test related items with the range operations supervisor.

c. After each shot recover as must test apparatus as possible. Remove

film cassettes (when applicable) and develop film prior to next shot. Determine depth of penetration, record data from foil switch electronics. Use polaroid camera to record pressure history from storage o'scope. Measure hole diameters. Recover slug, any pieces of the shaped charge and the MNW injector assembly for subsequent analysis at NPS. Utilize photo services during this phase.

d. Due to procedural unknowns and testing environment, the test officers are authorized to alter the test setup and procedures with the approval of the range safety supervisor.

TEST EQUIPMENT REQUIREMENTS

Item	Quantity	Responsibility	Comments
Storage oscilloscope (pressure)	1	EODF	Prefer dual beam, 10 MHz response or 1 MHz w/ x10 fcn, 1 μsec resolution
Digital timer/counter (velocity)	8	EODF/NPS	Two channel w/ "A-B" fcn μsec resolution
Polaroid camera ( )	1	EODF	w/ oscilloscope adaptor B/W film (40)
Pressure transducer	2	NPS	w/ couplers, pwr supply, cable
Foil switches	175	NPS	
Foil switch electronics	8	NPS	w/ pwr supply, cable
VOM	1	EODF/NPS	
Flash radiography equip	-	EODF	300 kv & 600 kv heads X-ray film and cassettes
Cordin pulser	-	EODF	

TABLE 2

Shaped Charge Research Team

MATERIALS REQUIREMENTS

Item	Quantity	Responsibility	Comments
Shaped charges	25	NPS	Jet Research Y-1076 prepositioned by 10 May 1978
Electric blasting caps	27	EODF	
Blasting machine	1	EODF	
Detonating cord		EODF	
Ordnance tape	5 rolls	EODF	
MNW assembly	9	NPS	see drawings
Foil switch mounting hardware	175	NPS/EODF	see drawings
Shaped charge mounting hardware			see drawings
stand-offs (1-10 CD)	1 ea	NPS/EODF	others to be specified
MNW	9	NPS/EODF	
Pressure transducer housing	4	NPS	see drawings including mounting hardware
Test rigs			
short	5	EODF	see drawings
long	6	EODF	see drawings
double garbage can		EODF	see drawings (req. 10 25-30 gal plastic garbage cans
Fiberglass screen		EODF	
Scissors (or tin snips)	2	EODF	to cut screen
Work gloves		EODF	

MATERIALS REQUIREMENTS (cont)

Item	Quantity	Responsibility	Comments
Sand		EODF	Beach or sand-box type
Shovels	2	EODF	
Carpenters level	1	EODF	
Meter stick	2	EODF	
Wooden pallets	25	EODF	Base for test rigs
Skip loader	1	EODF	For moving sand
Fork lift	1	EODF	For moving test rigs
Sand Bags	20	EODF	
Guy wire	100 ft	EODF	
Eye-bolts	80	EODF	
Spray paint	1 ea	EODF	Blue, yellow, red, green white, black
Staple gun	2	EODF	w/ staples
Caulking compound	24 tubes	EODF	
Caulking gun	2	EODF	
Banding machine	1	EODF	w/ supply of banding material

Shaped Charge Research Team

TEST MATRIX

Test No.	Date Planned	Standoff (CD)	Tgt	Injector (Y/N)	Instrumentation and Remarks
1	5/22	1	D	N	A,-,F,P
2	5/22	2	D	N	A,X,F,P
3	5/22	3	D	N	A,-,F,P
4	5/22	4	D	N	A,X,F,P
5	5/22	5	D	N	A,-,F,P
6	5/22	6	D	N	A,X,F,P
7	5/22	7	D	N	A,-,F,P
8	5/22	8	D	N	A,X,F,P
9	5/22	9	D	N	A,-,F,P
10	5/22	10	D	N	A,X,F,P
11	5/23	6	D	Y	B,X,F,P
12	5/23	6	D	Y	B,X,F,P
13	5/23	6	D	Y	B,X,F,P
14	5/24	6	W	Y	B,X,F,P
15	5/24	6	W	Y	B,X,F,P
16	5/24	6	W	Y	B,X,F,P
17	5/25	6	D/W	Y	B,X,F,P

KEY:
Test Rigs --
A Short box
B Long Box
C Double Garbage Cans
Other codes:
X Flash radiography
F Foil switches
P Pressure transducer
T Teller screens
Target material:
D Dry sand
W Wet sand

TEST MATRIX (cont)

Test No.	Date Planned	Standoff (CD)	Tgt	Injector (Y/N)	Instrumentation and Remarks
18	5/25	6	D/W	Y	B,X,F,P
19	5/25	6	D/W	Y	B,X,F,P
20	5/26	To be determined	D	N	C,P, F &/or T
21	5/26		D	N	C,P, F &/or T
22	5/26		D	N	C,P, F &/or T
23	5/26		D	N	C,P, F &/or T
24	5/26		D	N	C,P, F &/or T
25	5/26		D	N	C,P, F &/or T

Activation Energies for Elevated Temperature Deformation in an Aluminum 17.5 Weight Percent Copper Alloy

T. R. McNelley* and A. L. Cipriani**

ABSTRACT***

An Aluminum - 17.5 weight percent copper alloy was prepared by casting and subsequent hot rolling to produce a fine dispersion of Al_2Cu in a fine grained aluminum-copper solid solution matrix. This material was evaluated by compressive stress-strain testing from 300K to 760K, employing strain rates from $1.9 \times 10^{-4} S^{-1}$ to $9.5 \times 10^{-3} S^{-1}$. Three regimes in stress were observed for the activation energy: between modulus - compensated stresses of $\sim 0.3 \times 10^{-3}$ and $\sim 0.6 \times 10^{-3}$, the activation continuously increased from a value of 107 KJ/mol to 157 KJ/mol; for modulus compensated stress on up to $\sim 1.0 \times 10^{-3}$, the activation energy was approximately constant at 157 KJ/mol; above 1.0×10^{-3}, the activation energy decreased to a constant value of 116 KJ/mol. The strain rate sensitivity coefficient in the low stress regime was determined to be 0.4, suggesting a tendency toward superplastic deformation in this material; the observation, however, of the continuous varying activation energy in this regime is inconsistent with any existing theories for such a deformation process. The activation energies at higher stresses, and the stress exponents as well, suggest dislocation - climb control of deformation.

*T. R. McNelley, Assistant Professor of Materials Science, Naval Postgraduate School, Monterey, California 93940

**A. L. Cirpiani, LT, USN, formerly a graduate student, Naval Postgraduate School, Monterey, California 93940

***This paper is being prepared for submission to Acta Metallurgi

Title: Warm Working of Eutectic and Eutectoid alloys to produce Fine Microstructures

Investigator: T. R. McNelley, Assistant Professor of Mechanical Engineering

Sponsor: Foundation Research Program (6.2)

Objective: Production of fine, spherodized microstructures in Al-Cu and Al-Mg alloys by warm working and evaluation of subsequent mechanical properties, especially ambient temperature properties.

Summary: The objectives of this research were met. Fine two-phase microstructures were produced in both an Al-Cu alloy and several Al-Mg alloys. Subsequent mechanical characterization of these alloys has revealed that excellent mechanical properties can be developed in Aluminum-Magnesium alloys containing 11% to 14% magnesium by weight. Flow-stress values up to 90,000 psi were obtained, for example, in one alloy. This has lead to a focusing of this research primarily on the Al-Mg system; some additional practical factors concerning this system are that these alloys are less dense than conventional aluminum alloys, and the alloys being studied also exhibit superplasticity at warm temperatures. Research will now focus on the processing and related mechanical properties of these Al-Mg alloys.

Publications:

Theses Directed:

A. L. Cipriani, "An Investigation of the Mechanical Properties of Warm-Rolled Aluminum-17.5 weight percent Copper alloy," Master's Thesis, December 1976

F. G. Ness, "High Strength to Weight Aluminum-18 Weight percent Magnesium alloy through Thermomechanical Processing," Master's Thesis, December 1976

T. L. Glover, "Effects of Thermo Mechanical Processing on Aluminum-Magnesium, Alloys Containing High Weight Percentage Magnesium" Master's Thesis, December 1977

C. P. Bingay, "Microstructural Response of Aluminum-Magnesium Alloys to Thermomechanical Processing," Masters Thesis, December 1977

ABSTRACT

Mechanisms of microstructural refinement in aluminum-magnesium alloys were investigated. Alloys containing from 15% to 19% magnesium were exposed to various processing schemes, and the resulting microstructures were examined. Isothermal forging resulted in some refinement depending on the temperature, strain and strain rate. However, all samples isothermally forged exhibited microstructures having relatively large amounts of the brittle intermetallic phase present. The addition of tertiary alloy elements resulted in little improvement in microstructure. Non-isothermal schemes resulted in the most promising microstructures. A process of high temperature soaking followed by deformation at relatively lower temperatures is concluded to be the most promising means for processing alloys with magnesium contents below 15% by weight.

Bingay, Charles P., "Microstructural Response of Aluminum-Magnesium Alloys to Thermomechanical Processing," Master's Thesis, December 1977.

ABSTRACT

An aluminum-17.5 weight percent copper alloy was warm rolled to achieve refinement of the microstructure. This refined microstructure consisted of finely dispersed intermetallic Al_2Cu particles with an average size of 1.2 microns in an aluminum matrix. This led to improved room temperature properties as well as the onset of superplasticity at elevated temperatures. Ductility and toughness were increased almost six-fold, despite slightly decreased yield strength and maximum compressive strength. It would appear that warm working can be used to enhance both room temperature properties and superplasticity, and that further research could maximize this improvement.

Cipriani, Alfred L., "An Investigation of the Mechanical Properties of Warm Rolled Aluminum-17.5 Weight Percent Copper Alloy," Master's Thesis, December 1976.

ABSTRACT

Cast ingots of three Al-Mg alloys, containing 7% Mg, 15% Mg, and 19% Mg, were obtained from Kaiser Aluminum and Chemical Corporation Center for Technology. Billets were machined from these castings and upset forged at .9% of either the solvus or eutectic temperature, as appropriate, to a true strain of 1.5. These alloys were then evaluated by compressive stress-strain testing at various temperatures and strain rates to determine both the ambient and elevated temperature characteristics. Due to the inability to produce very fine second phase particles by upset forging, the 15% and 19% Mg alloys were high in strength but brittle at room temperature. However, refinement of the second phase particles during compression testing led to superplastic behavior at elevated temperature, with a strain rate sensitivity coefficient of .43 being achieved in the 19% Mg alloy. It was further observed that the addition of Mg increased strain rate sensitivity in these Al-Mg alloys at all temperatures, especially at elevated temperatures.

Glover, T. L., "Effects of Thermo-Mechanical Processing on Aluminum-Magnesium Alloys Containing High Weight Percentage Magnesium," Master's Thesis, December 1977.

ABSTRACT

An Aluminum-18 weight percent Magnesium alloy was prepared by casting and then warm rolling at 425° C to 94% true strain. This alloy was compression tested at six strain rates from 0.00664 per minute to 0.332 per minute and at eight temperatures ranging from 25° C to 425° C. The most significant result is that a warm rolled Aluminum-18 weight percent Magnesium alloy can exhibit compressive strengths in excess of 95 ksi, in a material of 10% lower density than commercial high strength Aluminum alloys. Furthermore, one can envision a thermal mechanical process involving warm working followed by cold working at room temperature whereby one can attain an ultimate tensile strength greater than 90 ksi. Additionally, superplastic behavior at elevated temperatures was manifest in the relatively high value of strain rate sensitivity and by the value of the activation energy for deformation.

Ness, Frank G. Jr., "High Strength To Weight Aluminum-18 Weight Percent Magnesium Alloy Through Thermal Mechanical Processing," Master's Thesis, December 1976.

Title: Interval Modulation of a Sinusoidal Carrier

Investigator: G. Myers, Associate Professor of Electrical Engineering

Sponsor: Foundation Research Program (6.2)

Objective: This work is concerned with radio communications. This research investigated use of a modulated interval, between bursts of a sinusoidal carrier, to convey the information in a message waveform. We called this form of modulation "interval modulation" (IM).

Summary: IM produces a delay between bursts of the sinusoidal carrier. This delay is proportional to the amplitude of the modulating message waveform. Thus, IM is unlike AM, PM, or PM where the amplitude, frequency or phase of a sine wave vary in accordance with a message waveform.

The research resulted in mathematical descriptions of the IM carrier in the time domain and frequency domain. Early in the investigation, we completed the design of an IM modulator and demodulator. We then built and tested these circuits. The technique was then verified by transmitting voice and music signals through the system. The quality of the recovered signals was excellent.

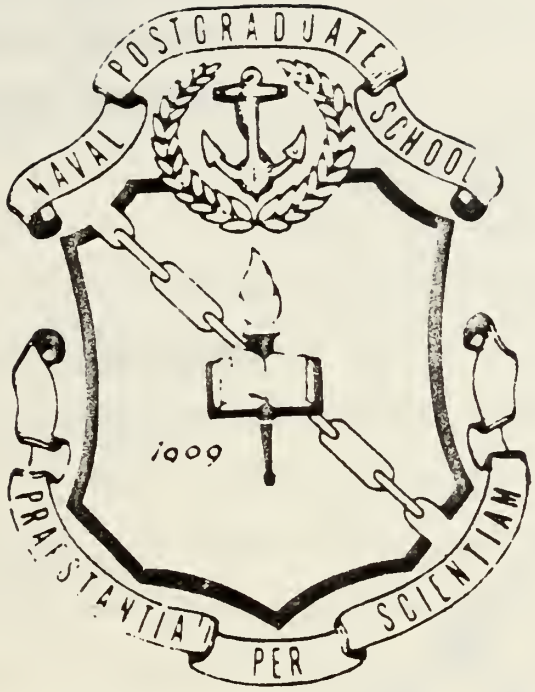
Conference Presentations:

G. Myers, "Interval Modulation (IM) of a Sinusoidal Carrier", Technical Conference, SIEEM 77, Monterey, Mexico, September 77.

G. Myers, "Symposium on Spread Spectrum Communications", Naval Postgraduate School, September 1976.

Publications: G. Myers and E. L. Kilborn, Jr., "Interval Modulation (IM) of a Sinusoidal Carrier", Technical Report, NPS62Mv77031, February 1977.

Thesis Directed: E. L. Kilborn, Jr., "Interval Modulation (IM) of a Sinusoidal Carrier", Master's Thesis, December 1976.



ADHOC SYMPOSIUM ON SPREAD SPECTRUM COMMUNICATIONS

7 - 9 SEPTEMBER 1976

SPONSORED BY

NATIONAL SECURITY AGENCY (W-3)
FT. GEORGE G. MEADE, MD 20755

NAVAL ELECTRONICS SYSTEMS COMMAND (PME 107)
NAVAL ELECTRONICS SYSTEMS COMMAND HEADQUARTERS
WASHINGTON, DC 20360

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93940

ABSTRACT

INTERVAL MODULATION (IM) CARRIER SIGNAL

Glen A. Myers
Naval Postgraduate School
Code 62Mv
Monterey, CA 93940

Edgar L. Kilborn, Jr.
Naval Postgraduate School
SMC Box #1657
Monterey, CA 93940

In classical radio communications, either the amplitude A, the phase $\phi(t)$, or the frequency $\frac{d\phi(t)}{dt}$ of a carrier wave A cos $\phi(t)$ is caused to vary in accordance with the message m(t) transmitted. This work uses a time delay to modulate a carrier in accordance with m(t) rather than amplitude (AM) or phase (PM) or frequency (FM) modulation of the carrier.

Operation of the time delay or interval modulation (IM) is as follows: the message m(t) is sampled and after a delay of τ seconds, where τ is proportional to that sample value of m(t), a burst of sinusoidal carrier is transmitted until the next sample of m(t) is taken. The sequence then repeats to define the IM carrier signal. Each burst of the carrier signal is of fixed duration (constant number of cycles transmitted per burst) the by making the sampling of m(t) non-periodic.

The IM carrier signal is useful only if the message (samp can be recovered from the intervals τ between bursts in a rece Present work involves the use of a phase-locked loop (PLL) as IM carrier signal demodulator. To date, successful generation and detection of IM carrier signals has been accomplished usin voice and music as the modulating (message) signal.

IEEE

IMPOSIUM SOBRE
PLICACION DE LA INGENIERIA
LECTRICA Y ELECTRONICA
RGANIZADO POR
A SECCION MONTERREY
EL IEEE.

MONTERREY, N. L., MEXICO
22 Y 23 DE SEPT. 1977

INTERVAL MODULATION (IM) OF SINUSOIDAL CARRIER WAVES

by

Glen A. Myers
U.S. Naval Postgraduate School
Monterey, California

In classical radio communications, either the amplitude A, the phase $\phi(t)$, or the frequency $\frac{d\phi(t)}{dt}$ of a carrier wave A cos $\phi(t)$ is caused to vary in accordance with the message m(t) transmitted. In this paper we consider use of a time delay between bursts of a pure sine wave to carry the message m(t). This technique, which is referred to as interval modulation (IM), is distinct from amplitude (AM), or phase (PM) or frequency (FM) carrier modulation.

The principle of IM is explained using waveforms, and parameters of the IM carrier are defined. Then a frequency description is considered, and the relation of modulated carrier bandwidth to system parameters is discussed.

A block diagram of the modulator used in the experimental work is provided, and its operation is briefly treated. A conventional phase-locked loop (PLL) is used to demodulate the IM carrier. Voice and music used to modulate the carrier were recovered using the PLL. The quality of the demodulated voice and music is very good.

Applications are briefly mentioned.

1. Time Waveforms

Interval Modulation (IM) of a sinusoidal carrier conveys the information of a modulating message waveform m(t) by sampling m(t) at the completion of burst of the sinusoidal carrier and varying the interval before the next burst by an amount proportional to the amplitude of the sampled value of m(t) [illegible] given DC voltage used to modulate the carrier, the intervals

between the bursts of the sinusoidal carrier are fixed and equal. Therefore, the IM signal modulated by a constant voltage is periodic. An example of the IM signal for a particular DC modulating voltage is shown in Fig. 1 where T_b is the duration of the burst and T_x is the interval between bursts.

For a time-varying (AC) modulating voltage m(t), the interval T_x between bursts of the sinusoidal carrier is caused to vary according to the amplitude A of the sample of this modulation or message voltage m(t). In all the experimental work, T_x is caused to vary linearly with A.

IM is closely related to pulse width modulation (PWM), which is also referred to as pulse duration modulation (PDM). Let $v_{PW}(t)$ represent PWM as shown in Fig. 2. Then, we can write the IM carrier which we call $v_{IM}(t)$ as

$$v_{IM}(t) = [\ 1 - v_{PW}(t)][V \sin 2\pi f_c t]$$

where V and f_c are the amplitude and frequency of the sinusoidal carrier.

2. Frequency Description

The Fourier transform of $v_{IM}(t)$ is the frequency description of interest. The equation for $v_{IM}(t)$ previously written in terms of $v_{PW}(t)$ shows that the spectrum of the IM carrier consists of a component at the carrier frequency and sidebands which are the frequency description of $v_{PW}(t)$, the PWM signal.

For an arbitrary message m(t), the Fourier transform of $v_{PW}(t)$ is not available. Results are available for particular periodic voltages m(t). For example, the Fourier transform of $v_{PW}(t)$ when $m(t) = M \cos 2\pi f_m t$ is presented in Black.* For even this simple case, however, the transform

*H. S. Black, Modulation Theory, D. Van Nostrand Company, Inc., 1953, p. 275, Eqs. 17-25.

is a complicated function containing many side frequencies separated by f_s Hz and having magnitudes given by the ordinary Bessel function. Here, f_s is the sampling rate.

Laboratory work confirms the presence of many side frequencies. Fig. 3 is the spectrum of the voltage shown in Fig. 1. In this case, m(t) is a constant (DC voltage).

For an arbitrary m(t), the side frequencies caused by the sampling are present. Each of these side frequencies as well as the carrier appears to be modulated by m(t). The local spectrum about the carrier and each side frequency is similar to that generated by FM. That is, the message m(t) appears to individually frequency modulate each spectral line.

An example of the spectrum of the IM carrier when m(t) is a tone (sine wave) of frequency f_m Hz is shown in Fig. 4. In this case, the carrier and four side frequencies are shown. The amplitude of m(t) is large, and this results in many lines separated by f_m Hz surrounding the carrier and each side frequency.

This nature of the spectrum of the IM carrier is of interest for at least two reasons. First, the broadband character may have applications as mentioned later. Second, the presence of an FM type of spectrum about each side frequency suggests a method of demodulating the IM carrier.

3. The Modulator

The IM signal modulator uses samples of the input message waveform m(t) to determine the length of the interval between the bursts of the sinusoidal carrier.

The form of the IM signal depends upon several parameters. These parameters include the number of cycles in a burst, the frequency of the sinusoidal carrier comprising the burst, the length of the basic interval

between bursts, and the range of variation (due to the amplitude of the modulating signal) of the interval between bursts. For this reason, the IM signal modulator is designed to permit, within certain limits, variation in the IM signal parameters.

Fig. 5 shows the basic block diagram for the IM signal modulator. Referring to Fig. 5, the frequency of the sinusoidal carrier is varied by changing the frequency of the Gated Sinusoidal Burst Generator. The number of cycles of the sinusoidal carrier in each burst is controlled by changing the length of the Burst Window Generator. The basic interval between bursts is controlled by changing the slope of the ramp output from the Gated Ramp Generator. And, the range of variation of the interval, due to modulation, is controlled by the amplitude of the modulator's input signal.

Fig. 6 is a simplified timing diagram for the IM signal modulator and shows the sequence of events which generate the IM signal. The operation of the modulator is as follows. Coincident with the completion of a burst of the sinusoidal carrier a sample command is issued to the Sample and Hold circuit and the Gated Ramp Generator is turned on. The outputs from the Sample and Hold circuit and the Gated Ramp Generator are fed through a Differential Amplifier having zero volts output when the output of the Sample and Hold circuit is equal to the output of the Gated Ramp Generator. The Zero Crossing Detector following the Differential Amplifier notes the zero crossing and turns on the Burst Window Generator. The Burst Window Generator then gates on the Gated Sinusoidal Burst Generator. The length of the burst from the Gated Sinusoidal Burst Generator is a function of the length of the Interval Modulated Gate signal produced by the Burst Window Generator. The length of the Interval Modulated Gate signal is

controlled by a timing sequence, the duration of which is variable. Upon completion of the timing sequence the Interval Modulated Gate signal is turned off, another sample is taken, and the sequence of events repeats.

4. The Demodulator

The time representation of the IM carrier indicates that samples of the message can be recovered with a circuit that converts the duration of the interval between bursts to a corresponding voltage. This task was of concern early in the work because of the microsecond variations expected for audio messages m(t). It was decided to investigate use of a phase-locked loop (PLL) as a possible demodulator. The notion was that the transient behavior of the PLL between bursts of the carrier might provide demodulation.

The PLL turned out to be a fortunate choice. A standard PLL chip recovered the message from the IM carrier. Further, it was discovered that the PLL could be "tuned" to many different discrete frequencies in the region about the carrier and recover the message.

The nature of the spectrum of the IM carrier indicates why a PLL can be used as a demodulator. Further, the reason demodulation is possible at many frequencies is explained by the presence of discrete side frequencies. In this application, then, the frequency description rather than the time description was found to be more useful in considering the demodulation process.

Limitations on the use of the PLL as a demodulator arise when the bandwidth of the modulated carrier (or a side frequency) exceeds the spacing between side frequencies (sampling rate). This occurs when the amplitude of the modulating signal is large. (This corresponds to large modulation index FM or PM.) To prevent such overmodulation, reduce the level of the modulating signal.

4. Applications

The work presented here was not motivated by any particular problem nor dictated by any specific application. The nature of the spectrum of the IM carrier suggests some possible uses.

The IM technique generates several "sub-carriers" of controlled spacing. Each of these sub-carriers is modulated (angle modulation) by the message m(t). This means a single transmitter can be used with several receivers each tuned to a different frequency.

Alternatively, any one receiver can select the sub-carrier which provides the best reception. The IM technique may be useful, then, where use of frequency diversity is desirable.

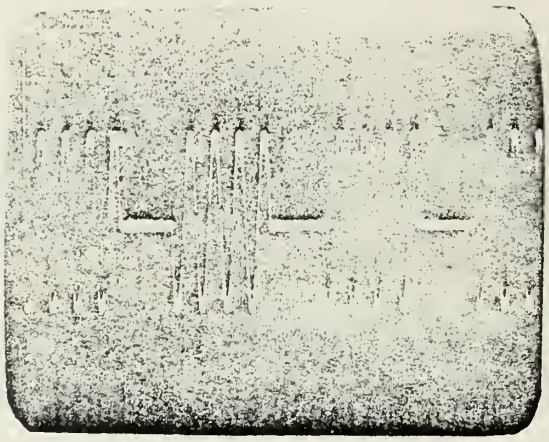


Figure 1. IM SIGNAL MODULATED BY A CONSTANT VOLTAGE. The carrier frequency is 200 kHz, T_x = 10μsec, and T_o = 20 sec.

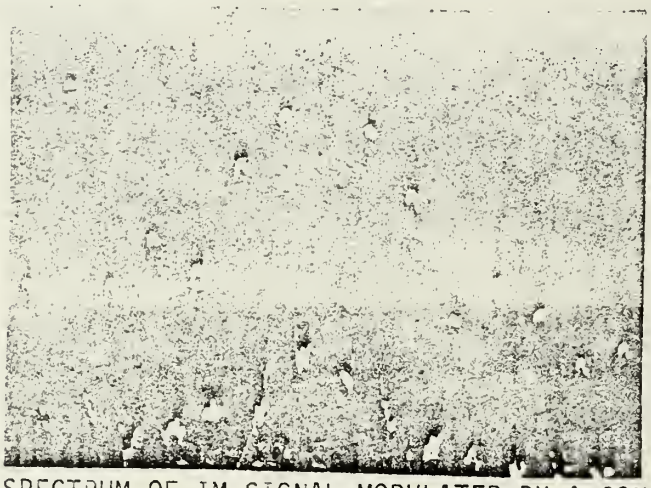


Figure 3. SPECTRUM OF IM SIGNAL MODULATED BY A CONSTANT VOLTAGE. The frequency range of 0 to 450 kHz is shown.

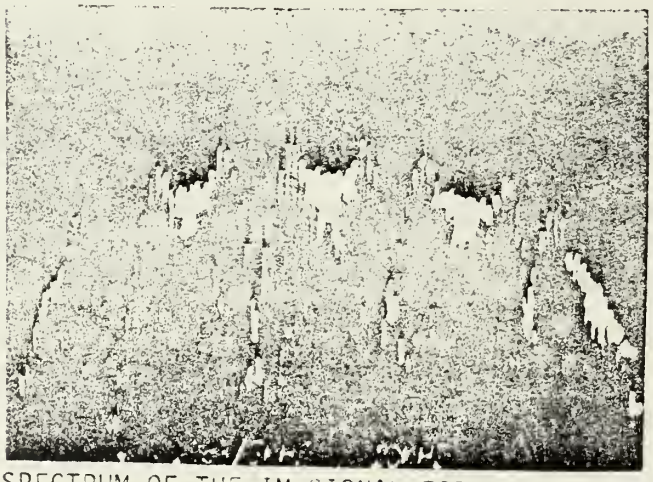


Figure 4. SPECTRUM OF THE IM SIGNAL FOR LARGE AMPLITUDE MODULATING SIGNAL. The modulating signal m(t) is a sine wave.

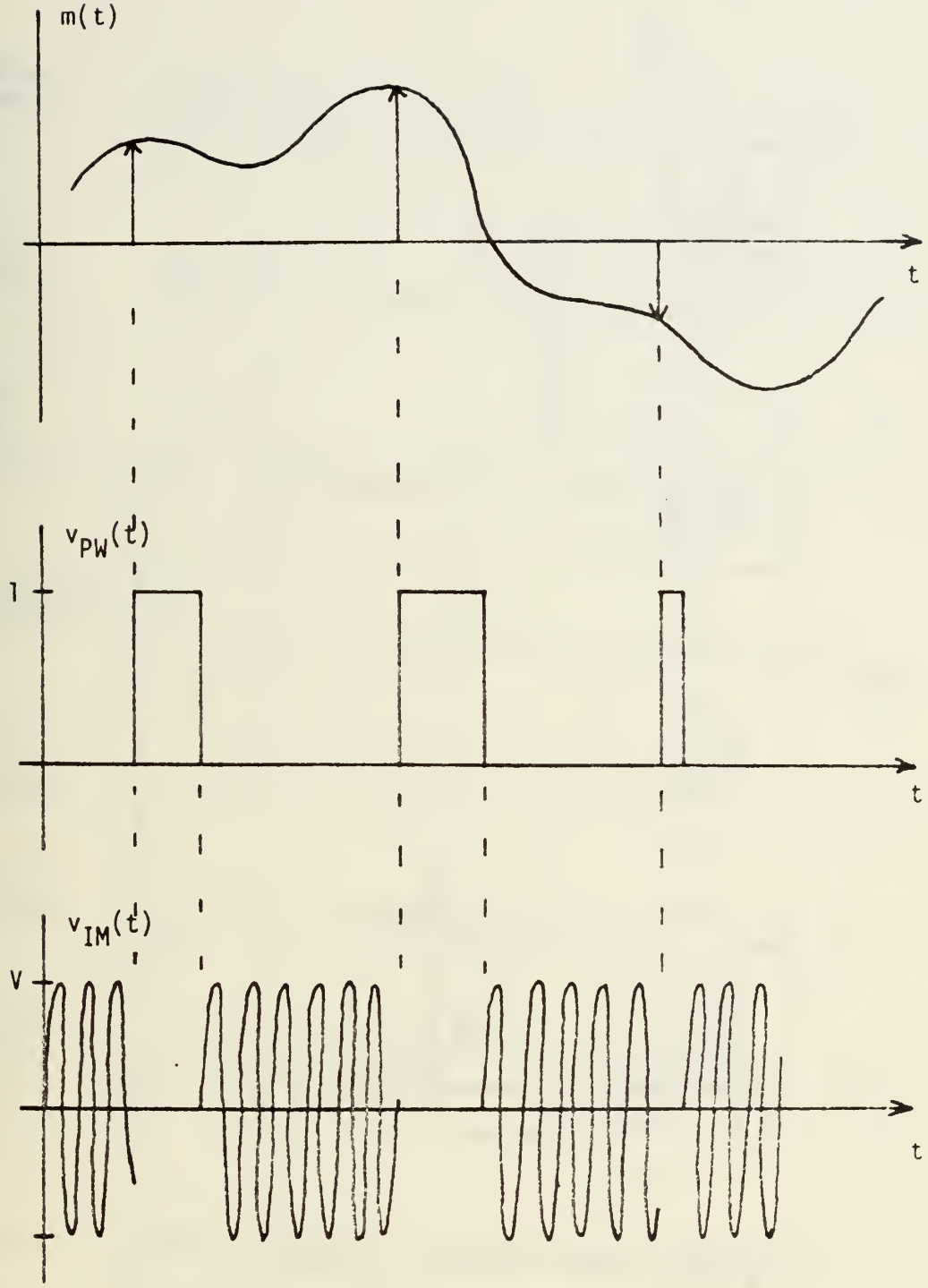


Figure 2. THE RELATION BETWEEN PWM AND IM.

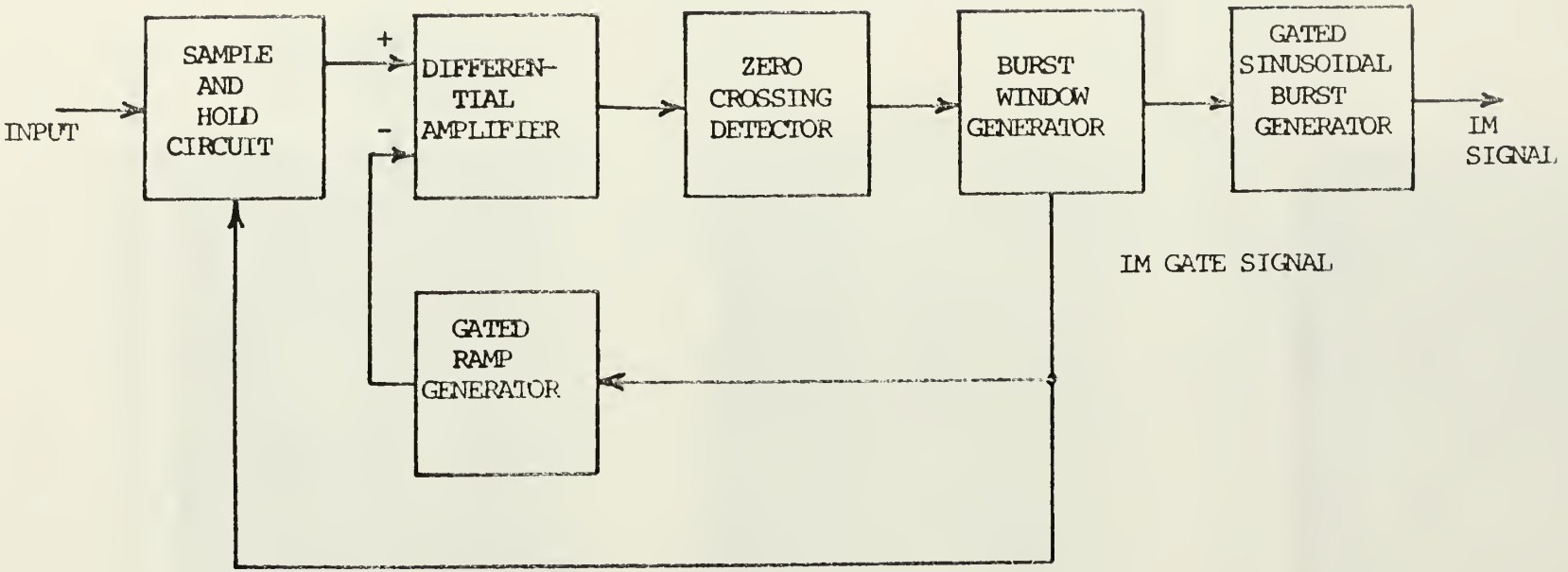


Figure 5. BLOCK DIAGRAM OF THE IM SIGNAL MODULATOR

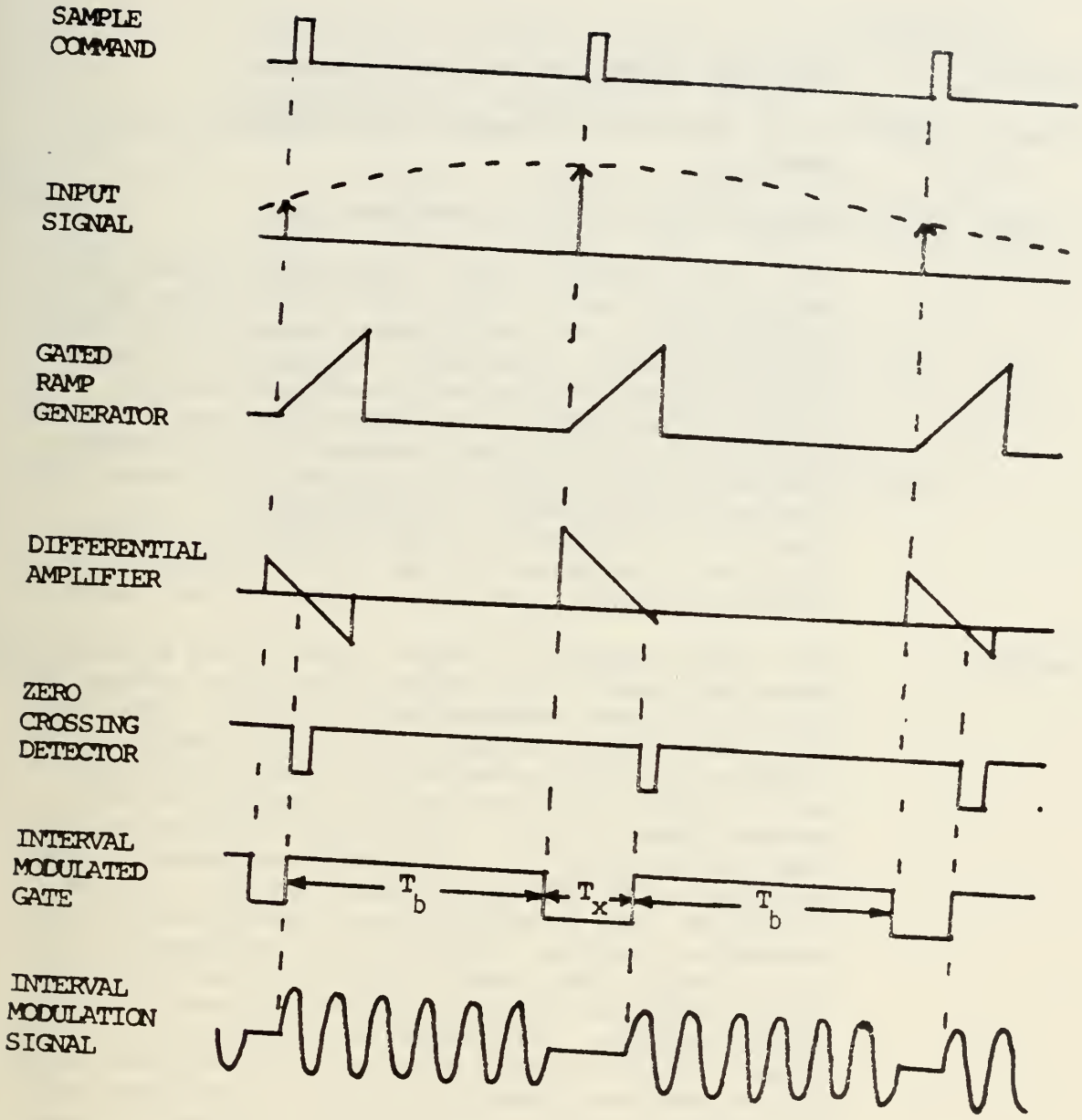


Figure 6. TIMING DIAGRAM OF THE IM SIGNAL MODULATOR.

Title: Computer Aided Acoustical Imaging

Investigator: J. P. Powers, Professor of Electrical Engineering

Sponsor: Foundation Research Program (6.2)

Objective: This work is part of a continuing effort to explore the use of computer processing of coherent (both amplitude and phase information present) data to generate images of objects from transmitted or reflected ultrasonic fields. The short term objective was to construct an experimental system to record the complex data fields of an ultrasonic wave for computer processing. Additionally it was desired to explore the presentation of the data on a computer driven video display for presentation of the data or resulting image.

Summary: The first acoustic images using this system have been obtained. The data aquisition system has been built and successfully tested. Preliminary acoustical images have been displayed on a Ramtek Gx-100 television display with 16 gray levels (limited by the display refresh memory capability). The data acquisition system consists of a logarithmic compressive amplifier with 60db of dynamic range, a phase detector, and a high accuracy positioning and recording system. Testing and calibration of the hardware implementation of these devices shows acceptably high degrees of accuracy and repeatability. The data is recorded on analog tape. After digitization of the recorded data and subsequent demultiplexing of the channels the data has been successfully stored in memory of PDP 11/40 computer for subsequent display or processing. Using the Ramtek Gx-100 display driven by a PDP 11/40, interactive display capability has been developed to allow the operator to call up displays of amplitude, phase or intensity data. He also has the capability to change the quantization levels, the gray scale assignment and to assign various pseudocolor values to the data. This interactive capability was designed and tested during the research period using both computer generated and experimentally recorded data. Future efforts include further expansion of this

interactive display capability, refinements in the data acquisition system and investigation of more experimental objects to assess the overall capabilities of the system.

Publications:

J. P. Powers, "Computer Simulation of Linear Acoustic Diffraction, Acoustical Holography," Vol. 7, L. W. Kessler, Ed., Plenum Press, New York, 1977, pp. 193-205.

J.P. Powers, Lt. R. T. O'Bryon, Lt. J. W. Patton, "Ultrasonic Imaging System Incorporating Computer Aided Coherent Processing", Abstracts, Ninety-fourth Meeting of the Acoustical Society of America, Miami Beach, 12-16 December 1977, in J. Acoustical Society of America, 62 (Supplement 1): S22, 1977.

COMPUTER SIMULATION OF LINEAR ACOUSTIC DIFFRACTION

John P. Powers

Department of Electrical Engineering
Naval Postgraduate School
Monterey, California 93940

ABSTRACT

Computer-aided acoustical imaging systems and computer simulations of other acoustic imaging techniques frequently require simulation of linear acoustic diffraction of large complex-valued data arrays. Computation efficiency requires the use of fast Fourier transform techniques. This paper compares two Fourier transform formulations of the propagation problem: the Fresnel integral and the spatial frequency domain approach. The following features are compared: restrictions on maximum and minimum propagation distances, sample sizes and number of samples required, adaptability to image processing techniques, and computational requirements.

INTRODUCTION

The use of computers in computer-aided acoustic imaging has become increasingly popular in recent years. The use of the computer in obtaining images by such techniques as backward wave propagation[1] offers such advantages as the elimination of the reconstruction wavelength scaling problem[2] obtained with optical reconstruction techniques, reference-free holography that uses the linear detection properties of piezoclectric transducers, and the possibility of incorporating image enhancement to improve the image obtained. Additionally computer simulation of the holographic process has been a useful tool in studying such novel techniques as phase-only holograms[3] and kinoforms[4]. In most computer-aided imaging techniques it is necessary to mathematically simulate the scalar wave diffraction process. This paper compares two techniques for this simulation,

describing their features and some of their advantages and disadvantages.

In representing the acoustic diffraction formulations we will implicitly assume that the propagation medium is linear and homogeneous. We also will require that the resulting diffraction integrals must be amenable to computer solution in a reasonable amount of time which at the present infers that the integrals must be in the form of Fourier transforms so that the speed and efficiency of the Fast Fourier Transform[5] (FFT) can be brought to bear on the problem. The general problem then is: given a complex scalar wave $\underline{U}_i(x_i, y_i, 0)$ at some input plane, find an expression for the wave $\underline{U}_o(x_o, y_o; z)$ at some parallel output plane a distance z away, subject to the wave equations. Two forms of the solution incorporate the Fourier transform and will be considered after a short review of the features of the analog Fourier transform and the discrete Fourier transform (DFT).

The two dimensional analog Fourier transform is defined by the relationships

$$\underline{U}(x,y) = \int\int_{-\infty}^{\infty} \underline{A}(u,v) e^{j2\pi(ux+vy)} du dv = \mathcal{F}^{-1}\{\underline{A}(u,v)\} \tag{1}$$

$$\underline{A}(u,v) = \int\int_{-\infty}^{\infty} \underline{U}(x,y) e^{-j2\pi(ux+vy)} dx dy = \mathcal{F}\{\underline{U}(x,y)\} \tag{2}$$

where $\underline{U}(x,y)$ is the complex function in the space domain; $\underline{A}(u,v)$ is the complex transform in the spatial frequency domain; u,v are spatial frequencies (dimensions of cycles/meter); and $\mathcal{F}, \mathcal{F}^{-1}$ are symbolic operators for the transform and inverse transform operations respectively.

The discrete version of the Fourier transform (assuming an equal number of samples and sample spacing in both dimensions) is given by

$$\underline{f}(ma,na) = \frac{1}{N^2} \sum_{k=o}^{N-1} \sum_{\ell=o}^{N-1} \underline{A}(k\Omega,\ell\Omega) e^{j2\pi\left[(ma\cdot k\Omega)+(na\cdot \ell\Omega)\right]}$$

$$= F^{-1}\{\underline{A}(k\Omega,\ell\Omega)\} \tag{3}$$

$$\underline{A}(k\Omega,\ell\Omega) = \sum_{m=o}^{N-1} \sum_{n=o}^{N-1} f(ma,na) e^{-j2\pi\left[(ma\cdot k\Omega)+(na\cdot \ell\Omega)\right]}$$

$$= F\{f(ma,na)\} \tag{4}$$

where $o \leq m,n \leq N-1$;

$o \leq k,\ell \leq N-1$;

$\underline{f}$ (ma,na) is a complex valued sequence of samples in the space domain;
$A(k\Omega,\ell\Omega)$ is a complex valued sequence of samples in the spatial frequency domain;
N is the total number of samples in one dimension in the space or frequency domain (N x N total sample values);
a is the sample spacing in the space domain;
Ω is the sample spacing in the spatial frequency domain (and is equal to 1/Na);
F, F^{-1} are symbolic operators for the discrete Fourier transform and the inverse transform operation respectively.

The fast Fourier transform (FFT) is an efficient algorithm that uses symmetry properties to compute this discrete transform. The efficiency of this algorithm for large $N(\geq 8)$ makes it the only practical method of processing large two-dimensional complex valued data arrays, as is required in computer-aided acoustic imaging. Several properties of the DFT are mentioned here as they have important repercussions later.

1. As mentioned above, if the sample size in the space domain is a, then the sample size in the spatial frequency domain is 1/Na where N is total number of samples in one dimension.

2. There are N x N samples in the space domain covering a region Na x Na; there are also N x N samples in the frequency domain covering a region 1/a x 1/a.

3. The DFT assumes that the input sequence is periodic in both the x and y dimensions. Hence the input is considered an infinite two-dimensional periodic array (with period Na). Similarly the inverse DFT requires that the sequence in the spatial frequency domain also be periodic in both dimensions (with a period of 1/a).

4. The scaling factor of $1/N^2$ in the inverse transform should be noted to ensure computational accuracy.

5. The wave fields in diffraction patterns are usually centered on the propagation axis (as in Fig 1a) to take full advantage of symmetry. The usual DFT algorithm however usually works on a wave-field that lies in the first quadrant and produces the spectrum also in the first quadrant with the (0,0) frequency component at the origin (as in Fig. 1b). In order to apply the usual DFT to the centered wave without getting a linear phase shift in the transform domain that accompanies simple translation, a data shuffle [10] can be used. Based on the assumed periodicity of the input wave

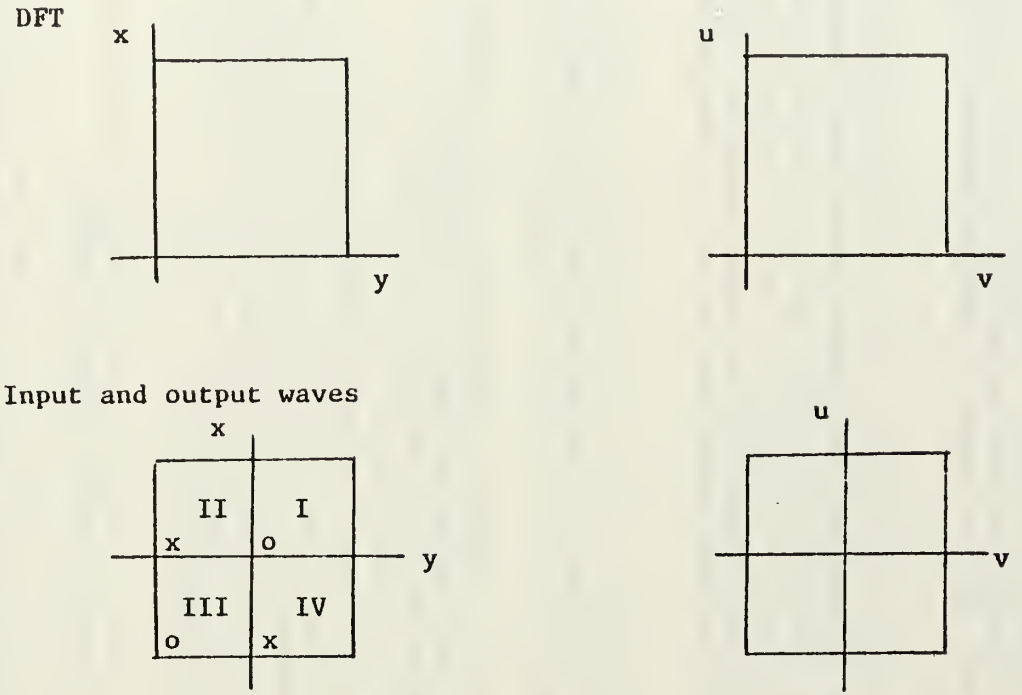


Figure 1 a) Geometric arrangement of input data and transform required by most discrete Fourier transform algorithms

b) Geometrical arrangement of input data and transform desired for diffraction problems. (Quadrant numbers and data location marking refer to data shuffle described in text.)

and the resulting symmetry, this shuffle applied both before and after taking any transform or inverse transform allows one to work with waves that are centered on the axis and FFT routines that work only in the first quadrant. The shuffle routine exchanges the data in quadrants II and IV (see Fig 1b) and the data in quadrants I and III. (The alignment of the shuffle is indicated by the fact that the data of position x in quadrant II of Fig 1b is exchanged with the data of position x in quadrant IV. Similarly the data of the positions marked by the o's of quadrants I and III are also exchanged. Since this data shuffle is a mere exchange of data locations, the additional computation time is minimal except for the largest of arrays.

THE FRESNEL INTEGRAL

The first form of the solution of the propagation problem that incorporates the Fourier transform representation is the Fresnel integral[6]. Using the notation of the discrete Fourier transform this expression is:

$$\underline{U}_o(k\Delta x_o, \ell\Delta y_o) = \frac{e^{j\frac{2\pi z}{\lambda}}}{j\lambda z} e^{j\frac{2\pi}{\lambda z}\left[(k\Delta x_o)^2+(\ell\Delta y_o)^2\right]} \cdot F\left\{\underline{U}_i(ma,na)\, e^{j\frac{\pi}{\lambda z}\left[(ma)^2+(na)^2\right]}\right\} \qquad (5)$$

$$k\Omega = k\Delta x_o/\lambda z$$
$$\ell\Omega = \ell\Delta y_o/\lambda z$$

where $\underline{U}_i(ma,na)$ is the sampled input function (sample spacing = a);

z is the propagation distance;
Δx_o, Δy_o are sample spacings of the output wave and are each equal to $\lambda z/Na$;
$\underline{U}_o(k\Delta x_o, \ell\Delta y_o)$ is the sampled value of the output wave at the plane a distance z from the plane of the input wave.

This expression for $\underline{U}(k\Delta x_o, \ell\Delta y_o)$ is valid only for propagation distances such that

$$z^3 >> \pi N^4 (\Delta x_o - a)^4 / 16\lambda \qquad (6)$$

The important properties of this formulation of the diffraction in-

tegrals are noted below.

1. There is a limitation on the minimum propagation distance due to the inequality of Eq. 6. There is no limitation on the maximum propagation distance; in fact the Fresnel integral becomes the much easier to solve Fraunhofer diffraction integral[6] at very large propagation distances.

2. The sample spacing of the output becomes larger with increasing propagation distance since $\Delta x_o = \lambda z / Na$. This is helpful for diverging waves since the sample spacing spreads at the same rate as the diverging wave ensuring complete coverage of the diverging beam with a minimum number of sample points.

3. The diffraction operation requires N^2 complex multiplications, one data shuffle, one FFT operation, another data shuffle, and another N^2 complex multiplications.

4. The required sample spacing is determined by frequency aliasing considerations[7]. Since the exact derivation of the number of samples depends on the object, we choose a representative one-dimensional object, a slit of width 2a, and find an estimate of the sample number as a rough guideline. The object is a slit of width 2a in a region 2W wide (see Fig 2a); the transform of this object is 2a sinc 2au (see Fig 2b). The spacing in the frequency domain will be $\Delta u = 1/2W$ and, if there are N samples, the maximum frequency will be

$$Umax = \frac{N}{2} \Delta u = \frac{N}{4W} \tag{7}$$

The energy contained in the frequencies above Umax will be aliased back into the frequencies below Umax. The fraction, ε, of the total energy that is aliased is

$$\varepsilon = \frac{1}{a} \int_{Umax}^{\infty} |A(u)|^2 du \simeq \frac{2W}{\pi^2 Na} \tag{8}$$

and hence the sample width is

$$\Delta x = \frac{2W}{N} = \pi^2 a \varepsilon \tag{9}$$

and the number of samples is

$$N = \frac{2W}{\pi^2 a \varepsilon} \tag{10}$$

Typically $\varepsilon \sim$ 5% or 10% is used. Hence the amount of aliasing tol-

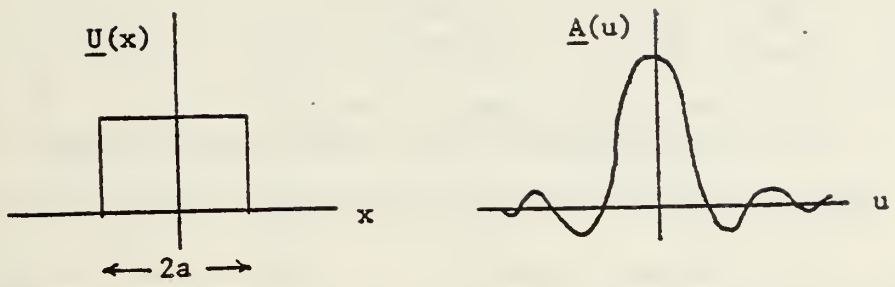


Figure 2 a) Hypothetical one dimensional object (a slit)
b) Transform of object

erated determines the sample spacing and the number of samples. As indicated previously this result is based on a particular input object (i.e. a slit). Objects with smaller scale discontinuities or perturbations might require finer sample spacing while objects with more gradual discontinuities can be analyzed with coarser sampling. However the results of Eqs. 9 and 10 give approximate values for sample spacing and total number of samples.

SPATIAL FREQUENCY DOMAIN APPROACH

The spatial frequency domain solution[6] to the diffraction problem relates the transform of the output wave to the transform of the input wave by a simple complex multiplication:

$$\underline{A}_o(k\Omega,\ell\Omega;z) = \underline{A}_i(k\Omega,\ell\Omega;0)e^{j\frac{2\pi z}{\lambda}\sqrt{1-(\lambda k\Omega)^2-(\lambda k\Omega)^2}} \qquad (11)$$

Realizing that the square root expression is imaginary for $(k\Omega)^2 + (\ell\Omega)^2 \geq 1/\lambda^2$ and that the contribution from these terms (the "evanescent waves") will be negligible if the propagation distance is more than several wavelengths we can simplify this expression to

$$\underline{A}_o(k\Omega,\ell\Omega;z) = \begin{cases} \underline{A}_i(k\Omega,\ell\Omega;o)\,e^{j\frac{2\pi z}{\lambda}\sqrt{1-(\lambda k\Omega)^2-(\lambda\ell\Omega)^2}} & \text{when } (k\Omega)^2+(\ell\Omega)^2 \leq \frac{1}{\lambda^2} \\ 0 & \text{when } (k\Omega)^2+(\ell\Omega)^2 \geq \frac{1}{\lambda^2} \end{cases} \qquad (12)$$

where

$$\underline{A}_i(k\Omega, \ell\Omega; o) = F\{\underline{U}_i(ma,na)\};$$
$$\underline{A}_o(k\Omega, \ell\Omega; z) = F\{U_o(ma,na)\};$$

a is the spatial sample spacing;
Ω is the frequency domain sample spacing (equals 1/Na).

The properties of this frequency domain appraoch are as follows:

1. The sample spacing in the output plane is the same as that of the input wave x ($\Delta x_o = \Delta x_i = a$). For a diverging wave we would need many more samples in the output plane to adequately describe the wave because of its larger size, so we must either have numerous zero valued samples at the input plane to adequately cover the output wave or we must restrict our coverage of the output wave to only a small portion of its breadth. Fortunately a remedy has been found to this dilemma by Sziklas and Siegman.[7] Reference 7 presents a wave transformation that converts a diverging wave diffraction problem into a collimated wave problem. The collimated wave problem is adequately handled in the frequency domain approach by the coordinate systems having equal spacings in the input and output planes. The solution to the collimated beam problem may then be used to easily find the solution to the diverging beam problem. The net effect is to obtain an effective output plane sample spacing that expands with propagation distance so that a conservative number of sample points can adequately describe both the input wave and the output wave.

2. Because the DFT assumes that the input wave samples are repeated in a periodic two-dimensional array, at some propagation distance L, the waves from the other "objects" will overlap the wave from the original object, thereby limiting the maximum propagation distance for which the frequency domain approach can be used. Figure 3 and the following analysis uses the one-dimensional slit as an example of this effect and an estimate of the maximum propagation distance. Using an approximation to the Fresnel integral for this specific object, Sziklas and Siegman[7] show that allowing ε_1% of the total wave energy in the overlapping fields at a propagation distance L requires a guard band or region of zero valued samples (as in Fig. 3) of value

$$G \geq 1 + \frac{L\lambda}{2\pi^2 a^2 \varepsilon_1} \qquad (13)$$

This equation can also be used to find the maximum propagation distance L, given an object with a certain guard band value G and an allowed amount of energy overlap (e.g. 5%). Again this result is based on the slit object and would have to be increased for objects

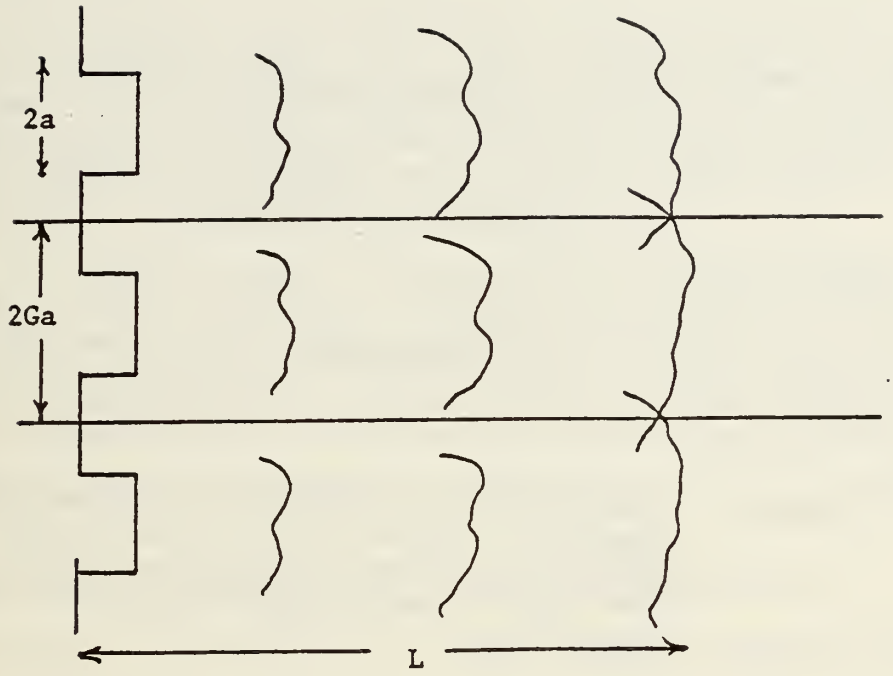


Figure 3 Hypothetical one dimensional slit object and its nearest periodic neighbors. At distance L the diffracted wavefronts are significantly overlapped.

with smaller features that would have to be resolved or decreased for objects with more tapered features than a slit.

3. It is noted from Eq. 12 that the output frequency spectrum is band-limited i.e. all frequency samples lying outside of a circle in the frequency domain with a radius of $1/\lambda$ are equal to zero (for propagation distances longer than several wavelengths). Hence only those samples of the input wave spectrum that lie within this same circle must be taken. This leads to a determination of the optimum spatial sample spacing and the fact that there is no frequency aliasing in the spatial frequency approach to the diffraction problem. Rather than restricting our frequency samples to those lying within the circle of radius $1/\lambda$ it is geometrically simpler to consider those lying in the rectangle $|u|<1/\lambda$ and $|v|<1/\lambda$ (as in Fig 4). This leads to an oversampling by 12% of the absolute minimum number of samples.

Choosing this sampling limit gives a maximum frequency in the

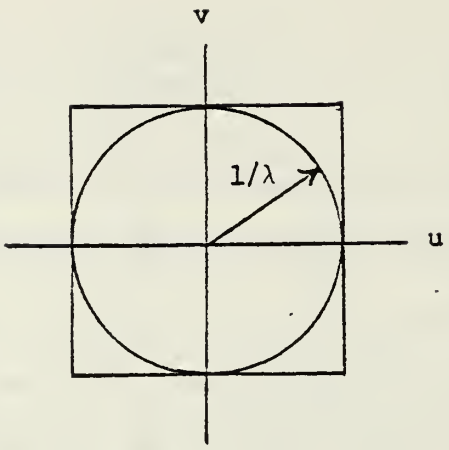


Figure 4 Frequency domain representation of bandlimited propagation showing circular limit (radius equals $1/\lambda$). Circumscribed rectangle shows geometrically simpler bandlimiting (with 12% oversampling).

U direction of

$$u_{max} = \frac{N}{2} \Delta U = \frac{1}{\lambda} \tag{14}$$

where Δu is the spatial frequency sample spacing. Hence,

$$\Delta u = \frac{2}{N\lambda} \tag{15}$$

and is also given by

$$\Delta u = \frac{1}{N \Delta x} \tag{16}$$

where Δx is the spatial sample spacing. Therefore

$$\Delta x = \frac{\lambda}{2} \tag{17}$$

is the optimum sample spacing of the object. It is noted that this sample spacing is that required to give resolution of $\lambda/2$, the "diffraction limited" resolution.

4. Knowing the required quard band size and the optimum sample spacing from properties 2 and 3 above it is now possible to compute the total number of samples by dividing the total object width (including the guard band) by the optimum spacing:

$$N = \frac{2Ga}{\Delta x} = \frac{4Ga}{\lambda} \qquad (18)$$

It should be noted that many FFT routines require that the number of samples be an integral power of 2. Hence the final value of N may be that power of two above or below the calculated value of Eq. 18. If larger the field will be oversampled (with a resulting loss of computing efficiency and longer running times) or undersampled (with shorter running time but a loss of some resolution).

5. Another property of working in the frequency domain to handle the diffraction problem is that this approach is easily amenable to frequency domain image processing techniques[8] such as Weiner filtering, edge enhancement, deconvolution of the point spread function of the receiver, etc.). Here the spectrum of the diffracted wave can be manipulated by multiplication with a filter function:

$$\underline{A}_o'(u,v) = \underline{A}i(u,v)\ \underline{H}\ \text{prop}(u,v)\ \underline{H}\ \text{filter}(u,v) \qquad (19)$$

where

$\underline{A}_o'(u,v)$ is the spectrum of the processed output wave;

$\underline{H}\ \text{prop}(u,v) = A_o(u,v)\,|\,Ai(u,v)$ is the "transfer function" for linear scalar diffraction and is found by dividing Eq. 12 by $Ai(u,v)$;

$\underline{H}\ \text{filter}(u,v)$ is the filter function to perform the desired operation.[9]

Since one of the primary advantages of computer-aided imaging is the flexibility and capability to enhance the image and extract information, the ease of incorporating this class of frequency domain operations is a major advantage of the frequency domain approach.

SUMMARY

Table I summarizes the relative advantages and disadvantages of the two diffraction approaches. With an awareness of these strengths and weaknesses the researcher attempting computer-aided imaging will be able to choose the technique most suitable for his application.

ACKNOWLEDGEMENTS

This research was supported by the Foundation Research Program of the Naval Postgraduate School and the U.S. - France Scientific Exchange program administered by the National Science Foundation and the Centre National de la Recherche Scientifique. The author would also like to acknowledge the hospitality of Professor Pierre Alais and his colleagues of the Laboratoire de Mecanique Physique of the University of Paris VI during his work there.

TABLE I Summary of comparison between diffraction techniques

FRESNEL INTEGRAL	FREQUENCY DOMAIN APPROACH
1. Expanding sample spacing for diverging wave.	1. Expanding sample spacing can be made to occur by wave transformation.
2. Limited minimum diffraction distance.	2. Unlimited minimum diffraction distance.
3. Unlimited maximum diffraction distance.	3. Requires large guard band (and more samples) for longer diffraction distances.
4. Does not predict diffraction limited diffraction.	4. Correctly predicts diffraction limited resolution.
5. Some frequency aliasing.	5. No frequency aliasing due to bandlimited nature.
5. Image processing is separate operation.	6. Frequency domain filtering techniques are easily incorporated.
7. Requires two N^2complex multiplications, two data shuffles, and one FFT operation.	7. Requires N^2 multiplications, four data shuffles, and two FFT operations.

REFERENCES

1. M.M. Sondhi, "Reconstruction of objects from their sound diffraction patterns", J. of the Acoustical Society of America, 46 (5):1158-1164, 1969.

2. F.L. Thurstone and A.M. Sherwood, "Three dimensional visualization using acoustical fields", Acoustical Holography, Vol 3, A.F. Metherell, Ed., Plenum Press, New York, pp. 317-331, 1971.

3. J. Powers, J. Landry and G. Wade, "Computed reconstructions from phase-only and amplitude-only holograms", Acoustical Holography, Vol 2, A.F. Metherell and L. Larmore, Ed., Plenum Press, New York, pp. 185-202, 1970.

4. L. B. Lesem, P.M. Hirsch and J.A. Jordan, Jr., "The kinoform: a new wavefront reconstruction device", IBM Journal of Research and Development, 13:150, 1969.

5. E. O. Brigham, The Fast Fourier Transform, Prentice Hall, Englewood Cliffs, New Jersey, 1974.

6. J. W. Goodman, Introduction to Fourier Optics, Chapts 3 and 4, McGray-Hill, New York, 1968.

7. E. A. Sziklas and A. E. Siegman, "Mode calculations in unstable resonators with flowing saturable gain. 2:Fast Fourier transform method", Applied Optics, 14(18):1874-1889, 1975.

8. M. Takagi, et al, "Image enhancement of acoustic images", Acoustical Holography, Vol 5, P.S. Green, Ed. Plenum Press, New York, pp. 541-550, 1974.

9. H. Andrews, Computer Techniques in Image Processing, Academic Press, New York, 1970.

10. D. E. Mueller, "A computerized acoustic imaging technique incorporating automatic object recognition", Engineer's degree thesis (unpublished), Naval Postgraduate School, Monterey, California, 1973.

J. P. Powers, LT R. T. O'Bryon, LT J. W. Patton, "Ultrasonic Imaging System Incorporating Computer Aided Coherent Processing" Abstracts, Ninety-fourth Meeting of the Acoustical Society of America, Miami Beach, 12-16 December 1977, in J. Acoustical Society of America, 62 (Supplement 1): S22, 1977.

ABSTRACT

11. Ultrasonic imaging system incorporating computer-aided coherent processing. J.P. Powers, Lt. R.T. O'Bryon, and Lt. J.W. Patton (Department of Electrical Engineering, Naval Postgraduate School, Monterey, CA 93940)

An ultrasonic imaging system incorporating two-dimensional coherent data processing and computer image processing has been built and tested. The system consists of two parts: data acquisition, and computer processing and display. The data acquisition portion coherently detects both the amplitude and phase of an ultrasonic diffraction pattern emitted by or reflected from an object. This diffraction pattern is then quantized and entered into the computer memory. The computer portion of the system processes the data and displays the resulting image on a high-resolution computer-controlled graphics terminal. The presence of both amplitude and phase information as the input to the processing portion of the system offers advantages in that coherent processing techniques can be used to process the data as well as incoherent techniques. This processing capability allows one to perform such operations as matched filtering, removal of the receiving transducer directivity pattern, edge enhancement, etc. Application of the system to ultrasonic imaging, acoustic transducer calibration, and studies of scalar wave propagation and phenomena will be given

Title: Sampled Analog Signal Processing

Investigator: T. F. Tao, Professor of Electrical Engineering

Sponsor: Foundation Research Program (6.2)

Objective: To develop theory, design procedure and applications of sampled analog recursive filters using charge transfer devices.

To investigate the limitations and sensitivities of sampled analog signal processing.

Summary: Signal processing can be generally separated into four categories:

Filter	Recursive Nonrecursive
Spectral Analyzers	Fourier Type Non-Fourier Type

They can be implemented either digitally using computers, integrated circuits or in sampled analog ways using charge transfer devices. Sampled analog signals differ from digital signals because their amplitudes are not digitized. They are similar because both types of signals are sampled. This study is concerned with two aspects of sampled analog signal processing.

(1) Sampled Analog Recursive Comb Filters:

Theory, design procedure and applications of sampled analog recursive filters have been investigated. It was found that recursive filters using sampled analog delay lines behave in many ways simliar to the digital recursive filters but with one major difference. Because a charge transfer delay line usually has delay of multiple sampling periods, the frequency characteristics of a sampled analog recursive filter has N/2 comb teeth below the Nyquist frequency. It is a comb filter. It was found that theories and design procedure

developed for digital recursive comb filters can be used for sampled analog recursive comb filters if two modifications are made: filter coefficients vary with frequency, charge transfer delay lines have delays of multiple sampling periods. Two types of comb filters have been studied: canceller type and integrator type. Their applications as MTI radar canceller and pulse-repetition-frequency sorter have been demonstrated.

(2) Sampled Analog Discrete Fourier Transforms:

Two algorithms have been used to implement sampled analog discrete Fourier transforms: Chirp Z Transform and Prime Transform. Both algorithms use the transversal filter which can be easily implemented using charge transfer devices. This study investigated the advantages and limitations of these two sampled analog discrete Fourier transforms in comparison with the digital FFT (fast-Fourier-transform) algorithm.

Publications:

S. Holmes, "Hardware Signal Processor Development-Session Summary", Proceedings 9th Asilomar Conference on Circuits, Systems, and Computers, pp. 289-299 (1976).

S. Holmes, M. Pollack and J. Campbell, "Sensitivity Study of Sampled Analog Signal Processing", Proceedings 19th IEEE Midwest Symposium on Circuits and Systems, pp. 430-439, (1976)

L. Saetre, S. Holmes and A. Ejaz, "Recursive CTD Comb Filter and Its Application to MTI Radar Signal Processing", Proceedings 1976 GOMAC Conference, (Government Microcircuit Applications Conference), pp. 246-249 (1976).

J. Campbell, T. F. Tao and M. Pollack, "Sensitivity Study of Chirp Z Transform and Prime Transform as Sampled Analog Discrete Fourier Transform Algorithms", Proceedings of 10th Asilomar Conference on Circuits, Systems and Computers, pp. 372-38- (1976)

T. F. Tao, S. Holmes A. Ejaz, F. Piazza, L. Saetre and B. Freund, "Theory and Applications of Sampled Analog Recursive CTD Comb Filters", Proceedings NASA/JPL Conference on Charge Coupled Device Technology and Applications, pp. 1-10, (1976)

Theses Directed:

J. Campbell, "Sensitivity Study of the Chirp Z Transform and the Prime Transform as Sampled Analog DFT Algorithms", Master's Thesis, December 1976

A. Ejaz, "Theory of Sampled Analog Recursive Comb Filters and Their Canceller Applications", Engineer Degree, December, 1976.

F. Piazza, "Theory and Applications of CTD Recursive Comb Filters", Master's Thesis, December 1976.

Sklivanos, "Sensitivity Study of Sampled Analog Tapped Delay Line Transversal Filters," Master's Thesis, September 1976.

ABSTRACT

Two different sampled analog Discrete Fourier Transform (DFT) algorithms, the Chirp-Z Transform (CZT) and the Prime Transform, are examined in a computer simulation to evaluate their sensitivity to implementation errors of various kinds. Methods of sensitivity study in analog and digital signal processing are summarized to provide a framework for the sampled analog study. The CZT and the Prime Transform are compared with one another and with the digital Fast Fourier Transform, and relative accuracies and design tradeoffs are noted.

Campbell, James G., "A Sensitivity Study of the Chirp-Z Transform and the Prime Transform as Sampled Analog DFT Algorithms," Master's Thesis, December 1976.

ABSTRACT

Sampled analog comb filter using a recursive filter implementation is studied. SAD-100 delay lines are used as the delay elements. The similarities and differences between sampled analog recursive filter and digital recursive filter are pointed out. Canceller type comb filters using the first order or the second order canonical circuits are studied. Experimental results of frequency response are presented to show their difference from that of digital filters. A theoretical analysis is presented based essentially on the digital filter theory but modified to take into account the differences described above. MTI-simulation was performed, and the ability of the CTD-filter to cancel clutter and pass doppler frequencies was demonstrated. The characteristic curves of the Fairchild CCD-321 have been obtained.

Ejaz, Arif, "Theory of Sampled Analog Recursive Comb Filters and Their Canceller Application," Master's Thesis, December, 1976.

ABSTRACT

Both integrator and canceller types of sampled analog comb filters implemented by a canonical second order recursive circuit have been studied experimentally and theoretically. Two 96 stage CTD (charge transfer device) delay lines are used in the experimental study. Theoretically, an analysis based on a modification of the digital recursive filter theory was developed and accounted for the major features of the experimental results.

Three comb filter applications have been demonstrated. The first two used the integrator type comb filter for the improvement of the signal-to-noise ratio of a periodic signal contaminated by noise and for the sorting of a periodic signal of selected PRF from other periodic signals. The third application used a canceller type comb filter to eliminate a sinusoidal interference from a desirable signal.

Piazza, Frank, "Theory and Applications of CTD Recursive Comb Filters," Master's Thesis, December 1976.

ABSTRACT

This thesis consists of two parts, one theoretical and one experimental. Theoretically, sensitivity of the transversal filter using a sampled analog tapped delay line is studied. Low pass filters of different delay stages, cut off frequencies and transition bands are designed by the Parks-McClellan procedure (which minimizes the maximum error). Tapping errors are modeled by uniformly distributed random variables. The effects of these errors on the filter's performance are analyzed by a computer simulation program. It was found that although longer TDL filters yield better perfromance, under certain conditions they have poorer sensitivity. In the second part a new 32 TADs sampled analog Bucket Brigade Device (BBD) tapped delay line is experimentally evaluated to characterize its errors.

Sklivanos, George S., "Sensitivity Study of Sampled Analog Tapped Delay Line Transversal Filters," Master's Thesis, September 1976.

Papers presented by T. F. Tao and students from 1973 to 1976 in the areas of Sampled Analog Signal Processing and Charge Coupled Devices based on students' theses results and partially supported by the Naval Postgraduate School Research Foundation

LIST OF PRESENTATIONS AND PAPERS

*1. "Feasibility Study of PbTe and PbSnTe Infrared Charge Coupled Imager"

T. F. Tao, J. R. Ellis, L. Kost and A. Doshier
Proceedings 1973 International Charge Coupled Device Applications Conference, September, 1973, San Diego, California, pp.259-268

2. "Sampled Analog CCD Recursive Comb Filter"

T. F. Tao, V. Iamsaad, S. Holmes, B. Freund, L. Saetre and T. Zimmermann
Proceedings 1975 International Charge Coupled Device Applications Conference, October, 1975, San Diego, California, pp. 257-266

3. "Hardware Signal Processor Development — Session Summary"

T. F. Tao and S. Holmes
Proceedings 9th Asilomar Conference on Circuits, Systems and Computers, November, 1975, Asilomar, California, pp. 289-299

4. "Sensitivity Study of Sampled Analog Signal Processing"

T. F. Tao, S. Holmes, M. Pollack and J. Campbell
Proceedings 19th IEEE Midwest Symposium on Circuits And Systems, August, 1976, Milwaukee, Wisconsin, pp. 430-439

5. "Present Status of Real Time Signal Processor Development" (not available)

T. F. Tao, S. Holmes and K. Pfennig
1976 Military Defense Electronics Expo Conference
Weisbaden, West Germany, October 6-8, 1976

6. "Recursive CTD Comb Filter and Its Applications to MTI Radar Signal Processing"

T. F. Tao, L. Saetre, S. Holmes and A. Ejaz
Proceedings 1976 GOMAC Conference (Government Microcircuit Applications Conference),
November, 1976, Orlando, Florida, pp. 246-249

7. "Sensitivity Study of the Chirp Z Transform and the Prime Transform as Sampled Analog Discrete Fourier Transform Algorithms"

J. Campbell, T. F. Tao and M. Pollack
Proceedings 10th Asilomar Conference on Circuits, Systems and Computers, November, 1976, Asilomar, California, pp. (Proceedings not ready yet)

8. "Theory and Applications of Sampled Analog Recursive CTD Comb Filters"

T. F. Tao, S. Holmes, A. Ejaz, F. Piazza, L. Saetre and B. Freund
Proceedings NASA/JPL Conference on Charge Coupled Device Technology and Applications,
December, 1976, Washington, D. C.

* not supported by Research Foundation

PAPER NO. 1

1973 INTERNATIONAL CHARGE COUPLED DEVICE APPLICATIONS
CONFERENCE, SAN DIEGO, CALIFORNIA
SEPTEMBER 18-20, 1973

FEASIBILITY STUDY OF PbTe AND $Pb_{0.76}Sn_{0.24}Te$ INFRARED CHARGE COUPLED IMAGER*

T. F. Tao, J. R. Ellis, L. Kost and A. Doshier

Naval Postgraduate School, Monterey, California

ABSTRACT

The feasibility of using MIS structures of PbTe and $Pb_{0.76}Sn_{0.24}Te$ for infrared charge coupled imaging in the 3 to 5 and the 8 to 14 microns regions have been examined. First, the storage capacity, dark current and storage time of the MIS capacitors of these two semiconductors were calculated at T=85°K based on the MOS theory developed for Si. P-type semiconductors of doping density $5x10^{16}cm^{-3}$ and a minority carrier lifetime of 10^{-7} sec were used. Four types of insulators were considered : SiO_2, Al_2O_3, Ta_2O_5 and TiO_2. TiO_2 was found to be better suited for CCI applications because of its high dielectric constant. Based on these MIS results, the feasibility of IRCCI for 300°K scene was examined. The background photon flux was used to produce a 50% fat zero which determined the exposure time. It was assumed that 10% of the storage time can be used for CCI operation. At a clock frequency of 2 MHz and without considering the transfer inefficiency,20,000 transfers can be made in PbTe CCI within this usable storage time. However, only 20 transfers can be made in $Pb_{0.76}Sn_{0.24}Te$ CCI. The 300°K background current was calculated and was found to be approximately 3000 times larger than the signal current for a 0.1°K temperature resolution for both semiconductors. Preliminary experimental studies of PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS capacitors were also made using Hg drops as metal contacts on MIS samples consisting of Al_2O_3 or SiO_2 layers e-gun evaporated on thin film semiconductors. C-V measurements indicated that accumulation, depletion and inversion layers existed at the interface and can be controlled by the gate voltage in a manner qualitatively similar to that of Si MOS. However, measured and calculated C-V characteristics have not been in close agreement.

It is concluded that PbTe IRCCI is feasible. For $Pb_{0.76}Sn_{0.24}Te$, CCI is questionable unless improvement in minority carrier life time to a value significantly above 10^{-7} sec can be achieved.

I. INTRODUCTION

Recent progress[1-5] has confirmed the expectation that Si charge coupled devices[6] will be one of the most important visible imagers. This paper deals with the question if CCD can also be developed for infrared imaging. Several approaches have been mentioned[7] for developing the Infrared Charge Coupled Imager (IRCCI). Some of them are hybrid circuits in which the infrared detectors are combined with Si CCD. This study considered the monolithic approach in which a metal-insulator-semiconductor array of narrow gap semiconductor is used both as the sensor and the transfer device. Its structure is similar to the Si CCD except that a semiconductor of smaller energy gap is used instead of Si. This paper will present the preliminary results of the feasibility study of using MIS structures of two 4-6 narrow gap semiconductors as IRCCI. The semiconductors considered are PbTe for a 3 to 5 microns imager and $Pb_{0.76}Sn_{0.24}Te$ for an 8 to 14 microns imager , both operated at a temperature of T=85°K. Since the MIS properties of these semiconductors have not been studied before, this feasibility

study did not examine the aspects of CCI which require information on the interface states, such as the problems of transfer inefficiencies and noises, etc. Instead, the feasibility study was based on the theoretical calculation of the MIS properties of single PbTe and $Pb_{0.76}Sn_{0.24}Te$ capacitors. Experimental studies of these MIS have been started, however, to examine the validity of the theoretical calculations. Their preliminary results will be reported in this paper also.

In Section II, the storage capacity, dark current and storage time of both the PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS were calculated based on the MOS theory developed for Si. Four insulators, SiO_2, Al_2O_3, Ta_2O_5 and TiO_2 were considered. It was found that TiO_2 is better than the others because of its higher dielectric constant. Based on a sample MIS using 1000A of TiO_2, the feasibility of IRCCI for 300°K scene was examined by comparing the transfer time and the storage time for a clock frequency of 2 MHz. The background photon flux was used to produce a 50% fat zero which determined the exposure time. Both the 300°K background current and the signal current of a 0.1°K temperature resolution were calculated. In Section IV, the preliminary results of the experimental study of the PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS using Al_2O_3 and SiO_2 layers as insulators are presented. The measured C-V characteristics will be used as evidence of the existence of accumulation, depletion and inversion layers. They are also compared with the theoretical calculation to examine if the behaviors of the present PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS samples can be well described by the established MOS theory.

II. PROPERTIES OF PbTe AND $Pb_{0.76}Sn_{0.24}Te$ MIS

The basic building block of a charge coupled device is the MIS capacitor. A thorough understanding of its properties is mandatory before the feasibility of the IRCCI can be examined. In this section, the properties of the PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS relevant to the CCI operations will be presented. Specifically, the storage capacity Q_s, dark current I_d and the storage time τ_s were calculated based on the following theories. First, the MIS capacitances for the deep depletion and high frequency inversion states were calculated by the depletion approximation[9,10] which was shown to agree very closely to the exact theory if the gate voltage is considerably higher than the threshold. The storage capacity, Q_s, was calculated from the relation

$$Q_s = V_G (C_{HINV} - C_{DD})$$

where V_G = applied gate voltage
C_{HINV} = high frequency inversion capacitance
C_{DD} = deep depletion capacitance

Its basis is shown in Fig. 1. In (a), the normalized MIS capacitance C/C_I is presented as a function of the gate voltage, V_G, for a p type semiconductor. The CCD operation takes place between the deep depletion state and the high frequency inversion state. This region is described in a different way in (b) which presents C/C_I as a function of time. The MIS was originally biased in accumulation, then suddenly switched to deep depletion which is not stable and will be changed to inversion after a period of time, called the storage time τ_s, due to the appearance of minority carriers contributed by the dark currents.[11]

$$\tau_s = \frac{Q_s}{\sum_i I_{di}}$$

The dark currents consist of carriers generated from several sources: one from the depletion region, one from the bulk semiconductor beyond the depletion region, one from the surface generation and recombination and one from the tunneling between the bulk and the surface inversion layer. In this study, the last two components were not considered. The dark current due to the surface generation and recombination was neglected although it contributed a significant portion of dark current in the indirect gap semiconductor, Si. It is believed that in narrow gap semiconductors like PbTe and $Pb_{0.76}Sn_{0.24}Te$, the direct band transition will dominate the generation even at the surface. The dark current due to the tunneling was also neglected because at the bias voltage of interest, the depletion layer width is more than 4000 A which can be seen in Table I and II to be presented later. The tunneling current through this barrier width was much smaller compared with the other two dark currents which are given by the following relations.[11]

I_{gd} = generation within the depletion region

$$= \frac{n_i \sqrt{\epsilon_s q \phi_s}}{\tau \sqrt{2N}}$$

I_{gb} = generation in the bulk semiconductor

$$= \frac{q\, n_i^2}{N \tau} L = \frac{q\, n_i^2}{N \tau} \sqrt{\mu \frac{kT}{q} \tau}$$

where n_i = intrinsic carrier con centration
τ = minority carrier life time
N = doping density
μ = mobility
L = diffusion length
ϕ_s = surface potential

In Table I and II, Q_s , x_{dd} , I_d , τ_s of the PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS are presented at a gate voltage of ten volts. x_{dd} is the deep depletion width. Four insulators of two thicknesses, 1000 and 2000 A, were considered. The insulator types and their dielectric constants are the following:

SiO_2	3.9	Al_2O_3	8.8
Ta_2O_5	50	TiO_2	75

The temperature was 85°K. p type semiconductors of doping density of 5×10^{16} cm^{-3} were used. For PbTe and $Pb_{0.76}Sn_{0.24}Te$, the following semiconductor properties are used.[8,12]

E_g = energy gap in ev
$= 0.181 + 4.52 \times 10^{-4} T - 0.568 X + 5.8 X^4$

m_d = density of states effective mass
$= (1.12 E_g) m_0$

Dielectric constant = 400
μ = mobility = 2×10^4 $cm^2/v\text{-}sec$
τ = minority carrier life time
$= 10^{-7}$ sec

where X is the percentage of Sn content. From the Table, it can be seen that for MIS having 2000 A of SiO_2 as an insulating layer, a gate voltage of 10 volts i s not even large enough to reach inversion. In MIS having 1000 A of SiO_2 , 10 volt is not very far above the threshold. Consequently, the Q_s is only 3×10^9 electrons/cm^2 for PbTe MIS. This is the result of a large difference in the dielectric constants of the insulator and the semiconductor. In the case of Al_2O_3, although its dielectric constant is higher at 8.8, the increase in Q_s is not dramatic. On the other hand, MIS using either Ta_2O_5 or TiO_2 as insulating layers showed definitely superior properties for CCI applications because their dielectric constants are considerably higher. In these cases, although the dark current was increased somewhat due to the increase of depletion width, the storage capacity is considerably larger which results in longer storage time. Such combination of larger Q_s and longer τ_s is desirable for CCI operation. Consequently, for the feasibility study of IRCCI to be presented in the next section, a MIS capacitor having 1000 A of TiO_2 was used as the standard sample. The storage capacity is approximately 1×10^{13} electrons/cm^2 for both PbTe and $Pb_{0.76}Sn_{0.24}Te$ at a V_G of 10 volt. The storage time is approximately 10^{-1} sec for PbTe and 10^{-4} sec for $Pb_{0.76}Sn_{0.24}Te$.

III. FEASIBILITY OF PbTe AND $Pb_{0.76}Sn_{0.24}Te$ IRCCI

Based on these properties of a single MIS capacitor, the feasibility of PbTe and $Pb_{0.76}Sn_{0.24}Te$ IRCCI for 300°K scene were examined by comparing the transfer time with the storage time and by comparing the background and signal currents. The exposure time was selected such that the 300°K background photons will provide a 50% fat zero. Since the properties of interface states can not be theoretically predicted, the feasibility study did not examine other performances of CCI which require the knowledge of interface states, such as the transfer inefficiency and noise etc.

The IRCCI was assumed to have the following band pass characteristics. The cutoff on the longwavelength side was determined by the photoresponse threshold of the semiconductor. At 85°K, it is 5.66 microns for PbTe and 12.16 microns for $Pb_{0.76}Sn_{0.24}Te$. Infrared filters were used to provide the cutoff on the short wavelength side. Together, the bandpass was from 3.5 to 5.66 microns for PbTe and from 9 to 12.16 microns for $Pb_{0.76}Sn_{0.24}Te$. The photon flux of the 300°K background and of a 0.1°K temperature resolution were calculated from the standard black body formula. It was found that the 300°K background photon flux is much greater than the signal photon flux. In order that the backgound photons not flood the storage capacity, the exposure time must be limited. It was proposed that the 300°K background photons provide a 50% fat zero. Since the storage capacity of the standard MIS sample used in this study is 1×10^{13} electrons/cm^2, the required exposure time is 1.54×10^{-4} sec

for PbTe and 9.83×10^{-6} sec for $Pb_{0.76}Sn_{0.24}Te$. Both exposure times are too short for practical applications although it probably is tolerable in the case of PbTe. For PbTe, the storage time is then 425 times longer than the exposure time. For $Pb_{0.76}Sn_{0.24}Te$, it is only six times longer.

It is further assumed that CCI operation can take place only within one tenth of the storage time. Beyond this usable fraction, the dark current will contribute too many minority carriers and smear out the signal carriers. Therefore, the usable storage time is approximately 10^{-2} sec for PbTe and only 10^{-5} sec for $Pb_{0.76}Sn_{0.24}Te$. If a clock frequency of 2 MHz is used, the transfer time per gate is 0.5×10^{-6} sec. Based on the consideration of time allowed for transfer alone, 20,000 transfers can be made within this usable storage time for PbTe. It should be noted that if transfer inefficiency is considered, the signal will deteriorate before this large number of transfers. In the case of $Pb_{0.76}Sn_{0.24}Te$, only 20 transfers can be made which is not enough for practical applications.

At the same clock frequency of 2 MHz, the current due to the 300K background is 1.6 amp/cm^2. The signal current of a 0.1°K temperature resolution of the 300°K scene is 4.8×10^{-3} amp/cm^2 for PbTe and 3.88×10^{-3} amp/cm^2 for $Pb_{0.76}Sn_{0.24}Te$. The ratio of the background current to signal current is therefore approximately 3,000 in both cases. Such a ratio, although quite high, can be handled by different signal processing schemes, such as a.c. coupling or the potentially promising recycling capability of the CCD which might be used to cancel the background.[2]

IV. EXPERIMENTAL STUDIES OF PbTe AND $Pb_{0.76}Sn_{0.24}Te$ MIS

Preliminary experimental investigations of the PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS have also been carried out. MIS capacitors were fabricated using PbTe and $Pb_{0.76}Sn_{0.24}Te$ thin films deposited on either CaF_2 and BaF_2 substrates and Al_2O_3or SiO_2 thin insulator films deposited by the e-gun method. The metal contact was provided by a Hg drop. The semiconductor films were typically 3 to 8 microns thick. They were all p-type with carrier concentrations in the high 10^{17} to low 10^{18}/cm^3 range. Based on the theoretical calculation, in order to obtain a reasonably low threshold voltage around one to two volts, the thickness of Al_2O_3 or Si O_2 layer should be less than 500A. Thin layers from 100 to 450A have been made. Probably because of the small thickness, considerable leakages were found in these layers. However, in spite of the leakage, capacitance-voltage measurements have been obtained as shown in Figures 2 and 3. Both measurements were made at room temperature using a Bootan 72A capacitance meter. The signal frequency was 1 MHz. In Figure 2, the normalized C/C_I of four gates on a PbTe MIS are shown. The insulating layer was 450A of Si O_2. The general behaviors of the C-V characteristics are similar to that of Si MOS in their early years of development. The decrease in capacitance when the gate voltage was increased from negative values toward positive values can be interpreted as evidence of the change from accumulation to depletion and then to high frequency inversion. The small peak and slight increase near and above zero gate voltage could be caused by weak inversion. A comparison was made between the measured C/C_I and a theoretical calculation using a carrier concentration of $1 \times 10^{18}/cm^2$. The calculated result is presented as a dotted line in the same figure. It can be seen that they are not in close agreement. The same large discrepancy was also found in Figure 3 which presents the C/C_I-V_G characteristics of two Hg gates on a $Pb_{0.76}Sn_{0.24}Te$ MIS capacitor. 100A of Al_2O_3 was used as the insulator. The theoretical C/C_I-V_G characteristics is also presented as a dotted line in the same figure. The lack of close agreement is clearly seen.

In spite of these discrepancies, it is believed that the MOS theory developed for Si should be valid for these semiconductors also. The lack of agreement between measurements and calculations probably is caused by the values of dielectric constants used in the calculation which may not be correct for the sample. For the 4-6 semiconductors, it is possible that their dielectric constants are not as high as 400. For the imperfect and thin insulating layers, the dielectric constants are likely higher than the bulk values of 3.9 and 8.8 for SiO_2 and Al_2O_3, respectively.

V. CONCLUSION

The feasibility of using MIS structures of narrow gap semiconductors PbTe and $Pb_{0.76}Sn_{0.24}Te$ for infrared charge coupled

imaging has been studied. Because of the unusually large dielectric constants of these semiconductors, it was found that insulators of large dielectric constants such as TiO_2 and Ta_2O_5 are more suitable than SiO_2 and Al_2O_3. For an insulating layer with a large dielectric constant, adequate storage capacity can be obtained with moderately high gate voltage and an insulating layer thickness not too small. It was found that using 1000A of TiO_2, storage capacity of $1x10^{13}$ electrons/cm^2 can be obtained at a gate voltage of 10 volts. Assuming that the generation of minority carriers in depletion region and in the bulk semiconductor are the main contributions to the dark current, the storage time was calculated to be approximately 10^{-1} sec for PbTe and 10^{-4} sec for $Pb_{0.76}Sn_{0.24}Te$. Based on these properties, the feasibility of IRCCI using these semiconductors were examined. The exposure time was selected such that the 300°K background photons did not flood the storage capacity. They were used to provide a 50% fat zero. At a clock frequency of 2 MHz, and not considering the limitation imposed by the transfer inefficiency, 20,000 transfers can be made within the usable storage time for PbTe. On the other hand, only 20 transfers can be made in $Pb_{0.76}Sn_{0.24}Te$. It is concluded that PbTe IRCCI is feasible. For $Pb_{0.76}Sn_{0.24}Te$, IRCCI is questionable unless significant material improvements can be made to increase the minority carrier lifetime considerably above 10^{-7} sec.

Preliminary experimental studies of PbTe and $Pb_{0.76}Sn_{0.24}Te$ MIS have also been carried out to examine if MIS of these narrow gap semiconductors follow very well the standard MOS theory. Since the technology of TiO_2 has hardly been developed, e-gun evaporated Al_2O_3 or SiO_2 layers were used in the fabrication of MIS capacitors. The high frequency (1 MHz) C-V characteristics measured at room temperature behaved qualitatively like that of Si MOS indicating that the interface of these MIS showed accumulation, depletion and inversion behaviors very similar to Si MOS. However, measured and calculated C-V characteristics have not been in agreement. This lack of agreement does not imply that the MOS theory is not valid for these semiconductors. Instead, the discrepancy may be caused by the values of dielectric constants used in the calculation. It should be emphasized that the success in experimentally obtaining C-V characteristics qualitatively similar to that of Si MOS indicated that a fruitful experimental study can be carried out to develop MIS of these semiconductors and to realize their potential as IRCCI.

ACKNOWLEDGEMENT

The authors wish to thank Drs. N. Bottka and V. Rehn of Michelson Laboratory of NWC for depositing the Al_2O_3 and the SiO_2 layers and technical discussions, Dr. D. Barbe of Naval Research Laboratory for the computer program to calculate the C-V characteristics, Dr. T. Zimmerman of TRW Systems for review of the manuscript and Mr. R. Zahm for assistance in many aspects of this study.

REFERENCES

1. R. D. Melen and J. D. Meindl, ISSCC Digest of Technical Papers, P. 130-131; Feb., 1973.

2. P. K. Weimer, W. S. Pike, M. G. Kovac and F. V. Shallcross, ISSCC Digest of Technical Papers, P. 132-133; Feb., 1973.

3. M. H. White, D. R. Lampe, I. A. Mack and F. C. Blaha, ISSCC Digest of Technical Papers, P. 134-135; Feb., 1973.

4. M. Tompsett, W. Bertram, D. Sealer and C. Sequin, Electronics, P. 162-169; Jan., 1973.

5. C. Sequin, D. Sealer, W. Bertram, M M. Tompsett, R. Buckley, T. Sharkoff and W. McNamara, IEEE Trans. ED-20, P. 244-252; 1973.

6. W. Boyle and G. Smith, BSTJ, P. 587-593; Apr., 1970.

7. D. Barbe, IEEE CCD Work Shop, Feb. 12, 1973, New York City.

8. I. Melngailis and T. C. Harman, Ch. 4, "Semiconductors and Semimetals", Ed. R. K. Willardson and A. Beer, P. 111-174, Academic Press, 1970.

9. A. S. Grove, "Physics and Technology of Semiconductor Devices", Ch. 9, Wiley 1967.

10. A. S. Grove, B. E. Deal, E. H. Snow and C. T. Sah, Solid State Electron, 8, P. 145-163; 1965.

11. G. Amelio, W. Bertram and M. Tompsett, IEEE Trans. ED-18, P. 986-994; 1971.

12. T. F. Tao and C. C. Wang, Technical Report AFML-TR-71-238, Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio, Dec., 1971.

13. W. Opyd, M.Sc. Thesis, Naval Postgraduate School, June 1973.

*Supported in part by the Air Force Materials Laboratory and the Office Naval Research.

TABLE I. Properties of PbTe MIS at T=85°K and a Gate Voltage of 10 volts

Insulator	Thickness (A°)	Q_s (e/cm^2)	Xdd (μ)	Id (amp/cm^2)	τ_s (sec)
SiO_2	1000 2000*	3.07×10^9 ---	0.422 ---	4.9×10^{-6} ---	1×10^{-4} ---
Al_2O_3	1000 2000	3.94×10^{11} 1.64×10^{10}	0.89 0.47	1.03×10^{-5} 5.5×10^{-6}	6.15×10^{-3} 4.78×10^{-4}
Ta_2O_5	1000 2000	1.14×10^{13} 4.47×10^{12}	2.28 1.78	2.64×10^{-5} 2.06×10^{-5}	6.76×10^{-2} 3.48×10^{-2}
TiO_2	1000 2000	1.62×10^{13} 8×10^{12}	2.49 2.09	2.88×10^{-5} 2.43×10^{-5}	9×10^{-2} 5.29×10^{-2}

* Gate voltage of 10 volts is below threshold of inversion.

P type PbTe, $N_A = 5 \times 10^{16}$ cm^{-3}, minority carrier lifetime = 10^{-7}sec, $\mu = 2 \times 10^4$ $cm^2/v\text{-sec}$.

TABLE II. Properties of $Pb_{0.76}Sn_{0.24}Te$ MIS at T=85°K and a Gate Voltage of 10 volts

Insulator	Thickness (A°)	Q_s (e/cm^2)	Xdd (μ)	Id (amp/cm^2)	τ_s (sec)
SiO_2	1000 2000*	2.86×10^{10} ---	0.42 ---	6.86×10^{-3} ---	6.68×10^{-7} ---
Al_2O_3	1000 2000	5.14×10^{11} 4.87×10^{10}	0.89 0.47	1.19×10^{-2} 7.42×10^{-3}	6.92×10^{-6} 1.05×10^{-6}
Ta_2O_5	1000 2000	1.34×10^{13} 5.23×10^{12}	2.28 1.78	2.71×10^{-2} 2.16×10^{-2}	7.9×10^{-5} 3.88×10^{-5}
TiO_2	1000 2000	2×10^{13} 9.46×10^{12}	2.49 2.09	2.93×10^{-2} 2.5×10^{-2}	1.1×10^{-4} 6.05×10^{-5}

* Gate Voltage of 10 volts is below threshold of inversion.

P type $Pb_{0.76}Sn_{0.24}Te$, $N_A = 5 \times 10^{16} cm^{-3}$, minority carrier lifetime = 10^{-7} sec, $\mu = 2 \times 10^4$ $cm^2/v\text{-sec}$.

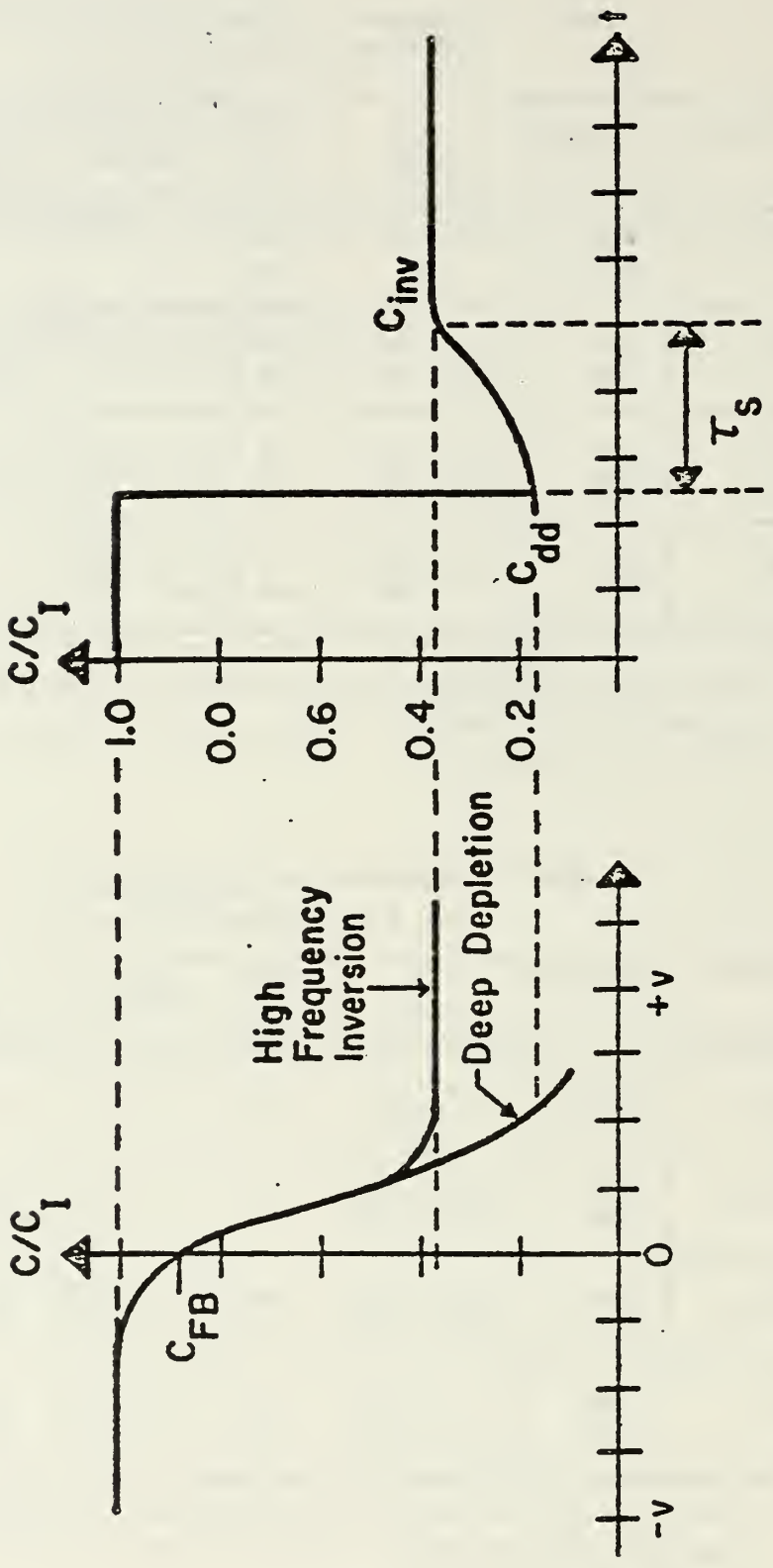


Fig. 1 C-V and C-t Charateristics of a MIS Capacitor

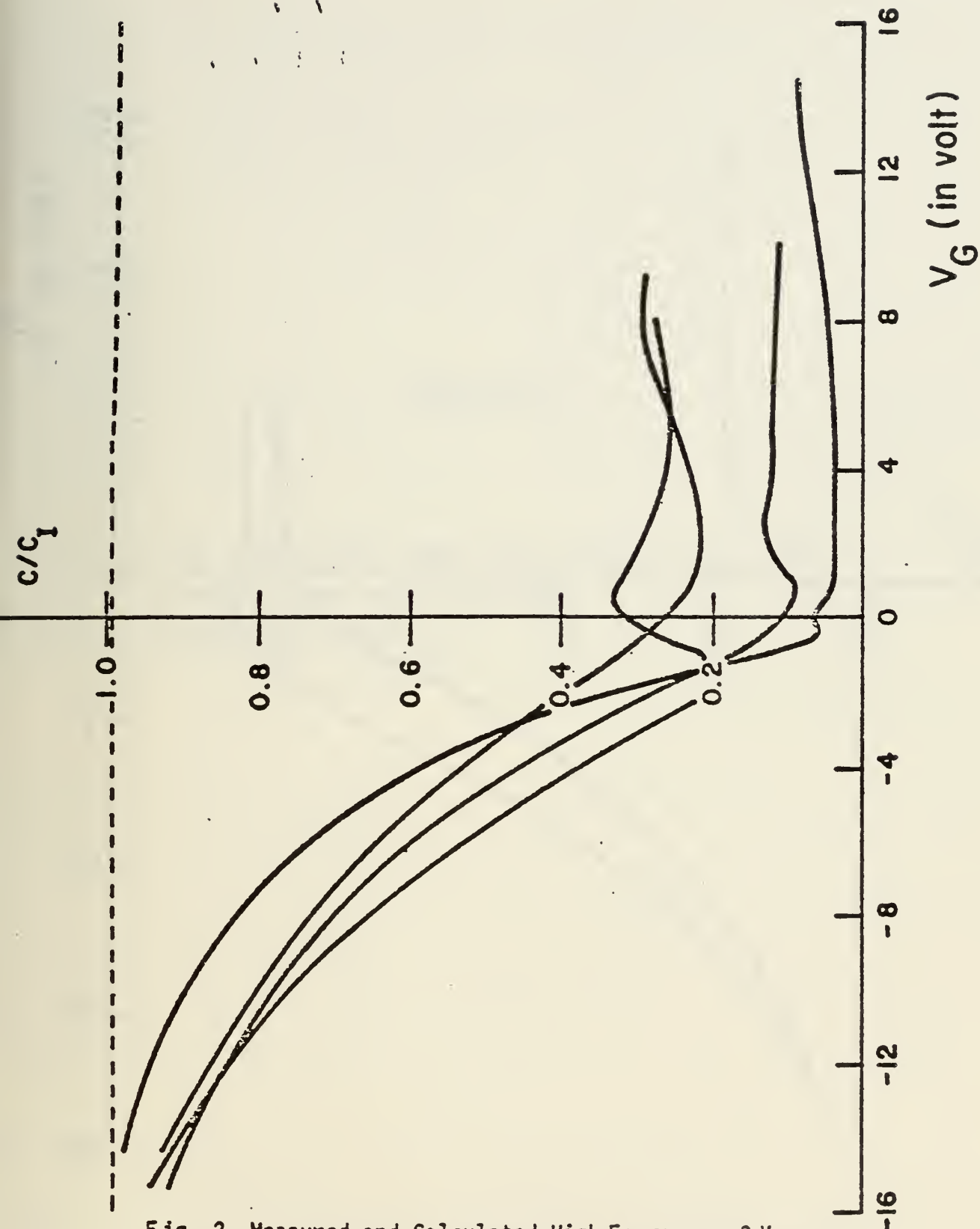


Fig. 2 Measured and Calculated High Frequency C-V Charateristics of PbTe MIS Capacitor
p type PbTe, $N_A = 1 \times 10^{18}/cm^3$
450 A SiO_2 , T = 300°K

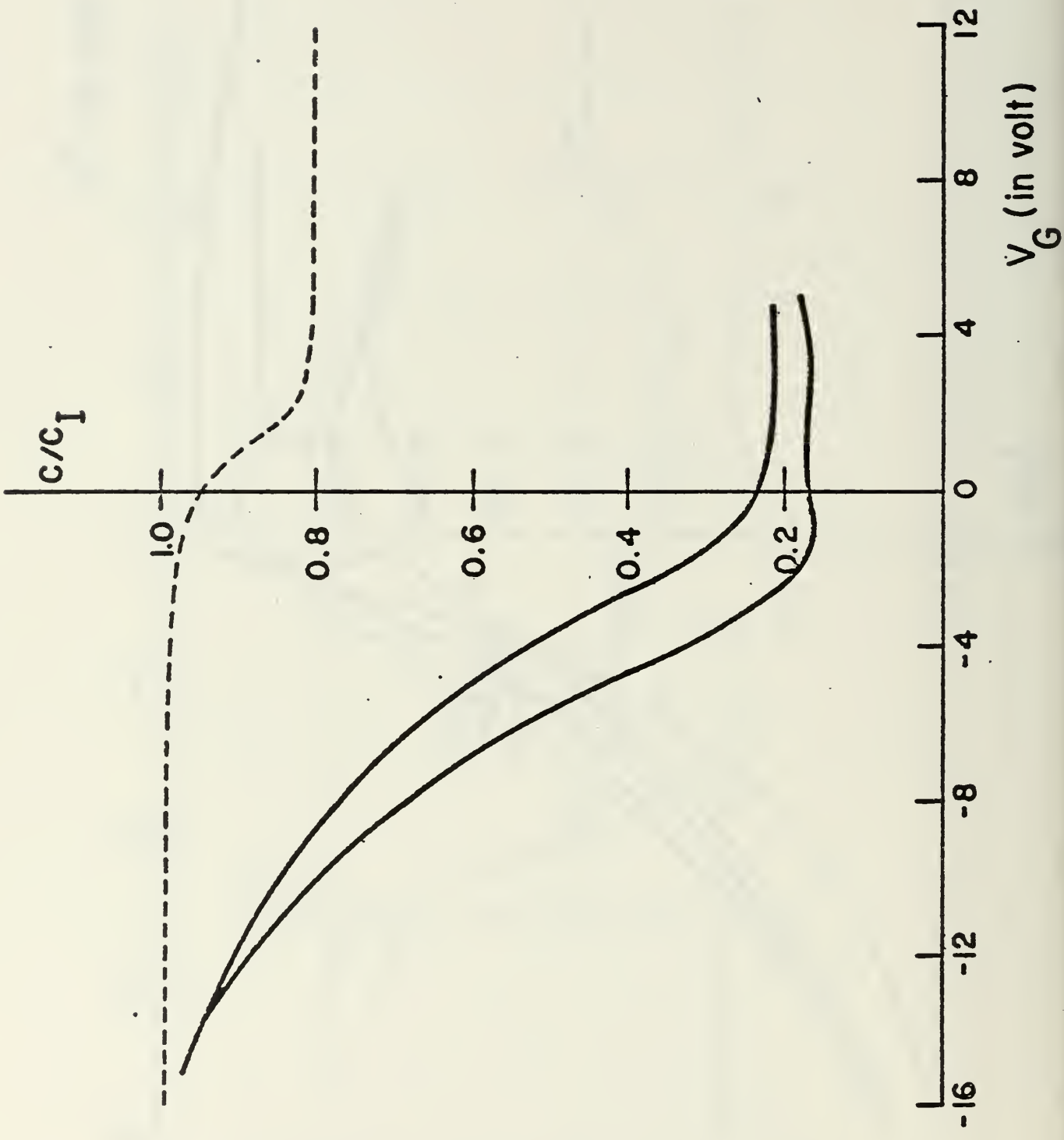


Fig. 3 Measured and Calculated High Frequency C-V Characteristics of $Pb_{0.76}Sn_{0.24}Te$ MIS Capacitor
p type $Pb_{0.76}Sn_{0.24}Te$, $N_A = 1 \times 10^{18}$ /cm^3
100 A Al_2O_3, T = 300°K

PAPER NO. 2

1975 INTERNATIONAL CHARGE COUPLED DEVICE APPLICATIONS
CONFERENCE, SAN DIEGO, CALIFORNIA
OCTOBER 29-31, 1975

SAMPLED ANALOG CCD RECURSIVE COMB FILTERS

T.F. Tao, V. Iamsaad, S. Holmes, B. Freund, L. Saetre
NAVAL POSTGRADUATE SCHOOL
Monterey, California

and

T.A. Zimmermann
TRW SYSTEMS
Redondo Beach, California

ABSTRACT - Sampled analog comb filters using a recursive implementation with CC the delay elements are studied. A theoretical analysis based essentially on digital re sive filters is developed with modifications taking into account that a CCD delay gene consists of more than one delay stage. Using the standard Z and the bilinear Z trans forms, transfer functions of a lowpass and a highpass continuous filter are transform into H(Z) forms suitable for CCD implementations. Design formulas for both the can celler type and integrator type comb filters using one CCD delay are obtained. Six di ent comb filters are implemented using an 8-bit two-phase P surface channel CCD clo at 20 kHz. The agreement between measurements and theoretical calculations is enc aging. Possible causes for the deviation from theory are discussed. The bilinear Z transform design procedure is shown to be superior because it provides a zero in the transfer function which leads to theoretically infinite attenuation at a series of period cally separated null frequencies. Applications of this type of comb filter to either suppress or to enhance a signal having periodic spectrum are suggested. The work p formed was partially supported by the Naval Electronics Systems Command and the N Postgraduate School Research Foundation.

1. INTRODUCTION

Designers of sampled analog CCD signal processors have concentrated mainly on transversal filters,[1,2,3] correlators,[4] chirp Z transformers,[5,6] and two-dimensional transforms.[7,8] This paper discusses sampled analog recursive filters, which have not received as much prior attention. They were first studied using a BBD as the delay elements[9,10] and, recently, a three-pole[11] and a two-pole/one zero[12] CCD recursive filter and a one-pole CCD recursive integrator[13] have been reported.

This paper addresses three aspects of sampled analog CCD recursive comb filters: theory, experimental results, and possible applications. The general characteristics of comb filters are highlighted in Section 2, and their recursive implementations are described. In Section 3, a theoretical analysis based essentially on digital recursive filter theory is presented. Modifications of the digita theory to fit the sampled analog recur filter case are provided. Measured r sults of recursive comb filters using 8-bit, two-phase, P surface channel C delay line are presented and compare with theoretical calculations in Sectio The agreement is generally close at l frequencies but deteriorates as the fr quency is increased or as the order o comb teeth is increased. However, t feasibility of using a CCD to implemen recursive comb filter is confirmed. sible applications using this type of co filter are discussed in Section 5.

2. COMB FILTER AND RECURSIVE IMPLEMENTATION

Comb filters are characterized by their periodic transfer characteristics the frequency domain. They can be cl sified into two general types. The fir type is the bandstop or canceller type

shown in Figure 1a. It has strong attenuation in a narrow neighborhood of a series of periodically separated frequencies and good transmission in between. The second type is the bandpass or integrator type shown in Figure 1b. It has good transmission in a narrow neighborhood of a series of periodically separated frequencies and strong attenuation in between.

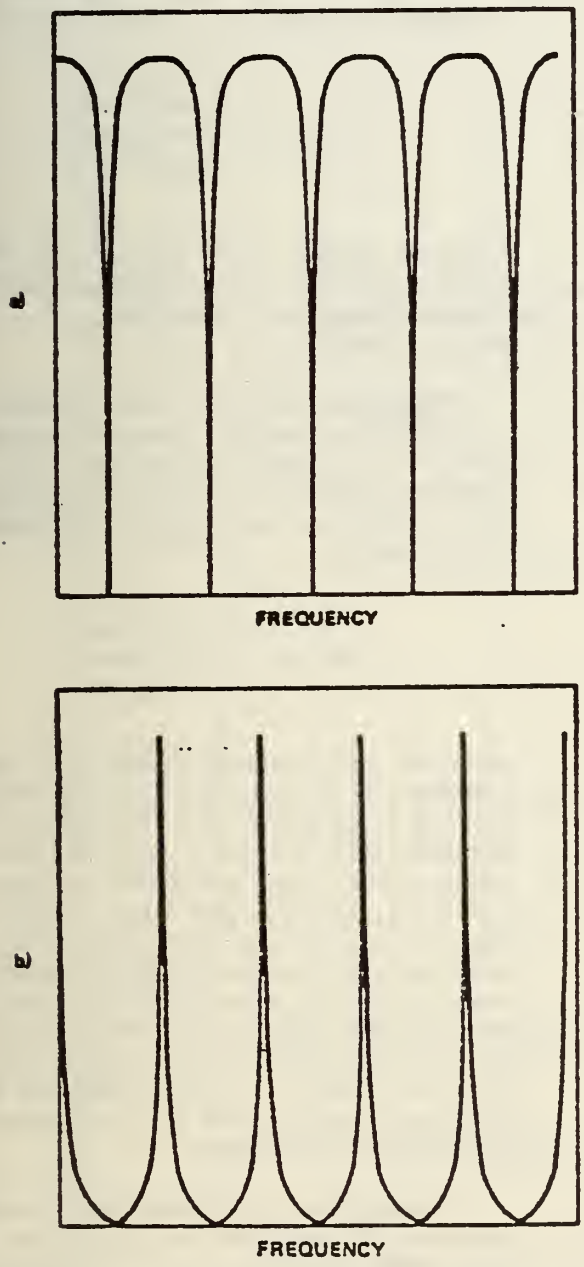


Figure 1. Frequency Characteristics of Two General Types of Comb Filters — Bandstop, Canceller Type, and Bandpass, Integrator Type

Both analog comb filters[17, 18, 19] and digital comb filters [20, 21] have been developed based on several different implementation techniques: feedforward, frequency sampling,[22] fast Fourier transform, and recursion. All use delay devices in some manner.

In the analog case, a quartz delay line has been the major candidate. For digital, the delay device is the shift register. However, a new family of comb filters is being developed using a different type of integrated circuit delay device, such as the BBD (bucket brigade device), CCD (charge coupled device), and SAD (serial analog delay).[23] In these, analog signals are first sampled and then delayed. As a result of the sampling, these devices not only have standard analog properties but also some digital properties such as aliasing and stability of delay.

Using these integrated circuit delay devices, sampled analog comb filters are being developed based on the feedforward circuit,[11, 14, 15] chirp Z transform, and recursive filter implementations.[9, 10, 11, 12]

We are studying the recursive implementation using the canonical circuit shown in Figure 2. The results presented use only one CCD delay device, i.e., $b_2 = 0$, $a_2 = 0$. However, the delay device consists of N delay stages (in the experimental study, N = 8.) As shown in later sections, the presence of N stages of delay in

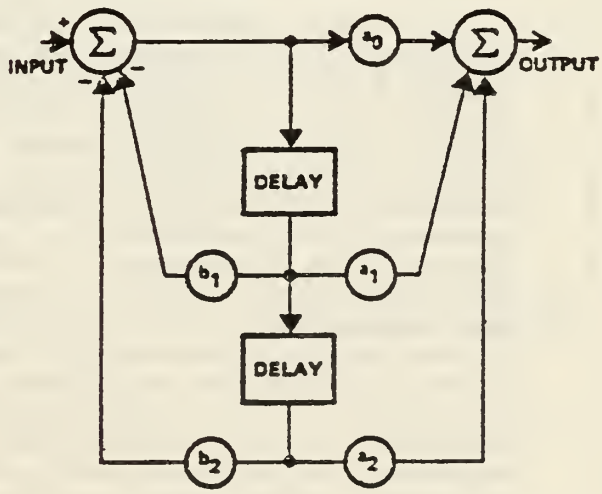


Figure 2. Canonical Form of a Second Order Recursive Filter

one delay device can be used to advantage in providing the comb feature of the frequency characteristics and also very large attenuation between these comb teeth.

The reason for this is that, since the signal is delayed N clock periods before it is fed back to the input, the frequency of recursive operation, f_r, is only one N^{th} of the clock frequency, f_c (or sampling frequency). Therefore, the frequency characteristics of the recursive filter are periodic with the frequency f_r. Since $f_r = f_c/N$, there are N/2 teeth within the Nyquist frequency range from 0 to $f_c/2$. This is in contrast to the usual digital recursive filters where the recursions take place after each delay stage, resulting in only one comb tooth within the Nyquist range. Since there are now N/2 teeth, a proper design procedure which introduces a zero at the recursion frequency f_r will provide infinite attenuation, in principle, at each of these teeth even when only one delay element is used. The procedure is known as the bilinear Z transform.

3. THEORY

Since the sampled analog recursive filters are implemented by the same circuit configurations used for digital recursive filters, we know intuitively that the theory of digital recursive filters should apply. However, proper modifications must be made to account for the fact that the signal is now sampled-analog and that the delay device now consists of N delay stages instead of one.

It is known that two general design approaches have been developed for the IIR (infinite impulse response) type digital recursive filter. The first approach is indirect. It starts with the transfer function in the Laplace transform variable S of a continuous filter and follows by digitizing the transfer function H(S) into a transfer function in the discrete time variable Z. Many different transforms are employed such as mapping by differential transform, impulse invariant (standard Z) transform, bilinear Z transform, and matched Z transform. The second approach is a direct digital design in either the frequency domain or the time domain using some type of computer aided design procedure.

We feel that the direct approach using the frequency domain is probably more instructive, and we will present it in this paper using two of the four transforms listed earlier (standard Z transform and bilinear Z transform).

The modification used to take into account the fact that one CCD delay device consists of N delay stages will be described. In conventional digital filter theory, Z^{-1} is used to represent a delay of one clock period. For a recursive filter using only one CCD device of N delay stages, the transfer function is

$$H(z) = \frac{a_N z^{-N} + a_0}{b_N z^{-N} + 1}$$

In other words, it is actually an N^{th} order recursive filter but with many of its coefficients, $a_{N-1}, \ldots a_1, b_{N-1}, \ldots b_1$, equal to zero.

The analysis can be simplified greatly if we consider the N clock delay stages as "one recursion" delay stage with its corresponding recursion frequency $f_r = f_c/N$, then the transfer function is simplified to

$$H(z) = \frac{a_1 z^{-1} + a_0}{b_1 z^{-1} + 1}$$

with the understanding that Z^{-1} actually consists of N clock periods of delay and that its frequency characteristics are periodic with respect to f_r and have N/2 comb teeth in the frequency range from 0 to $f_c/2$, the Nyquist limit. Because there is now more than one period in the frequency characteristics (or more than one tooth), the effects due to sampling and/or the sample and hold circuit should be taken into account in the theory. Only the effect due to the sample and hold circuit is considered in this paper, and is discussed in Section 4.

Using this straightforward modification and the well developed digital recursive filter theory, a set of design formula was derived for a first order lowpass filter

$$H(S) = \frac{\omega_x}{S + \omega_x}$$

and a highpass filter

$$H(S) = \frac{S}{S+\omega_x}$$

where ω_x = 3 dB corner frequency. The following two transforms are used:

Standard Z transform $Z = e^{ST}$

Bilinear Z transform $S = \frac{2}{T}\frac{Z-1}{Z+1}$

where $T = N\frac{1}{f_c}$, the CCD delay.

Without presenting the derivations in detail, the results are summarized in Table 1. It should be pointed out that an approximation, $\tan \omega/\omega_c \cong \omega/\omega_c$, was used. The cases of integrator type comb filters and cancellor type comb filters are listed separately. They are determined by the relative signs of the coefficients a_0, a_1, and b_1. The theoretical frequency characteristics of six special cases are presented as solid curves in Figures 3 through 7. The cases that have been implemented and measured experimentally are summarized in Table 2. The coefficients have all been normalized to unity. Their comb filter characteristics are also plotted in relative amplitudes in Figures 3 through 8. In Figure 8, the theoretical comb filter characteristics are modified by taking into account the effect of the sample and hold circuit. This introduces a factor of sin (π k/N)/(πk/N), where k = the order of the comb tooth. Discussions relating to this correction are given in Section 4.

Table 1. Summary of Design Formula of Sampled Analog Recursive Comb Filter Using One CCD Delay

Design Method	Filter Type	Coefficients		
		a_0	a_1	b_1
Standard Z	Lowpass, integrator			
	$\frac{\omega_x}{\omega_c} < \frac{1}{4}$	ω_x	0	$-e^{-\omega_x T}$
	Lowpass, cancellor			
	$\frac{\omega_x}{\omega_c} > \frac{1}{4}$	$1-\omega_x$	$-e^{-\omega_x T}$	$e^{-\omega_x T}$
	Highpass, cancellor			
	$\frac{\omega_x}{\omega_c} < \frac{1}{4}$	$1-\omega_x$	$-e^{-\omega_x T}$	$-e^{-\omega_x T}$
	Highpass, integrator			
	$\frac{\omega_x}{\omega_c} > \frac{1}{4}$	ω_x	0	$e^{-\omega_x T}$
Bilinear Z	Lowpass, integrator			
	$\frac{\omega_x}{\omega_c} < \frac{1}{4}$	$\frac{\omega_x T}{\omega_x T + 2}$	$\frac{\omega_x T}{\omega_x T + 2}$	$-\left\lvert\frac{\omega_x T - 2}{\omega_x T + 2}\right\rvert$
	Lowpass, cancellor			
	$\frac{\omega_x}{\omega_c} > \frac{1}{4}$	$\frac{\omega_x T}{\omega_x T + 2}$	$\frac{\omega_x T}{\omega_x T + 2}$	$\left\lvert\frac{\omega_x T - 2}{\omega_x T + 2}\right\rvert$
	Highpass, cancellor			
	$\frac{\omega_x}{\omega_c} < \frac{1}{4}$	$\frac{2}{\omega_x T + 2}$	$\frac{-2}{\omega_x T + 2}$	$-\left\lvert\frac{\omega_x T - 2}{\omega_x T + 2}\right\rvert$
	Highpass, integrator			
	$\frac{\omega_x}{\omega_c} > \frac{1}{4}$	$\frac{2}{\omega_x T + 2}$	$\frac{-2}{\omega_x T + 2}$	$\left\lvert\frac{\omega_x T - 2}{\omega_x T + 2}\right\rvert$

Table 2. Summary of Six Experimental Sampled Analog Recursive Filter Cases

Design Method	Filter Type	Coefficients			Figure
		a_0	a_1	b_1	
Standard Z	Highpass, integrator	1	0	0.7	3
	Lowpass, integrator	1	0	-0.3	4
	Lowpass, cancellor	1	1	0.7	5
	Lowpass, cancellor	1	1	0.3	6
Bilinear Z	Lowpass, integrator	1	1	-0.9	7
	Highpass, integrator	1	-1	0.7	8

NOTE: 8-bit, 2 phase, surface channel CCD (Figure 9)
f_c = 20 kHz
Circuit schematic (Figure 11)

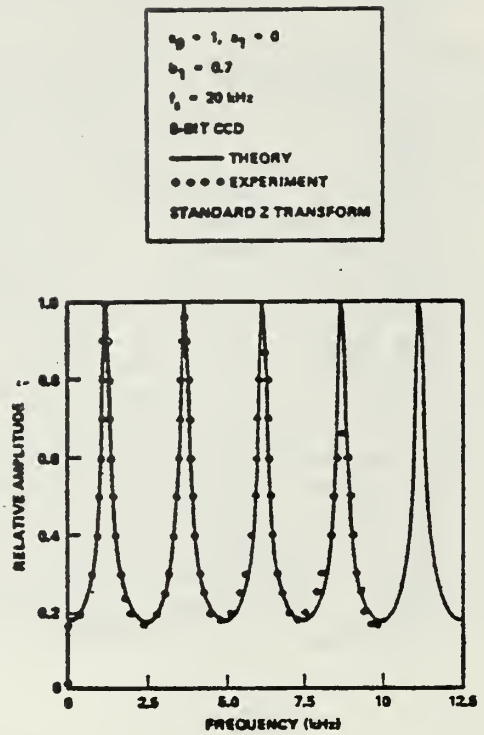


Figure 3. Highpass Integrator Filter — Frequency Characteristics

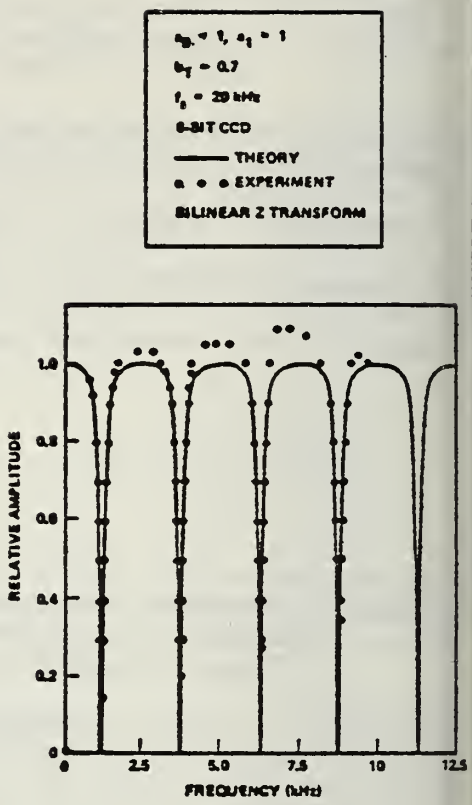


Figure 5. Lowpass Canceller Fi
Frequency Characteri

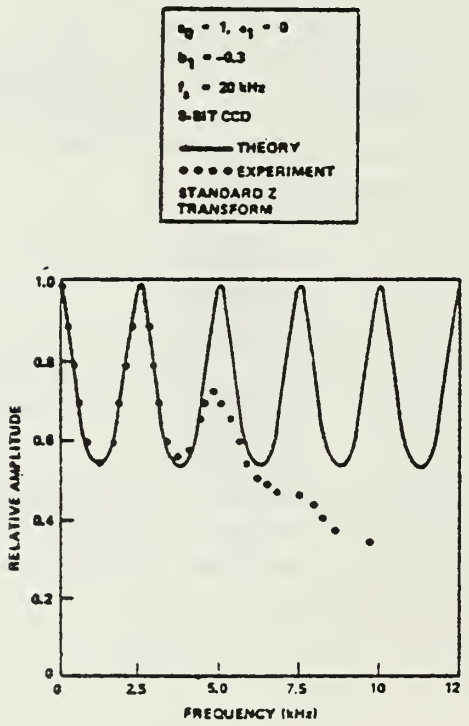


Figure 4. Lowpass Integrator Filter — Frequency Characteristics

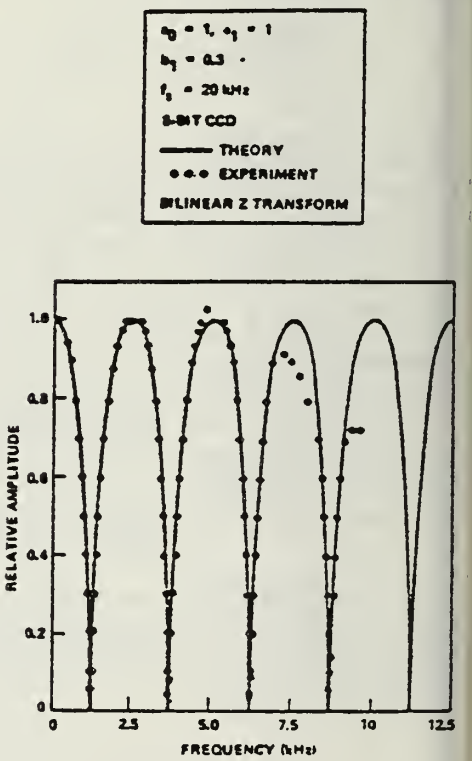


Figure 6. Lowpass Canceller Fi
Frequency Characteri

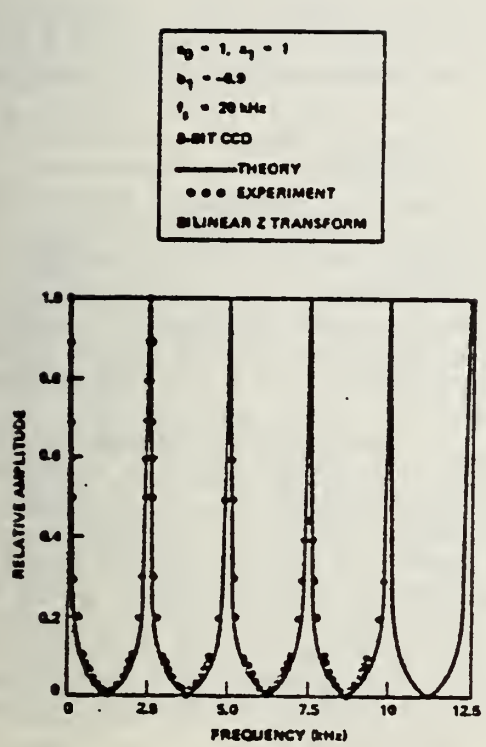


Figure 7. Lowpass Integrator Filter — Frequency Characteristics

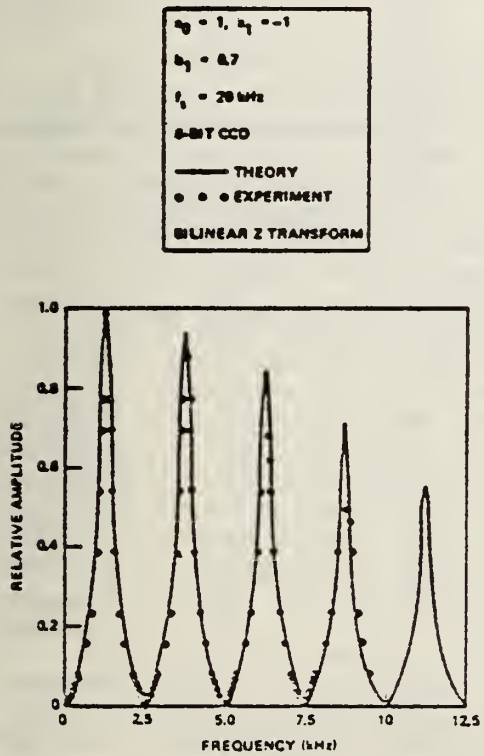


Figure 8. Highpass Integrator Filter — Frequency Characteristics

4. EXPERIMENTS, COMPARISONS, AND DISCUSSIONS

To demonstrate the feasibility of using CCD as delay elements for the implementation of sampled analog recursive comb filters and also to verify the validity of the theoretical analysis presented in Section 3, six cases of comb filters are implemented. They are listed in Table 2. The table includes both canceller type and integrator type comb filters whose Z transform transfer functions were obtained from both lowpass and highpass continuous filters. The CCD used is a two-phase, overlapping gate, P surface channel, 8-bit CCD. Its physical description is presented in Figure 9. Its output circuit consists of a reverse biased floating diode, gated by a reset MOSFET and followed by a MOSFET source follower. The load resistor is connected off the chip. Its input circuit consists of one diode, pulsed to a forward biased condition, and four gates with appropriate dc bias applied.

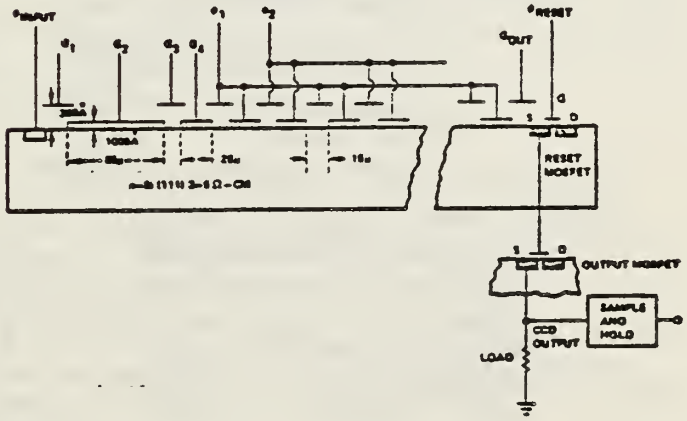


Figure 9. Cross-Sectional View of Two Phase, Overlapping Gate, P Surface Channel CCD

The electrical signal can be fed into the CCD either at the diode or at a gate. In this study, an input signal technique known variously as the surface potential equilibration method, the fill and drain method, or the scuppering method is used. To carry out this input procedure, a clock pulse is applied to the diode, and dc voltages of various levels are applied to the gates. The signal is applied to one of the gates. The combination of third and fourth

gates provides one of the widest dynamic ranges, as shown in Figure 10. The figure presents the output voltage measured after the sample and hold circuit as a function of the input signal expressed by a varying dc voltage level at the third gate. The bias voltage on the fourth gate is held at selected values. The operating conditions are described as follows:

- Four clock pulses:

  ϕ_{input} = input diode pulse level from -12.8 to -2.4 V

  ϕ_{reset} = reset MOSFET gate pulse level from -37.5 to -19 V

  ϕ_1, ϕ_2 = 50% duty cycle CCD clock pulses from -22.5 to -10 V

- Four dc gate biases:

  G_1, G_2 biased at -30 V

  G_{out} and V_{DD} at -35.5 V

  Load resistor = 50K ohms

  Clock frequency = 20 kHz.

From Figure 10, it can be seen that two operating modes exist. The first is a high gain mode occurring in a third gate voltage range from approximately -11.7 volts to values from -11.9 to -12.5 volts, depending on the bias voltage on the fourth gate. This mode is quite nonlinear. The second mode is characterized by a m lower gain but occurs over a much w bias range on G_3. The range is mor 8 V for V_{G4} = -15 V. For V_{G4} = -12 it can be seen that a section of the m characteristics is linear from -16.9 -20.5 V. Therefore, the combinatio V_{G3} = -18.7 V and V_{G3} = -12 V was as the electrical bias condition. The input signal was applied to the third g to carry out the frequency response m surements. In this way, a linear ra approximately 3.7 V was obtained.

The circuit schematic for the co filter measurements is shown in Figu The coefficient a_0 = 1 is easily imple mented by direct connection. The co cients a_1 = ±1 and the different value b_1 were implemented by using two po tiometers connected after the sample hold circuit and the level shifter. Op tional amplifiers with the proper feed were used to provide the two summat operations.

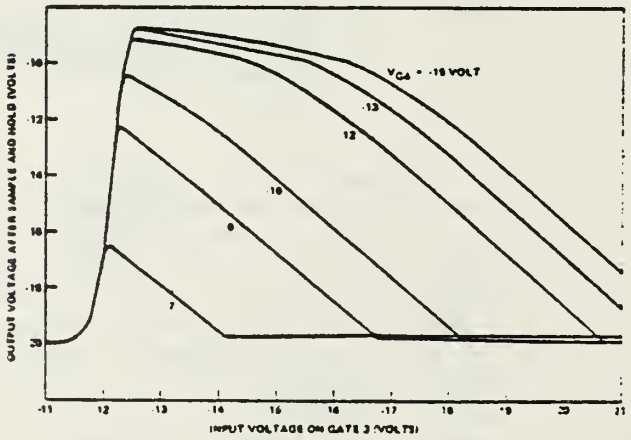


Figure 10. Input G_3 vs Output Characteristics of Two Phase Surface Channel 8 Bits CCD

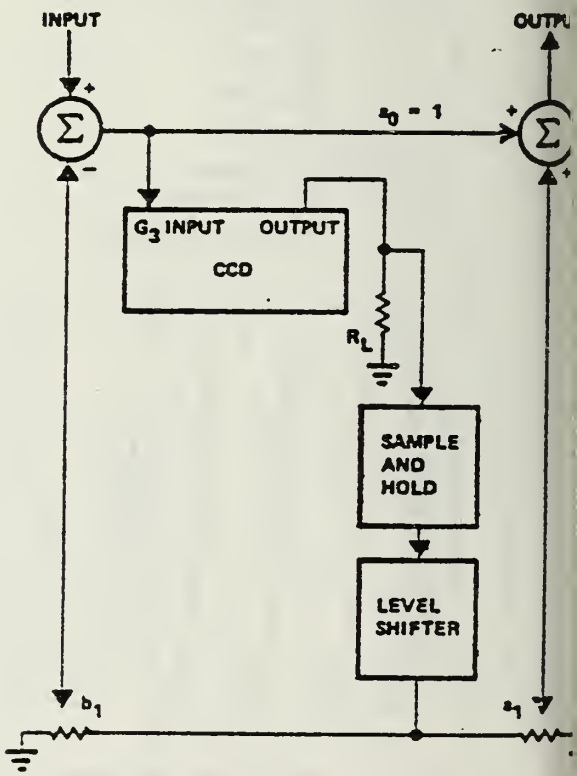


Figure 11. Circuit Schematic of CCD Recursive Filter Experime

essentially on the digital recursive filter theory is correct to the first order. Obviously, refinements are needed to extend the success of the theory at the first comb teeth to all higher order teeth.

5. APPLICATIONS

The range of applications of sampled analog recursive filters is obviously very broad. However, it is important to recognize that the performance of this type of filter is somewhere between that of analog and digital filters. Compared with regular analog filters, the delays of this family of integrated circuit electronic delay devices are controlled by the main clock pulse, and these IC filters do not have nearly the narrow bandwidth and time and temperature drift problems which plague other analog filters. They are as stable as the main clock. Compared with digital filters, they process signals in sampled analog form, thus avoiding the expense of A/D converters and digital multipliers, though there are penalties of somewhat lower performance and relatively more limited programmability. Another feature of sampled analog recursive filters is the possibility of time shared filtering because the delay elements generally have more than one stage.[15]

They are particularly useful as comb filters. Their periodic transfer characteristics make them well suited to process signals which are periodic in time and in frequency such as most radar and sonar signals. Canceller type comb filters can be used to reject periodic signals by aligning the null frequencies of the comb filter with the periodic frequency spectrum of the undesired signal. This is indeed the principle of both the feedforward and recursive filters[11] being developed to cancel clutter spectrum in MTI radars. Also, because the delay can be changed rather conveniently by changing the main clock frequency, the canceller type sampled analog comb filter may simplify the signal processing in either staggered PRF or jittered PRF radars.

In a different direction, integrator type comb filters can be used to enhance periodic signals contaminated by noise and/or interferences. A video integrator is being developed[13] based on this principle.

6. CONCLUSIONS

Sampled analog comb filters by recursive implementation have been studied. A theoretical analysis based essentially on digital recursive filter theory was used to calculate the frequency characteristics of both the canceller type and the integrator type comb filters, using only one delay element. The fact that the delay element has N delay stages is used to achieve the comb feature of the transfer characteristics. Using the zero introduced by the bilinear Z transform, infinite attenuations are achieved at null frequencies, which is a very desirable feature. Six calculated cases were selected for implementation using an 8-bit CCD as the delay element. The agreement between measurements and theory is excellent at the first set of comb teeth and varying degrees of closeness are achieved as the order of comb teeth is increased. The causes for this are under study. Studies are in progress to extend this work to cases where more than one delay element is used and to examine the reasons why the experimental results at higher order comb teeth are at variance with some of the theoretical predictions.

BIBLIOGRAPHY

1. D. D. Buss, D. R. Collins, W. Bailey, and C. Reeves, IEEE J. Solid States Circuits, SC-8, 1973, p. 134.

2. M. H. White, D. R. Lampe, F. J. Fagen, Int. Electronics Device Meeting, Tech. Digest, Washington, D. C., Dec. 1973, p. 130.

3. P. J. MacLennan, J. Mavor, G. Vanstone, and D. J. Windle, Proc. Int. Conf. Tech. and Appl. of CCD, Edinburgh, Sept. 1974, p. 221.

4. J. Tiemann, W. Engeler, R. Baertsch, IEEE J. Solid State Circuits, SC-9, 1974, p. 403.

5. R. W. Means, D. D. Buss, H. J. Whitehouse, Proc. CCD Appl. Conf., San Diego, Sept. 1973, p. 95.

6. R. W. Broderson, H. S. Fu, R. C. Frye, D. D. Buss, ISSCC Tech. Paper Digest, Philadelphia, Feb. 1975, p. 144.

7. I. Lagnado and H. J. Whitehouse, Proc. Int. Conf. Tech. and Appl. of CCD, Edinburgh, Sept. 1974, p. 198.

8. H. J. Whitehouse and R. W. Means, IEEE Int. Symp. on Ckts and Systems, Publication IEEE Advanced Solid State Comp. for Signal Processing, April 1975, p. 5.

9. D. A. Smith, C. M. Puckette, and W. J. Butler, IEEE J. Solid State Circuits, SC-7, 1972, p. 421.

10. D. A. Smith, W. J. Butler, and C. M. Puckett, IEEE Trans. Communications, COM-22, 1974, p. 921.

11. J. E. Bounden, R. Eames, and J. B. Roberts, Proc. Int. Conf. Tech. and Appl. of CCD, Edinburgh, Sept. 1974, p. 206.

12. J. Mattern and D. Lampe, ISSCC Tech. Paper Digest, Philadelphia, Feb. 1975, p. 148.

13. J. B. G. Roberts, M. Chesswas, and R. Eames, Electronics Letters, 10, 1974, p. 169.

14. W. J. Butler, W. E. Engeler, H. S. Goldberg, C. M. Puckette, and H. Lobenstein, IEEE Int. Symp. Ckts and Systems, Publication IEEE Advanced Solid State Comp. for Signal Processing, April 1975, p. 30.

15. M. F. Tompsett, A. M. Moshen, D. A. Sealer, and C. H. Sequin, IEEE Int. Symp. Ckts and Systems, Publication IEEE Advanced Solid State Comp. for Signal Processing, April 1975, p. 83.

16. W. H. Bailey, R. W. Broderson, W L. Eversole, M. M. Whatley, L. R Hite, D. D. Buss, and R. Sproat, Digest of Papers, 1974 Government Microcircuit Applications Conferenc 5, 1974, p. 76.

17. W. D. White and A. E. Ruvin, IRE Conv. Record, Pt. 2, 1957, p. 186.

18. A. G. J. MacFarlane, Proc. IEEE Paper 3121E, 1960, p. 39.

19. G. T. Flesher, Proceedings Nation Electronics Conference, 14, 1958, p. 282.

20. C. M. Rader, B. Gold, Proceeding IEEE, 55, 1967, p. 149.

21. S. A. White, IEEE Tran. Automatic Control, AC-14, 1969, p. 423.

22. R. E. Bogner, Bell Sys Tech Journa 48, 1969, p. 501.

23. Reticon Corporation, Sunnyvale, California.

24. L. Rabiner, B. Gold, Theory and Application of Digital Signal Proce sing, Prentiss Hall, 1975.

ACKNOWLEDGEMENTS

The authors are grateful to the follo ing persons for their technical advice an for their assistance in instrumentation a measurements: P. Kopp, S. R. Parker, and D. Hoisington of the Naval Postgrad uate School; J. L. Rogers and A. C. Schoening of TRW; and G. Temes of UCI

PAPER NO. 3

NINTH ASILOMAR CONFERENCE ON CIRCUITS, SYSTEMS AND
COMPUTERS, ASILOMAR, CALIFORNIA
NOVEMBER 3-5, 1975

HARDWARE SIGNAL PROCESSOR DEVELOPMENT --SESSION SUMMARY*

T. F. Tao and S. Holmes

Naval Postgraduate School, Monterey, Calif.

Abstract

After a brief summary on the relevance of the five papers presented in this session on "LSI Hardware Signal Processor Development" to the current digital signal processor developments, this paper discussed some developments of other types of hardware processors using sampled analog or analog signals and based on LSI electronics, acoustical and optical technologies.

I. INTRODUCTION

Signal processing was performed mainly by analog means before the advent of digital computers which promoted the intensive and successful development of digital signal processing in the past decade or so.[1-3] At present, much of the digital signal processing is carried out on general purpose digital computers using software . However, extensive progress has also been made in other types of signal processors using electronic firmware and hardware . Their activities are highlighted in Table I. Papers presented in this session on "Hardware LSI Signal Processor Developments" covered some of the important, current activities. Allen's paper discussed the present limitations of the glamorous bipolar integrated injection logic[4] when it is used to implement large scale digital circuits. Blankenship's paper discussed their applications of new LSIECL circuits to the developments of fast, special purpose digital computers.[5] Gehweiller and Pridgen's paper reported the status of their extensive hardware LSI signal processor development based on CMOS/SOS technology. Martinez's paper discussed the promises and requirements of microprocessors, currently one of the active, if not the most active LSI development. It will provide versatile low cost microcomputors adaptable to a wide variety of applications using firmware which can be programme by users and store special signal processing algorithms and programs in ROM or FPLA. White's paper commented on the basic considerations behind MOSLSI developments. However, signal processor developmen are being pursued in other directions also.[6] Using electronic technologies, a new class of sampled analog processors is being developed using charge transport devices.[7] Their signals are analog but the independent variables are discrete which can be either time variable, spatial variables or other transform variables. Acoustical[8] and optical[9] technologies have also been used for analog, sample analog and digital signal processing.

The purpose of this paper is to present some of the principles and highlights of these non-digital signal processor developments. Together with the pape on digital processors presented in this session, it is hoped that a bird's eye view of current hardware signal processor developments is presented. To cover such a broad area within its limited scope, this paper does not attempt to be a tutorial or a review. Instead, it plans to introduce some of the excellent review articles and special conference proceedings devoted to some of the individual topic but scattered over a diversified range of journals in an organized manner from the hardware viewpoint.

al processing covers an enormous range of
ations and applications. In this paper, they
organized and presented from the basic mathe-
cal operation viewpoint in order to present
the various hardware technologies are used for
r implementations. In section II, linear sig-
processing is grouped into two classes: filter-
and spectral analysis. They are then shown to
centered around four basic mathematical opera-
s: delay (or shift), summation, multiplication
generation of special functions. In sections
and IV, the basic principles of electronic,
stical and optical hardware technologies as
lied to signal processing implementations are
sented. In section V, some progress of the
dware developments are highlighted by presenting
present status of three important hardware
al processing modules: multipliers, correlators
convolvers, and discrete Fourier transformers
h include both fast Fourier transformers (FFT)
chirp Z transformers (CZT).

II. COMMON SIGNAL PROCESSING OPERATIONS

ar and nonlinear signal processing have been
vely developed. However, hardware developments
pursued mainly in linear processing applica-
s. Their basic operations can be introduced
g the general form of one dimensional linear
sform shown in Table II. The input signal i(u)
transformed into output signal o(v) by the trans-
kernel h(u,v). The independent variables u,v
be either time, spatial variables or transformed
ables such as frequency, spatial frequency and
rs. When the variables are continuous, as in
log signal processing, the transform relation
in integral form. When the variables are dis-
te, as in digital and sampled analog signal
cessing, the transform relation is in summation
as used in matrix operations. h(u,v) is often
rred to under different names. In some proces-
, h(u,v) is also known as the impulse response.
cases of two dimensional or image processing,
,v) is called the point spread function. In
relators, convolvers and matched filters, h(u,v)
the role of reference signals and are commonly
rred to by that name. Several special forms of
h(u,v) are listed in Table II to describe the mathematical operations of four of the most important signal processing functions: two filters--correlators and convolvers of which matched filter is a special case and two classes of spectral analysis--the Fourier type of transforms and other types such as Walsh transform, Kahunen-Loeve transforms, etc. It should be noted that in filters, u,v are the same type of variables related by a translation. It can be either time or spatial variables. In many transforms, u,v are not the same variables and are related by multiplication. For example, if u is the time variable t (or space variable x), then v is its transformed variable, frequency ω (or spatial frequency). In Table III, these signal processing operations in discrete variables are expressed in more detailed forms suitable for hardware implementations. For nonrecursive and recursive filters, expressions in terms of discrete time variables k and n and their transform variable Z are used. For spatial varying signals, spatial Z transform has not been commonly used. Three basic mathematical operations are involved: linear translation (delay or shift)e.g., $i(k-n), i(n-k)$; multiplication by constant, e.g., $a_n i(k-n)$; summation, e.g., $\Sigma a_n i(k-n)$. For spectral analysis, two algorithms for discrete Fourier transforms are given, FFT and CZT. It should be noted that in these two cases, n represents the discrete time (or space) variable while k represents its transformed time frequency (or spatial frequency) variable. It is noted that two additional mathematical operations are involved: multiplication of two variables nk(uv) and generation of special functions. For Fourier type of transforms, the special functions are the exponential or sinusoidal functions. For other transforms, it could either be other transcendental functions or matrices such as the matrix of elements ± 1 for the Walsh transform. Knowing the commonality of these basic mathematical operations will help to understand how different hardware technologies are used to implement a wide variety of signal processing operations.

III. HARDWARE TECHNOLOGIES

Three hardware technologies have been used to implement signal processing operations: electronic, acoustical and optical. In Table IV, the physical mechanisms of signal representations and input/output transducers are listed. Using electronic technology, signals are represented by voltage, current or charges. At present, most electronic signal processors are either digital or analog. It is important to note that a new class of electronic devices known as "charge transport devices" (CTD) is being actively developed in recent years which represent signals by number of charges. One of their most important new devices is the sampled analog delay line (or shift register) which is instrumental in making the new and promising sampled analog signal processing possible.[7] Using acoustical technology, signals are represented by lattice waves in piezoelectric materials, or simply acoustical waves. In most cases, surface acoustical waves (SAW) are used.[8,10] Recently, bulk acoustical wave devices are receiving some increasing attention because they may be used to develop programmable analog devices.[11] Although digital signals can be equally well represented, most acoustical signal processors today are analog and in some limited cases, sampled analog. The acoustical wave signals are launched and detected by two types of transducers: interdigital metallic fingers deposited on the surface of the piezoelectric device, or piezoelectric transducers such as ZnO. The former is much more frequently used while the latter is used sometimes for bulk acoutical waves. Using optical technology, signals are represented by electromagnetic waves. In contrast to other hardware signal processors which mostly deal with one dimensional time varying signals, optical processors perform operations in two dimensions. It makes them ideally suited to process spatially varying signals (image processing), or to perform many operations simultaneously in a manner that is commonly known as "parallel processing" which has much higher processing speed. Two types of optical processors are being developed: coherent and noncoherent. Coherent processors have received much more attention.[12] They are closely associated with holography and Fourier optics. Lasers and strong light source with a small pinhol output are used as monochromatic light sources. Signals are introduced by modulations: acousto-optical, electro-optical or transparency masks. The first two modulations (AO,EO) are often used to convert a time varying signal to a spatial vary ing signal suitable for optical processing. For noncoherent optical processors,[13] LED is often use as light source. Signals are applied at two place first at the LED by electrically modulating its radiation output and second, in the optical path by using mask. The signal pattern on the mask is spatially varying which is converted to a time varying signal by scanning the image of the patter across the detector plane. In both coherent and noncoherent processors, the output is achieved by using optical detectors or imagers such as vidicon or new CTD imagers. While most optical signal processors are analog, digital optical memories ar also being developed.[14] In image processing, when a picture is scanned over discrete numbers of detector elements, the signals become sampled analog due to the discrete nature of their spatial variables.

IV. HARDWARE IMPLEMENTATIONS OF MATHEMATICAL OPERATIONS

In Table V, implementations of basic mathematical operations using these three hardware technologies are presented.

(1) Electronic implementation:

In this paper, analog processors are not discussed. For digital processing, the delay is implemented by shift registers. Summation and multiplication are implemented either by logics, by table look-ups[15] or by hardware adders and multipliers. Generation of special function is usually achieved by table look-up using ROM. For sampled analog processing, delay is implemented by the new CTD delay lines. Summation and multiplication are usually implemented using linear integrated circuit devices such as op amp, multipliers, etc. Sometimes, summation can be accomplished by current or charge summing. In CTD transversal filters using

apped delay lines, multiplication by constants can e accomplished by using tapping weights. Generation of special functions has been achieved by sing tapped delay lines as convolvers. The nction to be generated is designed into the evice as tapping weights. An example of generating sinusoidal functions for implementing chirp Z ransform will be given in section V.

(2) Acoustical implementation

There are many similarities in implementing ignal processing operations using SAW devices and TD devices. The differences are in the signal andwidth W and delay time T. For SAW devices, is in the range from few MHZ to several hundred HZ, but T is inversely proportional and varies rom nanoseconds to microseconds with maximum chievable TW product around 1000. It is limited y the size of the delay device. For CTD devices, can be much longer and approach milliseconds. ut the bandwidth W is lower in the few MHZ range ith maximum TW product also around 1000 with the resent technology. Using acoustical waves, delay s implemented by acoustical delay line. Summation nd multiplication are usually implemented electronically using linear IC devices. In SAW ransversal filters using tapped delay lines, ultiplication by constants is accomplished using apping weights. Generation of special functions an be achieved by using SAW tapped delay lines as onvolvers. Multiplication of two variables can e accomplished in different ways.[13] For the pecial case of discrete Fourier transform which equires the product of uv, the multiplication is ccomplished indirectly using the chirp Z transorm algorithm which will be explained later in ection V. On the other hand, nonlinearities in coustical devices can be used to implement multilication of two signals.[11] In these devices, two coustical wave signals are launched in opposite irections, the nonlinearities will give rise to product signal at the sum frequency. Using this bility of multiplication in a delay line, convoutions of two signals can be accomplished. Its dvantage over the approach using tapped delay ine lies in the fact that both signals can be hanged easily. In the tapped delay line implementation, the reference signal is fixed by the tapping weights except when diodes are used in the taps which permitted programmable binary coded taps.[16]

(3) Optical implementation

The principles of coherent and noncoherent optical processors are quite different although they both process signals optically and thus possess parallel processing capability and fast multiplication rate. They will be presented separately.

Coherent optical processors[13] are capable of performing signal processing either in time domain, spatial domain or spatial frequency domain. In contrast, electronic analog processors operate in either time or frequency domain. Other hardware signal processors operate mostly in time domain. This is the result of the unique property that in a coherent optical system, the image at the back focal plane of a lens is the Fourier transform of the spatially varying signal at the front focal plane. Using this capability, two basic building blocks have been developed. The first is the "single lens, front and back focal planes" building block which performs the spatial Fourier transform (or spectral analysis). The second is a "two lenses, three focal planes" building block which performs the spatial filtering. A typical system is shown in Fig. 1. An additional lens is needed to collimate the monochromatic radiation. Input signal is introduced by placing a mask in the collimated optical path. It is located at the front focal plane of the "spatial Fourier transform lens." The Fourier transform of the spatial input signal is obtained at its back focal plane which is also the front focal plane of the next lens. Signal processing can be performed at this "transform plane" using masks commonly referred to as "spatial filters" because the signal at this plane is in spatial frequency domain. Next lens performs the inverse Fourier transform and delivers the processed spatial output signal at its back focal plane. Using this arrangement, low pass, high pass, band pass, band stop spatial filters and spectral analysis in spatial frequency domain can be carried out.

The same optical system can be used to perform time domain signal processing if the input

mask is replaced by an acousto-optical transducer which converts a time varying signal into a spatially varying phase modulation of the monochromatic radiation. By performing the Fourier transform, proper spatial filtering, which introduces the reference signal or impulse response, and the inverse Fourier transform, the spatially varying phase modulation can be converted back to a time varying amplitude signal at the output focal plane. Using this principle, correlation, convolution in time domain, spectral analysis of time varying signals have been demonstrated.

Coherent optical processor can also perform signal processing in spatial domain. An example is the optical correlator shown in Fig. 2 which consists of two building blocks--an optical filter followed by an optical spectral analyzer. The input signal is introduced by a mask at the front of the optical filter. The reference signal is introduced by another mask at the output of the optical filter which also serves as the input of the optical spectral analyzer. The spatial filter is necessary to block out the undiffracted light. The correlation is obtained at the center of the output focal plane of the analyzer section.

From the implementation of the four basic mathematical operations viewpoint, it can be readily explained that they are achieved in a natural way when coherent radiation is used. Exponential function is obtained because coherent radiation can be considered as a packet of plane waves which are mathematically described by exponential functions. Delay and shift are natural consequences of wave propagation. Summation of delayed and shifted signals follows the Huygen's principle directly. Multiplication is accomplished by using the spatial filters and masks.

Noncoherent optical processor[13] does not have the capability of obtaining a spatial Fourier transform by simply using a lens. However, the advantages of parallel processing and fast multiplication still exist. The basic system is described as follows: A time varying input signal is used to electrically modulate the output of a light emitting diode. A lens is used to collimate the radiation. A mask which contains the information of the transform kernel h(u,v) is inserted in the optical path. The light after the mask is imaged onto and scanned across the output detector device which is a CTD imager. In this system, time delay is accomplished in two steps. First, a spatial shift is formed on the mask. It is converted into time delay by scanning of its image across the CTD imager. Summation is accomplished because CTD imager is capable of integrating the signal charges over the exposure period. Multiplication is accomplished using the mask. Generation of special function is accomplished by designing the pattern on the mask using the special function. Its principle is similar to that of using tapped delay line as a convolver to generate special functions. Using the noncoherent processor, a wide variety of signal processing operations such as convolutions, correlations, Fourier transforms, Walsh transforms, etc. have been demonstrated using the same equipment simply by changing masks.

V. PROGRESS OF SOME HARDWARE SIGNAL PROCESSOR DEVELOPMENTS

To present a glimpse of the current hardware developments, some progress in three signal processing operations is briefly described: multiplication, correlation and convolution (matched filtering), and discrete Fourier transform. It should be pointed out that signal processor technologies are making rapid progress, data presented in this section probably do not represent their state of art and the survey is certainly not meant to be a complete one.

(1) Multiplication

Speed of signal processing is often limited by multiplication. In general, rate of multiplication is proportional to signal frequency. It is highest for optical processor followed by acoustic and electronic processors. For optical processors the computational rate is further increased by the capability of parallel processing. Using a one watt laser, it was calculated that 10^{14} multiplications/sec. can be made. Although practical coherent processors are all two orders slower, having computation rate less than 10^{12}/sec. In acoustical processors, if electronic multipliers are used, the computation rate will be limited by the electronic

vices which are of the order of 10^7/sec. How-
er, when multiplications by constants are carried
t using tapping weights in tapped delay lines
DL), the multiplication rate is equal to the
gnal bandwidth which can be as high as several
ndred MHZ. Furthermore, TDL has parallel pro-
ssing capability also because multiplication and
mation of all taps are performed at once. In
her words, there are TW multiplications being
rried out simultaneously. For a typical acous-
cal TDL of W = 100 MHZ and TW = 500, the compu-
tion rate is 5×10^{10}/sec. Multiplication rate
analog electronic processors is around 10^7/sec.
r sampled analog electronic TDL, the computing
te is approximately one order lower than that of
oustical TDL because its bandwidth is limited to
MHZ range today. A typical example is W = 10
Z, TW = 500 with computation rate at 5×10^9/sec.
ese technologies all process analog or sampled
alog signals. Their accuracies will be below
% which corresponds to 10 digital bits accuracy.
is obvious then, digital processing using more
an 10 bits must be used for higher accuracy.
ntinuous efforts are given in developing hardware
gital multipliers. Some typical performances are
ven in Table VI.

(2) Correlation and Convolution (Match Filter)[17]

Correlation and convolution are two of the
st important signal processing operations. Some
their outstanding properties are the following.
tput of a system is the convolution of input and
e system impulse response. Similarities between
o signals are measured by their cross correlation.
tocorrelation is a basic characteristic of random
gnal, because the Fourier transform of density
ction is the autocorrelation. Therefore, cor-
lators and convolvers are used in optimum detec-
n of signal in noise and interference environ-
t, pattern recognition and many others.

Mathematical operations of correlation and
volution are essentially the same except that
relative shift of variables is different, u-v
v-u. Their hardware implementation involves
shifted (delayed) signals, multipliers and a
mer. Optical correlators based on either co-
ent or noncoherent principles are explained in
section VI. Using acoustical and electronic technologies, two types of correlators and convolvers have been developed. In the first type, two shift registers are used to store the input and reference signals. For correlators, data are entered in the same direction. For convolvers, data are entered in opposite directions. All hardware digital correlators belong to this type. Because the data are binary, multiplications of each pair of bits are carried out by the simple "exclusive OR" logic gates. Their products are summed together simply by analog current summing. Correlators with digital summing are currently being developed. Performance data of five digital LSI correlators are listed in Table VII. The programmable nonlinear analog acoustical matched filter listed in Table VIII also belongs to this class. In the second type of hardware devices, only one shift register is used to store the input signal. They are all TDL used as matched filters. The second signal is stored as tapping weights and is therefore not programmable. Three analog electronic matched filters are listed in Table VII, all based on charge transport device technologies: CCD - charge coupled device, TAD - tapped analog delay, BBD - bucket brigade device. SCT (surface charge transistor) is a different version of CTD technology. It was used to develop the surface charge correlator which is claimed to be programmable. Three types of acoustical TDL, shown in Table VIII, have also been developed. Two of them are programmable. It is interesting to note that the figure of merit TW for the non-digital devices are all limited below 1000 at this time. Acoustical devices process IF signals and operate up to several hundred MHZ. Sampled analog electronic devices process baseband signals and operate in the low MHZ range. Digital correlators can be cascaded to yield very large TW product in principle. But their power consumption at present are too high to make long cascaded correlators of TW products much beyond 1000 practical.

(3) Discrete Fourier Transform

The importance of spectral analysis is obvious because many theories are formulated in transform domain such as the frequency analysis in electronics, harmonic analysis in wave phenomena,

plane wave expansion in applied mathematics, momentum space representation in physics, transform representations in image and voice processings and many others. For Fourier type of transforms, two algorithms have been invented to carry out discrete Fourier transforms: FFT (fast Fourier transform) and CZT (chirp Z transform). FFT has been the most frequently used algorithm for implementing digital Fourier transforms either on computers or dedicated FFT processors, which are essentially special purpose computers. However, LSI technologies are being used in developing hardware FFT on a few IC chips. Three examples are given in Table IX. CMOS/SOS FFT is described in Gehweiller and Pridgen's paper of this session. CCDFFT is currently being developed at TRW Systems. I^2L FFT is only a conceptual design at this time.[18]

However, CZT found new lease of life and is now being used to great advantage in developing sampled analog hardware Fourier transformers. The principle of implementation can be readily seen by examining the CZT algorithm shown in Table II. The second and third terms are a summation of shifted exponential functions in the form of a convolution. Therefore, the hardware implementation consists of (i) premultiplying the time varying signal with a linear FM signal (chirp), (ii) performing a convolution using a tapped delay line whose taps are designed by the exponential function and (iii) postmultiplying the signal with another linear FM signal (chirp). CTD and SAW tapped delay lines are ideally suited to implement the convolution. Therefore, both electronic and acoustical[19] CZT are being actively developed. 32 Bits and 512 bits CZT have been reported.

VI. CONCLUSIONS

Hardware signal processors are proliferating with a large number of LSI electronic, acoustical and optical devices. However, the principles of their implementations can be shown in an organized manner because most signal processing operations can be grouped into two categories: filtering and spectral analysis. Furthermore, they can be broken down into four basic mathematical operations. Based on these commonalities, digital, analog and sampled analog processors using electronic, acoustical an optical technologies are presented. A glimpse of their current developments is made by presenting some performance data of three important signal processor modules: multiplier, correlator and convolver, discrete Fourier transformer.

REFERENCES

1. L. R. Rabiner and C. M. Rader, "Digital Sign Processing," IEEE Press, 1972.
2. L. Rabiner and B. Gold, "Theory and Applicat of Digital Signal Processing," Prentice-Hall 1975.
3. A. Oppenheim and R. Schafer, "Digital Signal Processing," Prentice-Hall, 1975.
4. "Four Developers of I^2L Named Top Technology Achievers," Electronics, pp. 66-69, 16 Oct 7
5. J. Allen, "Computer Structure for Signal Processing," Proc. IEEE, 63, pp. 624-633, 1975.
6. "Signal Processing" NEREM 73 Record, Part 2, IEEE Catalog No. 73CH0840-9 NEREM (1973).
7. Special Session on Signal Processing, 1975 IEEE Symposium on Circuits and Systems, Boston, Mass., IEEE Catalog No.
8. "Component Performance and Systems Applicatio of Surface Acoustical Wave Devices," Sept 197 Aviemore, Scotland, Conference Publication N 109, Institute of Electrical Engineers, Engla
9. Special Issue on Optical Computing, IEEE Tra Computers, C-24, April 1975.
10. M. Holland and L. T. Claiborne, "Practical Surface Acoustic Wave Devices," Proc. IEEE 62, pp. 582-611 (1974).
11. G. S. Kino, S. Ludvik, H. Shaw, W. Shreve, J. White and D. Winslow, "Signal Processing b Parametric Interactions in Delay Line Devices IEEE Trans. MTT-21, pp. 244-255 (1973)
12. A. Vander Lugt, "Coherent Optical Processing" Proc. IEEE, 62, pp. 1300-1319 (1974)
13. M. Monahan, R. Bocker, K. Bromley and A. Loui, "Incoherent Electro-optical Processings with CCD's" Proc. Int'l Optical Computing Conference, April 1975, pp. 25-33, IEEE Catalog 75CH0941-5C.
14. O. Tufte and D. Chen, "Optical Techniques for Data Storage," IEEE Spectrum, pp. 26-32, Feb. (1973).
15. A. Peled and B. Liu, "A New Hardware Realization of Digital Filters" IEEE Trans. Acoust. Speech. Signal Processing, ASSP-22, pp. 456-462 (1974).
16. E. Staples and L. Claiborne, "A Review Of Device Technology for Programmable Surface Wave Filters" IEEE Trans. MTT-21, pp. 279-287 (1973).

7. T. F. Tao, "A Survey of Correlators and Matched Filters" Naval Postgraduate School Technical Report, to be released Feb. 1976

8. D. Buss, R. Veenkant, R. Brodersen and C. Hewes, "Comparison Between The CCD CZT and The Digital FFT", Proceedings International Conference on Applications of CCD, pp. 267-281, Oct. 1975.

9. J. Alsup, "Surface Acoustic Wave CZT Processor" Proceeding 1974 IEEE Ultrasonics Symposium, pp. 378-381, IEEE Catalog No.74CHO-866-1SU.

Supported in part by the Naval Postgraduate School Foundation Research and the Naval Electronic System Command

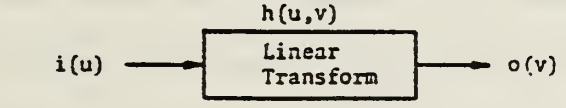
ACKNOWLEDGMENT

The authors wish to thank Messrs. J. Alsup, J. Spieser and H. Whitehouse of Naval Undersea Center for helpful discussions and A. Schow for her assistance in preparing this manuscript.

TABLE I. SIGNAL PROCESSING IMPLEMENTATIONS

SIGNAL	EQUIPMENT	IMPLEMENTATIONS
Analog---	General Purpose Computer	Software
	Special Purpose Computer	
	Mini/Microcomputer (Microprocessor)	Firmware
Digital		
Sampled--- Analog	Signal Processing Function Modules:	
	Multiplier	
	Correlator	
	Convolver (matched filter)	Hardware
	Filter	
	Discrete Fourier transformer	
	Frequency synthesizer	
	etc.	

TABLE II. LINEAR TRANSFORMATION



Continuous variables $o(v) = \int i(u)h(u,v)du$

Discrete variables $o_m = \sum_{0}^{N-1} i_n h_{mn}$

Examples of h(u,v):

Filter	Convolution	h(v-u)
	Cross-correlation	h(u-v)
	Autocorrelation	i(u-v)
Spectral Analysis	Fourier transform	$e^{-j2\pi uv}$
	Cosine transform	$\cos(2\pi uv)$
	Hilbert transform	$\frac{1}{u-v}$

TABLE III. SIGNAL PROCESSING OPERATIONS

A. FILTERS

Filters	Discrete Time Domain	Z Transform Domain
General Form		$H(Z) = \frac{O(Z)}{I(Z)} = \frac{\sum_{0}^{N} a_n Z^{-n}}{1 + \sum_{1}^{M} b_m Z^{-m}}$
Nonrecursive	$o(k) = \sum_{0}^{N} a_n i(k-n)$ Convolution $o(k) = \sum_{0}^{N} a_n i(n-k)$ Correlation	$O(Z) = \sum_{0}^{N} a_n I(Z) Z^{-n}$
Recursive	$o(k) = \sum_{0}^{N} a_n i(k-n) + \sum_{1}^{M} b_m o(k-m)$	$O(Z) = \sum_{0}^{N} a_n I(Z) Z^{-n} + \sum_{1}^{M} b_m O(Z) Z^{-m}$

B. SPECTRAL ANALYSIS

	Algorithm
Discrete Fourier Transforms	(FFT) $F_k = \sum_{0}^{N-1} f_n e^{-j2\pi nk/N}$ (CZT) $F_k = e^{-j\pi k^2/N} \sum_{0}^{N-1} f_n e^{-j\pi n^2/N} e^{j\pi(k-n)^2/N}$
Other Transforms	

TABLE IV. SIGNAL REPRESENTATION AND INTERFACE IN HARDWARE IMPLEMENTATIONS

Technology		Signal Representation	Interface & Transducer
Electronic		Voltage Current Charge	Electrical
Acoustical	Surface (SAW)	Lattice Wave	Interdigital metallic fingers
	Bulk		Piezoelectric transducers
Optical	Coherent Holography Fourier Optics	Electromagnetic Wave	Input: Optical modulation Acousto-optical Electro-optical Mask, transparency
	Noncoherent		Output: Detector Imager Photographic film

TABLE V. HARDWARE IMPLEMENTATIONS OF MATHEMATICAL OPERATIONS

Mathematical Operations	Electronic		Acoustical	Optical	
	Digital	Sampled Analog		Coherent	Noncoherent
Delay (shift)	Shift register	CTD Delay Line	Acoustical	Wave Propagation	Spatial shift on mask
Summation	Adder Accumulator	Op Amp Adder, or Current, Charge Summing		Hyygen's Principle	Integration in Imager
Multiplication by constant of two variables	Multiplier or Exclusive OR	Tapping weights or Multiplier		Spatial Filters	Mask
Special function Ex. Exponential	ROM or Look up tables	Design tap weights for Transversal filters used as Convolvers		Plane Wave	Pattern design on mask

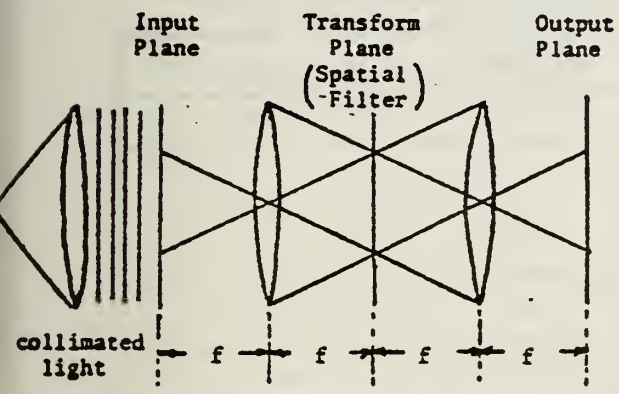


Fig.1 Coherent Optical Spectral Analyzer

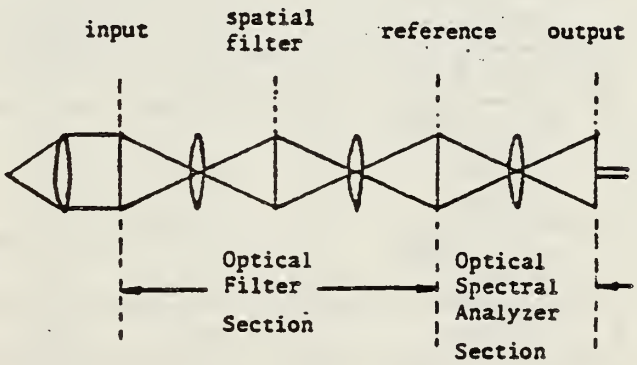


Fig.2 Coherent Optical Correlator

TABLE VI. DIGITAL MULTIPLIER PERFORMANCE

Technology	Clock Rate (MHZ)	Multiplication Time* (nsec)	Power Dissipation* (watt)
EFL	3	330	3.5
ECL 10K	87	184	14.4
LS-TTL	21	762	3.1
	30	1200	1.6
CMOS/SOS	10		0.72

* Required for 16 x 16 bit multiplication.

TABLE VII. PERFORMANCE OF ELECTRONIC CORRELATORS & MATCHED FILTERS

Characteristics	Digital LSI •					Sampled Analog CTD			
	EFL	ECL	CMOS	CMOS/SOS	CCD	CCD	SCT *	TAD	BBD
No. of Bits	64	16	50	32		32	32	12	104
Bit Rate (MHz)									
Maximum	25	150	14	40	10	5	4	5	3
Minimum	0	0	0	0	0.001	0.001			
Maximum Delay (sec)	Limited by:					1			
TW Product	Yield					10^3			
Dynamic Range	Cascading Capability					60 - 70 db			
Power Consumption (Watt/Bit)	10^{-2}	10^{-3}	10^{-2}	10^{-4}	10^{-3}	10^{-3}			

TABLE VIII. PERFORMANCE OF ACOUSTICAL CORRELATORS & MATCHED FILTERS

Characteristics	SAW	AlN/SOS *	ZnO/MOSFET *	Nonlinear SAW *
No. of Bits	256	64	32	600
Bit Rate (MHz)				
Maximum	200	20	10	30
Minimum	1	1	0	1
Maximum Delay (μ sec)	50	60	30	10
TW Product	1000	300	300	300
Dynamic Range (db)	∿60	∿60	∿60	∿40
Power Consumption (watt/bit)	0	10^{-3}	10^{-1}	0

* Programmable

TABLE IX. DIGITAL LSI FFT MODULES

Technology	Cycle Clock (MHZ)	Butterfly Time (μ sec)	Arithmatic	Coefficient Length	Data Length	Gate Delay-Power Product (pJ)
CCD	8	1	16 bits fixed point	16	16	0.2 pJ
IIL	2	4	16 bits floating point	8	16	0.5 - 1 pJ
CMOS/SOS	10		12 bits floating point over-flow	12	12	0.2 pJ

PAPER NO. 4

NINTEENTH IEEE MIDWEST SYMPOSIUM ON CIRCUITS AND SYSTEMS
MILWAUKEE, WISCONSIN
AUGUST 16-17, 1976

SENSITIVITY STUDY OF SAMPLED ANALOG SIGNAL PROCESSING*

T. F. Tao, S. V. Holmes, M. Pollack and J. Campbell
Naval Postgraduate School
Monterey, California 93940

Abstract

Sensitivity studies of three sampled analog signal processors are briefly surveyed and summarized. Processors considered are the nonrecursive tapped delay line filters, the recursive comb filters and the chirp Z transformers.

1. INTRODUCTION

In the past few years, a new signal processing technique different from both the digital processing and the analog processing has been actively developed. Its unique feature is that the signal being processed is in a sampled analog form. The independent variable of the signal is sampled (or discrete) similar to a digital signal. On the other hand, the signal itself is analog (or continuous) in contrast to a digital signal which is quantized. It is generally known as the sampled analog signal processing[1] and is the subject of this session. Its principle has a great deal in common with digital processing. Their signals are both sampled. The processing steps are both carried out in time domain using delays, or in spatial domain using shifts. Their theories are both based on the Z transform. In fact, the theory and algorithms of sampled analog processing follow closely the well developed digital processing theory. It has generated recent interest because the new LSI developments are making it possible to implement sampled analog processing using hardware LSI devices with low power consumption, light weight and other advantages explained by other papers in this session. However, sampled analog processing is different from digital processing in several ways. One is the sensitivity problem which is concerned with the change of processor performance due to parameter errors and variations. While digital processing is concerned with the quantization and truncation errors, sampled analog processing, like analog processing, is affected by drift, aging and nonuniformity and will be limited by lower accuracy less linearity and relatively more limited programmability. However, it is also different from the analog processing because its processing algorithms are essentially the same as the digital processing. Although they are implemented by hardware also, their circuits are very different from the circuits used in both the passive and the active filters. New sensitivity study is needed in order to assess the limitations and the relative merits of sampled analog signal processing. However, only preliminary studies have been made to date. Due to the incomplete state of these studies and their different approaches and considerations, it is not easy at this time to present them in a unified manner suitable for comparison. The purpose of this paper is to present the highlights of their results in order to give an impression of what have been done in regard to the sensitivity problem of sampled analog signal processing.

In section 2, three hardware signal processors are described: nonrecursive tapped delay line filters, recursive comb filters and chirp Z transformer. They are selected because some of their sensitivity problems have been studied. In section 3, the nature of parametric errors and variations are described. In section 4, some

results of the frequency domain sensitivity of the sampled analog signal processor are presented.

2. HARDWARE SAMPLED ANALOG SIGNAL PROCESSOR

The majority of signal processing operations can be grouped into four categories shown in Table I which includes two types of filters and two types of spectral analysis. To present their algorithms, only time varying signals are considered in this table. It describes the processing operations in two ways: first in discrete time domain, then in terms of expressions in Z domain. For the filters, only three mathematical operations are involved: delay (n - k or k - n), multiplication by a constant and summation. The constants are the filter coefficients. Both n and k are the discrete time variables. For the spectral analysis, two additional mathematical operations are needed: multiplication of two variables (n k) and the generation of special functions. The variables involved are usually not the same type. For example, in the case of discrete Fourier transform, n is the discrete time variable, k is its transform variable--the discrete frequency. The special function in this case is the sinusoidal function. It is different for other transforms: for example, a matrix for the Hadamard transform. For the implementation of sampled analog processing, all five mathematical operations can be carried out by electronic hardware as shown in Table II. One of the key devices is the CTD delay line (charge transfer device) which is a sampled analog shift register.

Nonrecursive Filters. Of the four types of processors, the nonrecursive filter has been more extensively developed than the other three. Two types of device structure have been used, as shown in Fig. 1: tapped delay line (TDL) and correlator. In the TDL structure, only one sampled analog shift register is used. The signal is nondestructively read out after every stage of delay. The multiplication of the delayed signal by the filter coefficient is not carried out by real multiplier but by using one of the following tapping methods:

- On-chip tapping:
  - 1. Fixed tap weight
    - split gate technique [2]
    - Voltage sensing diffusion technique [3]
  - 2. Programmable tap weight
    - charge sloshing and sequentially addressing technique [4]
    - MNOS variable conductance technique [5]
- Off-chip Tapping:
  - Programmable taps using external resistors [6,7]

The delayed and tap-weighted signals are then summed together by the op amp summers. For the correlator structure, two sampled analog shift registers are used: one for the signal, the other for the reference signal (or the filter coefficients).[8,9] The signals are also read out nondestructively after every stage of delay. The delayed signals corresponding to the same stage are multiplied together by the on-chip analog multipliers. Their products are then summed together by an op amp summer. Because the MOS on-chip analog multiplier has not been well developed, the TDL structure has been more extensively developed than the correlator structure. It is interesting to note that in the nonrecursive filter a time varying signal is converted into a time and spatially varying signal in the shift register. At any given instant, signals throughout the shift register are multiplied by the tapping weights and summed together simultaneously in a parallel processing manner which increases the speed of a serial processing by a factor of TW, where T = total delay time in the shift register, W = bandwidth = half of clocking frequency.

Recursive Filters. Relatively, sampled analog recursive filter has not been nearly as extensively developed as the nonrecursive filters. Several studies have been reported using the same canonical circuit for a digital filter shown in Fig. 2. However, there is a significant difference in the delay elements. While the delay is z^{-1} in the digital filter, i.e., recursion of signal takes place after delay of every sampling period, the delays used in the sampled analog recursive filters is z^{-N}, where N is the number of sampling period delay in each delay device. The filter coefficients are generally

set off chip by resistive potentiometers. Bucket brigade device (BBD) delay lines were first used[10]. More recently, recursive filters using CCD delay lines[11-13] have been reported. We have also studied recursive filters using Reticon delay lines[14]. In these recursive filter studies, because the delay devices all have more than one delay stage, their frequency characteristics are all periodic with a frequency f_c/N, where f_c = clock frequency, N = number of delay stages, and have N/2 periods within the Nyquist frequency range from 0 to $f_c/2$. They are comb filters with N/2 teeth[13] although in some applications only the first tooth is used.

Spectral Analysis--Chirp Z Transform. Several algorithms exist for performing a discrete Fourier transform (DFT). For digital processing, the fast Fourier transform (FFT) algorithm is the most popular one. For sampled analog processing, the chirp Z transform (CZT) algorithm is more attractive because the hardware tapped delay line can be used as a convolver to accomplish two of the five mathematical operations required in the DFT. The CZT algorithm is listed in Table I. Its hardware implementation is shown in Fig. 5.[20-21] The sinusoidal function is generated by designing the tapping weights of the TDL convolver according to the required function. The multiplication of two variables, n and k, is carried out by using the relation

$$2nk = k^2 - (n-k)^2 + n^2$$

It splits the DFT into three parts: premultiplication with a chirp, TDL convolver and postmultiplication with a chirp. All three parts can be implemented by hardware.

In this paper, the sensitivity problems of three signal processors are considered. They are the nonrecursive TDL filters, the recursive comb filters and the chirp Z transform. It should be noted that they only represent the standard form of their family of processors. Other variations and improved versions of their circuits will be developed as interest in sampled analog processing increases. Some of them are being presented in this session. Their sensitivities will undoubtedly be studied if the processors prove to be superior.

3. PARAMETRIC ERRORS AND VARIATIONS

Parametric errors and variations of analog devices can be modeled in different ways. Their magnitudes can be expressed either in terms of a relative error as a percentage of a parameter value or in terms of an absolute error as a percentage of a fixed maximum value. Variations either in time or in space are measured sometimes by a statistical distribution and other times by the worst case deviation. Time variations are important in noise studies. Spatial variations are of particular interest in some sensitivity studies such as the nonuniformity among the tap outputs on a TDL. Such spatial variations are also called the fixed pattern noise.

For sampled analog processing, it was shown that the processing algorithms can be broken down into five basic mathematical operations. Therefore, the parametric errors and variations of their hardware implementations should be the starting point of sensitivity study. The following sources of errors and variations can be identified.

(a) Inherent limit of accuracy of an analog parameter:

For an analog parameter, the best accuracy probably can not exceed 0.1% or extend beyond three significant numbers. In practice, errors much larger than 0.1% exist. For example, when resistors are used in the processors, errors up to 5% or even higher are not unrealistic. They are determined by the precision of a resistor. Examples of such cases are the recursive filters, the nonrecursive TDL with external tapping, and others. These errors can be modeled as relative errors. In split gate tapping, errors exist due to the truncation error incurred in the computer aided layout of the split gate. Errors in the order of 0.5-1% have been quoted. These errors should be modeled as an absolute error.

(b) Drift due to temperature variation and aging:

The parameters of both active and passive semiconductor devices change with temperature. Depending on the circuit details and the range of temperature drift, parametric variations up to 10%

is not inconceivable. Similar order of variations can be expected from aging. These variations can be appropriately modeled by the relative errors. Statistical consideration probably is not necessary.

(c) Nonuniformity and Fixed Pattern Noise:

For tapped delay lines, nonuniformity among taps is a major source of parametric variations. They are caused by the error during tapping design, the spatial variation of device characteristics during fabrication, the external loading effects, bandwidth limitation and other factors. Variations among split gates are reported to be less than 1%. They should be modeled by an absolute error. Variations among external taps are considerably higher. The error due to resistor precision should be modeled by a relative error. Fixed pattern noise should be modeled by an absolute error. Variations in other tapping methods have not been as well estimated. In all cases, statistical distributions should be considered.

(d) Errors Due to Nonlinearity:

Depending on the device and the signal amplitude, nonlinearity will introduce errors up to a few percent. They could be modeled as relative error. Statistical consideration probably is not important.

4. SENSITIVITY OF SAMPLED ANALOG SIGNAL PROCESSORS

Sensitivity study is concerned with the change of equipment performance when the device parameters deviate from their designed values. The performance can be measured in many ways depending on the applications, some in frequency domain, others in time domain, spatial domain, spatial frequency domain or in other transform domain. Due to the proliminary state of the sensitivity study, only a few performance characteristics have been examined. This paper considers the sensitivity of frequency domain characteristics only.

Nonrecursive Tapped Delay Line Filter. Two computer simulations and one theoretical analysis have been reported. The performance investigated is concentrated to the stop band (or side lobe) attenuation of a filter. In the first computer study,[15] a wide band low pass filter and a wide band bandpass filter were studied. Both used a 32 TDL. A Monte Carlo program was used to calculate the sensitivity of frequency characteristics with respect to random tap errors. The low pass filter case is shown in Fig. 4. The stop band attenuation is reduced approximately from -47 db to 27 db, a loss of 20 db for 2% of random tap error. In the second computer study,[16] narrow band filters were considered. The length of TDL is 128 taps. In Fig. 5, the sensitivity of a low pass filter with respect to a 5% random tap variation is shown. The stop band attenuation is reduced from -60 db to -40 db, a loss of 20 db for 5% of random tap error. It hinted that a wide band TDL filter is probably more sensitive to the tap error than a narrow band TDL filter. In Fig. 6, a different type of error, the truncation error, is considered. Any tap weight smaller than 2% of the maximum is set to zero. For this case, 72 out of the 128 taps were less than 2% and were set to zero. The frequency characteristics showed a ripple component. But the stop band attenuation was only deteriorated by 5 db. In the theoretical study,† much of the effort was concentrated on the worst case deviations. Both relative error and constant error were analyzed with no statistical considerations.

Recursive Comb Filter. A computer study of the sensitivity of second order sampled analog recursive comb filter was made.[17] A representative result is shown in Fig. 7 which presents the sensitivity of the cut-off frequency of a low pass comb filter for a relative 5% change in the feedback coefficients b_1 and b_2. No statistical consideration was applied. The feedforward coefficients a_0, a_1 and a_2 of this filter are fixed at 1, 2, 1 respectively because with this relation the frequency characteristic reached zero between neighboring comb teeth. In other words, the teeth are clearly separated from one another which is a very desirable characteristic in comb filter applications. It was found that for stable operations, b_1 and b_2 must be within the triangle shown in Fig. 7. In this figure, several contours representing a 2% shift in the cut-off frequency for a

† G. Temes, UCLA, private communication

5% variation either in b_1 or in b_2 are shown. The shaded area represents the region in which there will be a less than 2% shift in cut-off frequency when both b_1 and b_2 are changed by 5%. Similar studies have been carried out for the high pass comb filter cases.

Spectral Analysis--Chirp Z Transform. One computer study and one theoretical analysis of the sensitivity of CZT were made. In the computer study,[18] random variations of relative errors in the tapping coefficients and in the pre-and post-multipliers were considered. In Fig. 8, the frequency spectrum of the impulse response of a 512 point CZT is shown for both the ideal case and the case of ±4% relative error with constant random distributions. In the theoretical analysis,[19] the sensitivity of a 32 point CZT is estimated by first calculating the variance σ_k of the DFT coefficient G_k and then defining a relative accuracy as $A = 1.96\sigma_k/G_k$. Two cases were considered. In the first case, the errors in the premultiplier, the tap weights of the convolver, the nonlinearities in the differential current amplifier, in the post-multiplier and the mismatch of post-multipliers were all assumed to be ε. The relative accuracy is found to be $A = 1.27\varepsilon$ for the DC component and 2.54 for other spectral components. In the second case, errors for these five sources were different.

5. CONCLUSION

Sensitivity of sampled analog signal processing is of interest because it is needed in assessing the limitations and the relative merits of this signal processing technique. Furthermore, it should be used to develop and to compare new algorithms of sampled analog processing. This paper presented a glimpse of the preliminary studies of the sensitivity problems to date. It is difficult to organize them in a unified manner suitable for comparison because the processors studied, their errors and variations considered and the performance criteria applied have all been quite different. It is believed that a unified study is needed and will be attempted when the practical interest of sampled analog processing is more widely recognized. It is hoped that this paper will serve as the first stepping stone toward that objective.

ACKNOWLEDGEMENT

The authors wish to thank Prof. S. R. Parker for his support and encouragement, Mr. H. Whitehouse for his support of the summer visit of one of the authors (Tao) and A. Schow for her help in preparing the manuscript.

* Partially supported by the Naval Postgraduate School Research Foundation and by the Naval Electronics Systems Command.

REFERENCES

1. Proceedings of IEEE Advanced Solid State Components For Signal Processing, IEEE International Symposium on Circuits and Systems, Massachusetts, April 1975. IEEE Catalog No. 75CH0979-5CAS.

2. D. D. Buss, D. R. Collins, W. H. Bailey and C. R. Reeves, "Transversal Filtering Using Charge-Transfer Devices," IEEE Journal of Solid-State Circuits, SC-8, No. 2, p. 138, April 1973.

3. D. D. Buss, W. H. Bailey and D. R. Collins, "Matched Filtering Using Tapped Bucket-Brigade Delay Lines," Electronic Letters 8 No. 4, Jan., 1972.

4. J. J. Tiemann, W. E. Engeler, R. D. Baertsch and D. M. Brown, "Intracell Charge-Transfer Structures for Signal Processing," IEEE Trans. Elec. Dev., ED-21, pp. 300-308, May 1974.

5. M. H. White, D. R. Lampe and J. L. Fagan, "CCD MNOS Devices for Programmable Analog Signal Processing and Digital Nonvilatile Memory," 1973 Int'l Electron Devices Meeting, Tech. Dig., pp. 130-133, Washington, D.C., Dec. 3-5, 1973.

6. TAD-12 Tapped Analog Delay Device by Reticon Corp.

7. "Study of the Use of Charge Coupled Devices In Analog Signal Processing Systems" Second Report to Naval Research Laboratory, Westinghouse Defense and Electronics Systems Center, Feb. 1976.

8. J. Harp, G. Vanstone, D. McLennan and J. Mavor, Proceedings 1975 International CCD Applications Conference, San Diego, p. 229, 1975.

9. P. Bosshart, Tech. Digest 1976 International Solid State Circuits Conference p. 198, 1976.

10. D. A. Smith, C. M. Puckette and W. J. Butler, IEEE J. Solid State Circuits, SC-7, p. 421, 1972.

11. J. E. Bounden, R. Eames, and J. B. Roberts, Proc. Int. Conf. Tech. and Appl. of CCD, Edinburgh, p. 206, Sept. 1974.

12. J. Mattern and D. Lampe, ISSCC Tech. Paper Digest, p. 148, Feb. 1975.

13. T. F. Tao, V. Iamsaad, S. Holmes, B. Freund, L. Saetre, Proc. 1975 International CCD Applications Conference, San Diego, p. 257, 1975.

14. Unpublished Thesis by V. Iamsaad, B. Freund and L. Saetre, Naval Postgraduate School, Monterey, CA, 1975.

15. A. Ibrahim, J. Hupe and L. Sellers, Proceedings 1975 International CCD Applications Conference, San Diego, p. 245, 1975.

16. "Study of the Use of Charge Coupled Devices In Analog Signal Processing Systems" First Report to Naval Research Laboratory, Westinghouse Defense and Electronics Systems Center, May 1974.

17. Unpublished Thesis by S. V. Holmes, Naval Postgraduate School, Monterey, CA, 1976.

18. Unpublished Thesis by M. Pollack, Naval Postgraduate School, Monterey, CA. 1976.

19. E. Wrench, Naval Undersea Center, San Diego, Private Communication.

20. G. Bryan, J. Alsup, J. Speiser, H. Whitehouse, Proc. NATO Advanced Study Institute on Signal Processing, Aug. 1972.

21. H. Whitehouse and R. Means, Proc. IEEE Advanced Solid State Components For Signal Processing, p. 5, IEEE Catalog No. 75CH0979-5CAS, 1975.

TABLE I. SIGNAL PROCESSING OPERATIONS

Filtering

Filters	Discrete Time Domain	Z Transform Domain
General Forms		$H(Z) = \frac{O(Z)}{I(Z)} = \frac{\Sigma\, a_n Z^{-n}}{1 + \Sigma\, b_m Z^{-m}}$
Nonrecursive	$O(k) = \sum_{0}^{N} a_n\, i\,(k-n)$ (convolution) $O(k) = \sum_{0}^{N} a_n\, i\,(n-k)$ (correlation)	$O(Z) = \sum_{0}^{N} a_n Z^{-n}\, I(Z)$
Recursive	$O(k) = \sum_{0}^{N} a_n\, i\,(k-n) + \sum_{0}^{M} b_m\, o\,(k-n)$	$O(Z) = \sum_{0}^{N} a_n Z^{-n}\, I(Z) + \sum_{I}^{M} b_m Z^{-m} O(Z)$

Spectral Analysis, Transforms

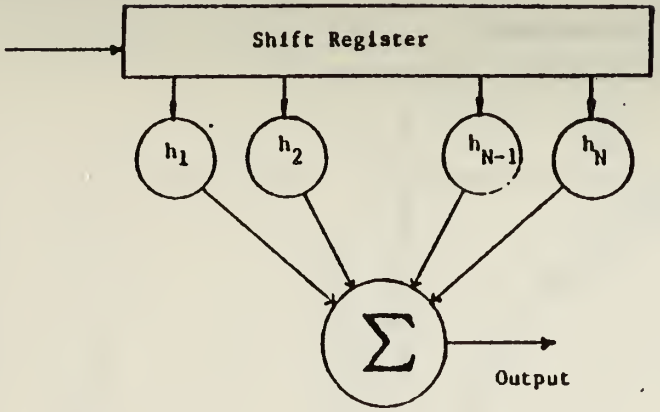
Discrete Fourier Transform	FFT: $F_k = \sum_{0}^{N-1} f_n\, e^{-j\frac{2\pi}{N}nk}$ CZT: $F_k = e^{-j\pi\frac{k^2}{N}} \left(\sum_{0}^{N-1} f_n\, e^{j\pi(k-n)^2/N}\, e^{-j\pi\frac{n^2}{N}} \right)$
Other Transforms	Karhunen-Loeve, Hadamard, Walsh, etc. Hilbert

Note: FFT is fast Fourier transform; CZT is Chirp Z transform

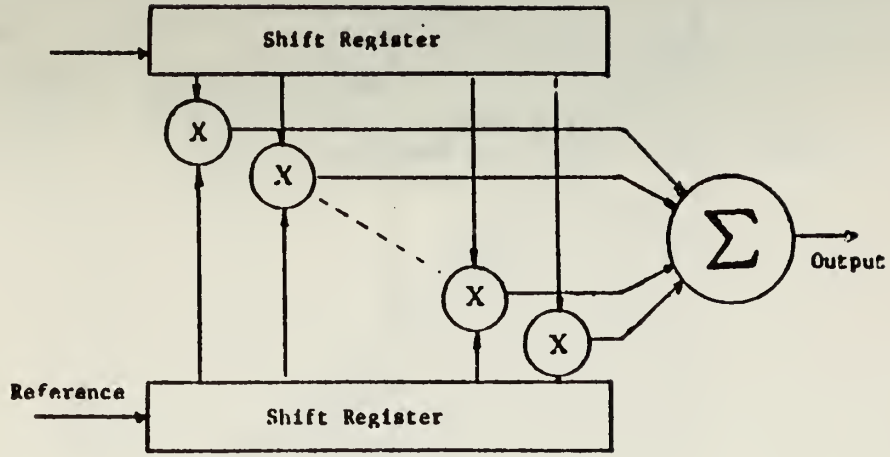
TABLE II. INTEGRATED CIRCUIT IMPLEMENTATION OF MATHEMATICAL OPERATIONS FOR SAMPLED ANALOG SIGNAL PROCESSING

Signal Processors	Delay	Multiplication By Constant	Summation	Multiplication of Two Variables	Generation of Function
Nonrecursive Filters					
Tapped Delay Line	CTD*	Tapping Weights	Op Amp Summer	-----	-----
Correlator	CTD	Analog Multipliers	Op Amp Summer	-----	-----
Recursive Filters	CTD	Potentiometers	Op Amp Summer or Charge Summing	-----	-----
Chirp Z Transformers	CTD	Tapping Weights	Op Amp Summer	Analog Multipliers and Convolver	Designed as Tapping Weights

* CTD = Charge Transfer Device



(a) Tapped Delay Line Structure



(b) Correlator Structure

Fig.1 Two Configurations of Nonrecursive Filters

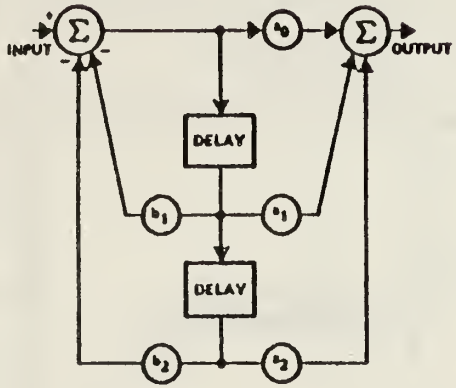


Fig.2 Canonical Form of a Second Order Recursive Filter

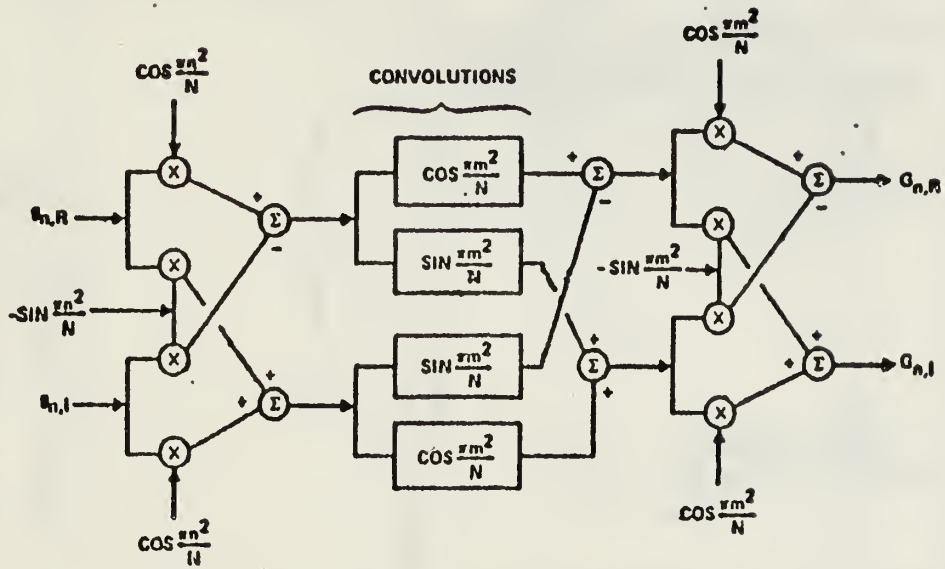


Fig.3 Implementation of Complex Discrete Fourier Transform by the Chirp Z Transform Algorithm

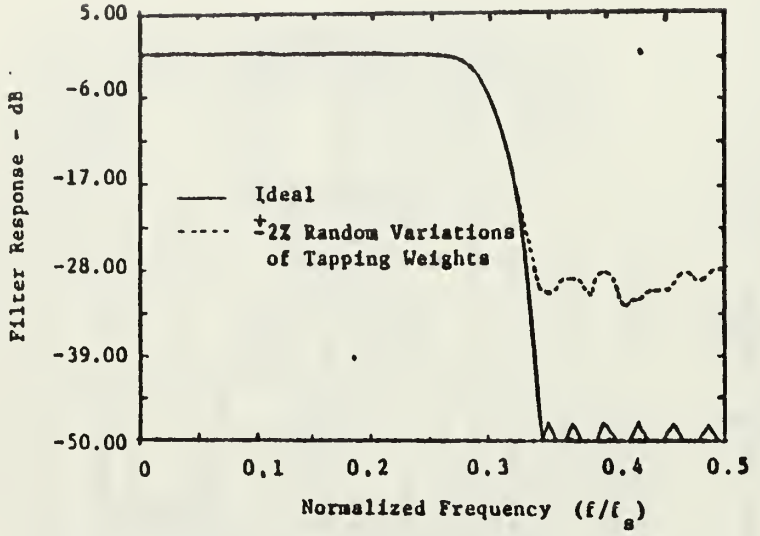


Fig.4 Frequency Characteristics of Wide Band Low Pass Filter

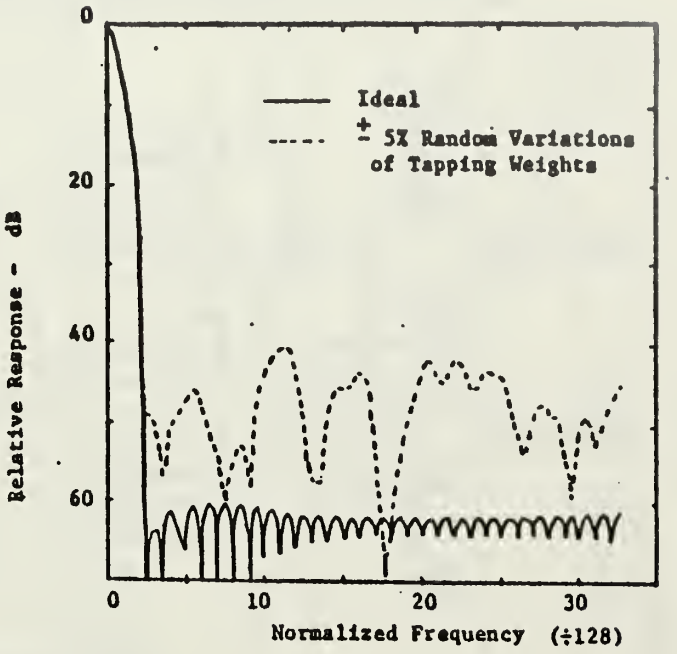


Fig.5 Frequency Characteristics of Narrow Band Low Pass Filter (128 Taps)

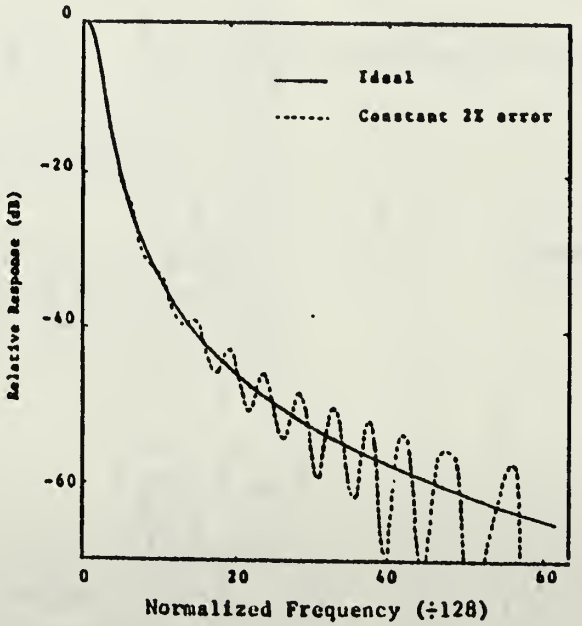


Fig.6 Frequency Characteristics of Narrow Band Low Pass Filter (128 points)

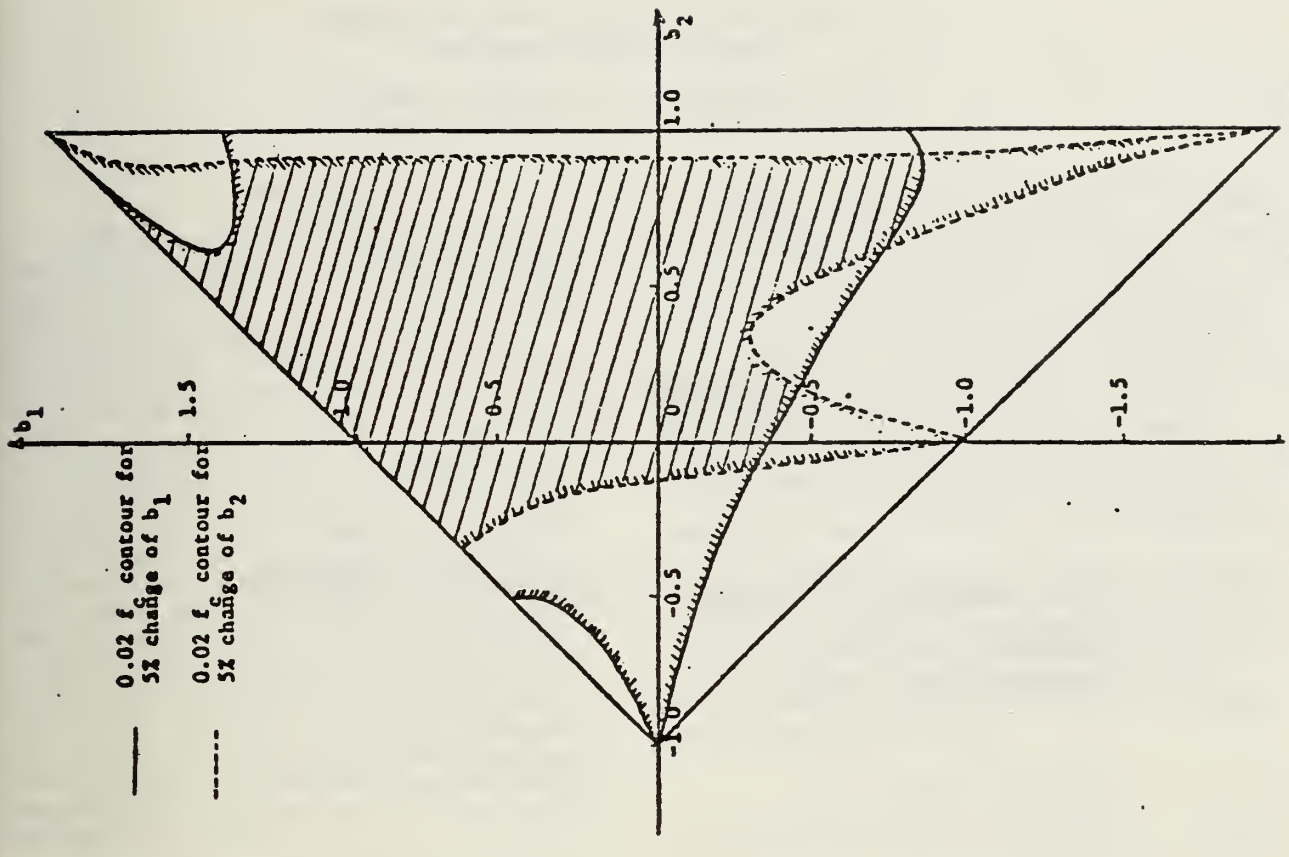


Fig.7 Cut-Off Frequency Sensitivity Contours of Low Pass Second Order Recursive Comb Filters (2% Change of f_c for 5% Change of b_1 and b_2)

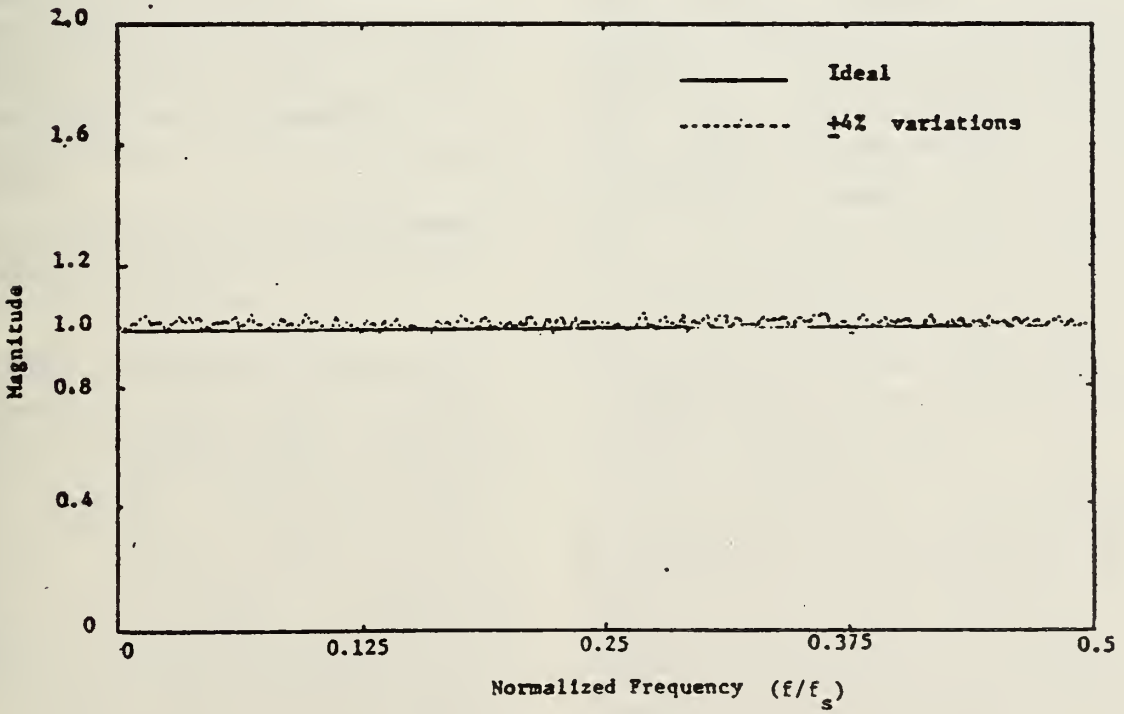


Fig.8 Frequency Spectrum of the Impulse Response of Chirp Z Transform (512 points)

PAPER NO. 6

1976 GOVERNMENT MICROCIRCUITS APPLICATIONS CONFERENCE
ORLANDO, FLORIDA
NOVEMBER 9-11, 1976

RECURSIVE CTD COMB FILTER AND ITS APPLICATION TO MTI RADAR SIGNAL PROCESSING*

T. F. Tao, L. Saetre, S. V. Holmes and A. Ejaz
Naval Postgraduate School, Monterey, California

I. Introduction

In MTI radar, moving targets are distinguished from clutters or stationary targets based on the Doppler shifts between their spectra in frequency domain, shown in Fig. 1(a). The clutter is characterized by a periodic spectra at D.C. and multiples of PRF. The moving target is characterized by another spectra of the same period but shifted by its Doppler frequency. The clutter spectra can be eliminated by a canceller type of filter of periodic (or comb) characteristics shown in Fig. 1(b). This type of comb filter for MTI radar applications was implemented first using quartz delay lines. In recent years, digital processing has been increasingly used for its implementation. More recently, a new sampled analog signal processing technique using charge transfer devices (CTD) is being developed which is different from the digital processing but can perform many of the digital processing functions. The basic difference lies in the fact that the signal itself remains in analog form although its time variable is sampled as in the digital case. Several approaches attempting to use the sampled analog charge transfer devices for radar processing have been reported.

- Nonrecursive Filter Approach -- Feedforward canceller (General Electric)[1]
- Spectral Analysis Approach -- Chirp Z transform comb filter (Texas Instrument)[2]
- Recursive Filter Approach -- Using interrupted clock (Royal Radar Establishment)[3]

  -- Recursive comb filter (Naval Postgraduate School)[4]

This paper describes the properties of a sampled analog recursive comb filter using CTD delay device and will demonstrate its feasibility as the canceller for an MTI radar.

II. Sampled Analog Recursive CTD Comb Filter

A second order sampled analog recursive CTD comb filter is used in this study. Its circuit is similar to the canonical circuit of the digital recursive filter shown in Fig. 2 except for one significant difference. The delay is now implemented by a CTD delay line which consists of N delay stages. For digital filters, the delay is accomplished either by memory or by shift register and consists of only one stage. The presence of N stages of delay is instrumental in providing the comb feature of its frequency characteristics with very large attenuation between neighboring comb teeth. The reason for this outstanding property can be explained qualitatively as follows. Since the signal is delayed N clock periods before it is fed back to the input, the frequency of recursive operation, f_r, is only one N^{th} of the sampling frequency, f_s. Therefore, frequency characteristics of this type of filter is periodic with the recursive frequency f_r. Since $f_r = f_s/N$, there are N/2 comb teeth within the Nyquist frequency range from 0 to $f_s/2$. This is in contrast to the usual digital recursive filters where the recursions take place after each delay stage, resulting in only one comb tooth within the Nyquist range. Since there are now N/2 teeth, a proper design procedure introducing a zero or zeroes at the recursion frequency will provide infinite attenuation between neighboring teeth in principle. The bilinear Z transform design procedure is able to provide this characteristic.

The theory of the recursive CTD comb filter is essentially the same as the digital recursive filter theory but with three modifications. The first modification takes into account the fact that a CTD delay line generally consists of N delay stages. The transfer function of a second order canonical circuit is:

$$H(Z) = \frac{a_0 + a_1(f)Z^{-N} + a_2(f)Z^{-2N}}{1 + b_1(f)Z^{-N} + b_2(f)Z^{-2N}} \qquad (1)$$

where Z^{-N} represents a delay of N clock periods. In digital recursive filters, N = 1. The second modification takes into account the frequency dependence of the filter coefficients caused by several factors: sampled and hold effect, charge transfer inefficiency effect, frequency roll-off effect of electronic components due to their bandwidth limitation and others. In digital filters, the coefficients are independent of frequency. Third, in order to introduce a zero (or multiple zeroes) at the recursion frequency to provide infinite attenuation between neighboring teeth, the feedforward coefficients a_0, a_1 and a_2 must be related by the ratio of 1:2:1. In digital filters, such a restriction is not necessary. An example of the calculated frequency response of a second order comb filter is shown in Fig. 3. An 8 bits CCD (charge coupled device) delay line was used. The sampling frequency was 20 KHz. The filter coefficients were: $a_0 = 1$, $a_1 = 2$, $a_2 = 1$, $b_1 = -0.3$, $b_2 = +0.3$. In the calculation, the sampled-and-hold effect and the charge transfer inefficiency effect were included for the filter coefficients. It can be seen that the filter response is periodic with respect to a frequency of 2.5 KHz which is equal to f_s/N. However, it deviated from the ideal comb filter characteristics in three ways as the frequency is increased or as the order of the comb teeth is increased. First, the attenuation at the null frequency deteriorated from over -50 db for the first tooth to -42 db for the second tooth and

to -34 db for the third tooth and so on. Second, the peaks of the teeth decrease. Third, the teeth shapes vary. These deviations are more pronounced in the filter of Fig. 4, which shows the experimental frequency response of another second order comb filter using two 96 bits CTD[5] delay lines. The sampling frequency was 400 KHz. The filter coefficients were: $a_0 = 1$, $a_1 = 2$, $a_2 = 1$, $b_0 = +0.2$, $b_1 = +0.2$. This comb filter has 48 teeth below the Nyquist frequency at 200 KHz. The deviations observed from the theoretical calculation shown in Fig. 3 are again seen in the experimental result of this filter. Theoretical calculation has not been included in this figure for its clarity. However, the status of the theoretical analysis will be summarized as follows. To a first approximation, the modified digital recursive filter theory has been able to analyze several major features of the sampled analog recursive comb filter: the comb nature, the loss of attenuation between neighboring teeth, the decrease of the tooth amplitude and the change of tooth shape. There are still some features unaccounted for by the theoretical analysis. One of such features is the rise of tooth amplitude, as indicated in Fig. 3. It can be seen also that perfect agreement between the theory and measurement has not been accomplished yet. From the synthesis point of view, how should the circuit be modified to achieve an ideal comb filter characteristic has not been answered. However, in spite of the less than perfect status of this comb filter development, an attempt was made to investigate the feasibility of using this type of comb filter as canceller for an MTI radar.

III. Application to MTI Radar Signal Processing

The principle of the application as canceller is to match the recursion frequency of the comb filter to the PRF (pulse repetition frequency) of the radar because the clutter spectra of a radar occur at the multiples of PRF and the comb filter response is zero at the multiples of f_r. The zero response will cancel the clutter spectra. In addition, bandwidth requirement must also be met. Altogether, three relations must be satisfied:

$$\begin{cases} \text{Canceller requirement:} & f_s/N = \text{PRF} \\ \text{Bandwidth requirement:} & BW \geq \frac{1}{2\tau} \\ \text{Sampling frequency requirement:} & f_s \geq 2BW = \frac{1}{\tau}, \\ & \text{i.e., } f_s \geq \frac{1}{\tau} \end{cases}$$

where τ = radar pulse width. Consider a practical MTI radar with PRF = 1000 Hz and τ = 1 microsec:

$$N \geq 1000 \text{ and } f_s \geq 1 \text{ MHZ} \qquad (2)$$

However, when the feasibility study was made the delay line available had only 96 delay stages.[5] If a PRF of 800 Hz is chosen, the CTD delay line can only be clocked at a sampling frequency of $96 \times 800 = 76.3$ KHz. Then the usable bandwidth is $f_s/2 = 38.4$ KHz which is only able to pass a pulse whose width is greater than 13 microseconds. Such a wide pulse width is not usually used in practical radars. Therefore, instead of using the signal in a real radar, an MTI radar signal using a wide pulse width was simulated to demonstrate the feasibility of using this type of comb filter as a canceller. First, a stationary target or a clutter signal is simulated by a pulse, as shown in Fig. 5(a). The output of the filter showed that the clutter signal has been cancelled except for its leading and trailing edges, indicating that some of its high frequency components had not been cancelled. This could be the consequence of the deteriorated nulls at multiples of f_r as the frequency is increased. In order to provide large enough bandwidth, a high PRF of 10 KHz was chosen which required a sampling frequency of 960 KHz. Next, a moving target signal is simulated by a train of pulses whose amplitudes were modulated by the Doppler frequency, producing the so-called "butterfly" or "bipolar" signal shown in Fig. 5(b). In this case, the Doppler frequency was 5 KHz. The filter output showed that the moving target signal passed through the filter with small insertion loss, except that some of its high frequency components were lost. Last, in Fig. 5(c), both the clutter and the bipolar moving target signal were superimposed. The filter output showed that the clutter was cancelled while the moving target passed through the filter.

IV. Conclusion

Although the simulated signal had a wide pulse width which is not used in practical radars, the filter response shown in Fig. 5 has demonstrated that this type of filter can be used as a canceller in MTI radar if a CTD delay line of enough delay stages and clocked at a high enough frequency can be found. Further investigation is being continued because such a CCD delay line has become available recently. It has 455 delay stages and can be clocked above 10 MHz.[6] By cascading two of these delay lines, it is approaching the requirements of 1000 stages and clock frequency above a few MHz. However, its dynamic range may be limited.

Acknowledgement

The authors wish to thank Professors S. R. Parker and R. Fossum for their support, P. Kopp for his technical assistance, V. Iamsaad, B. Freund, T. Zimmermann, and Professor G. Tomes for technical discussions and A. Schow for her help in preparing the manuscript.

References

1. Butler, W.J., Engeler, W.E., Goldberg, H.S., Puckett, C.M. and Lobenstein, H., IEEE International Symposium on Circuits and Systems, Boston, Mass., April 1975, IEEE Publication on Advanced Solid State Components for Signal Processing, p. 30.

2. Bailey, W.H., Broderson, R.W., Eversole, W.L., Whatley, M.M., Hite, R., Buss, D.D., and Sproat, R., 1974 Government Microcircuit Applications Conference, GOMAC Digest of Papers S, p. 76.

3. Bounden, J.E., Eames, R. and Roberts, J.B., Proceedings, International Conference on Technology and Applications of CCD, Edinburgh, Scotland, September 1974, p. 206.

4. Tao, T.F., Iamsaad, V., Holmes, S., Freund, B., Saetre, L. and Zimmermann, T., Proceedings, International Conference on CCD Applications, San Diego, California, October 1975, p.

5. Reticon SAD 100 Delay Device

6. Fairchild 321 CCD Delay Device.

* Supported by the Naval Postgraduate School Research Foundation and the Naval Electronics Systems Command.

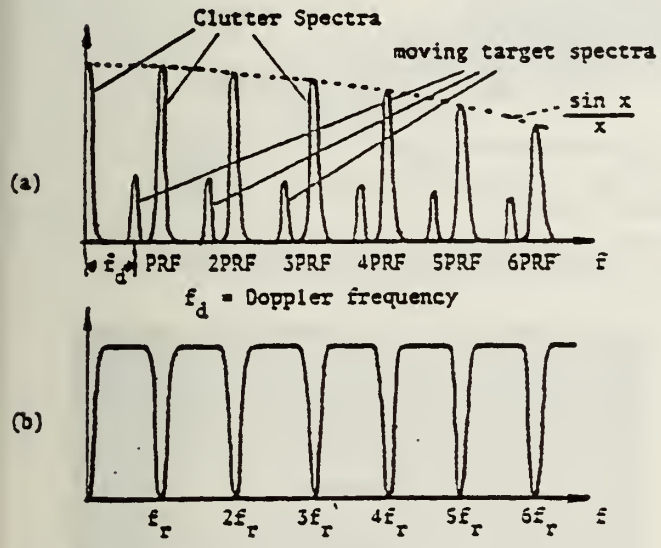


Fig. 1 FREQUENCY CHARACTERISTICS OF
(a) RADAR SIGNAL
(b) IDEAL CANCELLER TYPE COMB FILTER

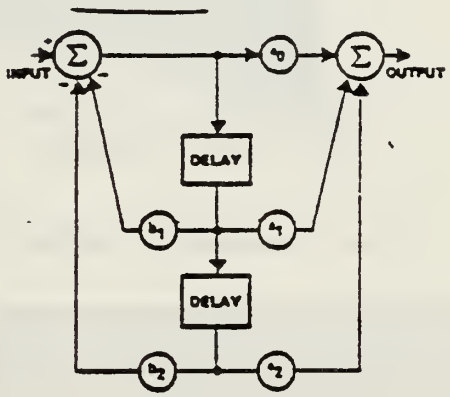


Fig. 2 CONFIGURATION OF SECOND ORDER CANONICAL RECURSIVE FILTER CIRCUIT

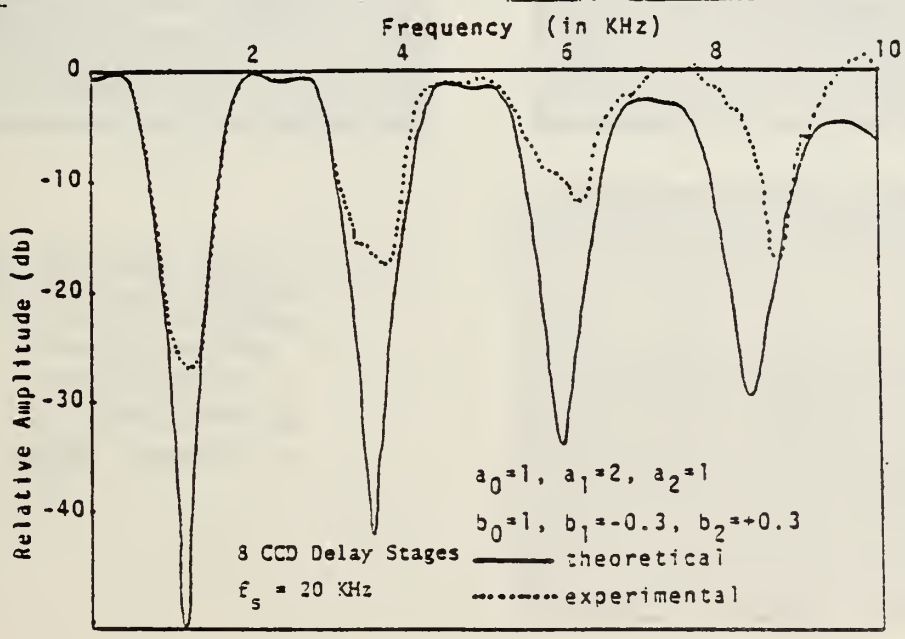


Fig.3 THEORETICAL AND EXPERIMENTAL TRANSFER FUNCTION OF A SECOND ORDER CANCELLER TYPE COMB FILTER

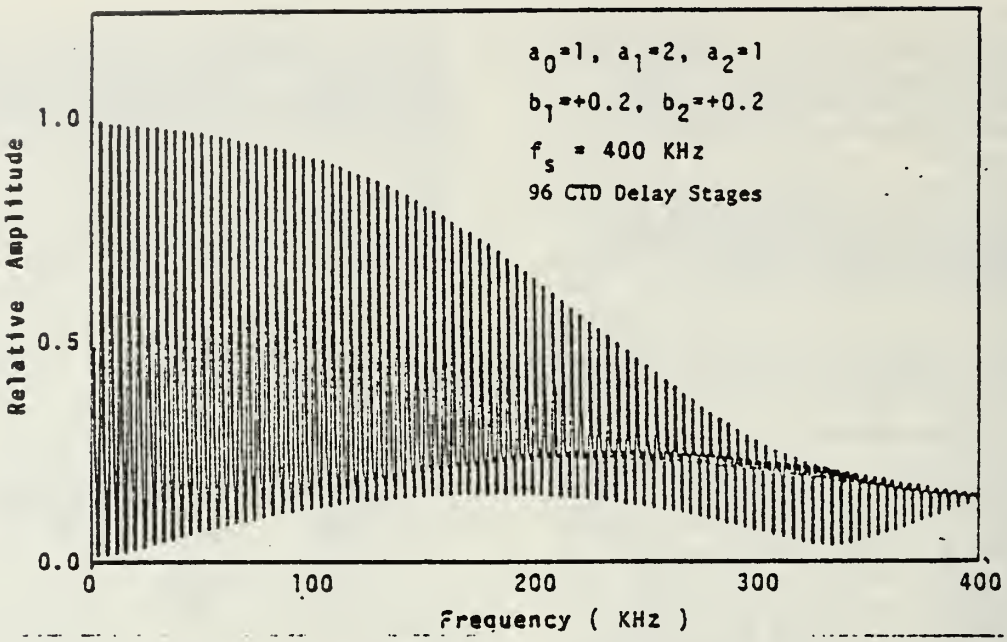
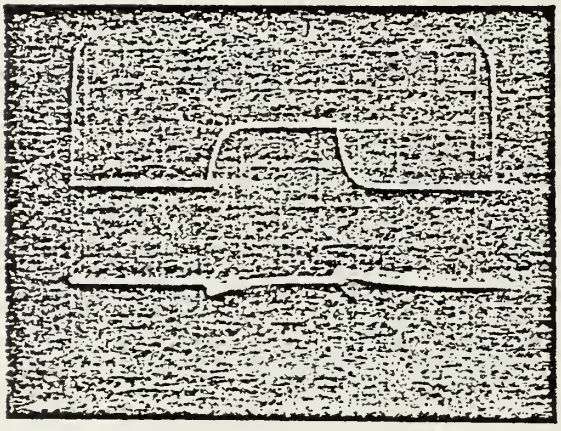
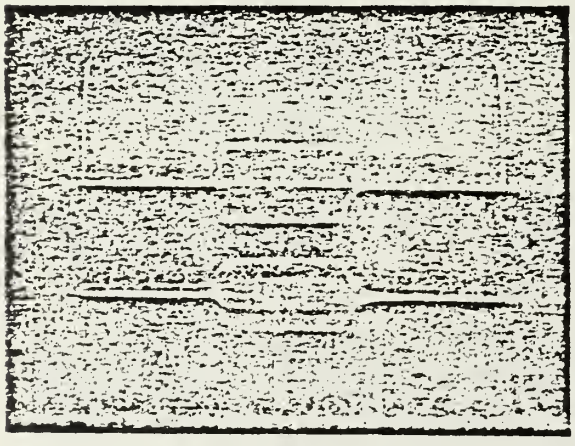


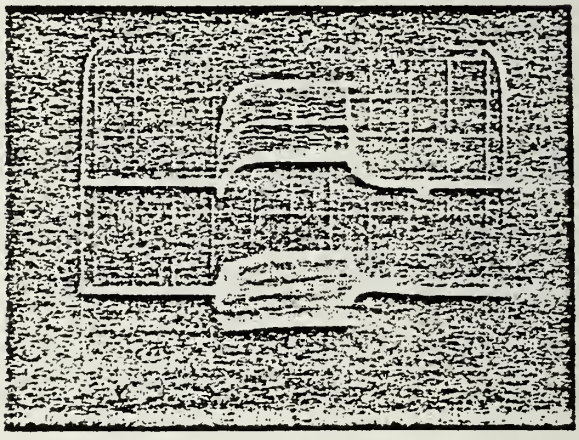
Fig. 4 EXPERIMENTAL TRANSFER FUNCTION OF A SECOND ORDER RECURSIVE COMB FILTER



(a)



(b)



(c)

Fig. 5 SIMULATED RADAR SIGNAL (TOP CURVE) AND RECURSIVE COMB FILTER OUTPUT (BOTTOM CURVE), PRF = 10 KHz
Horizontal: 50 μsec/cm

(a) STATIONARY TARGET

(b) MOVING TARGET (f_d = 5 KHz)

(c) STATIONARY + MOVING TARGETS (f_d = 5 KHz)

PAPER NO.7

Tenth Asilomar Conference on Circuits, Systems and Computers

November, 1976, Asilomar, California

SENSITIVITY STUDY OF THE CHIRP Z TRANSFORM AND THE PRIME TRANSFORM AS SAMPLED ANALOG DISCRETE FOURIER TRANSFORM ALGORITHMS*

J. G. Campbell, T. F. Tao and M. A. Pollack
Naval Postgraduate School
Monterey, California 93940

ABSTRACT

Sensitivity of sampled analog Discrete Fourier Transforms using the Chirp Z Transform and the Prime Transform algorithms to various hardware inaccuracies and errors are studied. The sensitivity results are presented in terms of the root mean square noise to signal ratio. Comparisons with digital Fast Fourier Transform are also made.

1. INTRODUCTION

In calculating the Fourier coefficients of a sampled signal, there are several algorithms for performing the Discrete Fourier Transform (DFT). The Fast Fourier Transform (FFT) is a highly efficient computational procedure which takes advantage of the fact that DFT coefficients can be calculated iteratively. The Chirp Z Transform (CZT) and the Prime Transform (PT) reduce the DFT to a sequence of three operations which center around the performance of a convolution in the CZT case and a correlation in the PT case. For digital DFT, time required in performing the convolution or correlation is quite long such that both CZT and PT showed no advantage over FFT. Consequently, FFT is by far the most popular algorithm for digital DFT. However if the signal is in a sampled analog form, convolution or correlation can be efficiently implemented using either surface acoustical wave (SAW) or charge transfer device (CTP) tapped delay line. Hence, CZT and PT become attractive algorithms for sampled analog DFT. Buss et. al [1] compared the performance of CZT and the digital FFT in terms of noise to signal ratio considering two sources of error in the CZT. Wrench [2] made an analytic sensitivity study of both CZT and PT and expressed the sensitivity results in a relative accuracy. In this paper, the sensitivities of both CZT and PT algorithms to four different sources of inaccuracies and errors in hardware implementations are studied by means of computer simulation. A single figure of merit, the root-mean-square noise to signal ratio is used to measure the sensitivity performance. CZT and PT are first compared between themselves and then are compared with digital FFT.

2. MEASUREMENT OF SENSITIVITY

In their sensitivity studies of digital FFT, Welch [3] and Weinstein [4] used the ratio of RMS error to RMS signal to measure the effect of roundoff or truncation error on fixed point and floating point digital FFT respectively. Because this ratio is based on the basic properties of the DFT, it is used also in this study to describe the sensitivities of sampled analog CZT and PT to inaccuracies and errors in their hardware implementations and to compare them with digital FFT. It is defined as follows:

Assume the coefficient outputs of a DFT of length N is an array containing N real and N imaginary numbers. Let A be the coefficient array for an ideal transform, A' be the coefficient array for the real transform containing device inaccuracies and errors; then the RMS error is the square root of the mean squared difference between A and A'.

Let K be the mean squared value of the input sequence. By Parseval's Theorem, the mean squared value of the complex coefficient output sequence is NK. The array A has mean squared value NK/2. The RMS signal is defined as $\sqrt{NK/2}$. Then the RMS error to RMS signal is a noise to signal ratio (N/S) which is normalized for signal amplitude and transform length. It will be expressed in db for presentation of sensitivity results and for comparisons among different algorithms and for different types of inaccuracies and errors.

Two types of real input signal are used in this study. One is a sequence of random numbers uniformly distributed between -1.0 and +1.0. In evaluating the sensitivity due to charge transfer inefficiency, a second type of input signal - sinusoids, is also used.

3. CHIRP Z TRANSFORM

The N-point DFT of a sequence of samples $\{x_n\}$ is defined as:

$$X_k = \sum_{n=0}^{N-1} x_n W^{nk}, \quad k = 0,1, \ N-1 \qquad (1)$$

where $W = \exp(-j2\pi/N)$

Bluestein [5] has shown that the substitution

$$2nk = n^2 + k^2 - (k-n)^2 \qquad (2)$$

allows the DFT to be rewritten in the form

$$X_k = W^{k^2/2} \sum_{n=0}^{N-1} (x_n W^{n^2/2}) W^{-(k-n)^2/2} \qquad (3)$$

Although this last expression appears more complicated than the original, the DFT has in fact been reduced to the following three operations:

(1) Premultiply the input sequence $\{x_n\}$ by the discrete chirp $\{W^{n^2/2}\}$.

(2) Convolve the resulting sequence with the discrete chirp $\{W^{-n^2/2}\}$.

(3) Postmultiply the result by the discrete chirp $\{W^{k^2/2}\}$.

The convolution of the step (2) may be performed in a surface acoustical wave (SAW) or charge transfer device (CTD) transversal filter of length 2N-1 [6]. For a real input sequence and transversal filters with real coefficients, the N-point CZT can be implemented as shown in Fig. 1. In this implementation, four errors should be considered:

- Errors in analog premultipliers and post multipliers
- Errors due to mismatch and nonlinearity in the two sets of differential amplifier summers
- Tap weight errors in the transversal filters
- Charge transfer inefficiency if CCD tapped delay lines are used as the transversal filter

(a) Multiplier Errors:

It is assumed that multiplication is performed by a multiplying D/A converter. The multiplier coefficients are stored digitally in a ROM. They are converted into analog numbers in the multiplying D/A converter. For a given multiplication, it is assumed that its error can be described by a percentage error uniformly distributed in the range from $-a_1$ to $+a_1$. Values of a_1 ranged from 0.001 to 0.02 (0.1% to 2%) which correspond roughly to 10 bits to 6 bits of ROM storage for the multiplier coefficients. A random number input signal is used in the computer simulation. Sensitivity results of two cases are shown in Fig. 2 for a 512 point CZT. In case A, errors are assumed in only one multiplier stage. In case B, both multipliers have the random errors. The N/S ratio of case B is 3db above that of case A, as expected.

(b) Differential Amplifier Errors:

It is assumed that errors in differential amplifiers can be described by percentage errors uniformly distributed in the range from $-a_2$ to $+a_2$. The range of a_2 considered in this study is also from 0.1% to 2%. It was found the sensitivity results of CZT to the differential amplifier

errors are exactly the same as in the case of multiplier errors. Therefore, Fig. 2 applies equally well to the case of differential amplifier errors.

(c) Transversal Filter Tap Weight Errors:

The model used describing the tap weight errors depends on the tapping method. If the tap weights are implemented by external resistors, then a percentage error would be appropriate. However, if a split gate weighting method is used [2], the metal gate across the channel width, w, has a samll gap cut in it. Outputs of the two sections are subtracted in a differential amplifier. The weight is determined by the position of the cut. Assume that the gap position in the photomask fabrication step is determined by a digital computer, it will introduce a quantized step, d, in the gap position. This corresponds to quantizing the tap weight to a step size of d/w which is the tap weight error. Values of d/w ranged from 0.001 to 0.004 are considered in this study. For example, if the channel width is 5 mil, the quantization step ranged from 5×10^{-3} mil to 20×10^{-3} mil.

Sensitivity of CZT to this type of tap weight error has been studied before [1] and was reported to be:

$$\frac{N}{S} = \frac{1}{\sqrt{6}} \frac{d}{w}$$

In this study, random input signal was used in the computer simulation for a 512 point CZT. The computed N/S ratios were all within 1 db of the above expression derived analytically. It is expected that the N/S ratio would remain approximately constant and is weakly dependent of N. Although increasing N might cause larger errors at the output, the RMS signal is also increased and keep the N/S ratio approximately constant. This is confirmed by computer simulation which found N/S ratio within 2 db for N from 32 to 512. This is in contrast with the digital FFT case, whose roundoff error will be accumulated. Its N/S ratio will increase with N.

(d) Charge Transfer Inefficiency:

Charge transfer inefficiency is characterized by a fraction number ϵ which is the percentage of charge left behind at each transfer in a CCD. As a charge packet moves through the CCD, it will leave behind some charges but will also pick up some charges left in the well by previous charge packets. Therefore, there is a smearing effect which spreads one charge packet over several neighboring packets. Since the time varying output of the CZT transversal filter is uded to represent the frequency coefficients, the effect of charge transfer inefficiency is also to smear the frequency components into one another, leading to a degradation of frequency resolution for the CZT. This degradation is also expressed in terms of a N/S ratio. It was suggested that if ϵ is known, the filter tap weights can be modified to compensate for the loss of signal due to charge transfer inefficiency [8] by changing the tap weight h_k to h_k':

$$h_k' = \sum_{j=0}^{k=1} h_{k-j} \binom{k-1}{j} \epsilon^j (1-\epsilon)^{k-j} \qquad (4)$$

However, this compensation may be impractical because it may be quite difficult to accurately determine the value of ϵ. Therefore, the sensitivity of CZT to charge transfer inefficiency is also investigated in this study. A 64-point CZT is considered. Two types of input are used. One is the random number input. The other is the sinusoidal type of input. Two frequencies are used: $2f_f$ and $25f_f$ where $f_f = \frac{1}{T_w}$, T_w = sample window length in seconds. The N/S results in db are shown in Table I.

TABLE I

Sensitivity of CZT to Charge Transfer Inefficiency

ε	Noise to Signal Ratio (db)		
	Random	$f = 2f_f$	$f = 25f_f$
0.0001	-39.9	-39.8	-40.1
0.001	-20.5	-20.3	-20.6
0.01	-4.7	4.7	-4.6

It can be seen that the effect of charge transfer inefficiency on N/S ratio is approximately the same for these three types of input signals. In terms of frequency resolution degradation, however, the sensitivity is found to be three times worse for the higher frequency signal than for the lower frequency signal.

Since longer transforms require longer transversal filters, it is expected that increasing N would increase the N/S ratio due to charge transfer inefficiency. This is indeed found to be true. For a given ε, increasing N from 16 to 64 causes N/S ratio to increase by about 5 db for each doubling of N.

4. PRIME TRANSFORM

As is the case of CZT, the PT algorithm depends on the changing of the product nk in the exponent into a sum of three parts as shown in equation (2). However, the transform is implemented by using a transversal filter as a correlator instead of a convolver in the case of CZT. From equation (1),

$$X_o = \sum_{n=0}^{N-1} x_n \tag{5}$$

$$x_k = x_o + \sum_{n=1}^{N-1} x_n W^{nk}, \quad 0 < k \leq N-1 \tag{6}$$

Assume now that N is an odd prime number; there exists at least one number R, known as a primitive root of N, such that the following transformation maps the set $\{1,\cdots,N-1\}$ one to one onto itself:

$$n = R^k \text{ modulo } N, \quad k = 1,2,\cdots,N-1 \tag{7}$$

In other words, this transformation simply perform a specific permutation of the first N-1 positive integers. For convenience, a notation

$$((a)) = R^a \text{ modulo } N \tag{8}$$

is used. In this notation, two sequences are defined: {n} and {k} such that $n = ((n'))$ and $k = ((k'))$. Equation (7) can be rewritten as [7]:

$$X_{((k'))} = x_o + \sum_{n'=1}^{N-1} x_{((n'))} W^{((n'+k'))} \tag{9}$$

The summation in (9) is a circular correlation of a permuted sequence of input data with permuted values of the complex sinusoid W^n. It can be implemented by a transversal filter of 2N-3 taps. The left side is a permuted sequence of DFT coefficients to which the inverse permutation may now be applied to yield the desired DFT. A block diagram of the PT implementation is shown in Fig. 3. It should be noted that Fig. 3 is not the complete DFT. For example, the circuit for calculating X_o is not included. Comparing Fig. 1 and Fig. 3, the basic difference between the CZT and the PT is the replacement of analog multipliers by permuters. Such a permuter has been built as an analog random access memory (ARAM) by Reticon Corp.

In Fig. 3, three sources of inaccuracies and errors for PT should be considered:

- Permuter errors
- Transversal filter tap weight errors
- Charge transfer inefficiency

(a) Permuter Memory Error:

Recent experimental work [10] on the first generation ARAM indicated that the percentage error in this device is as high as ±9%. The sensitivity of a 31-point PT to this type of error is studied for permuter error ranged from ±1% to ±10%. Random signal input is used. The results of N/S

ratio is shown in Fig. 4.

(b) Transversal Filter Tap Weight Error:

It was found that the effect of tap weight error on PT is the same as that on CZT. Therefore, the same results for CZT apply also to Prime Transform.

(c) Charge Transfer Inefficiency:

In PT, the DFT coefficient output does not follow the order of frequency as in the case of CZT. Therefore, one may no longer characterize the effect of charge transfer inefficiency as a degradation in frequency resolution. In this study, the N/S ratio is again used as the measure of the sensitivity performance. Similar to the CZT case, both random input and sinusoidal inputs of two frequencies $f = 2f_f$ and $25f_f$ are used. However, a 67 point PT is analyzed and compared with the results of a 64 point CZT. "67" is chosen because it is a prime number. The N/S ratio is shown in Table II.

TABLE II

Sensitivity of PT to Charge Transfer Inefficiency

ε	Noise to Signal Ratio (db)		
	Random	$f = 2f_f$	$f = 25f_f$
0.0001	-40.2	-42	-38.8
0.001	-20.7	-22.4	-19.3
0.01	-4.5	-6	-3.7

This table should be compared with the sensitivity of CZT in Table I. For random inputs, there is very little difference between CZT and PT.

However, for sinusoidal inputs, the PT is found to be more sensitive to charge transfer inefficiency at low frequency than at high frequency. Between $f = 2f_f$ and $25f_f$, the N/S ratios differ approximately by 3 db. The reason for this difference has not be identified.

As for the dependence on N, it was found that the sensitivity of PT to charge transfer inefficiency increases with N just as it did for CZT. When N is doubled, the N/S ratio increases by 5 to 6 db.

5. COMPARISON OF CZT AND PT

The PT has some advantages over the CZT in hardware implementation. For real input signals, the PT requires only two transversal filters with real coefficients while the CZT requires four. However, the new ARAM permuter imposes severe limitations currently on the usefulness of PT. Its $\pm 9\%$ inaccuracy can no doubt be improved. But it is too early to forecast by how much. Furthermore, the permuter must be large enough to hold the entire input data sequence. Longer transform requires larger permuters operating at higher speeds. It is not yet clear whether there will be fundamental limits to the size, speed and accuracy. However, to compare the sensitivities of CZT and PT, it is assumed that inaccuracies of future permuters can be reduced below $\pm 2\%$. It is also assumed that the tap weights have been properly compensated to eliminate the effect of charge transfer inefficiency. A 64 point CZT and a 67 point PT are compared. Random input signal is used in the computer simulation study. N/S ratio of four cases are shown in Table III.

TABLE III

Comparison of N/S Ratio of CZT and PT

Analog Inaccuracy	Tap Weight	CZT	PT
$\pm 0.1\%$	0.001	-57.5 db	-60.8
$\pm 0.5\%$	0.002	-44.4	-47.9
$\pm 1\%$	0.003	-39.2	-42.3
$\pm 2\%$	0.004	-33.3	-35.7

Analog inaccuracy refers to the analog multiplier and differential amplifier errors in the CZT case and the permuter error in the PT case. It can be seen that the PT is approximately 3 db better than the CZT. However, the errors used in the calculation of the PT sensitivity are estimated values which have not been accomplished yet in today's

devices.

In an analytical study. Wrench [2] found that the prime transform is an order of magnitude more accurate than the chirp Z transform based on his sensitivity analysis for errors of ±1% for all analog components including the tap weights. For the same errors, the computer simulation of this study showed that the PT is only approximately 3 db better. However, it should be noted that there are several differences between these two studies. Wrench considered a straight percentage tap weight error instead of the error caused by quantization of the split gate mask. To present the sensitivity results, Wrench defined a relative accuracy of the output transform coefficient as the ratio of the 95% confidence bound on the error divided by the maximum coefficient output. It is different from the N/S ratio used in this study.

6. COMPARISON WITH DIGITAL FFT

Using the noise to signal ratio versus the transform length plots, sampled analog chirp Z transform and prime transform are compared with digital fast Fourier transform. In Fig. 5, two CZT cases are shown together with two digital FFT cases:

- Sampled analog CZT:
  It is assumed that the effect of charge transform inefficiency can be compensated by changing the tap weights.
  - Case A: Analog error ±0.5%, tap error due to quantization step 0.2% (second case of Table III)
  - Case B: Multiplier error due to quantization of multiplying chirps to 7 bits plus sign and tap error due to quantization step of 0.2% [1].
- Digital FFT: [3]
  - Case C: Upper bound of 13 bits fixed point
  - Case D: Upper bound of 15 bits fixed point

Prime transform results are not included in this figure because it was shown in the last section that the N/S ratio of PT is approximately 3 db better than that of CZT.

It can be seen that the N/S ratio of sampled analog CZT (and PT) are approximately constant with respect to N. For digital FFT, the major source of error is the round off or truncation error which will accumulate with N. In the case of fixed point FFT, it increases with the square root of N [3] and is shown in Fig. 5. In the case of floating point FFT, it increases with the square root of $\log_2 N$ [4] but is not included in the Figure.

As is commonly known, digital processing has higher accuracy (or lower N/S ratio) than analog processing as shown in Fig. 5. The N/S ratio of CZT is better than that of a 13 bit digital FFT only for very long transform length. But by increasing the number of digital bits, digital FFT can be made to have better N/S ratio than the best CZT. In this case, for an increase of 2 bits to 15 bits, the N/S ratio is reduced by 12 db.

However, sampled analog transforms have their potential advantages [1]: faster processing speed lower power consumption, lower parts count and lower cost. When fully realized and developed, these advantages will be traded off with the somewhat lower performance as analyzed in this study.

ACKNOWLEDGEMENT

The authors wish to thank Professors S. R. Parker and R. Fossum for their support and encouragement, Messrs. J. Alsup, J. Speiser, E. Wrench and H. Whitehouse for illuminating technical discussions and Ms. D. Martinez for the preparation of the manuscript.

REFERENCES

1. Buss, D.D. and others, "Comparison Between the CCD, CZT and the Digital FFT," Proc. CCD Applications Conf., San Diego, pp. 267-280, Oct 1975.
2. Wrench, E.H., "The Effects of Finite Computational Accuracies on Several Cosine Transform Architectures," SPIE Conference Proceedings, Aug 1976.
3. Welch, P.D., "A Fixed-Point Fast Fourier Transform Error Analysis," IEEE Trans. Audio Electroacoust., Vol. AU-17, pp. 151-157, Jun 1969.
4. Weinstein, C.J., "Roundoff Noise in Floating Point Fast Fourier Transform Computation," IEEE Trans. Audio Electroacoust., Vol. AU-17, pp. 209-215, Sep 1969.
5. Bluestein, L.I., "A Linear Filtering Approach to the Computation of the Discrete Fourier Transform," 1968 Northeast Electronics Research and Engineering Meeting Record, 10, pp. 218-219, Nov 1968.
6. Whitehouse, H.J., R. Means and J. Speiser, Proceedings 1975 IEEE International Symposium on Circuits and Systems, Boston, April 1975.
7. Gersho, A., "Charge-Coupled Devices: The Analog Shift Register Comes of Age," Communications Society, Vol. 13, No. 6, pp. 27-32, Nov 1975.
8. Buss, D.D. and Bailey, W.H., "Application of Charge Transfer Devices to Communication," Proc. CCD Applications Conf., San Diego, pp. 83-93, Sep 1973.
9. Rader, C.M. "Discrete Fourier Transforms When the Number of Data Samples is Prime," Proc. IEEE, V. 56, No. 6, pp. 1107-1108, Jun 1968.
10. Kerber, G. L., "Analog Implementation of the Prime Cosine Transform," Proc. Government Microcircuit Applications Conf.

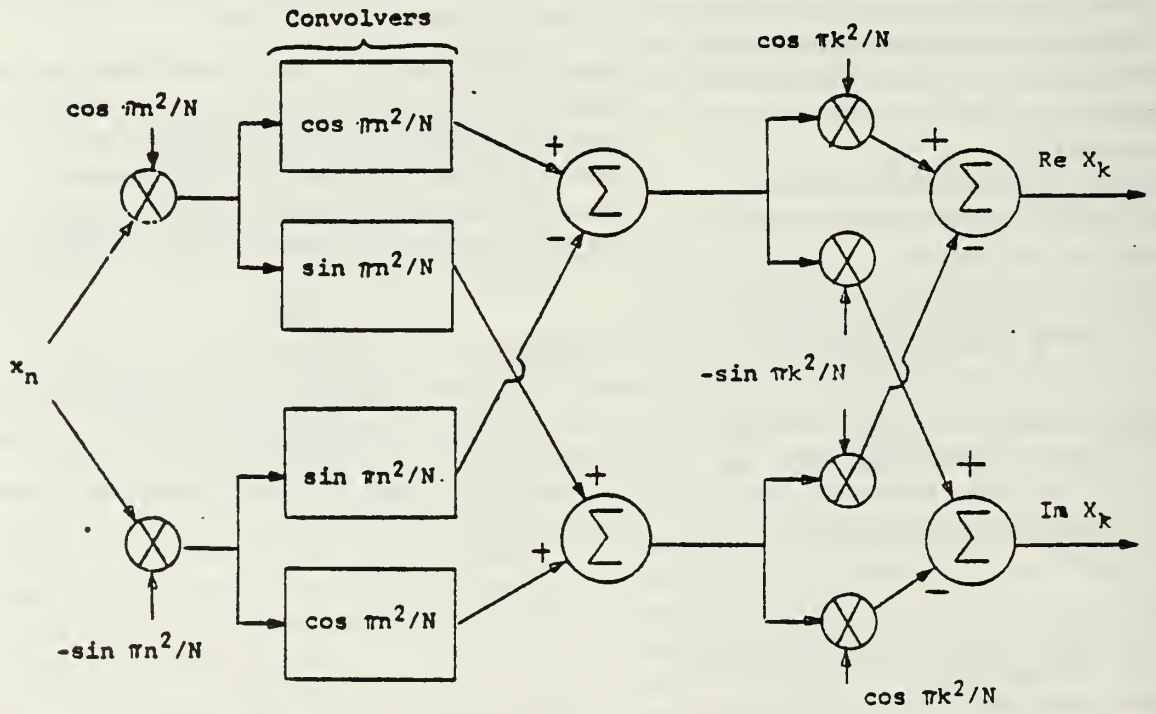


Figure 1. CZT Implementation

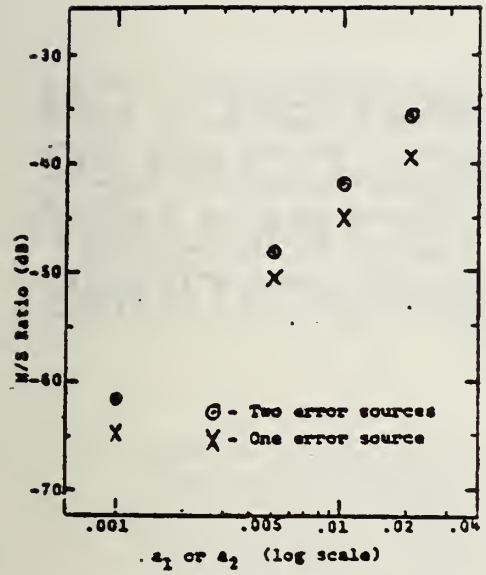


Figure 2. Multiplier/Amplifier Errors

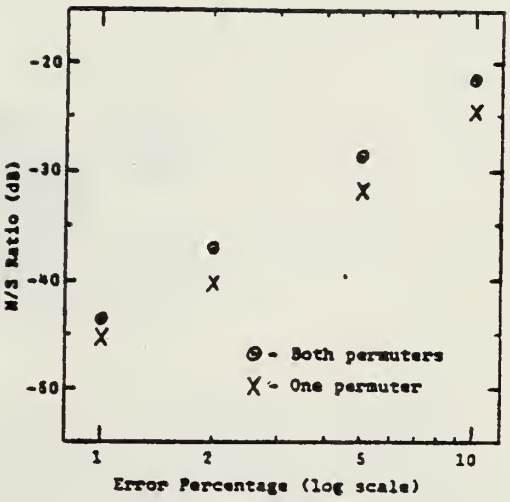


Figure 4. Permuter Errors

Figure 3. Prime Transform Block Diagram

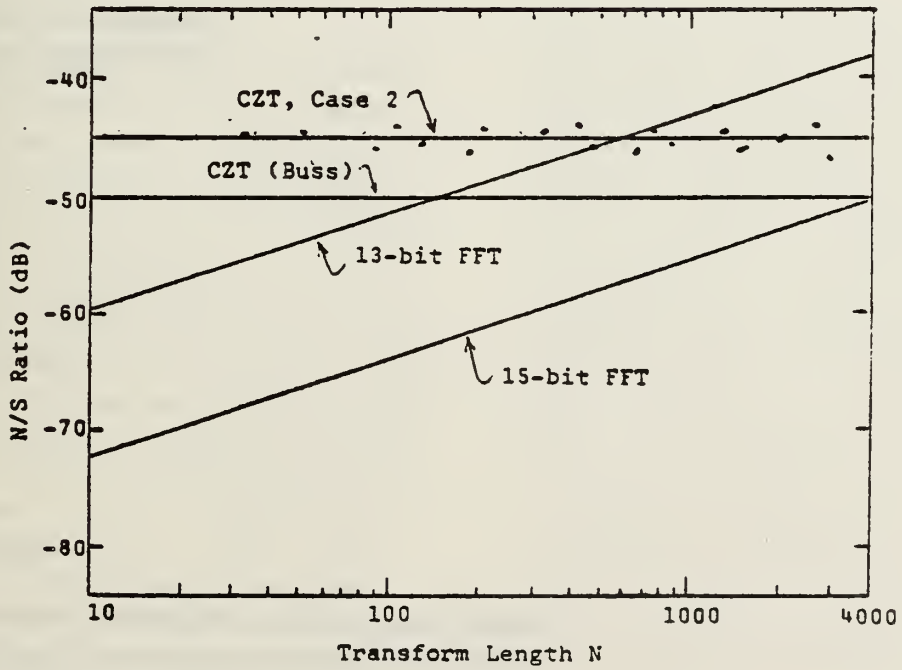


Figure 5. Comparison of CZT with FFT

PAPER NO. 8

CONFERENCE ON CHARGE-COUPLED DEVICE TECHNOLOGY AND APPLICATIONS

Nov. 30 - Dec. 2, 1976

Sheraton Park Hotel
Washington, D. C.

Sponsored by

National Aeronautics and Space Administration
Office of Aeronautics and Space Technology and
Jet Propulsion Laboratory
California Institute of Technology

THEORY AND APPLICATIONS OF SAMPLED ANALOG RECURSIVE CTD COMB FILTERS

T. F. Tao, S. Holmes, A. Ejaz,
F. Piazza, L. Saetre, and B. Freund
Naval Postgraduate School
Monterey, California 93940

Sampled analog comb filter using a recursive filter implementation is studied. Charge transfer device delay lines are used as the delay elements. The similarities and differences between sampled analog recursive filter and digital recursive filter are pointed out. Both canceller type and integrator type comb filter using the first order or the second order canonical circuits are studied. Experimental results of frequency response are presented to show their difference from that of digital filters. A theoretical analysis is presented based essentially on the digital filter theory but modified to take into account the differences described above. The usefulness of this theory in analyzing the comb filter characteristics is discussed. Several applications of both the canceller type comb filter and the integrator type comb filter are demonstrated.

1. INTRODUCTION

Several sampled analog recursive filter studies using charge transfer devices (CTD) as delay elements have been reported. BBD recursive filters were first studied.[1-2] Later, a three poles[3] and a two poles/one zero[4] CTD recursive filters and a one pole CTD recursive integrator[5] were reported. We presented the results of a different type of recursive filter which emphasized the comb feature of its frequency characteristics.[6] A first order comb filter was studied using one CTD delay line of eight delay stages. A modified digital recursive filter theory was used to account for the major features of the sampled analog comb filter. However, several deviations between experimental results and theoretical calculations were reported. This paper will present more recent results of our continuing study of the recursive comb filters.

Second order comb filters have been more extensively studied. The theory was improved and can analyze some of the characteristics unaccounted for by the earlier analysis. In section 2, the principle and general characteristics of comb filters are presented. The difference between sampled analog recursive filters and digital recursive filters is pointed out. In section 3, experimental results are presented to highlight the major features of the frequency characteristics. In section 4, a theory is presented which is able to analyze most of the features. It is based essentially on the digital recursive filter theory but modified to take into account the differences pointed out in section 2. In section 5, several applications of both the canceller type comb filter and the integrator type comb filter are demonstrated.

2. COMB FILTER AND SAMPLED ANALOG RECURSIVE IMPLEMENTATION

Comb filters are characterized by their periodic transfer characteristics in the frequency domain. They can be classified into two general types. The first one is the canceller or bandstop type shown in Figure 1(a). It has large attenuation in a narrow neighborhood of a series of periodically separated frequencies and good transmission in between. The second one is the integrator or bandpass type shown in Figure 1(b). It has good transmission in a narrow neighborhood of a series of periodically separated frequencies and strong attenuation in between. In order for the comb teeth to separate from each other, it is desirable to have zero transmission at certain frequencies between neighboring teeth.

Both analog comb filters and digital comb filters have been developed before using several different implementation techniques such as the feedforward circuits, frequency sampling, fast Fourier transform, recursive filter circuits, etc. All use delay devices in some manner. In the analog case, quartz delay lines are the major candidates. In the digital case, implementation has been mainly carried out by software programs on computers. However, using the family of charge transfer device delay lines, sampled analog comb filters are being developed using several implementation techniques: feedforward circuits,[7] chirp Z transform[8] and recursive filter implementations.[3,4,6,9]

The recursive filter approach is used in our study. The circuit configuration is shown in Figure 2 for the special case of second order filters. It is similar to the canonical circuit used in the digital recursive filters but with four basic differences.

(1) Delay elements of multiple delay stages:

The delay element used in the sampled analog recursive filter circuit is implemented by CTD delay lines which usually consists of N delay stages. In digital filters, the delay is accomplished either by memory or by shift registers and consists of one delay stage. The presence of N delay stages in one delay line is instrumental in providing the comb feature of its frequency characteristics. The reason for this important property can be explained qualitatively as follows. Since the signal is delayed by N clock periods before it is processed, the frequency of recursive filter operation (f_r) is not the sampling frequency (f_s) but is only one Nth of f_s. Therefore, frequency characteristics of this type of filter are not only periodic with respect to f_s but also periodic with respect to f_r. Since $f_r = f_s/N$, there are $N/2$ comb teeth within the Nyequist frequency range from 0 to $f_s/2$. This is in contrast to the situation for digital recursive filters where the recursive filter operations take place after each delay stage, resulting in only one comb tooth within the Nyquist range.

(2) Special relation among the feedforward coefficients a_0, a_1 and a_2:

A good comb filter requires that two neighboring comb teeth should be clearly separated from each other in frequency domain. In other words, the transfer function should be zero at a series of periodically separated frequencies starting either at f_r or $f_r/2$ and repeating at its multiples. These zeroes provide ideal separation between two neighboring teeth. Since there are now N/2 periods within the Nyquist frequency range for this type of filters, a proper design of the "zero" locations of the singularities of H(Z) can meet that requirement. This is accomplished by designing the values of the feedforward coefficients a_0, a_1 and a_2 according to the ratio of 1:±2:1. Consequently, a series of "zeroes" of the transfer function will occur at frequencies periodically separated by the period of f_s/N and started either at f_r or $f_r/2$. In digital filters, the delay element has only one delay stage. The frequency characteristics will also become zero at a series of periodically separated frequencies. But they are separated by the period f_s and started either at f_s or $f_s/2$ which is beyond the useful signal frequency range. Therefore, in digital filters, the requirement of keeping a_0, a_1 and a_2 at the ratio of 1:±2:1 never did take

place.

(3) Frequency dependence of filter coefficients:

In sampled analog comb filters, the filter coefficients are implemented by potentiometers. The summation operations are implemented by operational amplifiers with feedback. Both components have finite bandwidth and introduce frequency dependence. The signal is sampled and delayed by the CTD device which also introduces other frequency dependences due to two effects:

{ sampled and hold effect
{ charge transfer inefficiency effect.

Consequently, it is difficult to maintain a set of properly designed values of filter coefficients throughout the frequency range from 0 to $f_s/2$. In digital filters, the coefficients are set in memory and do not vary with frequency.

(4) Mixed processing of sampled and unsampled signals:

The input signal in this type of filter is usually analog. It is sampled by the CTD device. When the delayed signal is fedback and processed at the input summer, a mixing of the continuous and analog input signal with the sampled analog feedback signal takes place. Because a sampled signal has a $\sin x/x$ type of frequency dependence while an unsampled signal does not, the same processing operation can not be maintained throughout the Nyquist frequency range.

Because of these four differences, the frequency characteristics of the sampled analog comb filter are substantially different from that of the digital recursive filter even when their circuit configurations are the same canonical type. This will be shown by some experimental results in the next section.

3. EXPERIMENTAL RESULTS

Two types of CTD delay lines are used in the experimental studies:

{ 8 bits CCD delay line
{ 96 bits Reticon SAD-100 delay line

In Fig. 3, the frequency characteristics of a second order canceller type comb filter are shown. A theoretical calculation is included in the same figure and will be explained in the next section. The filter information is:

{ 8 bits CCD delay line
{ f_s = 20 kHz
{ Filter coefficients: $a_o=1$, $a_1=2$, $a_2=1$
$b_1=-0.3$, $b_2=+0.3$

The major features of a comb filter are clearly shown:

- Frequency characterists is periodic with respect to a frequency of 2.5 kHz which is equal to f_s/N.
- There are four comb teeth in the Nyquist frequency range from 0 to 10 kHz.

However, it is different from the ideal comb filter characteristics in three ways:

- Two neighboring teeth are not separated by infinite attenuation as in an ideal comb filter. The attenuation is more than -50 db at the first null frequency, but deteriorated to -42 db at the second null frequency and to -34 db and so on.
- The peak amplitude of each comb tooth does not remain constant as in an ideal comb filter. For this filter, it first decreases and then rises at the fourth tooth.
- The shape of the comb tooth changes with frequency.

These major features of sampled analog recursive comb filters are further demonstrated by the following frequency characteristics of comb filters using 96 bits Reticon delay line. The filter information is given as follows:

Fig. 4. { Canceller type comb filter
{ f_s = 400 kHz
{ Filter coefficients: $a_o=1$, $a_1=-2$, $a_2=1$
$b_1=-0.973$,
$b_2=+0.334$

Fig. 5. { Integrator type comb filter
{ f_s = 390 kHz
{ Filter coefficients: $a_o=1$, $a_1=2$, $a_2=1$
$b_1=-0.2$, $b_2=-0.6$

The periodic feature of the comb filter again is clearly shown. However, the variations among the first twelve to thirteen teeth are much smaller than among the first four teeth shown in Fig. 3. This difference can be qualitatively explained as follows. In the previous section, it was explained that in sampled analog recursive filters, several factors give rise to frequency dependences of the processing operations. They become more pronounced as the frequency is approaching the Nyquist frequency. In Fig. 3, an 8 bits CCD was used. There are only four comb teeth within the Nyquist range. Therefore, noticeable changes take place from the first tooth to the fourth tooth. On the other hand, a 96 bits recursive comb filter was studied in Fig. 4 and 5. There are 48 comb teeth within the Nyquist range. The first twelve to thirteen comb teeth cover only one-fourth of the Nyquist range. Their frequency dependence should be less pronounced. However, if the frequency range is extended to cover all 96 comb teeth, as shown in Fig. 6, strong frequency dependence is clearly seen. The filter information is:

96 bits Reticon SAD 100 delay line
f_s = 400 kHz
Filter coefficients: a_o=1, a_1=2, a_2=1
b_1=+0.2, b_2=+0.2

It is an integration type comb filter. In this case, all three types of deviations from an ideal comb filter can be detected. As the frequency or the order of comb teeth is increased, the attenuation between two neighboring teeth deteriorate, the peak amplitude decreases and the tooth shape changes. These behaviors do not exist in digital recursive filters. Therefore, digital recursive filter theory should be properly modified to account for these features.

4. THEORY

Since the circuit configuration of the sampled analog recursive comb filter is the same as the canonical circuit of digital recursive filter except for four differences pointed out in section 2, it is reasonable to expect that the digital recursive filter theory should apply to the sampled analog case if proper modifications are made.

The transfer function of a second order digital recursive filter is

$$H(Z) = \frac{a_o + a_1 Z^{-1} + a_2 Z^{-2}}{1 + b_1 Z^{-1} + b_2 Z^{-2}}$$

The modifications needed to account for the four differences are discussed as follows:

(1) To take into consideration the presence of N delay stages, Z^{-1} is changed to Z^{-N}.

(2) To provide zero between two neighboring teeth: $a_o = 1$, $a_1 = \pm 2$, $a_2 = 1$

- The "-" sign of a_1 is for the canceller type comb filter.
- The "+" sign of a_1 is for the integrator type comb filter.

(3) Filter coefficients are no longer constant and vary with frequency. The exact frequency dependence could be complicated because there are several factors contributing to the frequency dependence simultaneously.

(4) The effect due to the mixing of unsampled input signal and sampled feedback signal at the first summer is more difficult to model by a frequency dependence. Instead, a change of the circuit configuration should be considered to avoid this problem.

Combing these modifications, the transfer function for a sampled analog recursive comb filter is

$$H(Z) = \frac{a_o(f) + a_1(f) Z^{-N} + a_2(f) Z^{-2N}}{1 + b_1(f) Z^{-N} + b_2(f) Z^{-2N}}$$

where $a_o(f) = 1$, $a_1(f) = \pm 2$, $a_2(f) = 1$
for $f << f_s/2$

The agreement between this theoretical analysis and measured result has been investigated. For the 8 bits CCD recursive comb filter shown in Fig. 3, theoretical calculations have been made by considering two factors contributing to the frequency dependence.

- Sample and hold effect.
- Charge transfer inefficiency effect.

The calculated result is also plotted in Fig. 3. It can be seen that all three trends of deviation have been accounted for qualitatively:

- Deterioration of attenuation between neighboring comb teeth.
- Change of peak amplitude of each comb tooth.
- Change of shape of comb tooth.

However, quantitative agreement has not been accomplished. The same situation is found for the 96 bits CTD recursive comb filters shown in Fig. 4 through 6. A theoretical calculation considering only the effect of sampled and hold operation is shown in Fig. 7 for the filter of Fig. 6. Again, the major features of this comb filter have been accounted for only qualitatively.

It can be concluded that the basic consideration of the theoretical analysis is complete. Flaws in both calculations and measurements have contributed to the lack of perfect agreement. In the theoretical calculation, the frequency dependences of filter coefficients were only approximations. In the experinental results, measurement can be improved in high frequency range close to the Nyquist frequency. It is believed also that the circuit configuration can be improved, such as by the use of a sampled and hold circuit in front of the comb filter, to provide frequency characteristics closer to that of an ideal comb filter. However, major features of a comb filter have been accomplished such that meaningful investigation of their applications can be carried out. They will be presented in the next section.

5. APPLICATIONS

The periodic nature of the comb filters makes them well suited to process signals which are periodic in time and in frequency as in radar and sonar systems and many others. The canceller type comb filters can be used to reject periodic signals by aligning the null frequencies of the comb filter with the periodic frequency spectrum of the undesirable signal. Two applications have been investigated. The first is the cancellation of clutter spectrum in an MTI radar. The second is the cancellation of 60 Hz noise and its harmonics.

The integrator type comb filters can be used to enhance a periodic signal. Two applications have been investigated. The first is the separation of one periodic signal from another periodic signal of a different pulse repetition frequency, an application of interest to threat analysis. The second is the enhancement of signal to noise ratio of a periodic signal contaminated by noise. Some of their preliminary results will be presented.

(1) Canceller application:

In MTI radar, movilng targets can be distinguished from clutters or stationary targets by using the Doppler shifts between their spectra in frequency domain. The clutter is characterized by a periodic spectra at d.c. and multiples of PRF. The moving target is characterized by another spectrum of the same period but shifted by its Doppler frequency. The clutter spectrum can be eliminated by a canceller type comb filter by aligning the null frequencies of the filter with the PRF frequencies of the radar return signal. Therefore, f_s/N = PRF. In addition to this canceller requirement, there are two other requirements in order for the comb filter to pass the radar pulse signal without too much distortion:

- Bandwidth requirement = BW $\geq \frac{1}{2\tau}$
- Sampling frequency requirement: $f_s \geq \frac{1}{\tau}$

where τ = radar pulse width.

These considerations are reported in more details recently.[10] Because of the limited length of the CTD device used in this study, the sampling frequency cannot be raised high enough to provide adequate bandwidth for processing real radar signal. Simulated MTI radar signals are used to demonstrate the application of canceller type comb filters as shown in Fig. 8 and 9. A stationary target is simulated by a pulse. A moving target is simulated by a train of pulses whose amplitudes

are modulated by the Doppler frequency, producing the so-called "butterfly" or "bipolar" signal shown in Fig. 8. The filter information is:

Canceller type comb filter
96 bits Reticon SAD 10 delay line
f_s = 400 kHz
Filter coefficients: $a_0=1$, $a_1=-2$, $a_2=1$
$b_1=-0.334$, $b_2=+0.334$

It can be seen that the stationary target pulse is eliminated at the output while the moving target "butterfly" signal passed the filter with small amount of attenuation and distortion. However, some glitches can be seen suggesting that high frequency components have not been cancelled. This probably is the consequence of deteriorated attenuations of the comb filter at high frequency range as described in section 2. This effect is tested in Fig. 9 which presents the same experimental situation as in Fig. 8 except that the simulated stationary target pulse is much narrower and has more high frequency components. It can be seen that the cancellation is not as complete as in Fig. 8.

(2) Integrator application:

An integrator type comb filter will only pass a periodic signal whose PRF matches the recursion frequency of the filter, i.e., f_s/N = PRF. Other noises and/or undesirable signals will be suppressed and accomplish an overall improvement of signal to noise ratio or signal to interference ratio. Fig. 10 and 11 demonstrate this ability. The filter information is the same as in Fig. 5. In Fig. 10, a 0.3 volt peak to peak pulse train is contaminated by random noise of 0.4 rms volt. In Fig. 11, another pulse train of 0.1 volt peak to peak is contaminated by random noise of 0.2 rms volt. The improvement of S/N ratio is clearly seen.

6. CONCLUSION

The basic concept, circuit configuration and theory of a sampled analog recursive comb filter are presented. Two CTD delay lines have been used in experimental studies. The measured results agree fairly well with the calculation suggesting that the basic understanding and a usable theoretical analysis have been obtained. Feasibilities of several applications were demonstrated. However, other performance criteria important to practical applications such as the dynamic range, bandwidth limit, sensitivity, stability, noise properties, etc. have not been investigated. They should be investigated if practical applications in real systems are to be considered.

ACKNOWLEDGEMENT

The authors wish to thank Professors S. R. Parker, R. Fossum for encouragement and support, Mr. P. Kopp for laboratory assistance and Ms. A. Schow for preparing the manuscript.

* Supported by Naval Electronic Systems Command and Naval Postgraduate School Research Foundation.

REFERENCES

1. D.A. Smith, C.M. Puckett and W.J. Butler, IEEE J. Solid State Circuits, SC-7, p. 421 (1972).
2. D.A. Smith, W.J. Butler and C.M. Puckett, IEEE Trans. Communications, COM-22, p. 921 (1974).
3. J. Bounden, R. Eames and J. Roberts, Proceedings 1974 International Conference on Technology and Applications of CCD, Edinburgh, Scotland, September 1974, p. 206.
4. J. Mattern and D. Lampe, ISSCC Technical Paper Digest, Philadelphia, February 1975, p. 148.
5. J. Roberts, M. Chesswas and R. Eames, Electronic Letters, 10, p. 169 (1974).
6. T.F. Tao, V. Iamsaad, S. Holmes, B. Freund, L. Saetre and T. Zimmermann, Proceedings, 1975 International Conference on CCD Applications, October 1975, San Diego, California, p. 257.
7. W. Butler, W. Engeler, H. Goldberg, C. Puckett and H. Lobenstein, IEEE International Symposium on Circuits and Systems, Boston, April 1975, IEEE Publication Advanced Solid State Components for Signal Processing, p. 30.
8. W. Bailey, R. Broderson, W. Eversole, L. Hite, M. Whatley, D. Buss and R. Sproat, Digest of Papers, 1974 Government Microcircuits Applications Conference, 5, p. 76 (1974).
9. M. Tompsett, A. Moshen, D. Sealer and C. Sequin, IEEE International Symposium on Circuits and Systems, Boston, April 1975, IEEE Publication Advanced Solid State Components for Signal Processing, p. 83.
10. T. F. Tao, L. Saetre, S. Holmes and A. Ejaz, Digest of Papers, 1976 Government Microcircuits Applications Conference, Orlando, Florida, November 9-11, 1976.

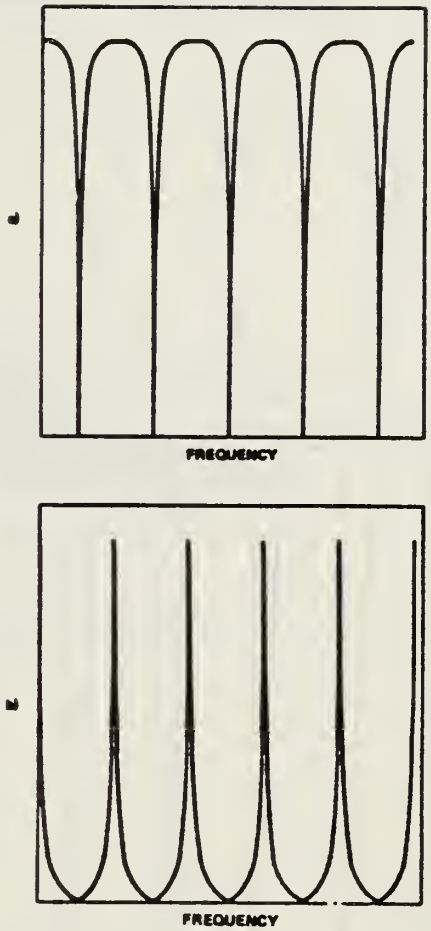


Fig. 1 Frequency Characteristics of (a) Canceller Type Comb Filter (b) Integrator Type Comb Filter

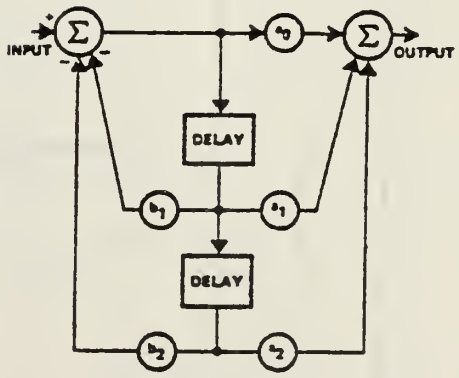


Fig. 2 Configuration of Second Order Canonical Circuit for Recursive Filter

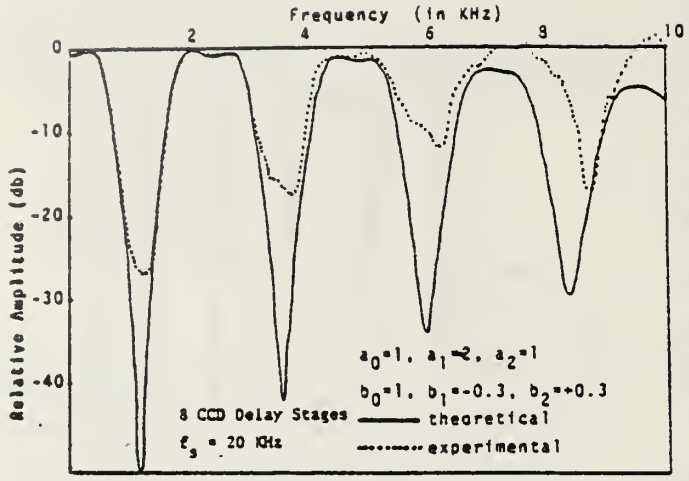


Fig. 3 Measured Frequency Characteristics of Canceller Type Comb Filter Using an 8 bits CCD Delay Line

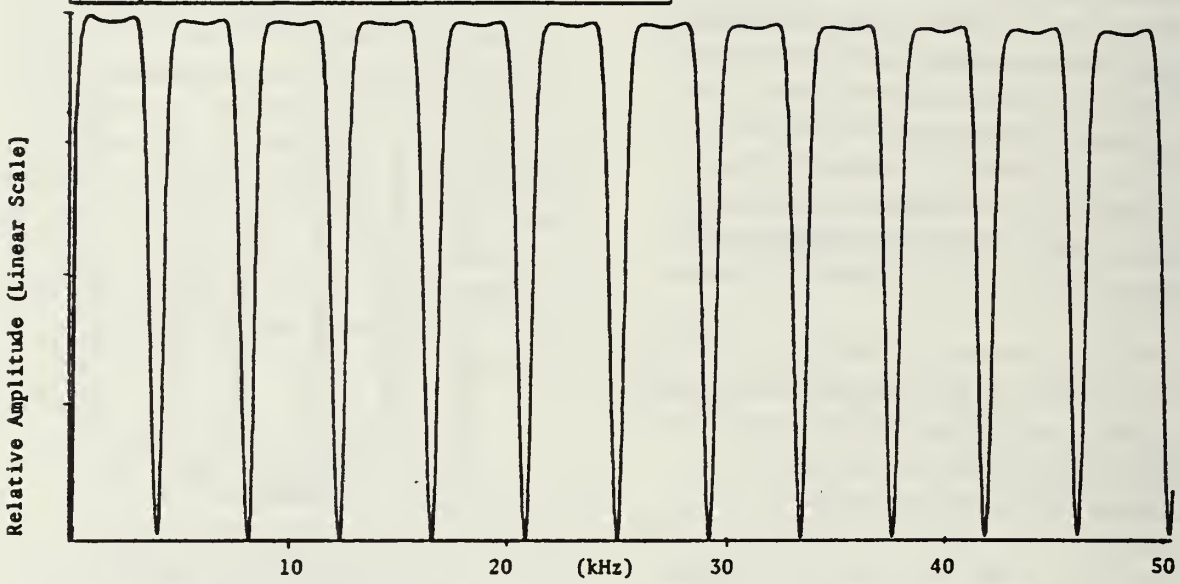


Fig. 4 Measured Frequency Characteristics of Canceller Type Comb Filter Using a 96 bits CTD Delay Line

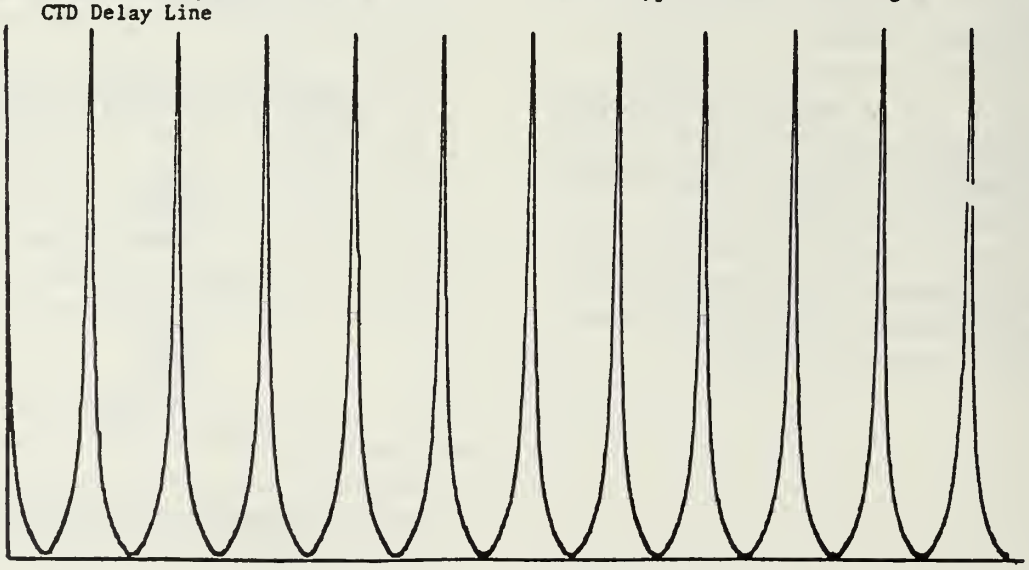


Fig. 5 Measured Frequency Characteristics of Integrator Type Comb Filter Using a 96 bits CTD Delay Line.

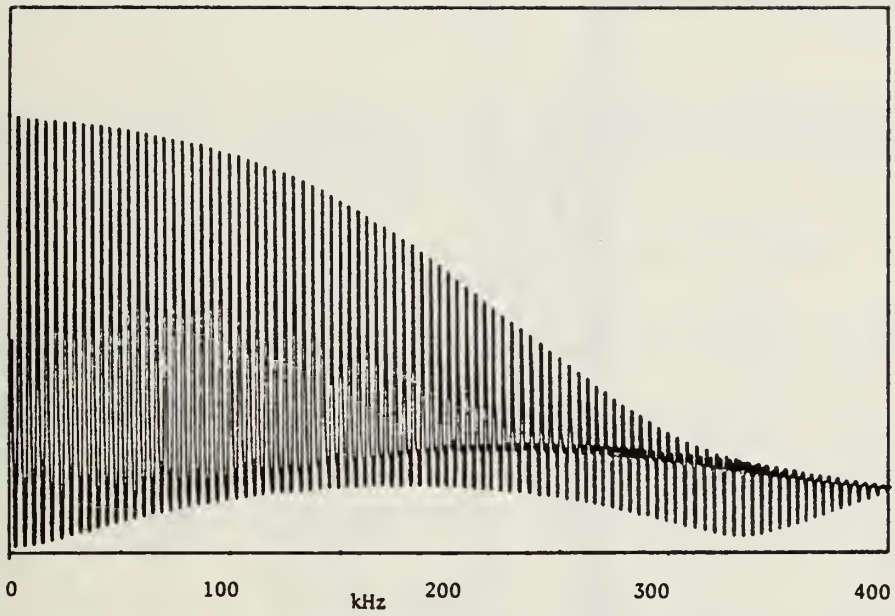


Fig. 6 Measured Frequency Characteristic of An Integrator Type Comb Filter Using 96 Bits CTD Delay Line

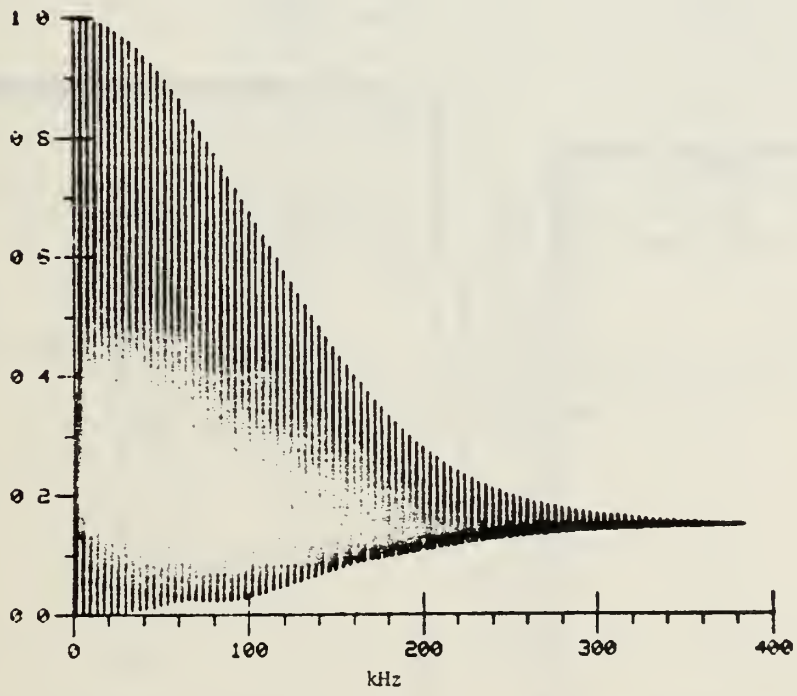


Fig. 7 Calculated Frequency Characteristic of An Integrator Type Comb Filter Using 96 Bits CTD Delay Line.

Fig. 8 Demonstration of Using Canceller Type Comb Filter for MTI Radar Applications

X scale: 10 μsec/division

Fig. 9 Demonstration of Using Canceller Type Comb Filter for MTI Radar Applications

X scale: 1 μsec/division

Fig. 10 Demonstration of Using Integrator Type Comb Filter for Enhancing Periodic Signal Contaminated by Noise

Signal = 0.3 volt P-P, Noise = 0.4 volt RMS

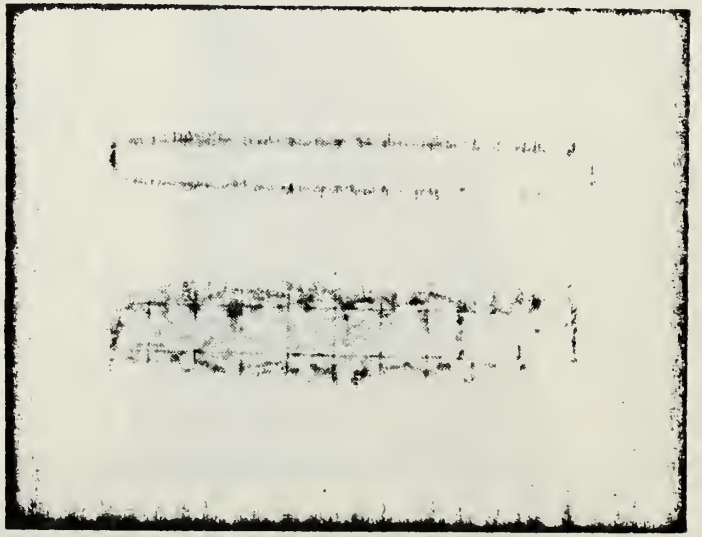


Fig. 11 Demonstration of Using Integrator Type Comb Filter for Enhancing Periodic Signal Contaminated by Noise

Signal = 0.1 volt P-P, Noise = 0.2 volt RMS

Title: Derivation of Critical Planning Properties of The Process Associated With Generating Demands For Health Care

Investigator: D. Whipple, Associate Professor of Administrative Sciences

Sponsor: Foundation Research Program (6.2)

Objectives: To develop analytically the major parameters involved in the generation of demands for health care resources given the significant agency relationship of the physician for the patient. These will be used to analyse the budgetary process used in the Military Health Services System (MHSS) under its present structure and to recommend changes where applicable.

Summary: An economic mathematical programming optimization model previously developed by the investigator will be modified and extended to develop testable hypotheses concerning the probable outcome of various parametric changes in the structure of the MHSS in terms of cost by the physician-agent and the patient. These outcomes will be analysed in light of present or suggested organizational change initiatives in the MHSS, and recommendations for empirical policy research mode.

Publications: Whipple, D., "The Relation of Demand Generation to Incentives in Prepaid Health Care Settings," (going to press 1 May 1978).

Title: Convexity in an Inventory Model

Investigator: P. W. Zehna, Professor of Operations Research

Sponsor: Foundation Research Program (6.2)

Objective: The Hadley-Whitin Q-r models in inventory theory have been widely used by practioners for some time now in spite of questions regarding their validity. The purpose of this research was to extend the author's investigations in these matters to hopefully characterize the proper solutions.

Summary: Beginning with some recently completed research in the backorders case, the author was able to successfully complete the research and give a full characterization of the solutions. Continuing with the lost sale case, one that had not been previously studied, the investigator was able to show that, while the situation was not as bad in this case, a similar anomaly existed and again, it was possible to completely characterize the solution in this new case also.

Publications: P. W. Zehna, "Solutions in Hadley-Whitin Q-r Models" submitted to the Naval Logistics Quarterly.

Lenna, Peter W., "Solutions in Hadley-Whitin Q-r Models:, Naval Postgraduate School Technical Report NPS 55-78-4, January, 1978.

NPS55-78-4

NAVAL POSTGRADUATE SCHOOL

Monterey, California

SOLUTIONS IN HADLEY-WHITIN Q-r MODELS

by

Peter W. Zehna

January 1978

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

Peter W. Zehna, Professor
Department of Operations Research

Reviewed by:

Michael G. Sovereign, Chairman
Department of Operations Research

Released by:

William M. Tolles
Acting Dean of Research

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-78-4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Solutions in Hadley-Whitin Q-r Models		5. TYPE OF REPORT & PERIOD COVERED Technical Report 1 Jul 77-30 Sep 77
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Peter W. Zehna		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, Ca. 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61152N, RR 000-01-01 N0001477WR70044
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE January 1978
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*)		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Inventory; convexity, Q-r models, optimal costs

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The Hadley-Whitin approximate Q-r models have been in existence for nearly fifteen years now and are rather widely used by practitioners because of their simplicity and rather general applicability. There are, however, some questions regarding the characterization of solutions in these models due to the fact that the objective function is not always convex as asserted by the authors. This paper gives a complete characterization of the solutions in both the backorders model and the lost sales case for normal leadtime

OVER

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. Continued
demand. It brings out the fact that the solution always exists in both cases but may in fact be different from the published one. A simple scheme is provided for isolating the solution in terms of the given valu of the parameters for any particular application.

1. INTRODUCTION

In what has become a standard reference in the study of inventory models, Hadley and Whitin [1] devote a chapter to lot size-reorder point, or Q-r models as they have come to be known. Of special interest to NavSup and others are the approximate models treated in Section 4-2 of the book and commonly referred to as the backorders and the lost sales cases. A typical three-component cost function comprised of set-up, holding and penalty costs is developed and depends directly on the probability distribution of lead time demand. Explicit results are derived in particular for the case when that distribution is normal (the only case to be discussed in this paper) with arbitrary parameters. An iterative scheme is provided for locating the optimal values in each case.

In the backorders case (the lost sales sales-case will be discussed in a later section of this paper) it is asserted that the objective function is strictly convex for normal lead time demand, that the partial derivatives vanish at a unique point thereby producing a global minimum, or else no solution exists at all. Not one of these statements is true! That the objective function is not in general convex has been known practically since the publication of the reference in a review given for JASA by Veinott [2]. As shall be seen shortly, the partial derivatives typically vanish at two points, not one. While the "other" point is a saddle point, hence not a candidate for a global minimum, it does give a clue to what is true, namely, that the function may take on values less than the asserted minimum. Finally, a solution always exists but may occur on the boundary of the region established by the authors' constraints on Q and r.

2. THE MODEL

A lot size-reorder point model is defined when a quantity Q is ordered each time an appropriate inventory level reaches a value r. Under a set of rather broad assumptions, the backorders case with normal lead time demand leads to an average annual cost function given by

$$(1) \qquad K \equiv K(Q,r) = \frac{\lambda A}{Q} + IC\left[\frac{Q}{2} + r - \mu\right] + \frac{\pi\lambda}{Q}\left[(\mu - r)\,\Phi\left(\frac{r-\mu}{\sigma}\right) + \sigma\phi\left(\frac{r-\mu}{\sigma}\right)\right]$$

where

A = set-up cost or cost of placing an order

λ = average annual demand

I = inventory carrying charge

C = unit cost

π = cost per unit backordered

μ = mean lead time demand

σ = standard deviation of lead time demand

$\phi(z)$ = standard normal density function evaluated at z.

$\Phi(z)$ = complementary cumulative distribution function for the standard normal density, i.e., $\Phi(z) = \int_z^\infty \phi(x)dx$.

All parameter values except μ are assumed positive.

Displayed in this form, K is seen to consist of the sum of three separate costs which are, respectively, ordering, holding and penalty for backordering. Incidentally, it should be pointed out that a term, precisely the expected number of backorders (the expression in brackets in the penalty cost), has been omitted in the holding cost expression by the authors. Hence, as they claim, the model only applies when the expected number of backorders is negligible, i.e. when backorder costs are high without, of course, being explicit as to what "negligible" and "high" means. As a

function of Q and r, K is to be defined by (1) only over the region $Q > 0$, $r \geq 0$ in the Q-r plane. Finally, as in all cases where normality is assumed for a nonnegative random variable (lead time demand in this case), some care has to be exercised in the choice of the parameters μ and σ. In particular, it should be the case that $\Phi(-\mu/\sigma) \doteq 1$. Before proceeding, since only the normal distribution will be involved, it is convenient to make a slight change of variable, namely, $z = (r-\mu)/\sigma$. In that case (1) may be re-written as a function of z and Q (in that order) as follows

$$(2) \qquad K \equiv K(z,Q) = \frac{\lambda A}{Q} + IC[\frac{Q}{2} + \sigma z] + \frac{\pi\lambda\sigma}{Q}[\phi(z) - z\Phi(z)]$$

Since the expression in the last brackets continually plays a crucial role in the mathematics to follow, it is designated $u(z) = \phi(z) - z\Phi(z)$ and K may be re-written

$$(3) \qquad K \equiv K(z,Q) = \frac{\lambda A}{Q} + IC[\frac{Q}{2} + \sigma z] + \frac{\pi\lambda\sigma u(z)}{Q}$$

In this form, K is defined over the region $z \geq -\mu/\sigma$, $Q > 0$ in the z-Q plane. The basic problem then is to find, if possible, a point (z^*,Q^*) in this region which minimizes K. The coordinates of such a point will be called optimal values of z and Q and immediately produce optimal values Q^* and $r^* = \sigma z^* + \mu$ of Q and r in the original formulation.

3. CONVEXITY OF K

To examine the convexity of K as well as the nature of its critical points, it is helpful to have an expression for the partial derivatives of K along with its Hessian matrix, Ω. Since $\phi'(z) = -z\phi(z)$ and

$\Phi'(z) = -\phi(z)$, it follows that $u'(z) = -\Phi(z) < 0$ and $u''(z) = \phi(z) > 0$ for all z. Thus, $u(z)$ is convex, strictly decreasing and, since $\lim_{z \to \infty} u(z) = 0$, $u(z) > 0$ for all z. Then it is a straightforward matter to verify that

$$\frac{\partial K}{\partial z} = IC\sigma - \frac{\pi\lambda\sigma\Phi(z)}{Q} \tag{4}$$

$$\frac{\partial K}{\partial Q} = \frac{IC}{2} - \frac{\lambda[A + \pi\sigma u(z)]}{Q^2} \tag{5}$$

and

$$\Omega = \begin{bmatrix} \frac{\pi\lambda\sigma\phi(z)}{Q} & \frac{\pi\lambda\sigma\phi(z)}{Q^2} \\ \frac{\pi\lambda\sigma\Phi(z)}{Q^2} & \frac{2\lambda[A + \pi\sigma u(z)]}{Q^3} \end{bmatrix} \tag{6}$$

Now the function K is convex over the region $z > -\mu/\sigma$, $Q > 0$ if and only if Ω is positive semi-definite [3] which will be true if and only if its principal minors are nonnegative. But all the terms of Ω are clearly positive in the region under consideration and,

$$\det \Omega = \frac{2\pi\lambda^2\sigma\phi(z)[A + \pi\sigma u(z)]}{Q^4} - \frac{\pi^2\lambda^2\sigma^2\Phi^2(z)}{Q^4} \tag{7}$$

$$= \frac{\pi\lambda^2\sigma}{Q^4} V(z)$$

where

$$V(z) = 2\phi(z)[A + \pi\sigma u(z)] - \pi\sigma\Phi^2(z) \tag{8}$$

Hence, K is convex if and only if $V(z) \geq 0$.

That K is not everywhere convex can be seen by an example using parameter values supplied by Hadley and Whitin themselves [1, page 173]. Taking $A = \$4000$, $\lambda = 1600$, $I = .20$, $C = \$50$, $\pi = \$2000$, $\mu = 750$ and $\sigma = 50$, it is easily verified that, for example, $V(-1) \doteq -16{,}410 < 0$.

In 1969, Brooks and Lu [4] established that, in the present context, the third term in the expression (3) for K is convex in z and Q provided $z > 0$ (or $r > \mu$). Since the other terms are obviously convex in z and Q for all values, K, being the sum of convex functions must itself be convex in the region $z > 0$, $Q > 0$. The point can only be of academic interest however since solutions can and do arise where $z < 0$. Convexity aside, the issue is the minimization of K. At the very least, a set of arguments different from those given by Hadley and Whitin then should be provided.

4. CRITICAL POINTS FOR K

For ease of reference, let $R = \{(z,Q);\ z \geq -\mu/\sigma,\ Q > 0\}$ be the domain of definition under consideration for K. Any point in the interior of R at which both partial derivatives (4) and (5) vanish will be called a critical point. If the optimal point (z^*,Q^*) is to be in the interior of R, then it must be such a point.

Now, when the partial derivatives (4) and (5) are set equal to zero, the respective equations, equivalent to those in Hadley and Whitin, become

$$Q = \sqrt{\frac{2\lambda[A + \pi\sigma u(z)]}{IC}} \tag{9}$$

$$\Phi(z) = \frac{QIC}{\pi\lambda} \tag{10}$$

Equating the corresponding expressions in Q^2 and solving for A, the set-up cost, one obtains

(11) $$A = \frac{\pi^2\lambda}{2IC}\phi^2(z) - \pi\sigma u(z)$$

Letting the right-hand side of (11) be denoted $A(z)$, it is thus seen that critical points are characterized completely by solutions of the equation $A(z) = A$ with corresponding values of Q coming from (10), say.

In the Appendix, the mathematical properties of $A(z)$ are developed in detail to reveal that the equation $A(z) = A$ typically has two solutions z_1 and z_2 where $z_2 < z_1$ and $z_2 < 0$. A typical graph of $A(z)$ is shown in Figure 1. In that figure (see the Appendix for details)

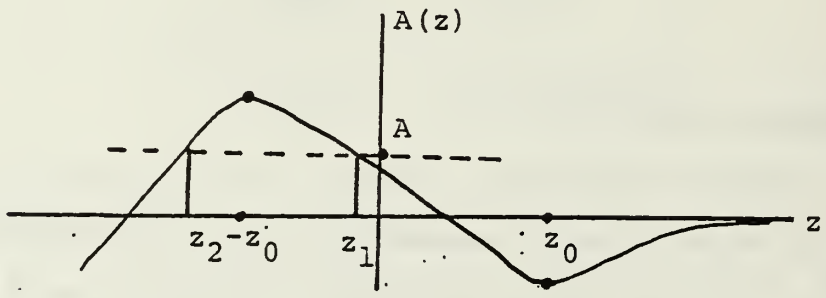


FIGURE 1. Typical Graph of $A(z)$

$\phi(z_0) = IC\sigma/\pi\lambda$ and $\pm z_0$ determine relative extrema for $A(z)$. As seen there, any value of A chosen between 0 and $A(-z_0)$ yields the two values z_1 and z_2. If $A > A(-z_0)$ of course there are no solutions to $A(z) = A$. Moreover, it is possible (for example, $I = .18$, $C = 120$, $\pi = 150$, $\lambda = 10$, $\sigma = 25$, $\mu = 125$) to have $A(-z_0) < 0$ in which case there will be no solution at all. When this happens, it follows that K cannot be minimized in the interior of R and one must then look to the boundary for solutions. This will also occur when $\phi(0) < IC\sigma/\pi\lambda$ for then z_0 will not exist at all. (Again, see the Appendix for details.)

When z_1 and z_2 do exist (with $-\mu/\sigma < z_2$), (10) may be used to solve for Q_1 and Q_2, respectively and then the determinant of the Hessian of K may be used to characterize the critical points (z_i,Q_i), $i = 1,2$. Indeed, once again one need only examine $V(z_i)$ from (8) since $(\pi\lambda^2\sigma)/Q_i^4 > 0$ will not affect the sign of $\det \Omega$. From (11), $A + \pi\sigma u(z) = (\pi^2\lambda)/(2IC)\,\Phi^2(z)$ and substituting this expression in (8) yields

$$V(z_i) = \frac{\pi^2\lambda}{IC}\,\Phi^2(z_i)\,[\phi(z_i) - \frac{IC\sigma}{\pi\lambda}] \tag{12}$$

In this way, the nature of the critical point (z_i,Q_i) is made to depend on a simple evaluation of $\phi(z_i)$, the standard normal density function. But, since $-z_0 < z_1 < z_0$, $\phi(z_1) > \phi(z_0) = IC\sigma/\pi\lambda$ so that $V(z_1) > 0$ and hence (z_1,Q_1) determines a relative minimum for K. On the other hand, since $z_2 < -z_0$, $\phi(z_2) < \phi(-z_0) = IC\sigma/\pi\lambda$ or $V(z_2) < 0$. Thus, (z_2,Q_2) is a saddle point.

Incidentally, this shows that the solution to (9) and (10) is not unique as claimed by Hadley and Whitin. Using these two equations, the authors provide a simple iterative scheme [1, Section 4-4] for solving. That scheme will always lead to (z_1,Q_1), the relative minimum point. It is a simple matter to derive a similar scheme for arriving at (z_2,Q_2) but that is of minor interest since it is not a competitor for a global minimum. It does, however, suggest that (z_1,Q_1), the relative minimum point, need not be a global one as asserted by Hadley and Whitin. For K must decrease in a direction away from the saddle point and hence may take on values less than $K(z_1,Q_1)$ before the boundary is reached.

A good example of this phenomenon is furnished by selecting parameter values $A = \$350$, $\lambda = 10$, $I = .18$, $C = \$12$, $\pi = \$25$, $\mu = 97.25$ and $\sigma = 25$.

Using the Hadley-Whitin scheme, it is easy to arrive at $(z_1,Q_1) = (-.97,96.5)$ with an associated cost of $K(-.98, 96.7) = \$156.17$. The saddle point occurs at $(z_2,Q_2) = (-1.21, 103)$ with a higher (always so) associated cost of $K(-1.21, 103) = \$156.64$. But if one arbitrarily selects $z = -2.2$ and $Q = 115$, a simple calculation reveals $K(-2.2, 115) = \$155.67$, a value less than the alleged minimum and still within the interior of R. Of course there are infinitely many policies that will do better still. The point is that a global minimum need not occur in the interior of R at all even when a solution (z_1,Q_1) exists, thereby creating a need for further investigation of the minimum solution. (For this particular problem, the absolute minimum happens to be $K(-3.89, 194.5) = \$143.00$.)

5. MINIMIZATION OF K

The entire problem of minimizing K can be reduced to the study of a function of the single variable z by observing the following facts. Let z be fixed and consider the corresponding function $K_z(Q) \equiv K(z,Q)$. Then dK_z/dQ is given by (5) and d^2K_z/dQ^2 in the second diagonal term in Ω. Hence, K_z is a convex function of Q with a unique minimum at

$$Q_1(z) = \sqrt{\frac{2\lambda[A + \pi\sigma u(z))]}{IC}} = \sqrt{Q_w^2 + \frac{2\pi\lambda\sigma}{IC}\, u(z)} \tag{13}$$

where $Q_w = \sqrt{(2\lambda A)/IC}$ is the familiar Wilson lot-size formula. Consequently, given any point (z,Q) in R, it follows that

$$K(z,Q) \geq K(z,Q_1(z))$$

and it suffices then to minimize the function

(14) $$K_1(z) = K(z, Q_1(z))$$

subject to the constraint $z \geq -\mu/\sigma$. Substituting $Q_1(z)$ in (3) and simplifying, (14) may be written

(15) $$K_1(z) = IC[Q_1(z) + \sigma z]$$

From this we have $K_1'(z) = IC[Q_1'(z) + \sigma]$. In the Appendix, it is shown that when the two solutions z_1, z_2 to $A(z) = A$ exist as in Section 4, then $Q_1'(z) = -\sigma$ precisely at $z = z_1$, z_2. Moreover, $Q_1'(z) < -\sigma$ if $z_2 < z < z_1$; otherwise, $Q_1'(z) > -\sigma$. That is to say,

$$K_1'(z) \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} 0 \quad \text{if and only if} \quad \left\{ \begin{array}{ll} z < z_2 & \text{or} \quad z > z_1 \\ z = z_2 & \text{or} \quad z = z_1 \\ z_2 < z < z_1 & \end{array} \right.$$

This means that K_1 has a relative maximum at z_2 and a relative minimum at z_1 and thus K_1 is minimized at z_1 or else at $z = -\mu/\sigma$, incorporating the constraint. When the two solutions z_1, z_2 do not exist, it is shown in the Appendix that $Q_1'(z) > -\sigma$ for all z. In that case K is an increasing function is thus minimized at $z = -\mu/\sigma$.

In summary then, K_1 is minimized in every case either at z_1 or $z = -\mu/\sigma$. Letting $Q_\mu = Q_1(-\mu/\sigma)$, the problem reduces to comparing the two values $K(z_1, Q_1)$ and $K(-\mu/\sigma, Q_\mu)$ to determine the minimum value of K. Since all of the above solutions are independent of μ, there is no way to reduce the problem further. There is, however, an additional complication arising from the fact that the expected number of backorders has been omitted in the holding cost, the second term in (3). Left unattended this could lead artificially to a negative holding cost thus unreasonably reducing

K to even a negative cost if carried to extremes. Thus, the omission of this term creates an additional implied constraint on the variables, namely,

$$\frac{Q}{2} + \sigma z \geq 0 \tag{16}$$

Fortunately, however, the constraint is never active except once again on the boundary $z = -\mu/\sigma$. In the Appendix once more, the details are provided to show that $Q_1/2 + \sigma z_1 > 0$ even when $z_1 < 0$. Hence only the condition $Q/2 + \sigma z = 0$ needs to be examined. But, on the line, $Q = -2\sigma z$, substituting in (3) and simplifying,

$$K_2(z) \equiv K(z,-2\sigma z) = -\frac{\lambda A + \pi\lambda\sigma u(z)}{2\sigma z}$$

Differentiating and simplifying,

$$K_2'(z) = \frac{\lambda A + \pi\lambda\sigma\phi(z)}{2\sigma z^2}$$

Since $K_2'(z) > 0$ for all z, it follows that K_2 is minimized at $z = -\mu/\sigma$.

This unfortunately adds another candidate on the boundary but it is a simple matter to select the appropriate point. Since $K_z(Q)$ is convex with a unique minimum at Q_μ when $z = -\mu/\sigma$, K is minimized at $(-\mu/\sigma, Q_\mu)$ if $Q_\mu > 2\mu$; otherwise, its minimum occurs at $(-\mu/\sigma, 2\mu)$. It is easy to construct examples yielding both solutions so that it is not possible to reduce the solution on the boundary further.

The entire minimization of K may now be summarized by a relatively simple computational scheme as follows. Given a set of parameters, first calculate $IC\sigma/\pi\lambda$.

(1) If $\phi(0) \leq IC\sigma/\pi\lambda$, calculate $Q_\mu = Q_1(-\mu/\sigma)$ using (13). Let $Q_2 = \max\{Q_\mu, 2\mu\}$ and set $(z^*,Q^*) = (-\mu/\sigma, Q_2)$.

(2) If $\phi(0) > IC\phi/\pi\lambda$, calculate $A(-z_0)$ where $\phi(\pm z_0) = IC\sigma/\pi\lambda$.

(a) If $A(-z_0) \leq 0$ or $A(-z_0) > 0$ but $A > A(-z_0)$ proceed as in Step (1) to locate the solution on the boundary.

(b) If $A(-z_0) > 0$ and $0 < A < A(-z_0)$ then determine (z_1,Q_1) using Hadley and Whitin's iterative scheme. Calculate Q_μ and let $Q_2 = \max\{Q_\mu, 2\mu\}$ as before. Compare $K(-\mu/\sigma,Q_2)$ to $K(z_1,Q_1)$. If $K(z_1,Q_1) < K(-\mu/\sigma,Q_2)$ and $z_1 \geq -\mu/\sigma$, then $(z_1,Q_1) = (z^*,Q^*)$; otherwise, $(z^*,Q^*) = (-\mu/\sigma,Q_2)$.

(3) In all cases, calculate $r^* = \sigma z^* + \mu$ to determine the optimal Q-r policy.

6. THE LOST SALES CASE

The lost sales case as treated in [1] differs mathematically from the backorders case only in the inclusion of an extra holding cost term corresponding to the neglected expected number of backorders, now lost sales. The objective function becomes, in place of (3),

$$K \equiv K(z,Q) = \frac{\lambda A}{2} + IC\left[\frac{Q}{2} + \sigma z + \sigma u(z)\right] + \frac{\pi\lambda\sigma u(z)}{Q} \tag{17}$$

where, this time, π is the cost of a lost sale. The authors claim that K is convex and that a unique solution always exists in the interior of R, the domian of definition as previously defined for the backorders model. But K is not convex and in fact the solution once again may be located on the boundary $z = -\mu/\sigma$ even though no saddle point arises.

Since K has been modified, some of the related functions also are changed. Equation (5) is still valid but (4) now becomes

$$\frac{\partial K}{\partial z} = IC\sigma[1 - \Phi(z)] - \frac{\pi\lambda\sigma\phi(z)}{Q} \tag{18}$$

and the only change in the Hessian matrix Ω is in the first diagonal term which now becomes

$$\frac{\partial^2 K}{\partial z^2} = \frac{\pi\lambda\sigma\phi(z)}{Q} + IC\sigma\phi(z) \tag{19}$$

All terms of the Hessian are positive once more and convexity again depends on the sign of $\det \Omega$ which may be written

$$\det \Omega = \frac{\lambda\sigma}{Q^4} V(z,Q)$$

where

$$V(z,Q) = 2\phi(z)\,[A + \pi\sigma u(z)][\pi\lambda + ICQ] - \pi^2\lambda\sigma\phi^2(z)\ . \tag{20}$$

This is the expression that replaces $V(z)$ in (8) and, unfortunately cannot be written as a function of z alone. Nevertheless, a counterexample to convexity can be supplied by selecting an appropriate set of parameter values.

Let $A = \$4000$, $\lambda = 1600$, $I = .20$, $C = \$50$, $\pi = \$2000$, $\mu = 750$, $\sigma = 50$ and consider the operating policy

$$z = -1.0 \quad (r = 700) \qquad Q = 10$$

Substitutng these values into (20) it can easily be verified that $V(z,Q) < 0$ and hence K is not everywhere convex.

When the partial derivatives are equated to zero the two equations

(21) $$Q = \sqrt{\frac{2\lambda[A + \pi\sigma u(z)]}{IC}}$$

as before and (instead of (10))

(22) $$\Phi(z) = \frac{QIC}{\pi\lambda + QIC}$$

must be satisfied simultaneously. Hadley and Whitin are quite right in their assertion that a solution (z_1,Q_1) to these equations always exists, is unique and constitutes a relative minimum for K. One way to show this is through a modified argument as in the backorders case where $A(z)$ is now given by

(23) $$A(z) = \frac{\pi^2\lambda\phi^2(z)}{2IC[1 - \Phi(z)]^2} - \pi\sigma u(z)$$

With slightly more detail it may be shown once again that $A(z)$ has a unique minimum at z_0 where this time z_0 is determined by

$$\frac{\phi(z_0)}{[1 - \Phi(z_0)]^3} = \frac{IC\sigma}{\pi\lambda}$$

and always exists. Moreover, $A(z_0) < 0$ as before but in contrast to the backorders case, A has no relative maximum and the solution to $A(z) = A$ always exists and is unique. The details are given at length by Trapnell [6].

As to the problem of minimizing K, the problem can be reduced again to a single function of z as in the backorders case. Equation (13) is still a valid formula in this case, but now

(24) $$K_1(z) \equiv K(z,Q_1(z)) = IC[Q_1(z) + \sigma z + \sigma u(z)]$$

For any pair (z,Q) in the interior of R, then

$$K(z,Q) \geq K_1(z)$$

Modifications of the arguments given in the appendix can then be provided to prove that K_1 is a decreasing function on $(-\infty, z_1)$ and is increasing (z_1, ∞) with a unique minimum at z_1. Thus,

$$K_1(z) \geq K_1(z_1) = K(z_1, Q_1)$$

where $Q_1 = Q_1(z_1)$. In this way, it may be seen that (z_1, Q_1), which may be computed using the Hadley-Whitin iterative scheme, does constitute an unconstrained global minimum, convexity notwithstanding.

It should be pointed out, however, that the solution in this case is still independent of μ and so it is conceivable that the solution does not lie in R. That this is indeed the case is shown by the following example.

Let $A = \$25{,}000$, $\lambda = 5$, $I = .20$, $C = \$49{,}500$, $\pi = \$10$, $\mu = 30$ and $\sigma = 10$. For this choice of parameters, the iterative solution yields $(z_1, Q_1) = (-3.09, 5.06)$. But $-\mu/\sigma = -3$, so that the solution lies outside R. Again, the boundary constraint must be noted although it is easily managed. In this case $K(-3, 5.055) = \$50{,}084.43$ is the optimal solution located on the boundary. In other words, one has only to calculate (z_1, Q_1). If $z_1 < -\mu/\sigma$ then let $z^* = -\mu/\sigma$, $Q^* = Q_1(-\mu/\sigma)$; otherwise $z^* = z_1$, $Q^* = Q_1$.

7. APPENDIX

In this appendix, the mathematical details to support claims in different sections of the paper are provided. Additional details may be found in [5] and [6].

From Section 4, the function $A(z)$, was defined by the expression

$$A(z) = \frac{\pi^2 \lambda}{2IC} \Phi^2(z) - \pi\sigma u(z)$$

Using the properties of $u(z)$ in Section 3,

$$A'(z) = \frac{\pi^2 \lambda}{IC} \Phi(z) \left[\frac{IC\sigma}{\pi\lambda} - \phi(z)\right]$$

and

$$A''(z) = \pi\sigma\phi(z) \left[\frac{\pi\lambda}{IC\sigma} (\phi(z) + z\Phi(z)) - 1\right] ,$$

are formulas that can be easily verified with straightforward algebraic simplification. Moreover,

$$\lim_{z \to -\infty} A(z) = -\infty \quad \text{and} \quad \lim_{z \to +\infty} A(z) = 0$$

Now, if it should happen that $IC\sigma/\pi\lambda \geq \phi(0) \doteq .3989$ at the outset, then for all z, $IC\sigma/\pi\lambda \geq \phi(z)$ and hence $A'(z) > 0$ or A is an increasing function. In that case there can be no solution to the equation $A(z) = A$ for any $A > 0$ because of the limits for $A(z)$.

On the other hand, if $IC\sigma/\pi\lambda < \phi(0)$, then there exists a unique z_0 such that $\phi(\pm z_0) = IC\sigma/\pi\lambda$ from the nature of the standard normal density function ϕ. Then it follows that $A'(z) = 0$ if and only if $z = \pm z_0$. But

$$A''(z_0) = \pi\sigma \cdot \frac{IC\sigma}{\pi\lambda} \left[\frac{\pi\lambda}{IC\sigma} \left(\frac{IC\sigma}{\pi\lambda} + z_0\Phi(z_0)\right) - 1\right] = \pi\sigma z_0\Phi(z_0) > 0$$

while

$$A''(-z_0) = -\pi\sigma z_0\Phi(z_0) < 0 .$$

Thus, z_0 determines a relative minimum for $A(z)$ while $-z_0$ determines a relative maximum. Hence $A(z_0) < 0$ and, further, since $\phi(z) > \phi(z_0)$

for $|z| < z_0$ while $\phi(z) < \phi(z_0)$ for $|z| > z_0$, A is increasing on $(-\infty, -z_0)$, is decreasing on $(-z_0, z_0)$ and again increasing on (z_0,∞). Hence, if $A(-z_0) > 0$, then for any $0 < A < A(-z_0)$, there exist two solutions to the equation $A(z) = A$, say z_2 and z_1 (see Figure 1 in Section 4 for visual reference) where,

$$z_2 < -z_0 < z_1 < z_0$$

In summary, then, the equation $A(z) = A$ has two unique solutions if (1) $\phi(0) > IC\sigma/\pi\lambda$, (2) $A(-z_0) > 0$ and (3) $A < A(-z_0)$; otherwise, there are no solutions.

In Section 5 the function $Q_1(z)$ was defined by the formula

$$Q_1(z) = \sqrt{Q_w^2 + \frac{2\pi\lambda\sigma}{IC}\, u(z)}$$

Differentiating $Q_1^2(z)$ implicitly yields

$$2Q_1(z)\, Q_1'(z) = -\frac{2\pi\lambda\sigma}{IC}\, \Phi(z)$$

so that

$$Q_1'(z) = -\frac{\pi\lambda\sigma}{IC}\,\frac{\Phi(z)}{Q_1(z)} = -\sigma\,\frac{Q_2(z)}{Q_1(z)}$$

where

$$Q_2(z) = \frac{\pi\lambda}{IC}\,\Phi(z)$$

which happens, incidentally, to be equation (10) solved for Q.

Now let

$$D(z) = Q_1^2(z) - Q_2^2(z) = Q_w^2 + \frac{2\pi\lambda\sigma}{IC}\, u(z) - \left(\frac{\pi\lambda}{IC}\right)^2 \Phi^2(z)$$

$$= Q_w^2 - \frac{2\lambda}{IC}\, A(z) = \frac{2\lambda}{IC}\,[A - A(z)]$$

where $A(z)$ is as previously defined. Thus,

$$D'(z) = -\frac{2\lambda}{IC} A'(z) \quad \text{and} \quad D''(z) = -\frac{2\lambda}{IC} A''(z)$$

The properties of $A(z)$ just developed can thus be used to infer the following properties of $D(z)$.

If $A(z) = A$ has two solutions z_1, z_2, then $D(z) = 0$ if and only if $z = z_1, z_2$, $D'(z) = 0$ if and only if $z = \pm z_0$, $D''(z_0) < 0$ while $D''(-z_0) > 0$ and, finally, $\lim_{z \to -\infty} D(z) = +\infty$ while $\lim_{z \to -\infty} D(z) = Q_w^2$. A sketch of $D(z)$ appears in Figure 2 for ease of reference. When the solutions to $A(z) = A$ do not exist, D has no zeros and $A(z) < A$ or $D(z) > 0$ for all z.

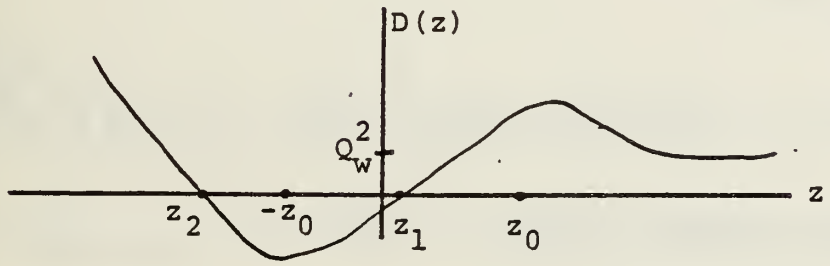


FIGURE 2. Graph of $D(z)$

These simple properties of $D(z)$ can then be related to those of $Q_1'(z)$ since $D(z) = [Q_1(z) - Q_2(z)][Q_1(z) + Q_2(z)]$ and $Q_1(z) > 0$, $Q_2(z) > 0$. Thus, if there are no solutions z_1, z_2, then $Q_1(z) > Q_2(z)$ so that $Q_1'(z) > -\sigma$ for all z. When z_1 and z_2 exist, then, for $z_2 < z < z_1$, $Q_1(z) < Q_2(z)$ or $Q_1'(z) < -\sigma$; otherwise, $Q_1(z) > Q_2(z)$ or $Q_1'(z) > -\sigma$. Moreover $Q_1'(z) = -\sigma$ precisely when $Q_1(z) = Q_2(z)$ or $D(z) = 0$, i.e. when $z = z_1$ or $z = z_2$.

Finally, as required in Section 5, it is necessary to show that $Q_1/2 + \sigma z_1 > 0$. To show this, consider the function

$$g(z) = z\phi(z) + \frac{1}{2}\Phi(z)$$

It is straightforward to establish that

$$g'(z) = \phi(z)\left(\frac{1}{2} - z^2\right)$$

$$g''(z) = z\phi(z)\left(z^2 - \frac{5}{2}\right)$$

Of special interest is the critical value $z = -1/\sqrt{2}$. Since

$$g''\left(-\frac{1}{\sqrt{2}}\right) = \frac{2}{\sqrt{2}}\,\phi\left(-\frac{1}{\sqrt{2}}\right)$$

it follows that

$$g\left(-\frac{1}{\sqrt{2}}\right) = -\frac{1}{\sqrt{2}}\,\phi\left(-\frac{1}{\sqrt{2}}\right) + \Phi\left(-\frac{1}{\sqrt{2}}\right) \doteq .16$$

is a relative minimum. But $g(0) = .25$ and

$$\lim_{z \to -\infty} g(z) = .5$$

since, applying L'Hospital's rule,

$$\lim_{z \to -\infty} z\phi(z) = -\lim_{z \to -\infty} \frac{-z}{1/\phi(z)} = -\lim_{z \to -\infty} \frac{1}{\frac{z\phi(z)}{\phi^2(z)}} = \lim_{z \to -\infty} \frac{\phi(z)}{z} = 0$$

Thus, for all $z > 0$, $g(z) > 0$ and in particular, if $z_1 < 0$, then, since $-z_0 < z_1$, $\phi(z_1) > \phi(-z_0) = IC\sigma/\pi\lambda$, $\Phi(z_1) = (Q_1 IC)/\pi\lambda$ (from (10) and then,

$$z_1\phi(z_1) + \frac{1}{2}\Phi(z_1) > 0$$

and

$$\frac{\pi\lambda}{IC}\,\phi(z_1) > \sigma \qquad \text{or} \qquad \frac{\pi\lambda}{IC}\,z_1\phi(z_1) < \sigma z_1 .$$

It follows that,

$$\frac{Q_1}{2} + \sigma z_1 > \frac{Q_1}{2} + \frac{\pi\lambda}{IC}\,z_1\phi(z_1) = \frac{\pi\lambda}{IC}\,[z_1\phi(z_1) + \frac{1}{2}\,\Phi(z_1)] > 0$$

Naturally this is already so if $z_1 \geq 0$.

REFERENCES

[1] Hadley, G. and T. M. Whitin, Analysis of Inventory Systems, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[2] Veinott, A. F., "Book Reviews," Journal of the American Statistical Association, Vol. 59, No. 305 (March, 1964), 283-285.

[3] Roberts, A. N. and D. E. Varberg, Convex Functions, Academic Press, New York, and London, 1973.

[4] Brooks, R. S. and J. Y. Lu, "On the Convexity of the Backorder Function for an E.O.Q. Policy," Management Science, Vol. 15, No. 7 (March, 1969), pp. 453-454.

[5] Minh, T.V., Convexity in a Hadley-Whitin Model, M.S. Thesis, Naval Postgraduate School, Monterey, Ca., 1975.

[6] Trapnell, R. G., On the Convexity of the Hadley-Whitin Lost Sales Inventory Model and the Characterization of its Minimum Solution, M.S. Thesis, Naval Postgraduate School, Monterey, Ca., 1977.

DISTRIBUTION LIST

	No. of Copies
Deputy Chief of Naval Material (Acquisition) Department of the Navy Washington, D. C. 20360	
Code 08	1
Code 08B	1
Code 08T	1
Code 08T1	1
Code 08T2	1
Code 08T2B	1
Library Code 0142 Naval Postgraduate School Monterey, CA 93940	2
Dean of Research Code 012 Naval Postgraduate School Monterey, CA 93940	15
Research Administration Code 012A Naval Postgraduate School Monterey, CA 93940	5