



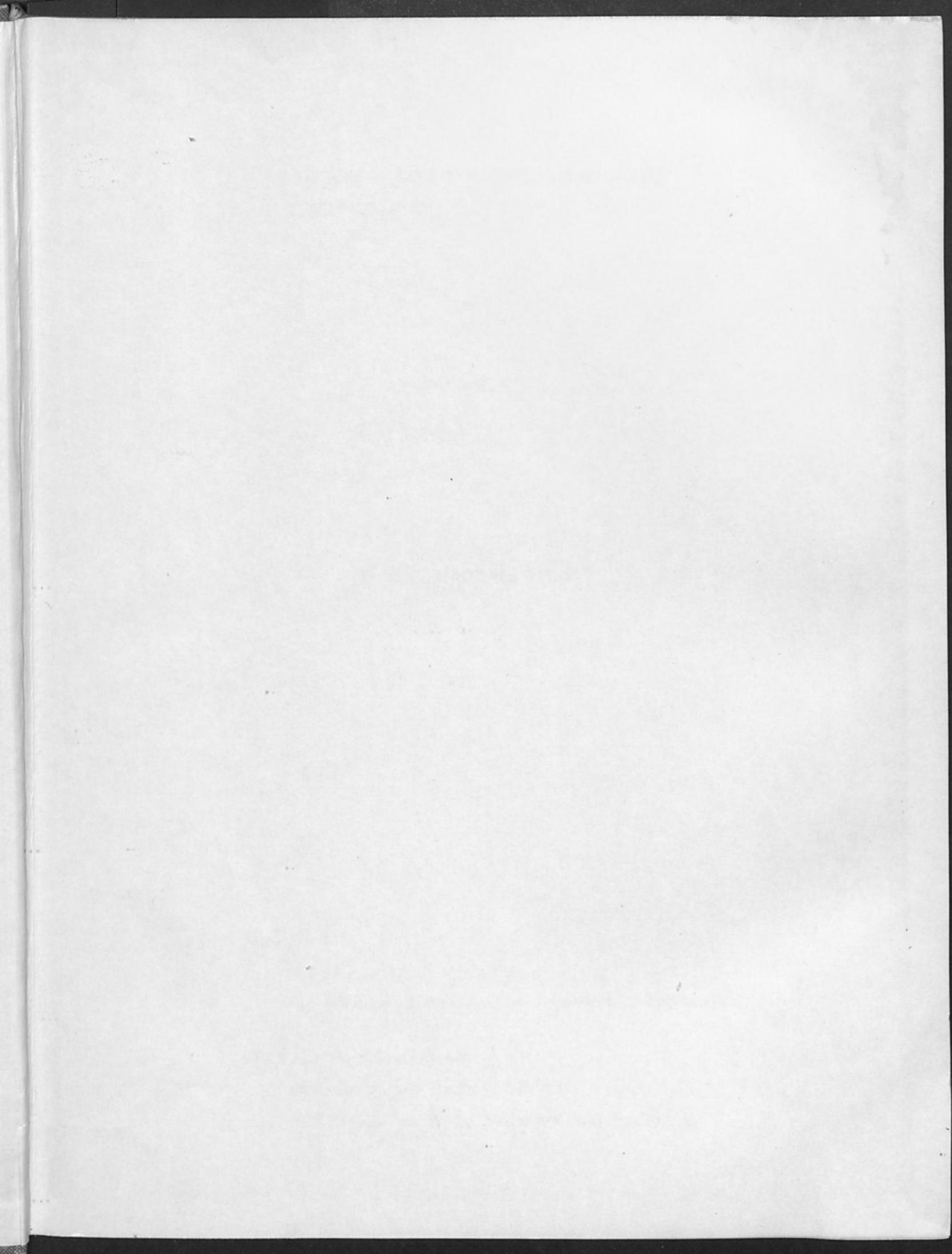
PROCEEDINGS OF THE 8TH ANNUAL
ELECTRON AND LASER BEAM SYMPOSIUM

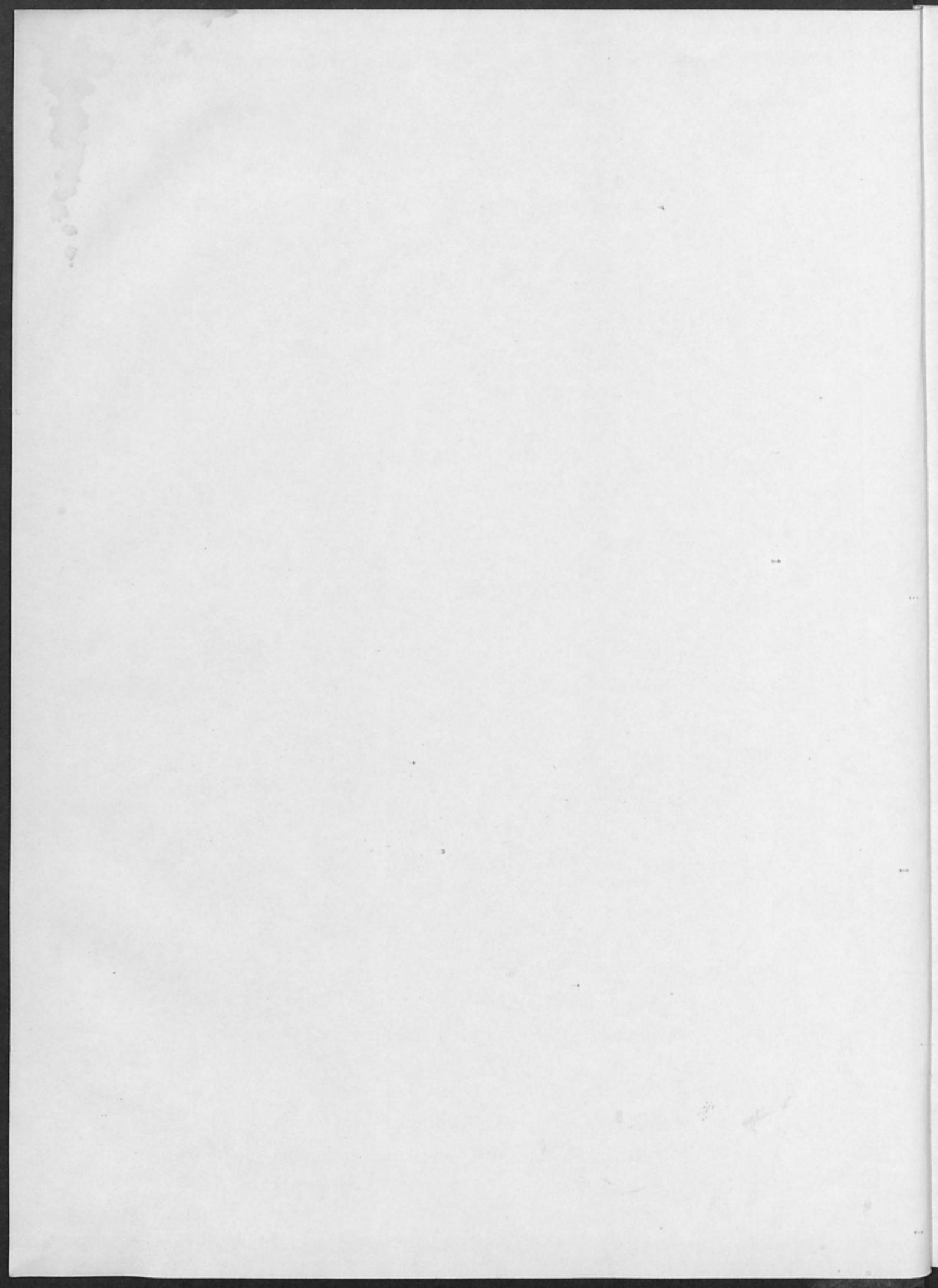
APRIL 6 - 8 1966

Editor - G. I. HADDAD

Sponsored by:

The UNIVERSITY OF MICHIGAN and
The INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS





PROCEEDINGS OF THE EIGHTH ANNUAL
ELECTRON AND LASER BEAM SYMPOSIUM

April 6-8, 1966

G. I. HADDAD, *Editor*

Sponsored by:

The University of Michigan

The Institute of Electrical and Electronics Engineers

With the Participation of:

IEEE Group on Electron Devices

IEEE Group on Parts, Materials and Packaging

PROCEEDINGS OF THE EIGHTH ANNUAL
ELECTRON AND LASER BEAM SYMPOSIUM

April 6-8, 1966

G. I. HADDAD, Editor

Sponsored by

The University of Michigan

The Institute of Electrical and Electronic Engineers

With the Participation of

IEEE Group on Electron Devices

IEEE Group on Part, Materials and Packaging

PREFACE

The first meeting on electron-beam technology was organized in Boston in 1959 by the Alloyd Corporation and at that time the discussions were primarily concerned with the application of electron beams to metal working i.e. melting and refining of metals. In subsequent symposia additional fields such as electron beam welding, machining and evaporating and recording became major topics of interest and discussion. In recent years the interactions of electron and laser beams with materials began to occupy a substantial portion of the symposia and this year the subject of "holography" was introduced for the first time. The principal topics of this years symposium are

1. The physics of electron, ion and light beams
2. The application of energy beams to
 - microminiaturization and thin films
 - welding and materials
 - melting and evaporation
 - microcircuit fabrication
 - materials study
3. New energy beam equipment and processes
4. Holography.

Additional developments in this important and fast moving field promise to broaden the scope of the symposium even further in the future.

The Eight Annual Symposium was held on the campus of The University of Michigan in Ann Arbor during the period April 6-8, 1966 under the guidance of the following symposium officers

Co-chairmen:

- Dr. J. E. Rowe, The University of Michigan
- Dr. A. B. El-Kareh, The Pennsylvania State University

Program Committee:

- Dr. G. I. Haddad, The University of Michigan, Chairman
- Dr. H. B. Law, RCA Laboratories
- Dr. L. L. Marton, National Bureau of Standards
- Dr. H. Moss, Westinghouse
- Dip. Phys. K. H. Steigerwald, Munich, Germany
- Dr. C. Susskind, University of California
- Dr. S. S. White, Texas Instruments

Local Arrangements:

- Dr. N. A. Masnari, Chairman

I wish to express my sincere appreciation to all of the Committee members who worked diligently to make the meeting a success and, of course, to all the authors for their excellent presentations. It is also a pleasure to acknowledge the valuable contributions of Messrs. R. E. Carroll and D. L. Danford and their staffs in the management of the conference.

Joseph E. Rowe
Co-chairman

Department of Electrical Engrg.
The University of Michigan
Ann Arbor, Michigan

FOREWORD

This volume contains a collection of papers presented at the Eighth Annual Symposium on Electron and Laser Beam Technology. A wide variety of subjects are discussed which indicate the importance of electron and laser beams in various applications. The subject on "Holography" was included for the first time in the program of this symposium and two papers by two of the most prominent people in this field are included. The first paper by Professor D. Gabor reviews the basic principles of Holography and some applications. Professor E. N. Leith discusses some of the recent results and applications of this important subject in the second paper. The very timely and important subject of optical data processing for radar and communications and the use of lasers for this purpose is discussed in a detailed paper by Professor L. J. Cutrona. Professor P. A. Franken, J. A. Jenney and D. M. Rank consider the use of laser beams for detection of Clear-Air Turbulence which is of great interest. The rest of the papers deal mainly with the properties of electron and laser beams and their application to machining, welding, evaporation and measurement of physical properties of materials. These papers indicate major advances in the practical realization of production-worthy systems for these applications.

I wish to express my sincere appreciation to the authors for submitting their papers on time. I also apologize for the delay in the publication of these proceedings. This was beyond my control.

The assistance of Mrs. Kathy Stewart in editing the manuscripts and of Mr. D. L. Danford and his staff in the preparation of this volume is greatly appreciated.

George I. Haddad, Editor
Associate Professor of
Electrical Engineering
University of Michigan

TABLE OF CONTENTS

	<u>Page</u>
PREFACE.....	iii
FOREWORD.....	v
WAVEFRONT RECONSTRUCTION OR "HOLOGRAPHY"	
Dennis Gabor.....	1
RECENT RESULTS IN HOLOGRAPHY	
E. N. Leith.....	21
THE USE OF LASERS IN SIGNAL PROCESSING FOR RADAR AND COMMUNICATIONS	
L. J. Cutrona.....	39
AIRBORNE INVESTIGATIONS OF CLEAR AIR TURBULENCE WITH LASER RADARS*	
P. A. Franken, J. A. Jenney, D. M. Rank.....	87
ZAPPING PAPER	
Jon H. Myer.....	105
THE APPLICATION OF LASERS IN THERMOPHYSICAL PROPERTIES MEASUREMENTS*	
M. M. Nakata.....	107
VANADIUM CHARGE COMPENSATOR SITE IN LASER CaWO_4	
C. Kikuchi, N. Mahootian, W. Viehmann and R. T. Farrar.....	117
THE EFFECT OF INTERNAL REFLECTION IN A CORNER CUBE UPON THE POLARIZATION OF A REFLECTED BEAM	
Peter J. Walsh and Irvin Krause.....	139
LASER MACHINING STUDY	
Warren V. Trammell.....	157
THE LASER NOW A PRODUCTION TOOL	
J. P. Epperson, R. W. Dyer and J. C. Grzywa.....	188
DESIGN OF A PRODUCTION-WORTHY LASER MICROWELDER	
Jon H. Myer.....	207

TABLE OF CONTENTS (CONT'D)

	<u>Page</u>
THE ANNULAR HOLLOW CATHODE; ITS OPERATION AND APPLICATIONS J. W. Davis, A. P. Walch, and E. A. Pinsley.....	217
PRELIMINARY EVALUATION OF A NONVACUUM WELDER FOR AEROSPACE APPLICATIONS J. Lempert, J. F. Lowry, C. S. Williams, and F. D. Seaman....	233
TRANSITION OF ELECTRON BEAM WELDING FROM HARD VACUUM TO SOFT VACUUM H. A. James and J. L. Solomon.....	259
PROPERTY CHANGES INDUCED BY ELECTRON BEAM EXPOSURE IN VARIOUS SEMICONDUCTOR AND DIELECTRIC MATERIALS* David R. Sivertsen, Olin B. Cecil, and Rolf R. Haberecht.....	279
ELECTRON-BEAM EVAPORATION OF SILICON DIOXIDE AS A STORAGE LAYER FOR THE RCA AUTOMATIC-PICTURE-TRANSMISSION VIDICON* D. W. Roe.....	319
A PULSED MACHINE-SCAN ELECTRON BEAM DEVICE A. B. El-Kareh, R. B. Fair and C. R. Marsh.....	331
APPLICATIONS OF THE BARRIER ELECTRON VOLTAIC EFFECT C. Munakata.....	357
A METHOD OF EXAMINATION OF SEMICONDUCTOR OXIDES WITH A SCANNING ELECTRON MICROSCOPE D. Green.....	375
ELECTRON BEAM SYNTHESIS AND CRYSTALLIZATION OF InSb FILMS N. M. Davis and H. H. Wieder.....	385
Al ₂ O ₃ SINGLE CRYSTAL GROWTH BY ELECTRON BEAM FLOAT-ZONE MELTING Walter Class and Harvey R. Nesor.....	399
DIFFRACTION-LIMITED SPOT SCANNING Samuel Bousky.....	415

TABLE OF CONTENTS (CONT'D)

	<u>Page</u>
DIFFRACTION LIMITED CONCENTRIC OPTICS	
E. K. Thorburn.....	429
THE PRODUCTION AND USE OF MONOENERGETIC ELECTRON BEAMS	
J. Arol Simpson.....	437
MECHANISM OF ENERGY RELAXATION IN ELECTRON BEAMS	
Kurt Ulmer and Bodo W. Zimmermann.....	449
TECHNIQUES FOR THE STUDY OF SELF-FOCUSING ELECTRON STREAMS*	
S. E. Graybill and S. V. Nablo.....	465
THE QUEST FOR SUPERDENSE ELECTRON BEAMS	
W. E. Waters.....	487
AN ELECTRON ACCELERATOR WITH A PEAK BEAM POWER OF 10^{10} WATTS	
W. P. Dyke, J. P. Barbour, F. J. Grundhauser and F. M. Charbonnier.....	505

WAVEFRONT RECONSTRUCTION OR "HOLOGRAPHY"

by

Dennis Gabor, F.R.S.
Professor of Applied Electron Physics
Imperial College of Science and Technology
University of London
London, England

ABSTRACT

So far the main interest in holography has arisen from the possibility of three-dimensional reconstruction of optical fields, and from the almost perfect, "noise-like" coding in holograms. But it deserves attention also because of the amplification principle contained in it, which makes it possible to see or to photograph objects with very reduced exposures. Other still new and unexplored possibilities are recoding (translating a sign A into a sign B) and applications to ultrasound, electromagnetic waves and X-rays. The purpose for which holography was originally invented, improvement of electron microscopy, is also still waiting for exploitation. The fundamentals of holography and various possible applications will be discussed.

I. INTRODUCTION

Holography is the art of freezing a light wave into a photographic emulsion, so that it can be revived by another light wave. There is only one condition for this somewhat unbelievable experiment to succeed: the light waves must be coherent. I must start therefore with a brief explanation of coherence.

Consider two point sources O_1 and O_2 , for instance two pinholes in a dark screen, illuminated by a common light source (Fig. 1). In order to avoid making the explanation too easy or too complicated, let the illumination be neither exactly monochromatic, nor entirely general, but approximately or "quasi" monochromatic; a small group of spectral lines centering on some wavelength λ . The superposition of the two waves emanating from O_1 and O_2 will form a set of standing waves, whose maxima will be on a family of hyperbolas, as these are the loci of constant path differences. The exact position of these maxima depends on the phase of the light from O_1 and O_2 ; for simplicity it has been assumed that these are equal.

This is sufficient to explain the concept of spatial coherence.

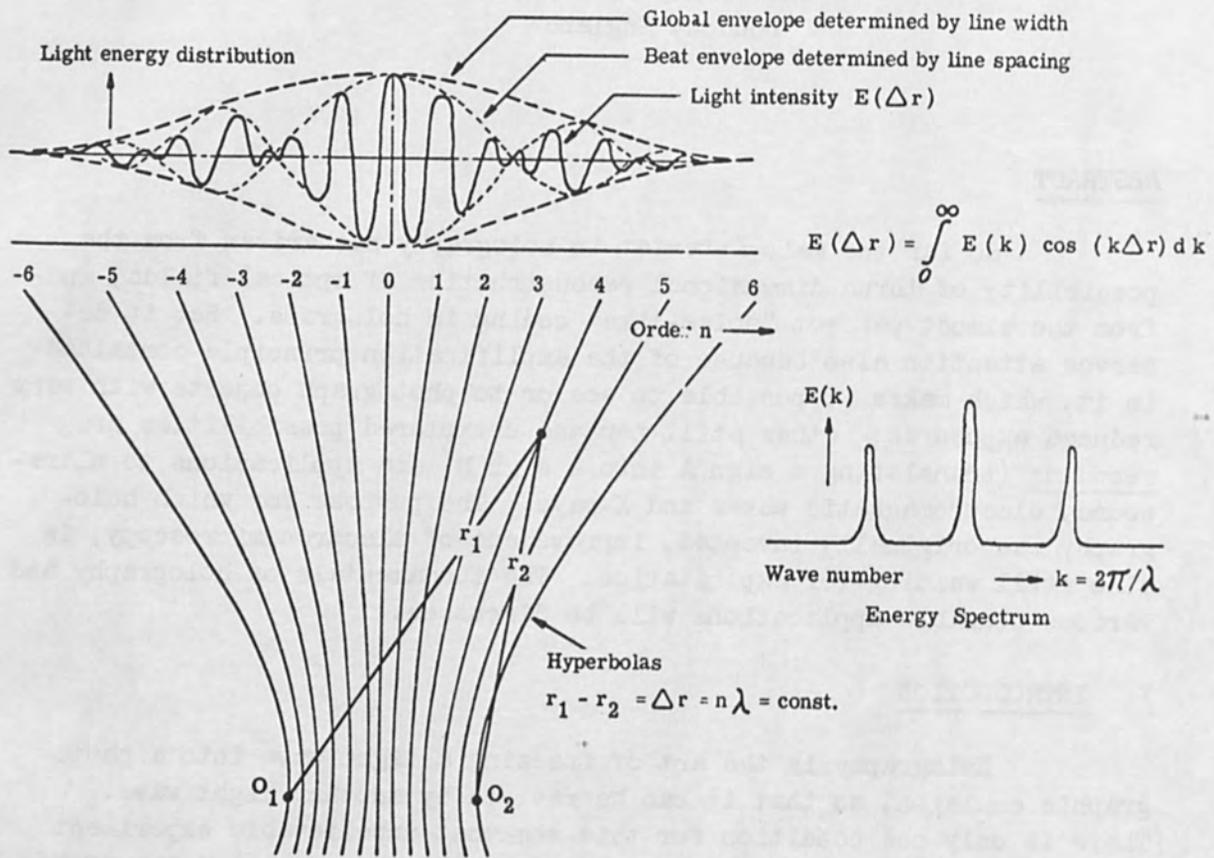


Figure 1. Coherence of Two Point Sources.

In any one of the zones between two neighboring fringes the two point sources are like one; they are optically indistinguishable by a lens system which does not extend beyond this zone. (More exactly one half of a zone.) If we approach the two points O_1 and O_2 toward one another, the interference fringes are gradually squeezed out. In the end, when the two points have approached to half a wavelength, only half a zone will extend over the whole hemisphere, and the two points have now become absolutely indistinguishable. If one wants to carry out experiments with coherent light within a certain cone of light, one must take a pinhole small enough to make it unresolvable for any optical system which views it within that cone. This was the basic difficulty of all interference experiments before the invention of the laser; the very small pinholes which had to serve as illuminators. The laser has made it possible to put enormous intensities through very fine pinholes.

Consider now the second aspect: chromatic coherence. If the light source is not absolutely monochromatic, with increasing order n , that is to say increasing path difference, the two light waves gradually get out of step. The quantitative aspect of this was first understood by Albert Michelson and Lord Rayleigh, and later greatly elucidated by a succession of distinguished authors: Van Cittert, Zernicke, Hopkins, Wolf, Blanc-Lapierre, Mandel. It was Michelson who first found the formula for the visibility of interference fringes, shown in Fig. 1. The intensity as a function of the path difference Δr is the cosine Fourier transform of the spectral energy distribution. The narrower the spectral distribution, the larger the path difference at which the interference fringes fade out and the light becomes incoherent. A single spectral line gives a broad global envelope. If several lines are present, these produce beats, and the coherence comes and goes repeatedly, until it finally fades out. With laser light containing a few modes, the first zero usually appears at a distance of a foot or two, but with a single mode it can extend to many miles.

Assume now that we have sufficient coherence within a reasonably wide angle, and we can now immediately understand the basis of wavefront reconstruction or "holography". Let us construct a sphere around one of the two point sources, now called O (Fig. 2), and make this transparent only at the maxima of the standing waves, i.e., the bright fringes. We have now obtained a hologram; a record of the other point source here called O' . The spherical wave from O arrives in phase at all the perforations. But a wave from O' would also arrive in phase, because it differs from the other only in an integer number of wavelengths. By Huygens' Principle, each of the gaps in the spherical screen emits spherical wavelets, but these add up only in their common envelope, where their phases agree. But this is equally true for the illuminating wave, centering on O , as for another wave apparently issuing from O' . We have therefore "reconstructed" the source O' . We have also obtained a little

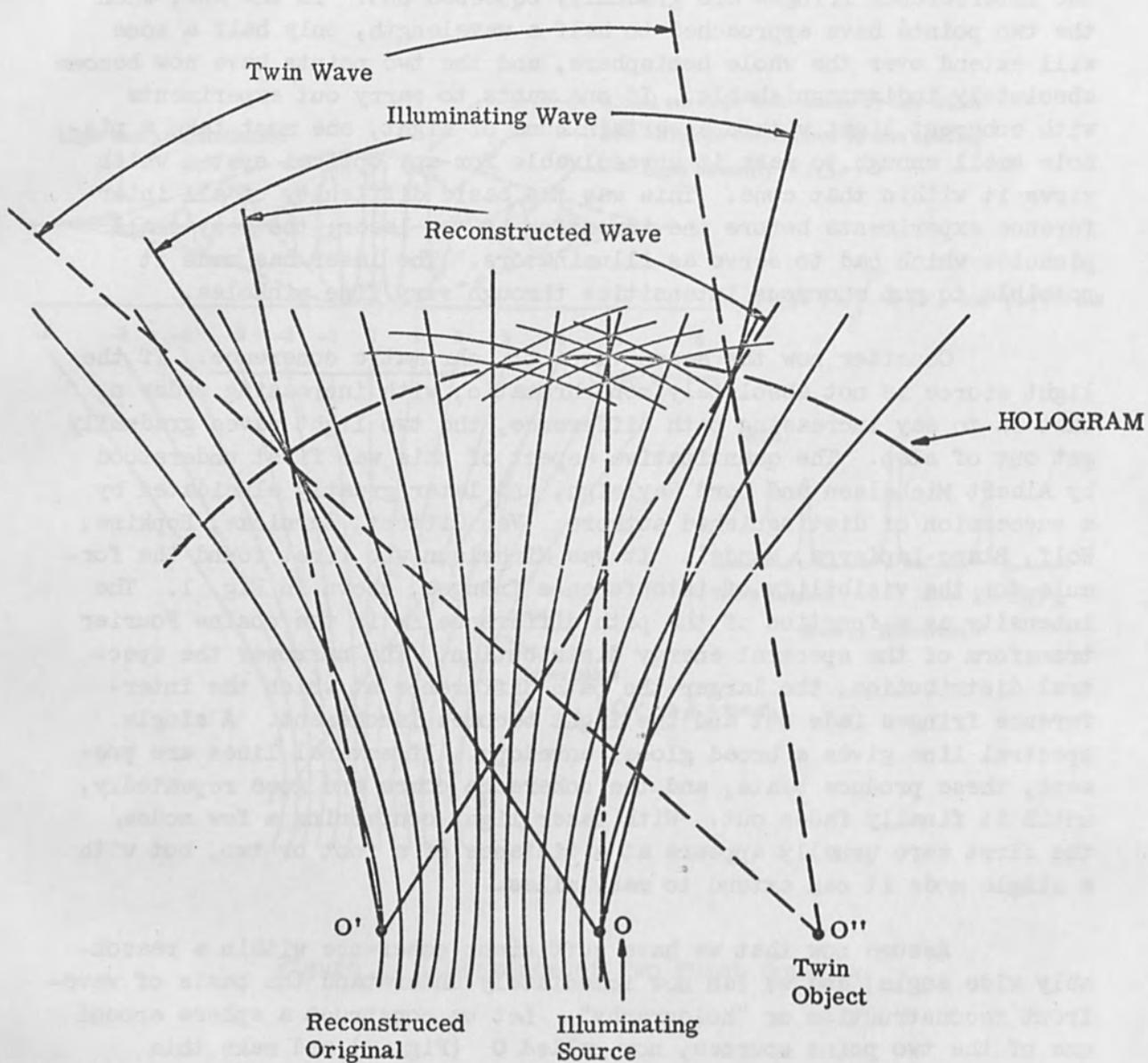


Figure 2. Wavefront Reconstruction.

more than we wanted, because the same is true also for another point, O'' , which is symmetrical to O' with respect to O . This is the "twin object", of which we will say more later. For the moment we note that O'' is the optical conjugate to O' with respect to the reconstructing wave-front, which is a spherical mirror; it will be seen that this is a general rule for the relation of the two objects to one another.

We can now at once see another curious property of holograms. Let us make the dark fringes transparent instead of the bright fringes; that is to say, let us take a negative photograph instead of a positive one. It makes no difference! The whole effect is only as if we had introduced a difference of π between the phases of the two sources, and this is unnoticeable; we obtain a bright reconstruction in either case. There are in fact some differences between positive and negative holograms, but these are rather subtle, and we need not discuss them here.

Let us now turn our attention to the region between O and O' . Can we produce a hologram by marking off the interference fringes on one of the hyperboloids? We cannot of course; there are no fringes. But let us now put a reflector into any one of these hyperboloids. We now obtain the reconstructed object; what is more we obtain it with 100% efficiency, because the hyperboloid is a perfect imaging system; all the light from O reflected at it will appear as emanating from the virtual source O' , and in this case there is no twin object.

Let us now put not one perfectly reflecting mirror into one of the hyperboloids, but partially reflecting mirrors (or scattering particles) into several of them. What we are doing is a combination of holography with Lippmann color photography. It was Gabriel Lippmann, who in 1891 first obtained photographs of standing light waves in an emulsion with very fine, colloidal silverbromide particles. Fine, scattering silver grains were precipitated in the maxima of the standing waves. The combination of Lippmann photography and holography was first proposed by Denysiuk (1962), and independently by Van Heerden (1963), but it was first successfully realized only in the course of the last year by Pennington and Lin, first alone, later in a very fruitful collaboration with Stroke and Labeyrie. The last named two authors first realized the experiment with light coming from the two sides of a plate and interfering in the emulsion. Such an interference, with light beams meeting at nearly 180° , was quite outside the reach of physics in Lippmann's time, when interferences were possible only at very acute angles, and is one of the miracles made possible only by the laser.

This "Lippmann holography" gives two new effects, both very welcome. First of all, there is no twin object. Second, one now obtains color selection, by Lippmann's principle (now better known as the Lippmann-Bragg principle), by which only those colors will be effectively

reflected whose wavelengths fit between the silver grain layers, because only these reflected wavelets will add up in phase; the others will destroy each other by interferences and will be transmitted, to be absorbed by a black background.

II. FUNDAMENTALS OF HOLOGRAPHY

I have started by explaining holography in geometrical terms, as a preparation to its mathematical formulation, which is also very simple. The basic idea is that one takes two coherent waves A and B; the first known, the other emanating from the unknown object which one wants to reconstruct, and one lets them fall together on a photographic plate. More exactly: let $A(x,y)$ and $B(x,y)$ be the complex amplitudes of the two waves in the plane x,y of the plate, and the resulting amplitude is $A + B$. The photographic plate is completely indifferent to the phase of a light wave; it responds only to its intensity which is the absolute square of $A + B$

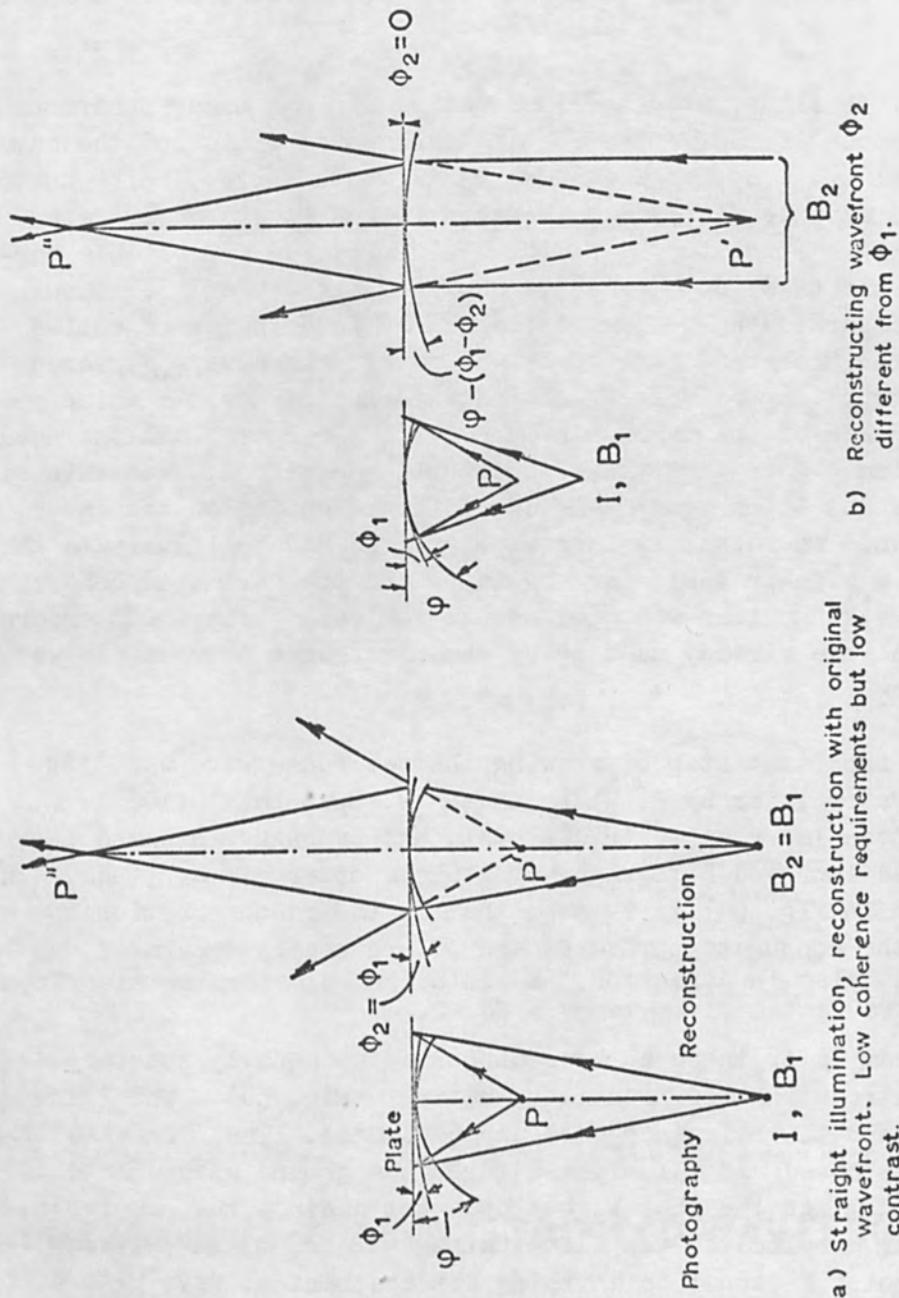
$$\text{Intensity } I = (A + B)(A^* + B^*) = (AA^* + BB^*) + A^*B + AB^* \quad (1)$$

where the asterisk means the complex conjugate. For simplicity let us assume that we process and print the photograph in such a way that the intensity transmission of the print becomes proportional to the square of I . Its amplitude transmission is then proportional to I itself. (We have seen previously that this does not matter much, but it greatly simplifies the explanation.) If now we take this print, and illuminate it with the original wave A alone, the transmitted (or reflected, or scattered) light in the plane of the plate will be proportional to

$$\begin{array}{l} A \cdot I = \quad (AA^* + BB^*)A \quad + \quad (AA^*)B \quad + \quad A^2B^* \\ \text{Illuminating wave} \quad \quad \quad \text{Reconstructed wave} \quad \quad \quad \text{Twin wave} \end{array} \quad (2).$$

If A is a plane wave, its intensity AA^* is a constant all over the hologram plane, and the second term in Eq. 2 is an exact reconstruction of the original wave, because by Huygens' Principle if two waves are identical at a surface, they are identical everywhere. The condition $AA^* = \text{const.}$ can be considerably relaxed; almost any background wave will do whose intensity varies slowly compared with the rapid phase change of B.

The last term, the "twin" or "conjugate" wave is better discussed in geometrical terms, as shown in Fig. 3. 3a is the simplest holographic arrangement, which I first realized in 1947. A pinhole I is the illuminator; this is at the same time the source of the coherent background B_1 . The spherical background wave ϕ_1 and the wavefront scattered by the point object P fall together on the plate, where they form a system of interference fringes. If after development the plate is



a) Straight illumination, reconstruction with original wavefront. Low coherence requirements but low contrast.

b) Reconstructing wavefront ϕ_2 different from ϕ_1 .

Figure 3. Wavefront Reconstruction, Gabor 1948.

replaced into its original position and is illuminated with the background wave alone, one wave diffracted by the hologram will appear to issue from the original point P', the other from its optical conjugate with respect to the reconstructing wavefront $\phi_2 = \phi_1$, considered as a spherical mirror.

Fig. 3b illustrates the interesting optical transformations which are produced if the reconstructing background ϕ_2 is not the same as the original ϕ_1 . In the case shown, B_2 is at infinity, the reconstructing reference wave is plane, and the twin images P' and P" are mirror-symmetrical with respect to the plate. This produced appreciable perturbations in the early holograms, as shown in Figs. 4 and 5. Though one tried to separate the conjugate images, by focusing an objective on one only, the "coherent wake" from the twin image always appeared in the reconstruction and caused for instance the dark spots which appear between the spokes of the protractor in Fig. 5. Of course all one needed to separate them was to skew the reconstructing wavefront, but this simple device was not at our disposal before the invention of the laser. In order to avoid inordinately long exposures we had to illuminate the pinholes with a high-pressure mercury lamp, and the chromatic coherence of any of its strong lines was good enough for only perhaps 200 orders, and 200 orders were already used up by the divergence between the wavefronts ϕ_1 and ϕ_2 .

The important step of skewing the reference wave was taken only fifteen years later by E. N. Leith and J. Upatnieks (1963), who had a helium-neon laser at their disposal, and by using this and skew reference waves improved holography by a whole order of magnitude. This is illustrated in Fig. 6a; it is seen that in the reconstruction the beams issuing from the conjugate points P' and P" are neatly separated, not only in position but also in direction, and both can be separated also from the reference wave; the first term in Eq. 2.

A year later these authors announced an equally spectacular progress by introducing diffused illumination (Fig. 6b). The illuminating wave and the reference wave are separated. The object (a transparency in this case) is illuminated through a ground glass plate. This plate randomizes the phases, but does not destroy the coherence, because however complicated the illuminating wavefront, the wavelet issuing from any point P cannot be anything but a spherical wave with a definite phase. However, as a consequence of the diffuse illumination the phase of the wave issuing from any point of the object is in no simple relation to the phase of the wave from a neighboring point. But this makes no difference in the reconstruction when the points are again separated so that only their intensity matters, not their phase. On the other hand the diffuse illumination has a very important advantage. In all the previous examples the information from an object was spread out in the hologram over an area equal to its diffraction figure. With



HOLOGRAM



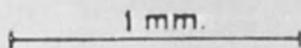
ORIGINAL



RECONSTRUCTION

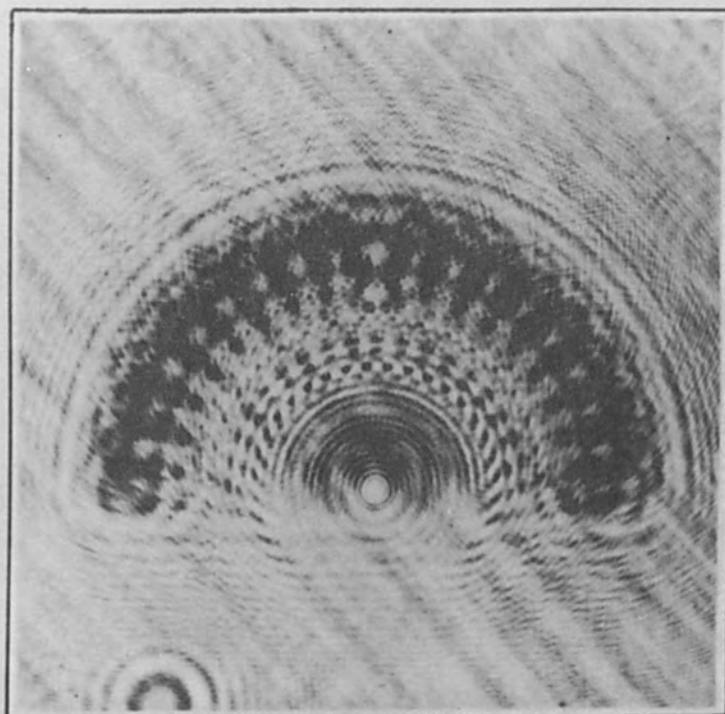
MICROSCOPY BY RECONSTRUCTED WAVEFRONTS

·003 mm. PINHOLE, $\lambda = 4358\mu$, 10 mm. OBJECT DISTANCE
180 mm. PLATE DISTANCE, 0.125 NUMERICAL APERTURE

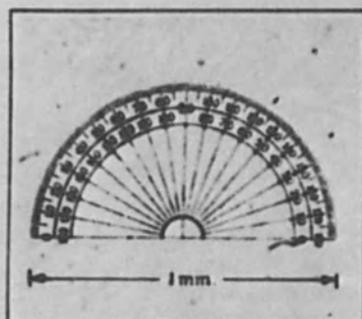


44589
22749

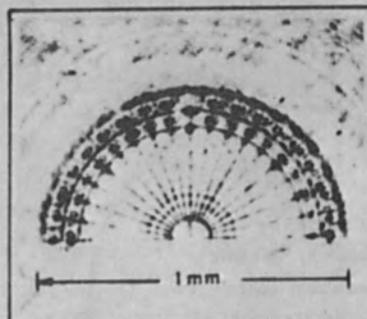
Figure 4. Original, Hologram and Reconstruction. D. Gabor, 1948.



HOLOGRAM



DIRECT PHOTOGRAPH



RECONSTRUCTION

MICROSCOPY BY RECONSTRUCTED WAVEFRONTS

·003 mm. PINHOLE, $\lambda = \cdot4358 \mu$, 10 mm. OBJECT DISTANCE

180 mm. PLATE DISTANCE, 0·125 NUMERAL APERTURE

(The original microphotograph of the protractor on Kodak Maximum Resolution plate was supplied by Messrs. Kodak Ltd.)

44 661
C. 1948

Figure 5. Original, Hologram and Reconstruction. D. Gabor, 1948

wide-angle diffused illumination the information can be spread out almost evenly over the whole area of the photographic plate.

The progress in this respect between the setups illustrated in Figs. 6a and 6b are shown in Fig. 7. The regular interferences in the hologram at the left are obtained between two plane waves. The middle figure is the hologram produced by the plane, skew reference wave and an irregular wave issuing from an object. It appears in the form of mixed amplitude and phase modulation of the first pattern. The third figure shows the effects of diffuse illumination again with the same plane, skew reference wave. The phases are completely mixed up, and the hologram has become "noise-like".

This is even more dramatically illustrated in Messrs. Leith and Upatnieks' hologram in Fig. 8 and its brilliant reconstruction in Fig. 9. It appears almost unbelievable that anything orderly could be contained in this randomness, and yet, the hologram in Fig. 7 need be only illuminated by a skew plane wave (which need not at all have the original direction or even wavelength) to bring out the structures encoded in it. In the same paper (1964), E. N. Leith and J. Upatnieks demonstrated the first good reconstructions of three-dimensional objects. Holography was of course three dimensional from the start, but it required the diffuse illumination to wipe out the disturbing "coherent wake" which every object point left even in planes far before or behind it.

III. APPLICATIONS

These methods are of course highly interesting from the point of view of the communication engineer, because the random encoding in Fig. 8 is far beyond anything achieved in electrical communications. Two questions arise: The first is, what have we to learn from the mathematical principles embodied in holography? The second is, how can we profit from the holographic techniques themselves?

My answer to the first question is short. There is nothing in the mathematics of holography of which communication-mathematicians have not been aware before. In brief, holography is an example of encoding by a linear integral transform, such that the resolving kernel is known and available. (A treatment of the type of integral transform most important in holography which I called "shadow transforms" and which I now rather would call "Fresnel transforms" is contained in my paper of 1949.) But the transformations which are so easily realized with coherent light are extremely difficult to achieve by circuitry. Communication engineers have for some time been aware of the problem, and have thought out ingenious methods of quasi-random coding, which can be comparatively easily realized. An example of this are the transformations of data

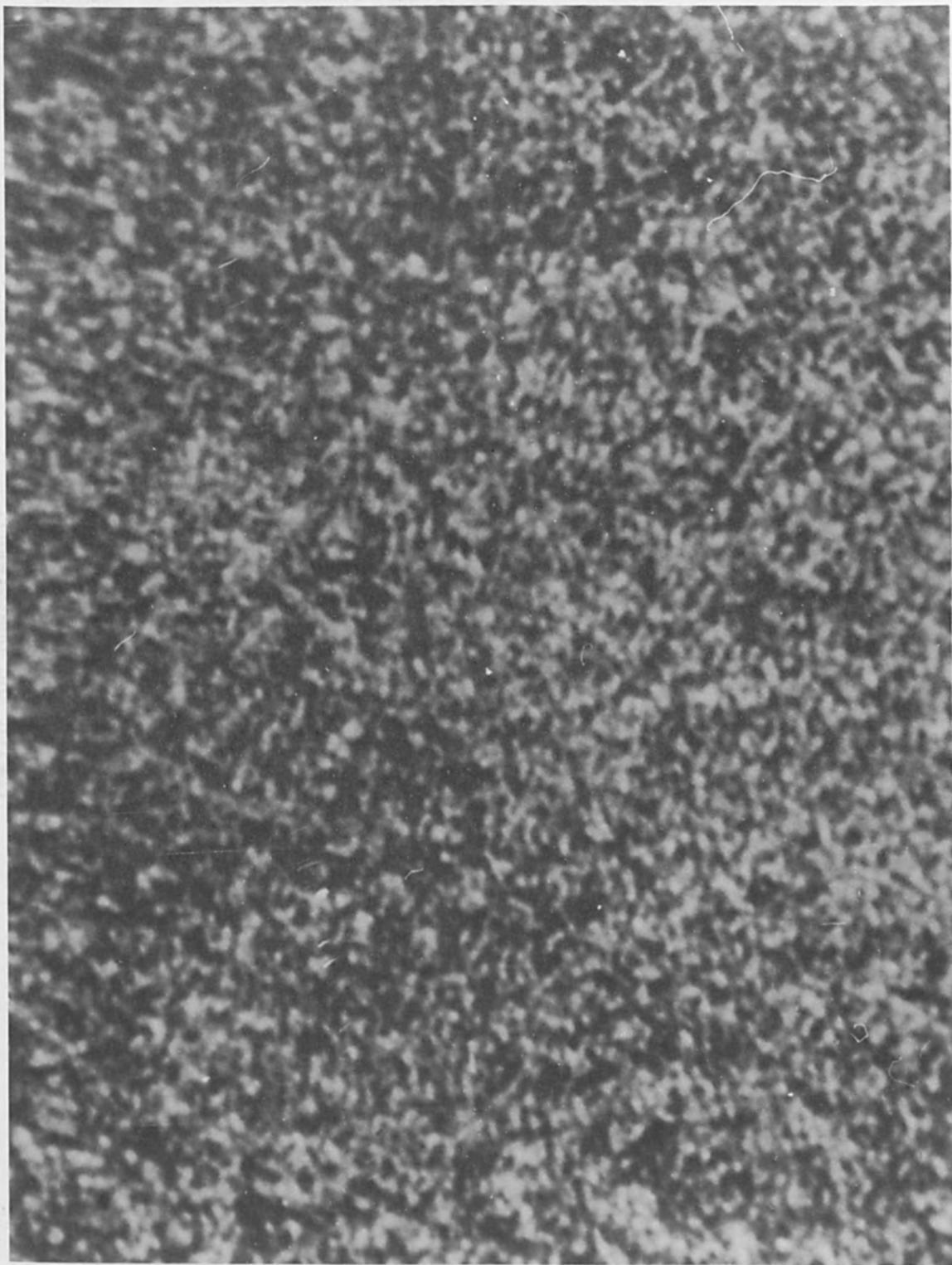


Figure 7. Hologram with Diffusing Illumination

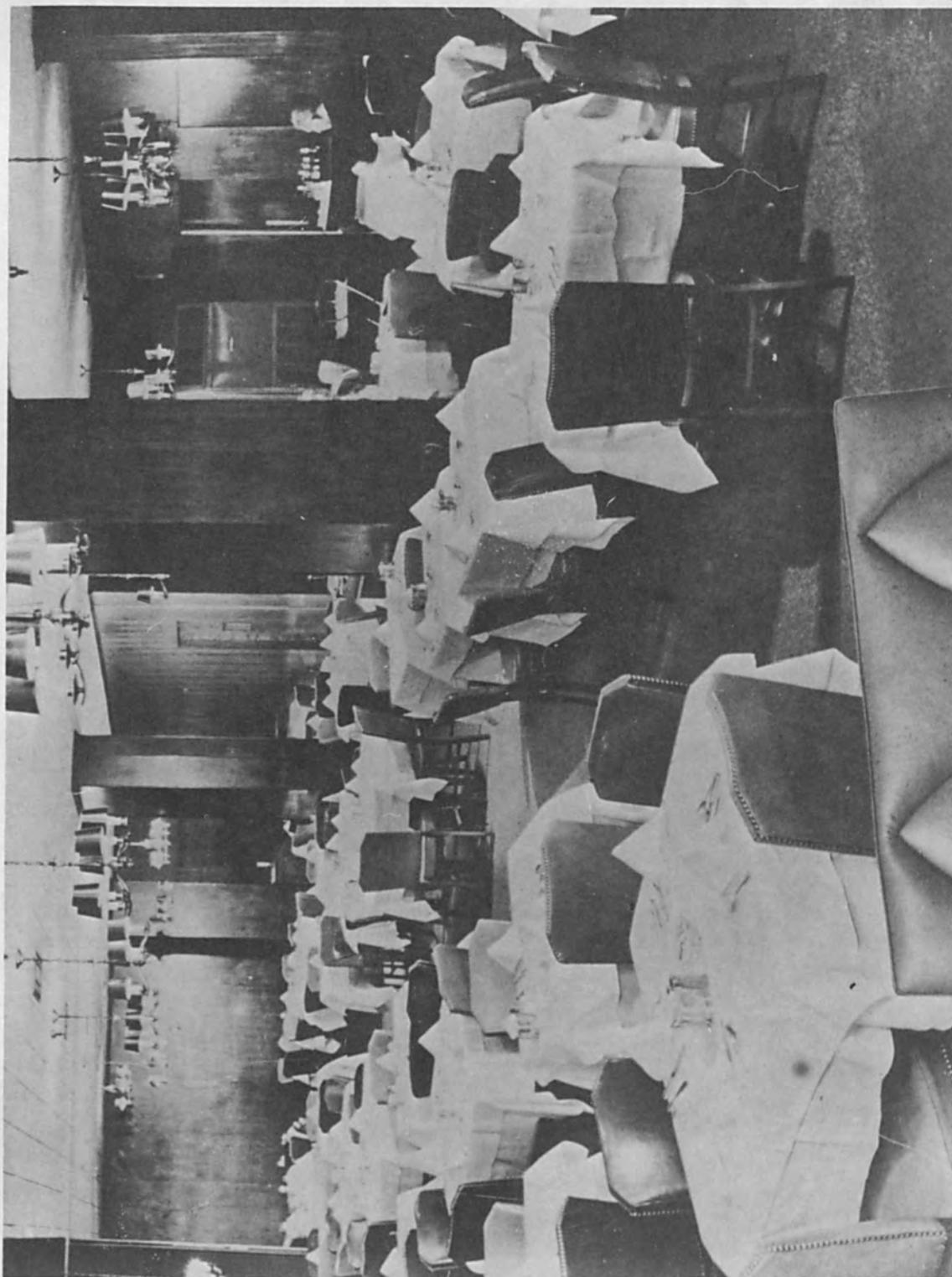


Figure 8. Reconstruction from a Hologram
(Leith and Upatnieks).

by means of Hadamard matrices, of which an example is shown in Fig. 9 . A 16-tuple of data is transformed by this matrix into another 16-tuple, which has no likeness to the original. A single datum (a "delta function") is transformed into one of the rows (or columns) and takes on a nearly noise-like character. The distinctive property of the Hadamard matrices is that the terms are either +1 or -1, so that no signal power is lost, moreover they are orthogonal and symmetrical, and hence self-reciprocal. A signal encoded by a Hadamard matrix is decoded by the same matrix. (See, e.g., W. W. Peterson, Error Correcting Codes, MIT Press, Cambridge 1961.) In my opinion Hadamard matrices are more likely to solve the problems of electrical communication than the transforms of holography. They are easy to realize (especially by digital techniques) and they produce equally noise-like transmission signals.

Coming now to the second question, what the communication engineer has to learn from holographic techniques, this is less easy to answer. To my regret, I find it easier to be negative than positive. Several notices have appeared in the press, that holography may be the key to three-dimensional television. I consider these as rather irresponsible. The trouble with three-dimensional holograms is that they are too good; they contain a whole dimension of redundant information. We have eyes only horizontally spaced, not vertically, but one cannot forbid a hologram to contain also the entirely redundant vertical information about depth. Leith and Upatnieks were right to point out the enormous extension in bandwidth which would be required to transmit three-dimensional holograms. The redundancy in the case of three-dimensional objects far outweighs the advantage of almost ideal coding, with maximum entropy.

Nevertheless, it cannot be gainsaid that the entirely random coding in holograms may be an attractive solution to waveband saving in picture transmission. (I wish to exclude television, because any holographic decoding would be far too complicated and expensive for the home set.) In the attempts so far made for waveband saving by variable speed scanning the first step was easy; all one had to do was to make the scanning spot run fast over the empty spaces. The difficulty came in making the information "queue up" to be transmitted at a uniform rate; and this part of the problem is solved by holography almost ideally. What we require now is devices which will allow transforming the holograms transmitted by television techniques into transparencies which can be reconstructed in coherent light - or else digital devices which of course can solve any problem. But I want to point out that it is hard to beat holography, which can perform, say, a hundred million summations each over a hundred million data, literally with the speed of light.

+	-	-	-	-	+	+	+	-	+	+	+	-	+	+	+
-	+	-	-	-	+	-	+	+	+	-	+	+	+	+	+
-	-	+	-	-	+	+	-	-	+	+	-	+	+	-	+
-	-	-	+	-	+	+	+	-	+	+	+	-	+	+	-
-	+	+	+	+	+	-	-	-	-	+	+	+	+	+	+
+	-	+	+	+	-	+	-	-	+	-	+	+	+	+	+
+	+	-	+	+	-	-	+	-	+	+	-	+	+	-	+
+	+	+	-	-	-	-	-	+	+	+	+	-	+	+	-
-	+	+	+	+	-	+	+	+	+	-	-	-	-	+	+
+	-	+	+	+	+	-	+	+	+	-	-	-	-	+	+
+	+	-	+	+	+	+	-	+	+	+	-	+	+	-	+
+	+	+	-	-	+	+	+	+	+	+	-	-	-	-	+

Figure 9. Example of a 16 x 16 Hadamard Matrix.

At the present time, I am told, over 100 (some people say 500!) laboratories are working on holography and related applications of coherent light in the United States. The expectations are very high that something communally important will come out of it; there is even some talk about new "billion dollar industries". I consider this as somewhat irresponsible talk, but I must admit that already much more has come out of holography than I could ever have imagined nineteen years ago, when I first conceived the idea, and perhaps I am as shortsighted now as I was then.

I did not invent holography at all with a view to a new light-optical process. My intention was to take a hologram with an electron microscope, and then to reconstruct the electron image with light, to see atoms. It is just this application of holography which has not progressed since 1953, when Dr. M. E. Haine and his collaborators in the AEI Research Laboratory at Aldermaston made a first spirited attempt toward it. This work brought many interesting results, but we could not see atoms, because it turned out that at that time the electron microscope was not limited by optical imperfections, but by secondary effects such as vibrations, stray magnetic fields, stage creep and object contamination. In the thirteen years since that time, patient development work has eliminated these secondary sources of trouble and now would be the time to start on it again.

The impressive progress in the last three years was in the main the exploitation of the undreamt-of possibilities opened up by the invention of the laser. I have already mentioned the neat separation of the reconstructed image made possible by skew reference beams and diffuse but coherent illumination. Another exciting possibility was soon realized; producing interferences between two light waves which have not existed at the same time. One can express this also as "an interferometer with one branch only" which sounds like the same sort of nonsense as "the clapping of one hand". But in fact, using a reference beam A one can "freeze" a wave B into a photographic plate and then, again with A, another wave C, and afterwards one can revive B and C together by means of A.

This idea was utilized by Powell and Stetson in Michigan and a little later, independently, by Burch at the N.P.L. for measuring small deformations of objects. One obtains Newton rings between the two surfaces, just as if one had a precise transparent replica of the original. The most astonishing feature of this method is that one can observe deformations of the order of a few fringes of an object which is optically rough, for instance turned on a lathe.

Figure 10 is a remarkable picture of a flying bullet taken by Heflinger, Wuerker and Brooks of TRW, Los Angeles, in which almost all the advantages of laser holography are combined. It was taken with a 30 nanosecond flash of a ruby laser, the bullet moved perhaps 1/1000 of an inch during the exposure. But it is not a simple photograph: it is reconstructed from a hologram-interferogram. A first flash (a plane wave) was sent into the clear background, then a second, simultaneously with the bullet. It is therefore an interferogram of the plane wave and the one disturbed by the bullet. But as it is also a hologram, it can be explored at leisure in three dimensions, cross section after cross section.

V. CONCLUSIONS

I have already briefly mentioned the greatest progress of the last year: "Lippmann-Holography" in natural color with the reconstruction visible in ordinary white light. This is very likely to lead to commercial applications as 3-color 3D pictures are an attractive proposition. All the more as reconstructions in laser light, though they look okay in photographs, look very disagreeable when viewed "live"; they seem to be crawling with ants or flies. (A not yet quite understood effect of coherent light.)

Of coming applications of holography, I want only to mention character recognition and vision by sound or ultrasound. These are as good as certain to come, but in a field which has already brought so many surprises and which is now so energetically exploited by hundreds of able workers I dare not take prophecy any further.

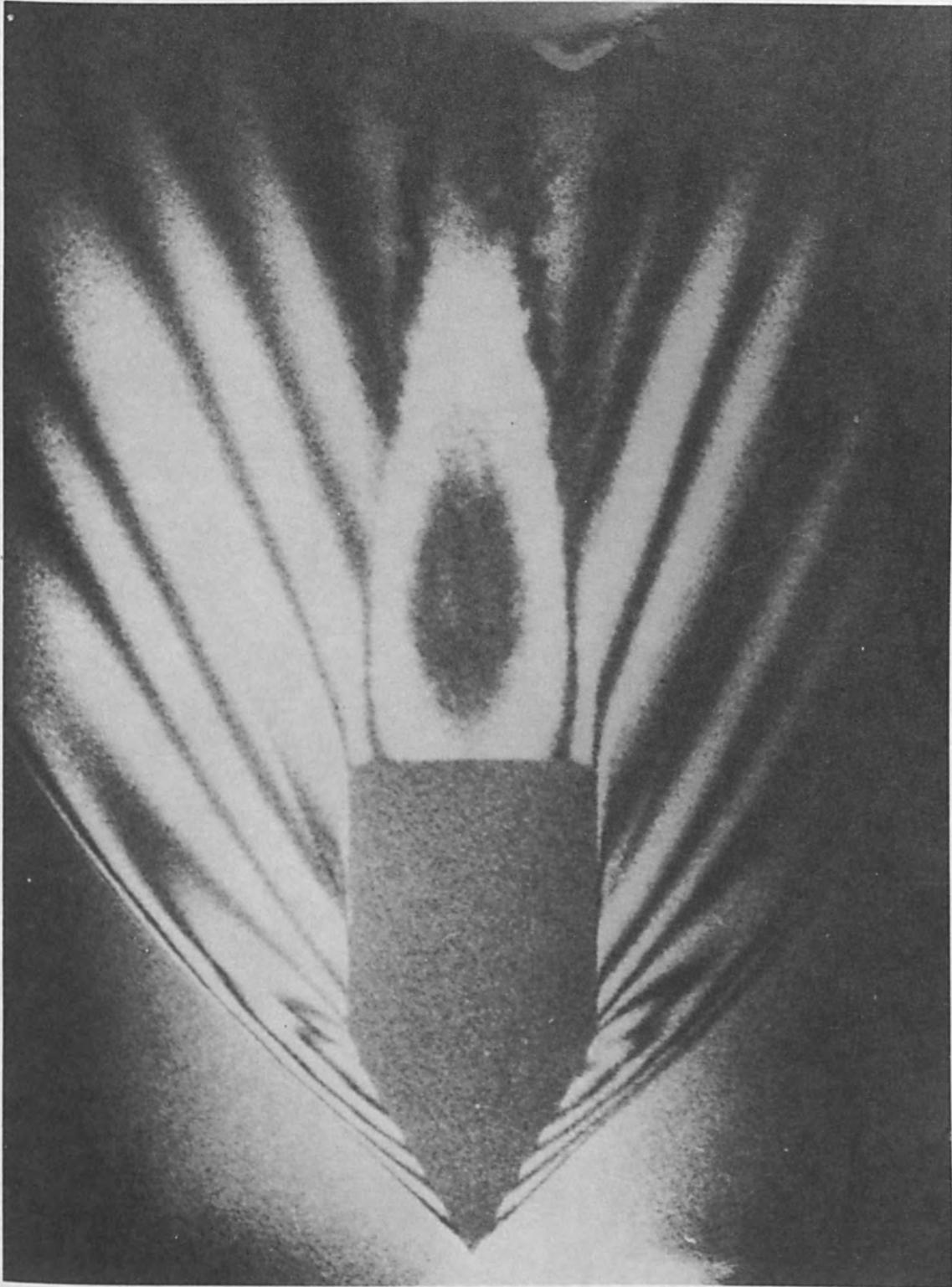
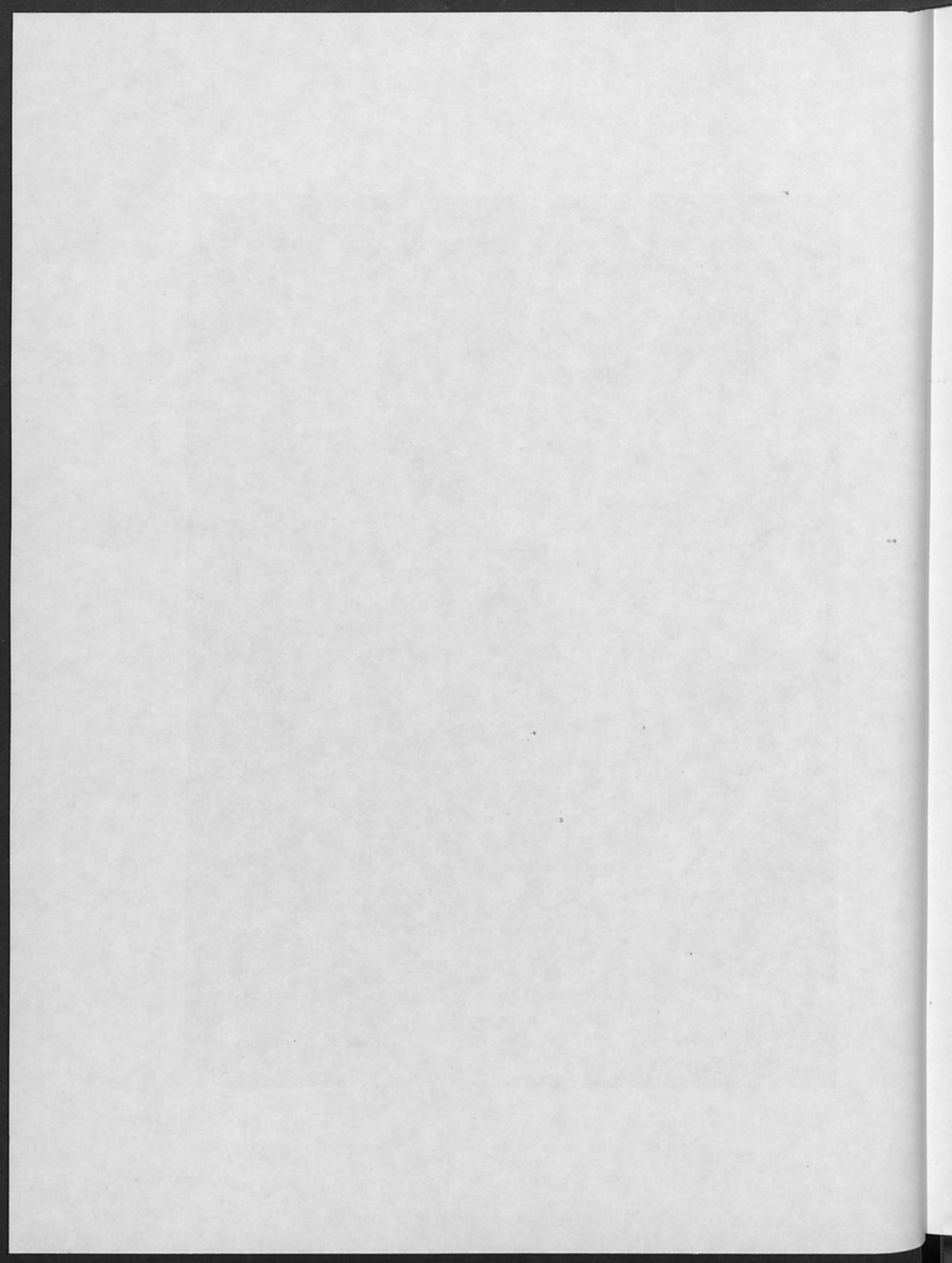


Figure 10. Holographic Interferogram of a Bullet. L.O. HeFlinger, R.F. Wuerker, and R. E. Brooks, (TRW).



RECENT RESULTS IN HOLOGRAPHY

by

E. N. Leith

Assoc. Prof. of Elec. Eng. and

Head, Optics Group of the Radar and Optics Lab.

Institute of Science and Technology

The University of Michigan

Ann Arbor, Michigan

ABSTRACT

Recent results obtained in our laboratory are described. These include a study of holographic storage on thick media, Lippman-type holograms that reconstruct in white light, imagery through diffusing media, a correlation technique for producing holograms with incoherent light, a method for implanting contour lines on holographic images, and the construction of holograms in full color.

I. INTRODUCTION

Since the recent resurgence of interest in holography, the Optics Group of the Radar and Optics Laboratory has conducted a broad research effort in this area, designed both to enlarge the scope of the field through basic research, and to find applications. In this paper we present a survey of our most recent work; the work thus described is that of perhaps six to ten members of our research staff, to whom I am indebted for allowing me to discuss their work.

Included in this paper are the following topics:

1. Imagery through diffuse media
2. Holographic storage in three-dimensional media, including storage of multiple images and "Lippman" type holograms, which produce an image in white light.
3. Holograms made with monochromatic, spatially incoherent light, along with a technique for eliminating the bias buildup problem encountered with this kind of hologram.
4. The use of holographic contouring, a technique by which a hologram becomes a three-dimensional "contour map" of the object.
5. The production of multicolor holograms

II. HOLOGRAPHIC IMAGERY THROUGH DIFFUSING MEDIA*

Using holographic methods, it is possible to obtain imagery through a diffusing medium, such as ground glass, opal glass or a chunk of translucent mineral matter. An experimental arrangement is shown in Fig. 1, where the diffuser is placed between the object and the hologram; the hologram is otherwise made in the usual manner, with a reference beam impinging on the recording plate at some oblique angle.

From the hologram thus produced, the wavefront from the original object cannot be reconstructed by ordinary methods; the hologram reconstructs only that which can be seen from the plane where the hologram was made. Thus, an image of the diffuser is readily reconstructed, but not of the object, since this is obscured by the diffuser.

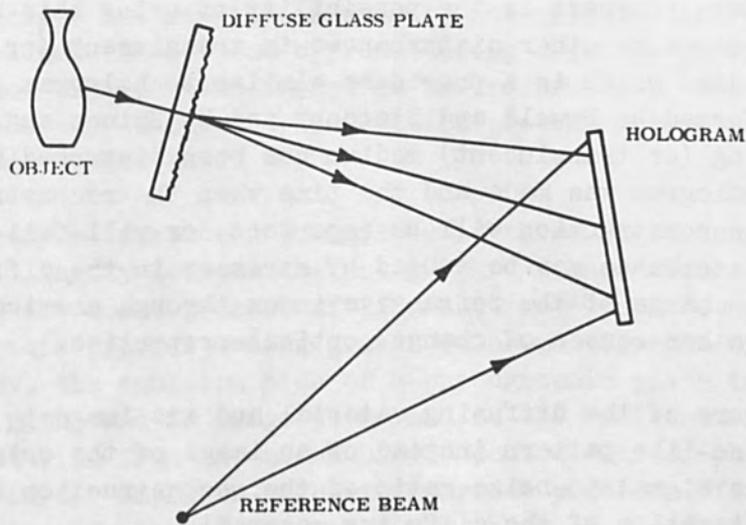
Suppose, however, that in the reconstruction process we place the diffuser in the path of the light emanating from the hologram, in such a position that the diffuser coincides with the hologram-produced real image of the diffuser. When this is done, an image of the original object is produced.

The explanation of the phenomenon is as follows: the diffuser can be regarded as a two-dimensional (or three-dimensional) phase function $e^{i\phi(x,y)}$, where ϕ is a random or noise-like function. The real image is phase conjugate to the object, i.e., has the form $e^{-i\phi(x,y)}$. When the diffuser and its real image superimpose, the phase effects of one are cancelled by those of the other, the diffuser effectively disappears, and the original subject, previously obscured by the diffuser, now is clearly produced.

The cancellation is incomplete unless the diffuser is brought into exact coincidence with its reconstructed image. Such alignment is difficult, since there are six degrees of freedom for the diffuser location (three each of translation and rotation). Also, the hologram must be illuminated with a beam having just the right divergence to yield a magnification of unity for the diffuser.

Since the wavefront from the original object can be reconstructed only by using the same diffuser that was used for making the

*This work was done as a collaborative effort with J. Upatnieks. It was first presented at the SPIE meeting in San Francisco, in August 1965, and later published in the SPIE Journal, October-November 1965.



STORAGE OF PICTURES BY ENCODED SIGNAL BEAM

Figure 1. Hologram System With Obstructing Diffuser.

A MAN'S REACH MUST EXCEED HIS GRASP OR WHAT'S A HEAVEN FOR

Figure 2. A Reconstruction Made Using the Obstructing Diffuser Setup.

hologram, we have here an interesting coding system.

Of more interest is the possibility of using this system for measuring stresses or other disturbances in translucent (or in general, non-opaque) media, which is a procedure similar to hologram interferometry, as performed by Powell and Stetson¹ and by Haines and Hildebrand.² If the diffusing (or translucent) medium has been disturbed between the time the hologram was made and the time when the reconstruction was made, the reconstruction will be imperfect, or will fail to materialize. This disturbance may be caused by stresses in the diffuser, thermal expansion, change of the refractive index through chemical or other means, or any other causes of changed optical properties.

Failure of the diffusing material and its image to coincide produces a noise-like pattern instead of an image of the original subject. Thus the signal-to-noise ratio of the reconstruction is a measure of the alteration of the diffusing material.

An example of a reconstruction obtained through a diffuser is shown in Fig. 2. The object was transparent lettering; the diffuser was a piece of ground glass. Note that the lettering has been recovered with good fidelity. In a further experiment, two pieces of ground glass were used in tandem, with about 1 inch separation. Currently, we are attempting to extend the experimental results to much thicker diffusers such as chunks of translucent mineral material.

We have also produced high-quality reconstructions when the diffuser is in the reference beam. This configuration is an example of a type described by van Heerden.³

NOTE: In private discussion with K. Pennington of Bell Laboratories, we have learned of similar work carried out at that organization.

III. HOLOGRAPHIC RECORDING IN THREE-DIMENSIONAL MEDIA*

Holographic theory, as developed originally by Gabor, assumed that the emulsion on which the hologram was recorded could be treated as a two-dimensional surface; all of the subsequent work (of which the authors are aware) on this subject proceeded on the same assumption, until Denisjuk⁴ and van Heerden⁵ described holographic recording in three-dimensional media.

*This material is a summary of a paper by E. Leith, A. Kozma, J. Upatnieks, J. Marks and N. Massey, which was presented at the Optical Society of America Meeting, March 1966 in Washington, D.C., and will shortly be published in Journal of Applied Optics.

Diffraction from three-dimensional structures has been widely studied in the form of diffraction of x-rays by crystals, and the basic concepts are readily adapted to holographic theory. Van Heerden approached the subject by considering crystals in which diffracting "color centers" are formed in the recording process; in the readout process the subsequent diffraction process recreates the originally-recorded wavefronts.

Denisyuk introduced holography in three-dimensional media by noting the similarity between the holographic process, as introduced by Gabor, and the color process of Lippman,⁶ developed late in the 19th century. In this process, which yields color photographs of excellent color fidelity, the emulsion side of a photographic plate is backed by a mercury bath, and an image is formed on the emulsion surface from light entering from the non-emulsion coated side of the plate. Reflection from the emulsion-mercury interface produces, for each wavelength component, a standing wave pattern which is recorded by the emulsion. Each recorded pattern then acts as an interference-filter resonant structure which, when illuminated with white light, reflects light of the same wavelength which produced the structure. The viewer sees, in reflected light, the image in its natural color. The emulsions used in Lippman photography must be capable of resolving with good contrast the standing wave pattern, and thus must resolve to within half the wavelength of light. The emulsion must, in addition, be sufficiently thick so that many fringes will be recorded, otherwise the Lippman photograph will lack selectivity, resulting in poor color rendition. Typically, the emulsion thickness of commercial high resolution plates is 10 to 20 microns, hence 40 to 80 standing wave periods are recorded. The similarity of the process to holography is apparent, and as shown by Denisyuk, one can substitute coherent light in the process, record the diffraction pattern of an object in the manner thus described, and thereby produce a hologram.

A holographic recording acquires the properties of a three-dimensional medium when the recorded fringe pattern has a spatial period comparable to the thickness of the emulsion. Since few of the commercial photographic emulsions have this kind of resolution, they do not give rise to holograms of the sort of concern here. Using Eastman Kodak type 649 emulsion, the authors have produced holograms in which the carrier frequency has been in the range of 200-2000 lines/mm; at these spatial frequencies the recordings must be treated as three-dimensional.

The major effect manifested by a thick emulsion is Bragg-angle diffraction, a phenomenon well known to x-ray crystallographers. The diffracted orders, other than the zero-order diffraction, are produced only for specific angles of incidence of the readout beam. Two conditions are imposed on the diffracted wave. First, the diffraction-

grating equation,

$$\sin \theta - \sin \phi = \frac{\lambda n}{d} \quad (1)$$

must be satisfied, where θ and ϕ are defined in Fig. 3, d is the grating spacing, λ the wavelength of the illuminating light, and n is the order of the diffraction. The second condition is readily explained by noting that the recorded fringes are now surfaces within the emulsion. The diffraction process becomes a process of reflection from these surfaces, and thus it must satisfy the fundamental condition, that the angle of incidence of illumination on the surface must equal the angle of reflection.

In the hologram construction, let the object wave be decomposed into a spectrum of plane waves, and consider a single component $u = a_1 e^{i\bar{K}_1 \cdot \bar{R}}$ where \bar{K}_1 is the propagation constant, $K_1 = k = \frac{2\pi}{\lambda}$, θ_1 is the angle of incidence of this wave, and \bar{R} is the vector from the origin to the point x_1, y_1 . Hence, $\bar{K}_1 \cdot \bar{R} = x \cos \theta_1 + y \sin \theta_1$. Similarly, let the reference beam (a planar wave) be $u = a_r e^{i\bar{K}_2 \cdot \bar{R}}$. Neglecting refraction effects as the waves enter the emulsion, we have, for the intensity,

$$I = |u + u_r|^2 = 2 + 2 \cos \phi \quad (2)$$

where $\phi = (\bar{K}_1 - \bar{K}_2) \cdot \bar{R}$.

The surfaces of constant phase set up within the emulsion are defined by

$$(\bar{K}_1 - \bar{K}_2) \cdot \bar{R} = k[x(\cos \theta_2 - \cos \theta_1) + y(\sin \theta_2 - \sin \theta_1)] = \text{constant.} \quad (3)$$

These surfaces, it can be shown, lie along the bisector of the propagation vectors \bar{K}_1 and \bar{K}_2 ; thus, the recorded waves satisfy the reflection law for the surfaces they create.

In the readout process, satisfaction of both the grating equation and the reflection law is required, thus four situations arise, as illustrated in Fig. 4. θ and ψ are the angles of the readout and diffracted waves, respectively.

- (a) The readout wave can duplicate the reference wave, whereupon the diffracted wave duplicates the object wave [$\theta = \theta_2, \psi = \theta_1$]. This produces a virtual image.
- (b) The readout wave can enter in the reverse direction, $\theta = \theta_2 + \pi$, whereupon the diffracted wave occurs at an angle $\theta_1 + \pi$, producing a real image.

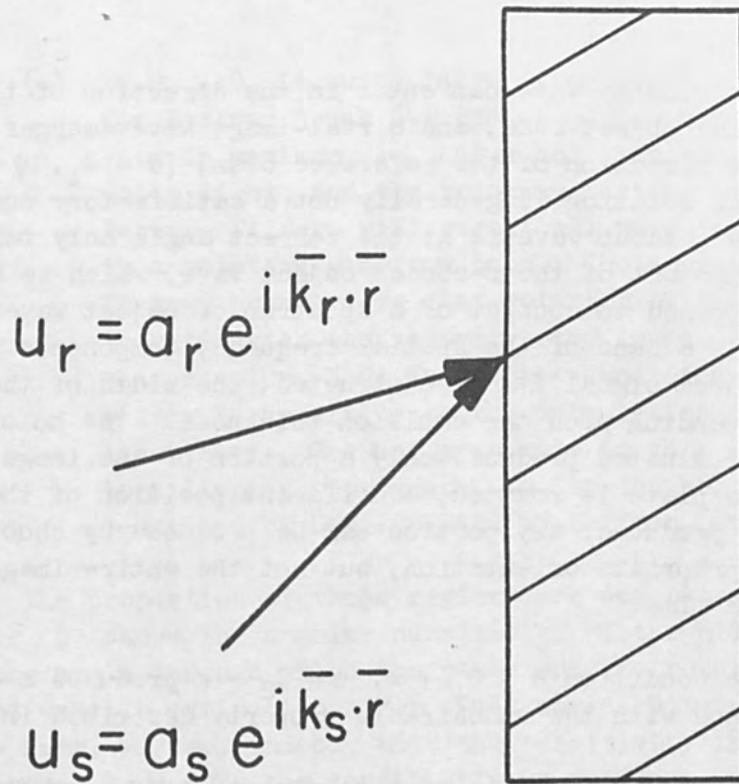


Figure 3. Recording Within a Thick Emulsion.

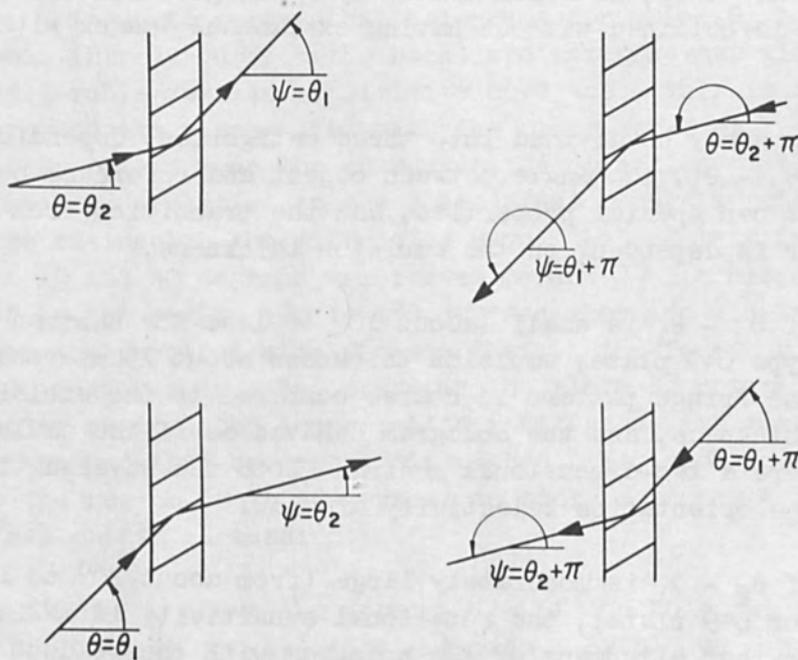


Figure 4. Four Ways to Read Out of a Thick Medium.

- (c) The readout wave can enter in the direction of the original object beam, and a real-image wave emerges, having the direction of the reference beam, $[\theta = \theta_1, \psi = \theta_2]$. This solution is generally not a satisfactory one, for the readout wave is at the correct angle only for one component of the recorded object wave, which we have supposed to consist of a spectrum of object waves. Thus, only a band of the spatial frequency components of the object signal is reconstructed, the width of the band depending upon the emulsion thickness. The hologram so illuminated produces only a portion of the image. As the plate is rotated, a different position of the image is produced; any portion can be produced by choosing the appropriate orientation, but not the entire image at one time.
- (d) The condition $\theta = \theta_1 + \pi, \psi = \theta_2 + \pi$ produces a virtual image with the undesirable property described in (c).

When the readout wave is of a wavelength different from that used for making the hologram, a mismatch may occur which prevents the generation of diffracted waves. Thus, holograms may be made using several wavelengths, and in the reconstruction the various read-in wavelengths will produce diffraction from only the fringe pattern generated by that wavelength. Thus, as demonstrated by Pennington and Lin⁷, color holography is obtained without having extraneous images with which to contend.

Holograms may be divided into three categories, depending on the value of $\theta_2 - \theta_1$, the angle between object and reference beam. Each type has its own special properties, and the transition from one kind to the other is dependent on the emulsion thickness.

- (a) If $\theta_2 - \theta_1$ is small (about 10° or less for Eastman Kodak type 649 plate, emulsion thickness about 15 microns), the fringe pattern is coarse compared to the emulsion thickness, and the hologram behaves as if the emulsion were a two-dimensional medium. Both the wavelength and the orientation selectivity are low.
- (b) If $\theta_2 - \theta_1$ is moderately large (from about 20° to 120° for 649 plate), the rotational sensitivity is maximum, so that alignment of the hologram with the readout beam is critical. Wavelength sensitivity is fair; thus one may obtain suppression of wavelengths that are separated by a few hundred angstroms, as for example, in the color hologram experiments of Pennington and Lin.

- (c) If $\theta_r - \theta_s$ is quite large (from about 135° to 180° for Eastman Kodak 649 plates), the wavelength sensitivity is maximum, so that a hologram can be observed in white light, and the hologram, acting now like an interference filter, will reject all wavelengths except those in a relatively narrow band. The angular sensitivity is low, however, so that rotation of the hologram fails to extinguish the reconstructed image, but causes it to change color, just as the passband of an interference filter is sensitive to the orientation of the incident light beam. The hologram made in this manner is similar to a Lippman photograph, and to the hologram experiments described by Denisjuk and by Hoffman, et. al.⁸

The properties of these regions are demonstrated in Figs. 5 and 6. Figure 5 shows the angular sensitivity of the plate. By this, we mean the angle through which the plate must be rotated to drive the diffracted wave to extinction. For small angles between signal and reference beam, we see, indeed, that the sensitivity is low. At higher angles, around $\theta_r = 90^\circ$, the sensitivity is greatest, while beyond this, the sensitivity decreases. Hence, we can store a multiplicity of pictures and read these out at different Bragg angles. This is shown in Fig. 7, in which a sequence of five pictures has been stored. These may be read out in succession, by rotating the hologram while it is illuminated.

It would appear that holograms for which $\theta_r > 90^\circ$ (the reference beam introduced from the back) are satisfactory also, since there is considerable angular sensitivity here too. This is not the case, however, and the reason lies with the other curve of Fig. 5, the $\Delta\psi$ curve. By $\Delta\psi$, we mean the change in the direction of the diffracted wave (measured relative to the plate) that occurs as the plate is rotated so as to extinguish the diffracted wave. We see that in the region between 10 and 40 degrees, the curves $\Delta\theta$ and $\Delta\psi$ lie close together, whereas in the region 120 to 180 degrees they are widely separated. This has the physical significance that in the former region the diffracted image remains stationary as the plate is rotated, whereas in the latter region, the image shifts position as its intensity builds to maximum and then becomes extinguished. It is obviously preferable to use the region 10 to 40 degrees to store a sequence of pictures and read them out in succession.

Figure 6 shows the sensitivity of the hologram to a change in the wavelength of the illuminating beam. By $\Delta\lambda$, we mean the change in wavelength needed to drive the diffracted beam from its maximum value to extinction. This sensitivity is greatest where the reference beam is introduced from the back. For 649F plate, the sensitivity is about 80 angstroms.

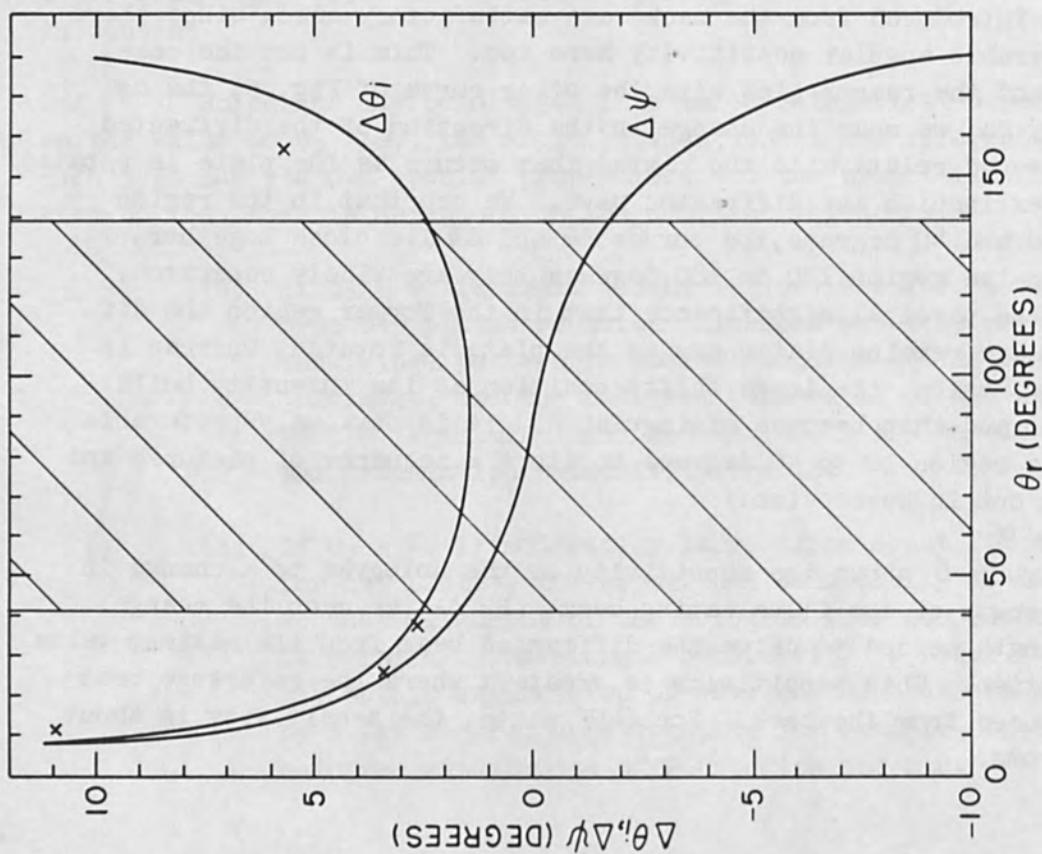


Figure 5. Angular Sensitivity of 649 F Plate, as a Function of the Angle of Incidence of the Reference Beam.

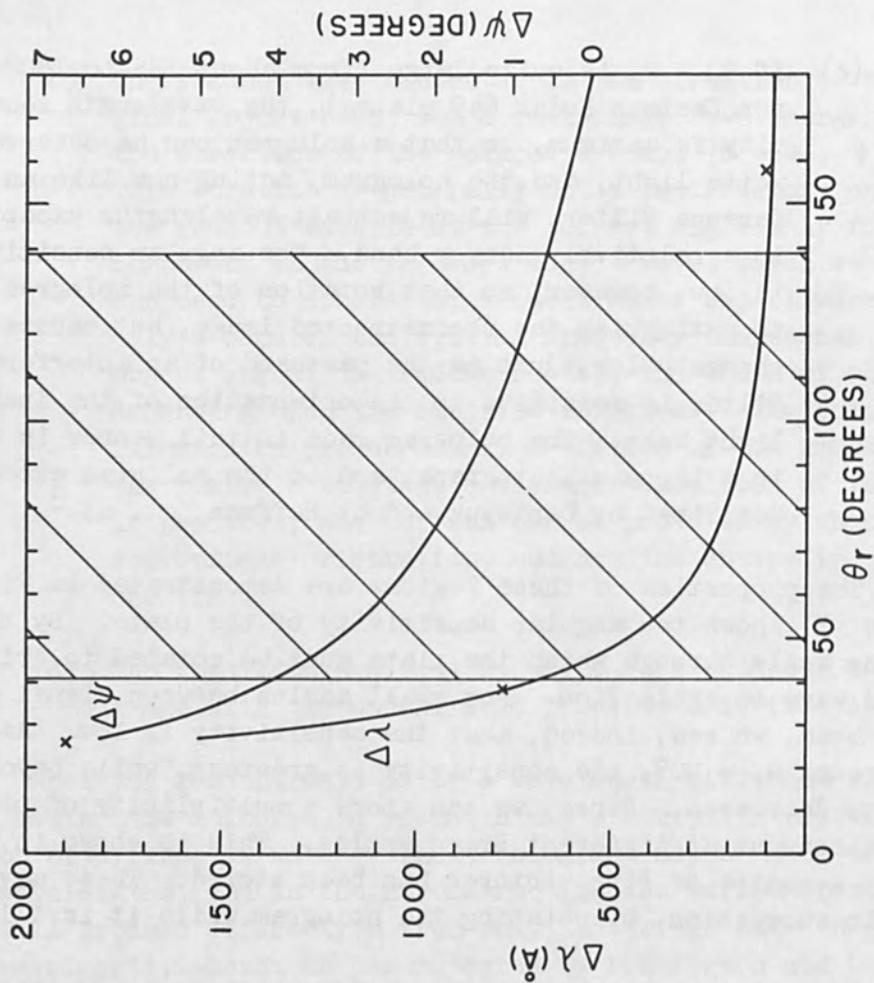


Figure 6. Wavelength Sensitivity of 9F Plate, as a Function of the Angle of Incidence of the Reference Beam.

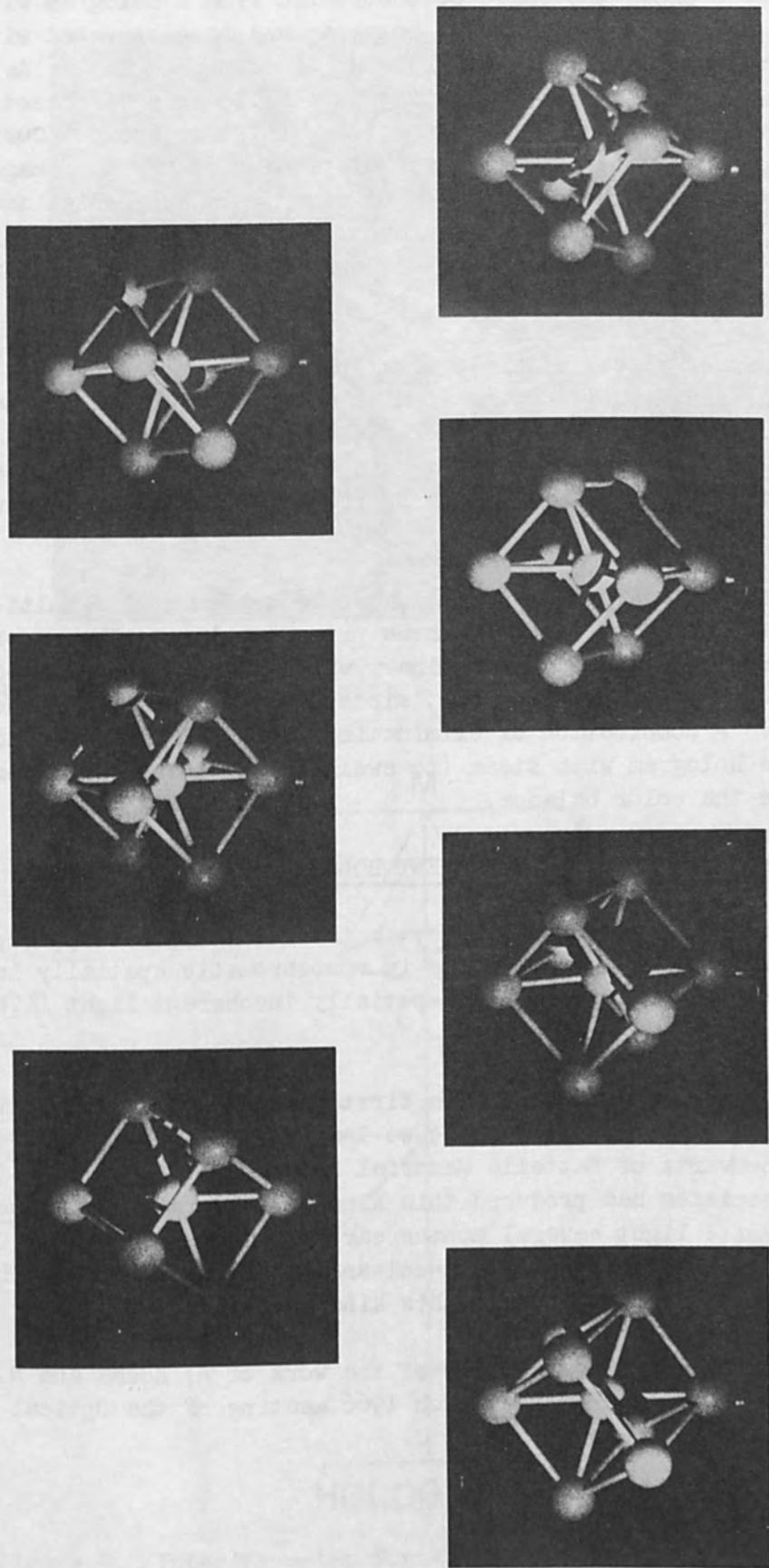


Figure 7. Sequence of Five Superimposed Pictures, Read Out With Bragg Diffraction While Rotating the Hologram.

Figure 8 shows the image reconstructed from a hologram with a back-introduced reference beam. The hologram was reconstructed with a white light point source, viz., a zirconium concentrated arc.* As the plate is rotated, and the incident light strikes it at a different angle, the color of the image changes, varying from violet to green. Curiously, the image does not reconstruct in red; this is due to the shrinkage of the emulsion during the developing and fixing process, which amounts to about 15%. By eliminating the fixing process, the shrinkage is reduced to about 3%, and the image then reconstructs over the color range green to orange-red.

Holograms placed in white light produce inferior images: the diffraction image must compete with an ambient background formed by light reflected from the surface of the plate, and the image lacks sharpness due to the inadequate wavelength selectivity of the plate. The latter fault could be cured by using thicker emulsions, which will increase the wavelength selectivity.

Holograms of the Lippman type can be produced with multi-color light, for example, light in three primary colors. The reconstruction will occur in full color in ordinary white light. The emulsion shrinkage problem is severe, however, since it causes a loss of the color balance. A combination of eliminating the fixing process and moistening the hologram with steam (to swell the emulsion) will momentarily restore the color balance.

IV. BIAS LEVEL REDUCTION OF INCOHERENT HOLOGRAMS**

Adam Kozma and Norman Massey, both of our laboratory, reported on a method of constructing holograms in monochromatic spatially incoherent light.⁹ Holograms made with spatially incoherent light (i.e.,

*White light holograms were first produced in our laboratory by J. Upatnieks. Shortly after this, we learned from private conversation with C. Schwartz of Battelle Memorial Institute (Nov. 1965) that he and his associates had produced this kind of hologram and reconstructed its image in white light several months earlier. More recently (Optical Society Meeting, Washington, D. C.), we learned that G. Stroke and his associates have similarly produced this kind of hologram.

**This section is a summary of the work of A. Kozma and N. Massey, and was presented at the March 1966 meeting of the Optical Society of America in Washington, D. C.

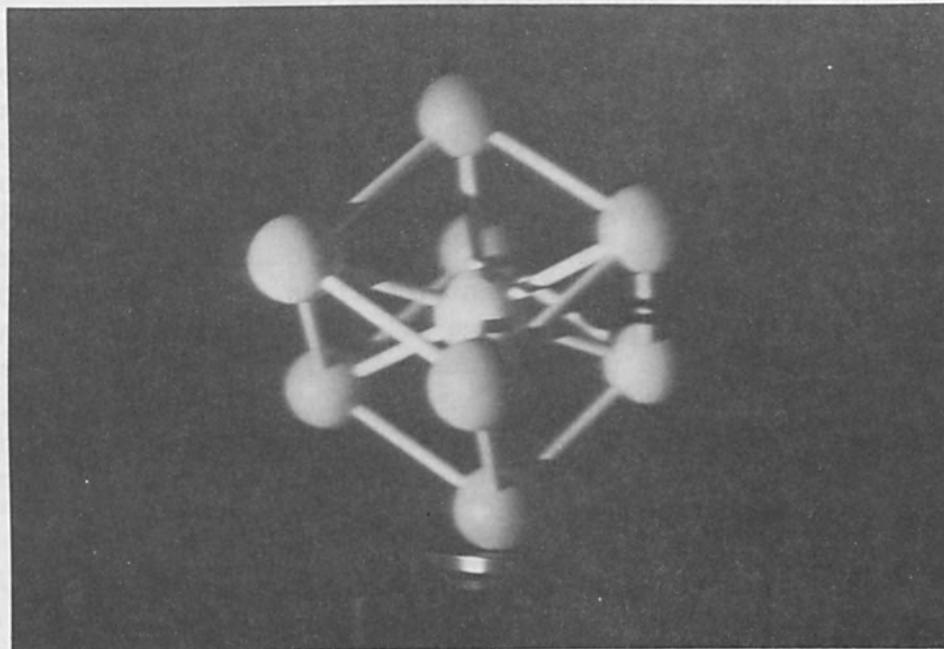


Figure 8. Image Reconstructed with White Light, from a Hologram Made with a Reference Beam Introduced from the Back.

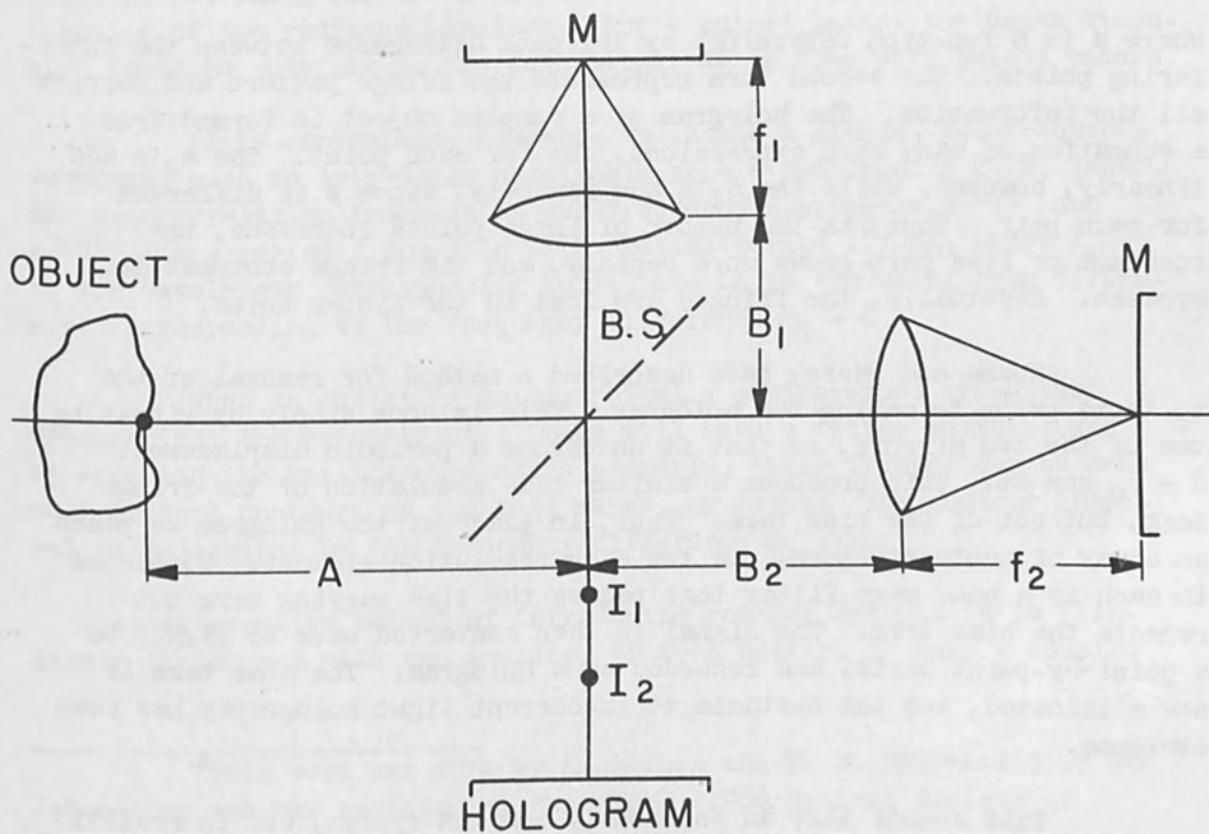


Figure 9. Interferometer for Making Holograms Using Incoherent, Monochromatic Light.

light from a broad source) were first described by Mertz and Young,¹⁰ and have more recently been investigated by others.

Several similar methods for producing incoherent holograms were described by Cochran and Lohmann; all involve the basic idea of forming two images, and interfering these images. One system for doing this is shown in Fig. 9. Light from the object passes through a beam splitter, and is divided into two portions, each portion passing through a lens to a mirror, and back again through the lens. Each portion forms an image, but in different locations. The essential point here is that, while within each image all the image points are incoherent with each other, each point of one image is coherent with the corresponding point of the other image. Thus, at the plane labeled H, each such coherent pair of points produces an interference pattern, and the recorded summation of all the patterns constitutes a hologram, from which an image can be reconstructed.

This works well for a simple object, consisting of only a few points, but breaks down as the object becomes more complex. The reason for this is that the light intensity from each coherent point pair is of the form

$$a_0 + a_1 \cos \phi(x,y)$$

where ϕ is a function determined by the path difference between the interfering points. The second term represents the fringe pattern and carries all the information. The hologram of a complex object is formed from a summation of many such expressions, one for each point. The a_0 's add linearly, however, while the a_1 's add randomly, since ϕ is different for each pair. Thus, as the number of image points increases, the constant or bias part grows more rapidly, and the fringe contrast decreases. Eventually, the fringes are lost in the system noise.

Kozma and Massey have described a method for removal of the a_0 terms prior to making the hologram. This is done simply by vibrating one of the two mirrors, so that it undergoes a periodic displacement $d = d_0 \cos \omega t$. This produces a similar time modulation of the fringe term, but not of the bias term. Thus, in place of the hologram we place an array of photodetectors, one for each resolution element. Connected to each is a band pass filter that passes the time varying term but rejects the bias term. The signal is then converted back to light, on a point-by-point basis, and recorded as a hologram. The bias term is now eliminated, and the obstacle to incoherent light holography has been overcome.

This sounds like an enormously complex system, but in reality the various photodetectors and filters can be realized using, for example, an image orthicon and associated electronics.

V. CONTOUR GENERATION*

A holographic image presents in a dramatic way information about the depth dimension of the object. This information, however, generally lacks accuracy, because of the difficulty of precise determination of object depth; the depth of focus is sufficiently great to preclude the measurement of this coordinate to the accuracy that is possible with the lateral coordinates. In short, the observer just can't tell precisely from the holographic image how far away each part of the image is.

Methods have been developed, however, for improving the accuracy of the depth dimension by holographically implanting depth contours on the holographic image. The holographic image then becomes a contour map. Haines and Hildebrand have described two methods for generating such contours automatically in the process of constructing a hologram. We shall discuss each of these, the two-frequency method and the two-source method.

As is well known, range information in a radar system is generally obtained by proper design of the spectrum of the radiated pulse: the broader the spectrum, the finer the range resolution. The range resolution obtainable is determined by the ambiguity function of the radar pulse, which for the range variable becomes the autocorrelation function of the radiated waveform. For a pulsed laser, the depth dimension could be obtained in the same manner as is done with pulsed radars.

It is convenient, however, to choose a simple, two-frequency waveform, such as is readily obtainable from a multimoding C. W. laser. The autocorrelation function of two-frequency components, $e^{j\omega_1 t}$ and $e^{j\omega_2 t}$, is a periodic function with spatial period $\lambda_1^2/2\Delta\lambda$, where λ_1 is the wavelength corresponding to ω_1 , and $\Delta\lambda$ is the wavelength difference corresponding to the frequency difference $\omega_1 - \omega_2$.

When an object is holographically photographed using two frequencies, the reconstructed image will have a fringe pattern, and by choosing the various parameters properly, these fringes can be made to represent contours of constant range, or depth. The contour interval is determined by the frequency difference $\omega_1 - \omega_2$.

An alternative method of introducing contour fringes is to illuminate the object with two sources, or even better, to make a double-

*This work was done by K. Haines and B. P. Hildebrand of our laboratory and was reported at the March, 1966 Optical Society of America Meeting in Washington, D. C.

exposure hologram, with the reference source moved to a different location for the second exposure. The images produced in this manner will, in the reconstruction, add coherently and thereby produce interference fringes. Again, by proper design of the experiment, the fringes can be made to represent depth contours.

VI. COLOR HOLOGRAMS*

Holograms can be made in color; various methods for doing this have been proposed. Although we have proposed several, the one proposed by Pennington and Lin appears to us the most feasible, and our efforts to date have been confined to their method. This method consists of using various colors (e.g., the three primary colors) that are available from lasers, and relying on the wavelength discrimination of thick emulsions (as described in Sec. II) to prevent interaction between the various recorded components in the reconstruction process; that is, we wish the red light incident on the hologram during the reconstruction step to interact only with the fringe pattern recorded with red light, etc. Color holograms are readily made, with either a front surface or a back surface reference beam. In the latter case, the color hologram reconstructs in full color in ordinary white light, although we have found that the front surface reference beam results in superior holograms, for the reasons noted earlier.

VII. CONCLUSION

The present paper has of necessity been confined to a survey. The work touched upon here will be described in greater depth in journal papers by members of the staff of the Radar and Optics Laboratory.

*This work was performed by A. Friesem and R. Fedorowicz of our laboratory

REFERENCES

1. K. Stetson and R. Powell, J. Opt. Soc. Am., 56, (1966).
2. B. P. Hildebrand and K. A. Haines, J. Appl. Opt., 5, (1966).
3. P.J. van Heerden, J. Appl. Opt., 2, 393(1963).
4. Y. N. Denisyuk, Opt. and Spect., 15, 279 (1963).
5. P. J. van Heerden, Appl. Optics, 2, 393 (1963).
6. G. Lippman, J. Phys., 3, 97 (1894).
7. K. S. Pennington and L. A. Lin, Appl. Phys. Letters, 7, 56 (1963).
8. A. S. Hoffman, J. G. Doidge, and D. G. Mooney, J. Opt. Soc. Am., 55, 1559 (1965).
9. A. Kozma and N. Massey, "Bias Level Reduction of Incoherent Holograms", paper delivered at the Optical Society of America Meeting, Washington, D. C., 18 March 1966.
10. L. Mertz and N. Young, Proc. ICO Conf. Opt. Instr. (London). p. 305, 1961.
11. G. Cochran, J. Opt. Soc. Am., 55, 615 (1965) Abstract.
12. A. Lohmann, J. Opt. Soc. Am., 55, 1556 (1965).

- 1. J. G. ...
- 2. J. F. ...
- 3. J. T. ...
- 4. J. W. ...

The following is a list of the names of the authors of the papers in this volume. The names are arranged in alphabetical order of the authors' surnames. The names of the authors of the papers are given in full, including their first names and initials. The names of the authors of the papers are given in full, including their first names and initials. The names of the authors of the papers are given in full, including their first names and initials.

- 5. J. G. ...
- 6. J. F. ...
- 7. J. T. ...
- 8. J. W. ...

THE USE OF LASERS IN SIGNAL PROCESSING

FOR RADAR AND COMMUNICATIONS

by

L. J. Cutrona

Electrical Engineering Department

The University of Michigan and Conductron Corporation

Ann Arbor, Michigan

ABSTRACT

Lasers have been used in a large variety of signal processing optical equipments. The basic ideas will be reviewed and then applied to a number of problems in radar processing and communications signal processing.

The major ideas involving the use of optics for radar will be applied to the process of generating a synthetic antenna whereby a resolution significantly finer than that of a radiated beamwidth is achieved. In the application of optics to communications, it is shown that many of the coding-decoding and synchronization operations can be mechanized optically.

I. INTRODUCTION

In this paper a number of applications of laser light to signal processing functions are described. The basic diffraction problems and the Fourier transform relations are described and illustrated. The flexibility and power of optical technology for performing a large number of linear operations is described. Optical configurations for such linear operations are given. The primary topics discussed are those dealing with the use of an optical configuration for radar signal processing in which a synthetic antenna is generated and the use of optical techniques in communications for a variety of coding, decoding and synchronization purposes. It is shown that these operations can be performed using relatively simple optical configurations.

There are two important optical configurations which occur repeatedly in optical computing configurations. These configurations are shown in Figures 1 and 2. Figure 1 shows a configuration for performing two-dimensional spectrum analysis. Figure 2 shows a modification of this configuration (by addition of a cylindrical lens) to produce a multi-channel one-dimensional spectrum analyzer. The operations of these configurations are given by Equations 1 and 2.

The capability of optical equipment as a computing tool arises in part from diffraction phenomena. For this reason two of the basic configurations which appear repeatedly in computing configurations will be described. The two configurations referred to are:

- a) A configuration using a spherical lens which produces two-dimensional diffraction, and
- b) A configuration consisting of a spherical lens in conjunction with a cylindrical lens which produces a multiplicity of one-dimensional diffraction patterns.

The basic configuration for obtaining two-dimensional diffraction patterns is shown in Figure 1:

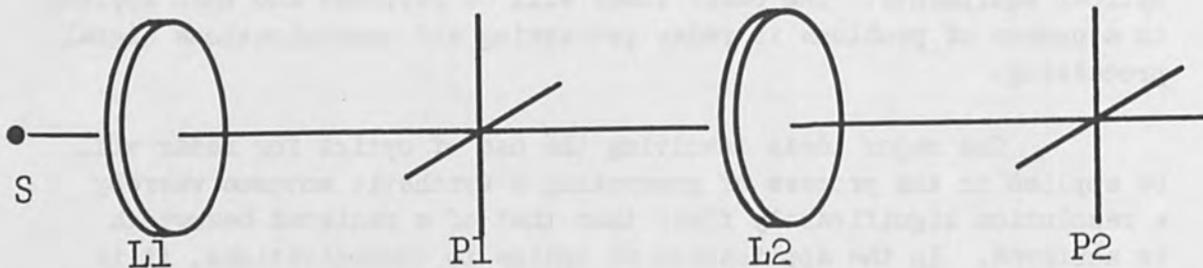


Figure 1

Configuration For Two-Dimensional Spectrum Analysis

In this figure S represents a source of light, L_1 represents a collimating lens, P_1 represents the input plane in which a transparency is placed, lens L_2 is the spherical lens which is the essential element for producing a two-dimensional diffraction pattern. Plane P_2 is the plane in which the two-dimensional spectrum (of the transparency in plane P_1) is exhibited.

In order that the distribution of light in plane P_2 be the two-dimensional spectrum analysis of the density distribution of the transparency in plane P_1 , it is necessary that planes P_1 and P_2 be spaced a focal length on either side of lens L_2 . If $f(x,y)$ represents the amplitude of light emerging from plane P_1 , then the distribution of light amplitude in plane P_2 is given by

$$F(\alpha, \beta) = \iint f(x,y) e^{jk(\alpha x + \beta y)} dx dy \quad (1)$$

In Eq. (1), the amplitude of the light in plane P_2 is given by $F(\alpha, \beta)$. Here k represents the wave number of the light while α and β represent the direction cosines of the diffracted beam with respect to the x and y axes.

Equation (1) is the fundamental relationship for two-dimensional diffraction. The configuration shown in Figure 1 is used in a number of applications.

The configuration in Figure 1 produces a single two-dimensional diffraction pattern. This equipment can be converted to a multi-channel one-dimensional diffraction equipment by the addition of a cylindrical lens to the configuration of Figure 1. This cylindrical lens is placed between planes P_1 and P_2 to give the configuration shown in Figure 2:

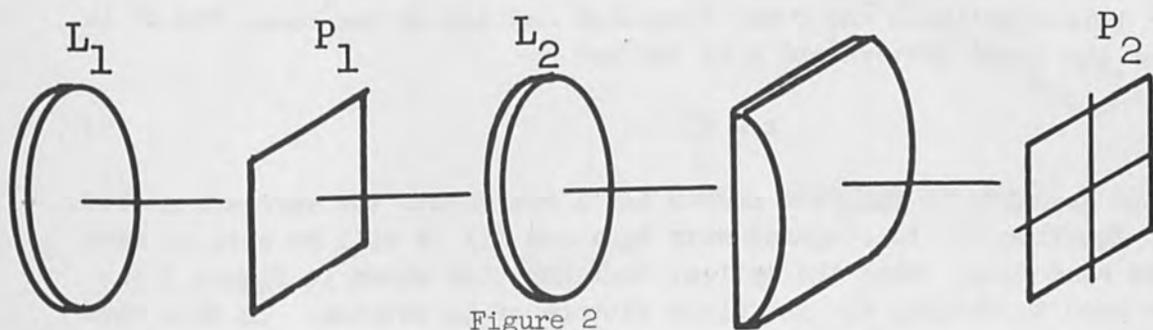


Figure 2

In this case the distribution of light in plane P_2 is given by Eq. (2). It will be noted that this expression indicates distribution of light corresponding to a multi-channel spectrum analysis. The parameter y is an index referring to a given channel. The other parameters have been previously defined.

$$F(\alpha, y) = \int f(x, y) e^{jkx} dx \quad (2)$$

It will be noted that if α and β in Eq. (1) are set equal to zero, two-dimensional integration can be performed, whereas setting $\alpha = 0$ in Eq. (2) performs a multiplicity of one-dimensional integrations. In the former case, a hole near the optical axis accepts the light. In the second case, a slit contains the multiplicity of integrations. This operation is important in the radar configurations. Specifically, one has for this case

$$\left(\int f(x, y) e^{-jk\alpha x} dx \right)_{\alpha=0} = \int f(x, y) dx \quad (3)$$

II. TECHNICAL DISCUSSION

A. Antenna Simulation

As an example of the two-dimensional spectrum analysis capability of optical configurations, one can consider antenna simulation.

The far field pattern of an antenna can be computed from its illumination function $f(x,y)$ by the use of

$$F(\alpha, \beta) = \iint_{\substack{\text{over} \\ \text{aperture}}} f(x,y) e^{-jk(\alpha x + \beta y)} dx dy \quad (4)$$

In this equation α and β are direction cosines of the beam, $F(\alpha, \beta)$ is the far field pattern and k is defined by

$$k = \frac{2\pi}{\lambda} \quad (5)$$

This quantity is the wave number and λ represents the wavelength used. If Equation (4) is compared with Equation (1) it will be seen to have the same form. Thus the optical configuration shown in Figure 3 can be used to display the far field pattern of an antenna. In this case a transparency containing the aperture function $f(x,y)$ is placed in plane P_1 and the far field pattern is observed in plane P_2 .

While the configuration of Figure 1 is usually satisfactory, in some cases, the far field pattern found in plane P_2 may be too small. In this case it is desirable to use another lens to magnify the image found in plane P_2 . Such a configuration is shown in Figure 3 where an enlarged image of the field in plane P_2 is displayed in plane P_3 .

In Figure 3 one finds that an image of plane P_1 occurs between lenses L_3 and plane P_3 at plane P_4 . Thus between planes P_4 and P_3 one has an opportunity to observe the pattern as it emerges from the illumination function through the near field until at plane P_3 the far field pattern is obtained. With this configuration, studies of the relationship between near field and far field can be made. In addition, by inserting perturbations into the regions between planes P_4 and P_3 , for example, by simulating a non-homogeneous medium, it is possible to observe the effects of perturbations on the far field pattern.

A series of photographs showing the far field developing from the illumination function for an array antenna¹ is shown in Figures 5, 6, 7, 8, 9, and 10. This set of photographs is the optical simulation of an antenna array being built by J. P. Wild in Australia. The array consists of 96 parabolas arranged on a circle where diameter is about 1-1/2 miles. Each parabola has a diameter of 45 feet. The frequency

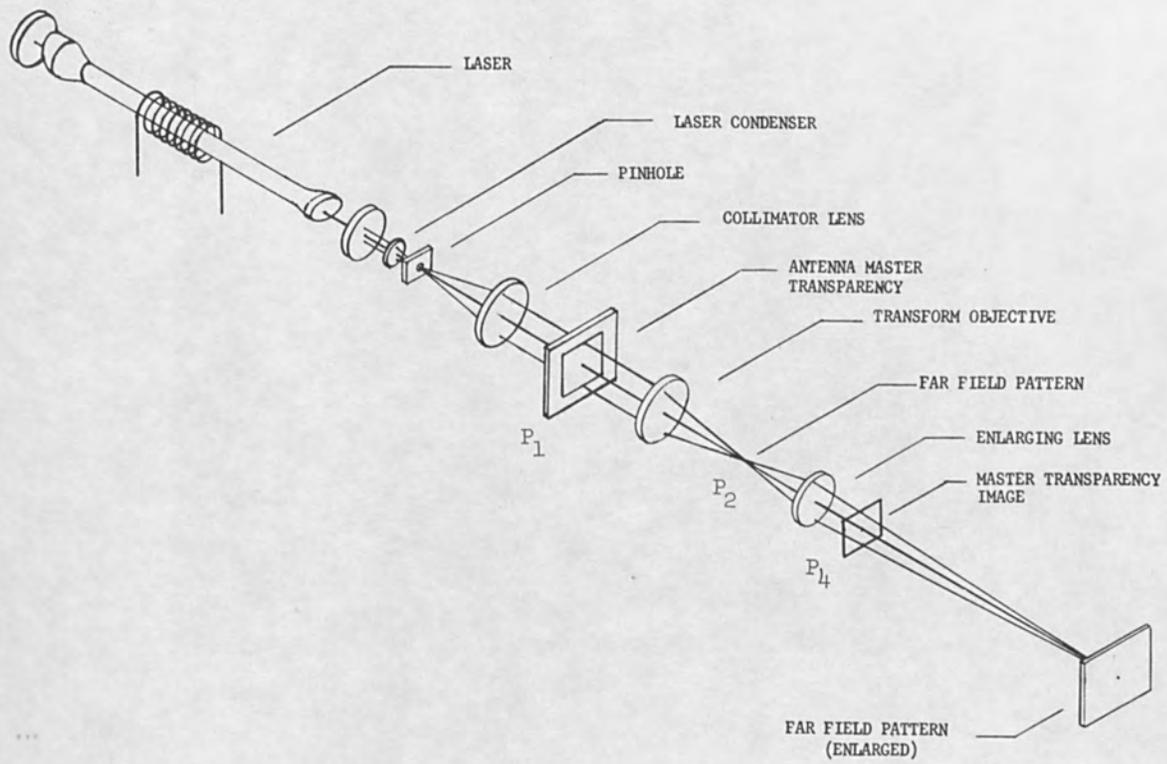


Figure 3.

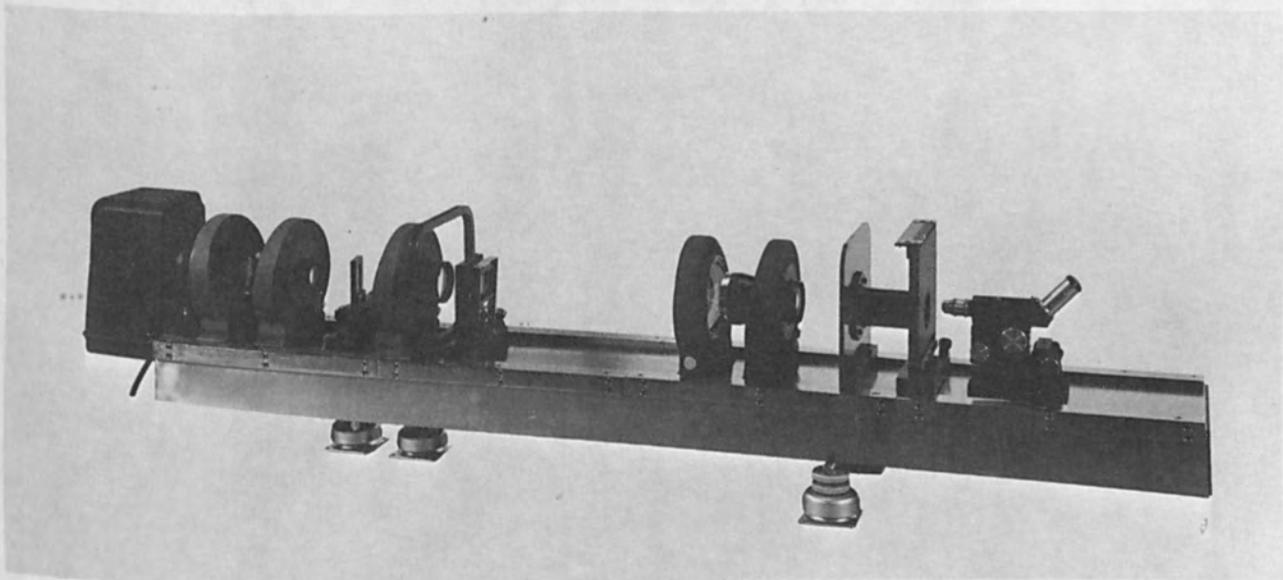


Figure 4.

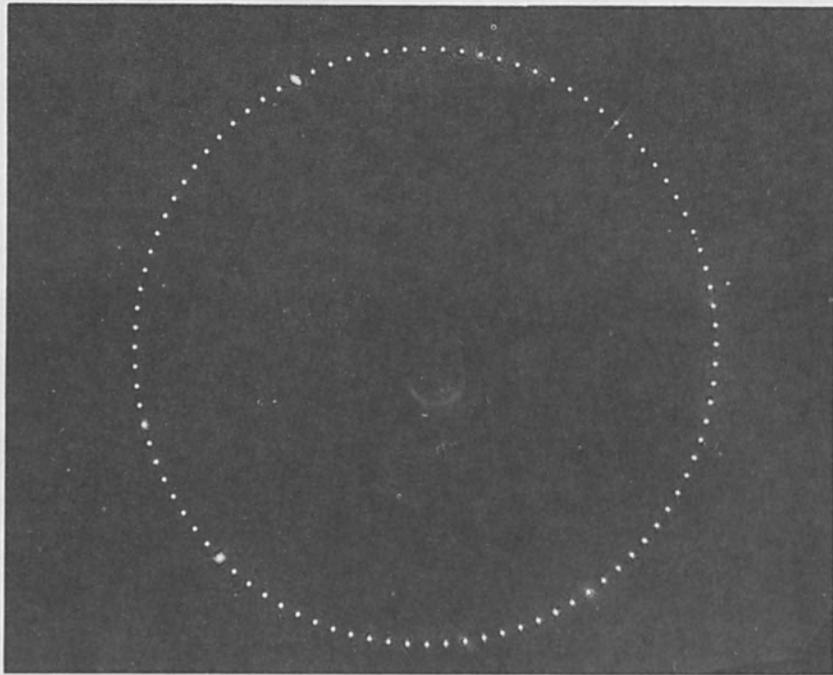


Figure 5.

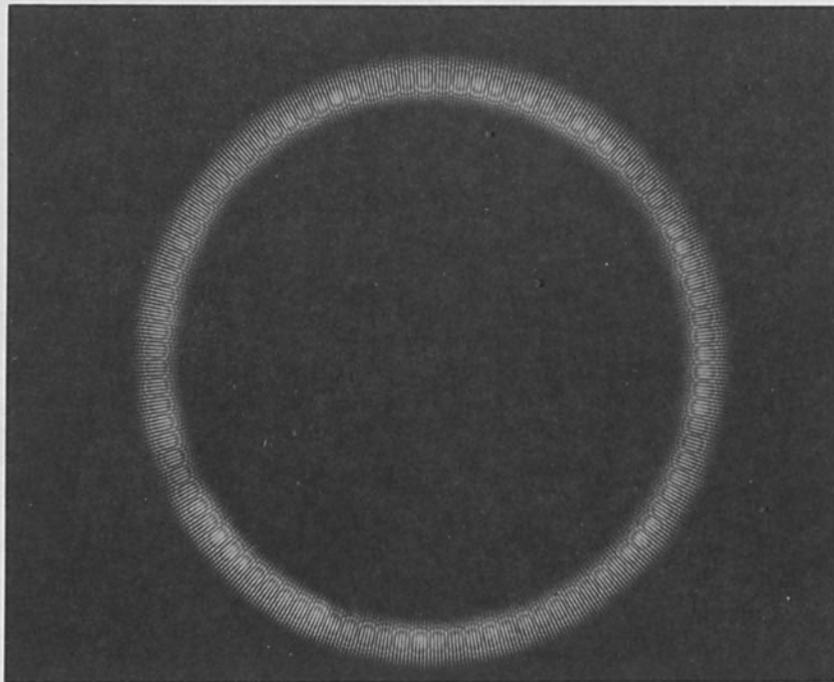


Figure 6.

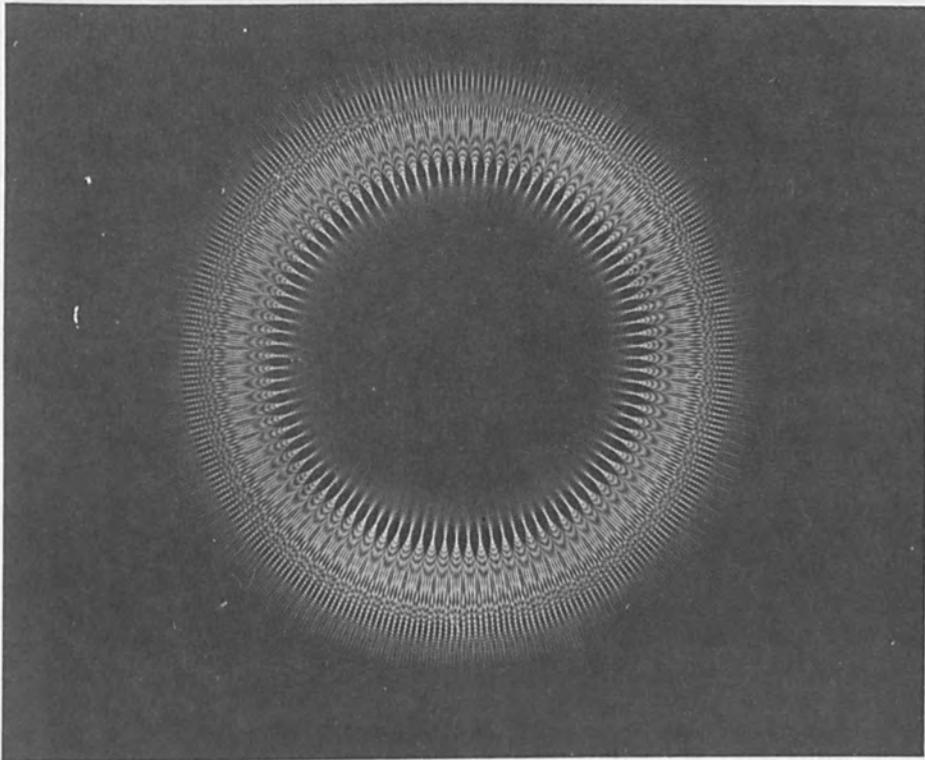


Figure 7.

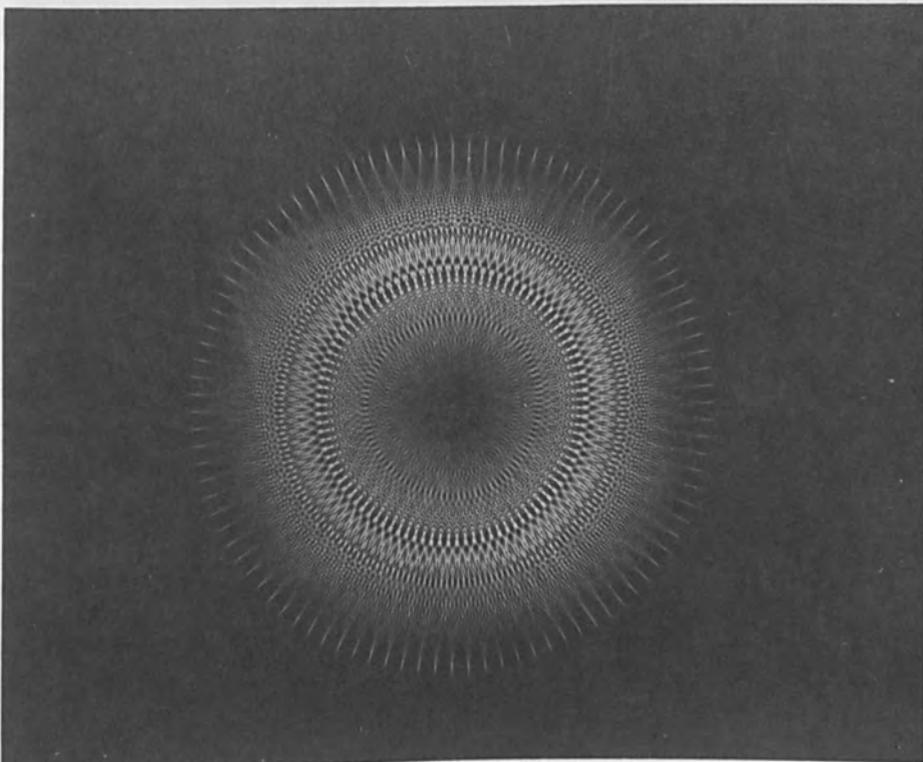


Figure 8.

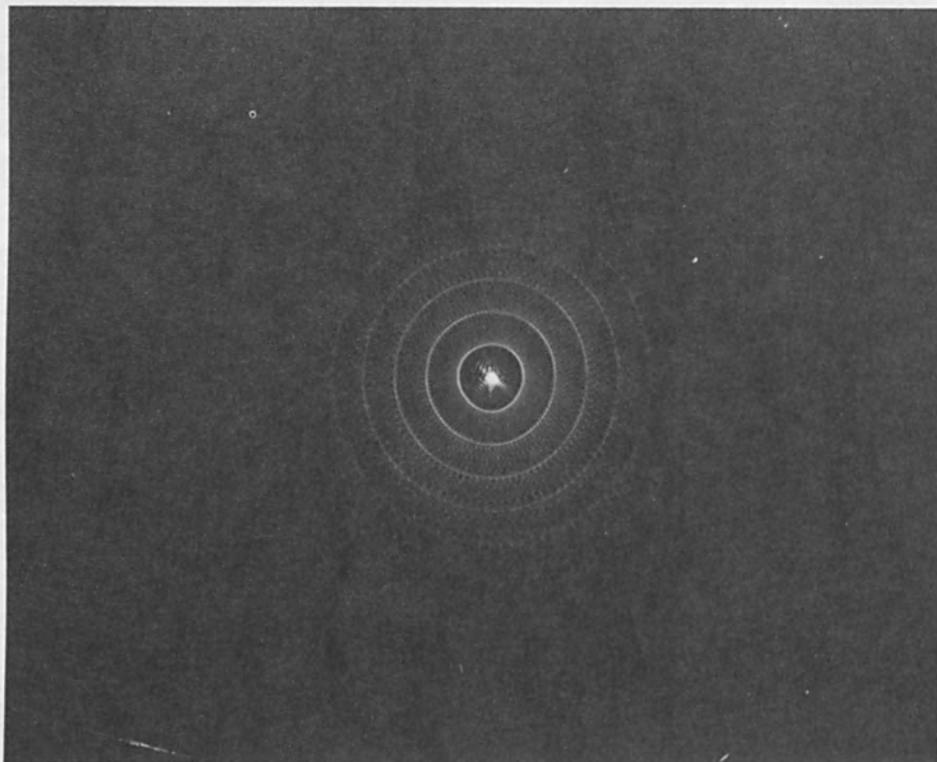


Figure 9.

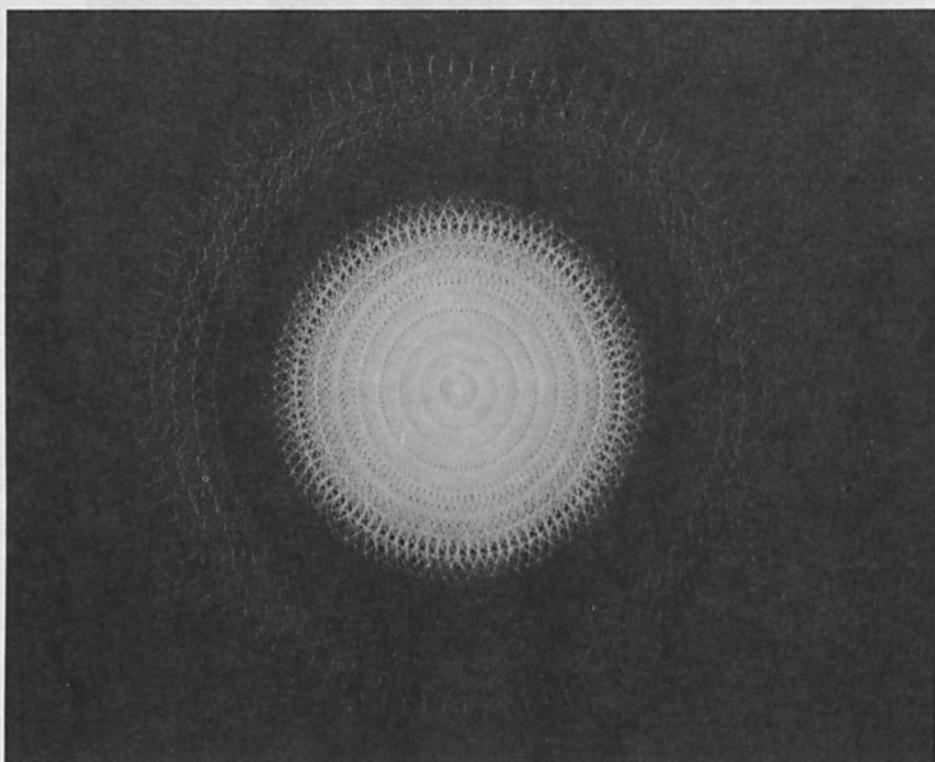


Figure 10.

of operation is 80 megacycles.

Assuming that the far field pattern begins at a distance $\frac{2D^2}{\lambda}$, the far field pattern begins approximately 2,400 miles from earth. This poses real problems for measurement of the far field pattern. The configuration of Figure 3, however, enables one to obtain not only the far field pattern shown in Figures 9 and 10 but also views of the near field pattern are obtained. Additional examples of illumination functions and the far field patterns are shown in Figures 11, 12, 13, and 14.

In the first figure a one-dimensional pattern is shown and in the second case a two-dimensional pattern and its array are shown. For the one-dimensional pattern the digital computer computation of the far field pattern is shown in Figure 15.

Thus antenna simulation is another demonstration of the versatility of optical equipments.

B. Synthetic Antenna Radar

Perhaps the most important application of coherent optical signal processing to date has been the use of such techniques for synthetic antenna generation.² In these techniques a radar resolution independent of both range and frequency is predicted. Furthermore, finer resolution is associated with smaller antennas. This provided the initial impetus for work in optics at the Institute of Science and Technology of the University of Michigan. The considerations leading to this mechanization follow.

C. Factors Affecting Resolution of a Radar System

In this section a brief comparison of the conventional antenna, the unfocused synthetic antenna, and the focused antenna will be given. The language of synthetic antennas is used and a comparison of the resolution capability of three cases is given.

The three cases are compared for their azimuth resolution capability.

1. The conventional technique: In this technique azimuth resolution depends upon the width of the radiated beam.

2. The unfocused synthetic antenna technique: In this case the synthetic antenna length is made as long as the unfocused technique permits.



Figure 11.

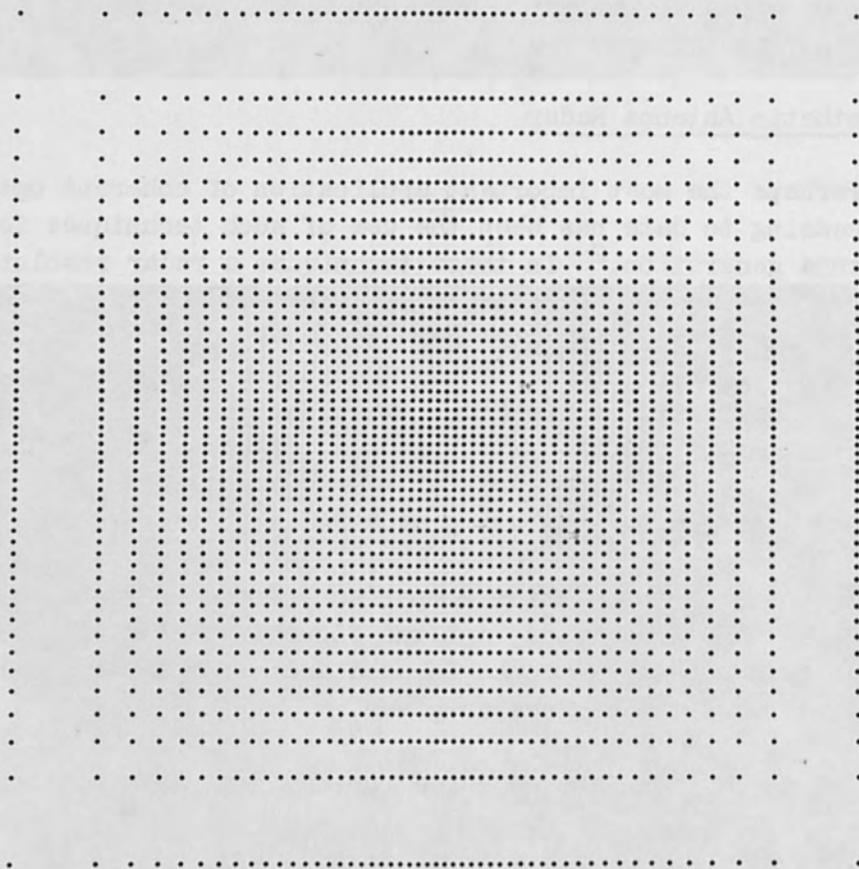


Figure 12.

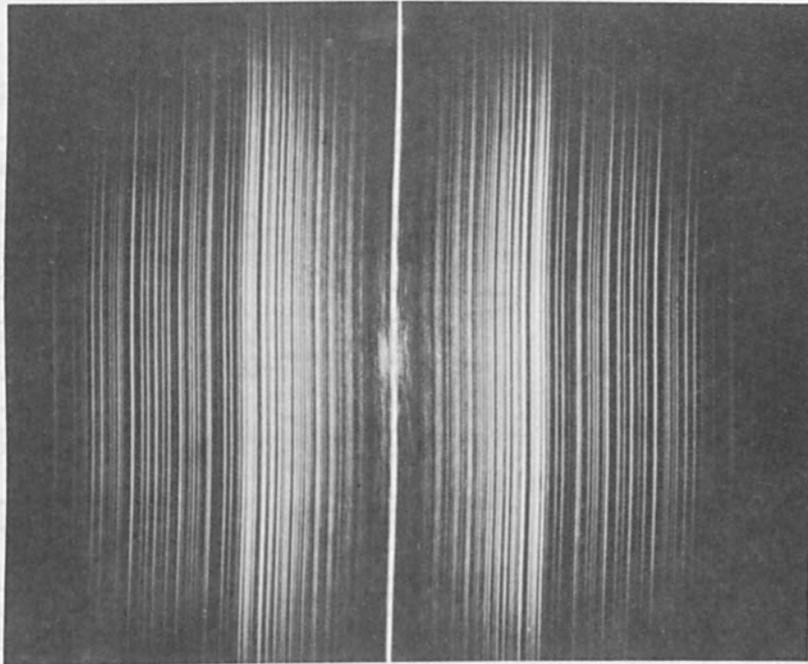


Figure 13.

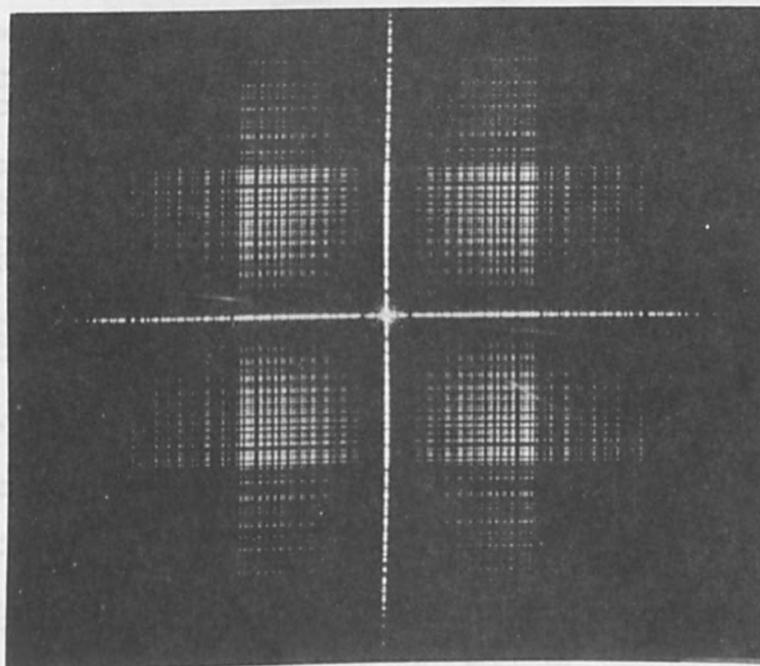


Figure 14.

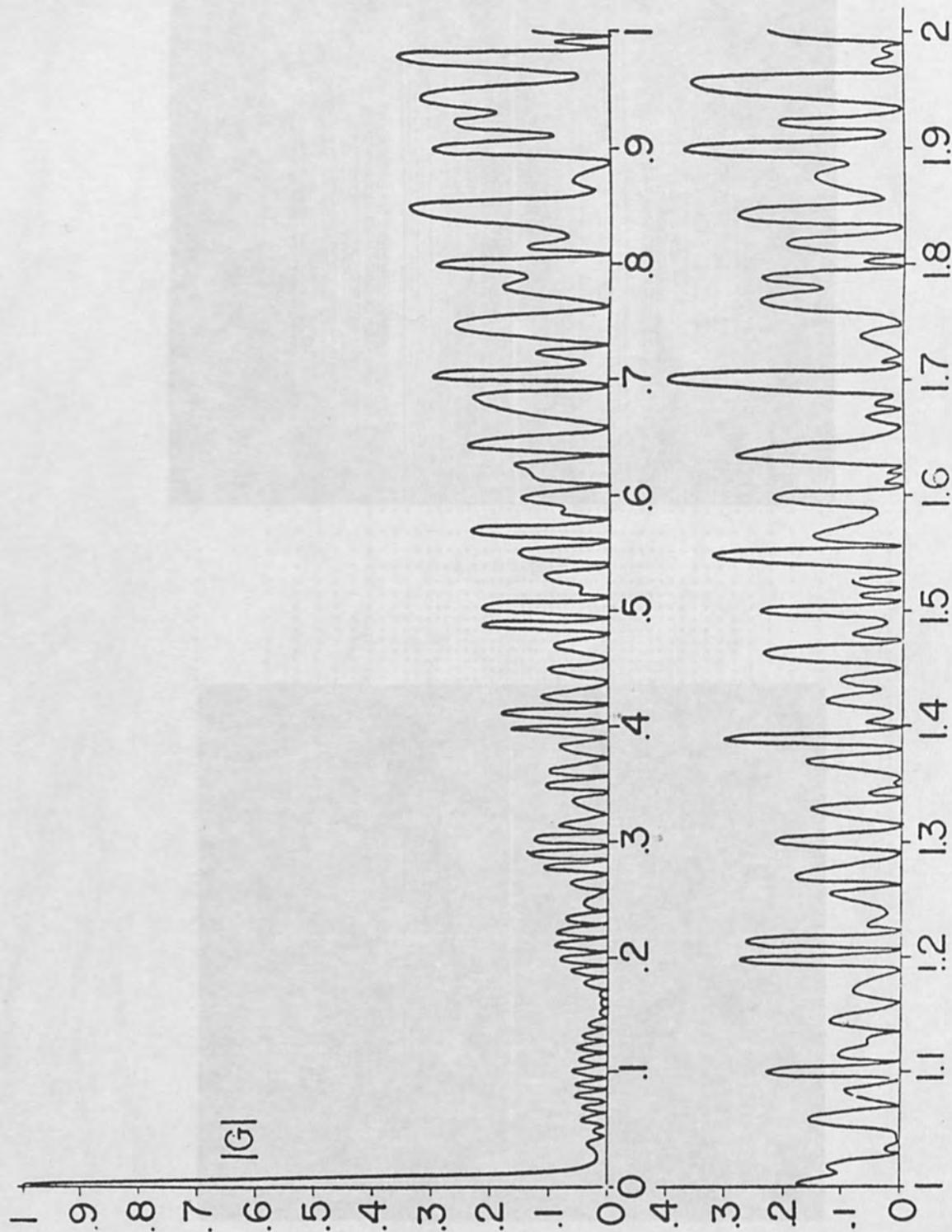


Figure 15.

3. The focused synthetic antenna technique: In this case the synthetic antenna length is made equal to the linear width of the radiated beam at each range.

As is shown below, the linear transverse resolution for the conventional case is given by

$$\text{Resolution}_{\text{conv}} = \frac{\lambda R}{D} \quad (6)$$

For the unfocused case, the resolution is given by

$$\text{Resolution}_{\text{unf}} = \frac{1}{2} \sqrt{\lambda R} \quad (7)$$

whereas for the focused case, the resolution is given by

$$\text{Resolution}_{\text{foc}} = \frac{D}{2} \quad (8)$$

In the above expressions, λ is the wavelength of the radar signal transmitted, D is the horizontal aperture of the antenna, and R is the radar range.

Figure 16 is a plot of the resolution for each of these cases as a function of radar range. This plot is for an antenna aperture of 5 feet and a wavelength of 0.1 feet.

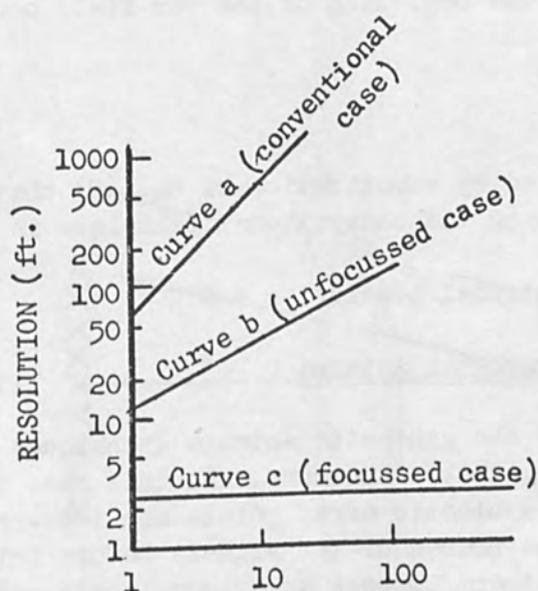


Figure 16

Azimuth Resolution for Three Cases:
(a) conventional, (b) unfocused, (c) focused.

D. Conventional Technique

The conventional technique for achieving azimuth resolution has been that of radiating a narrow beam. In this case the resolution of a target depends upon whether the target is included within the radiated beam. While some techniques exist for resolving targets somewhat less than a beamwidth apart, it will be assumed for the purposes of this paper that the resolving capability of an antenna is equal to the width of the radiated beam at each range.

The computation of the linear azimuth resolution for the conventional case is well known. The appropriate expression is obtained by noting that the width of the radiated beam in radians is given by the ratio λ/D , whereas the linear width of the beam at range R is the product of this beamwidth by range. These considerations lead to the result already written as Eq. (6).

It is noted that for a given wavelength and antenna aperture the fineness of resolution deteriorates in direct proportion to range. For a specified antenna aperture and range the linear resolution can be improved by using a shorter wavelength. Conversely, with the wavelength fixed, an increase in antenna aperture is required to improve resolution.

A consideration not evident from the above discussion but pertinent from antenna theory is that Eq. (6) applies only to the far field pattern of an antenna. The beginning of the far field occurs at a distance R_{\min} for which

$$R_{\min} \simeq D^2/\lambda \quad (9)$$

It will be noted by substitution of Eq. (9) that the finest resolution achievable using the conventional technique is given by

$$\text{Minimum Conventional Resolution} = D \quad (10)$$

E. The Unfocused Synthetic Antenna

The simpler of the synthetic antenna techniques is that which generates an unfocused synthetic aperture. In this case the coherent signals received at the synthetic array points are integrated with no attempt made to shift the phases of the signals before integration. This lack of phase adjustment imposes a maximum length upon the synthetic antenna length which can be generated. This maximum synthetic antenna length occurs at a given range when the round-trip distance from a radar target to the center of the synthetic array differs by $\lambda/4$ from the round-trip distance between the radar target and the extremities of the synthetic antenna array.

The pertinent geometry is shown in Figure 17. In this figure R_0 represents the range from a radar target to the center of the array, L_{eff} represents the maximum synthetic antenna length for which the distance from the target to the extremities of the synthetic antenna does not exceed $R_0 + \lambda/8$.

It is evident from this geometry that

$$\left(R_0 + \frac{\lambda}{8}\right)^2 = \frac{(L_{eff})^2}{4} + R_0^2. \quad (11)$$

If this expression is solved for L_{eff} , subject to the assumption that $\lambda/16$ is small compared to R_0 , the result is

$$L_{eff} = \sqrt{R_0 \lambda}. \quad (12)$$

Because of coherence, the relative phase of signal received from a target depends on the round-trip distance to it. Consequently, the beamwidth of the synthetic antenna is half that of a conventional array of the same length, i.e.,

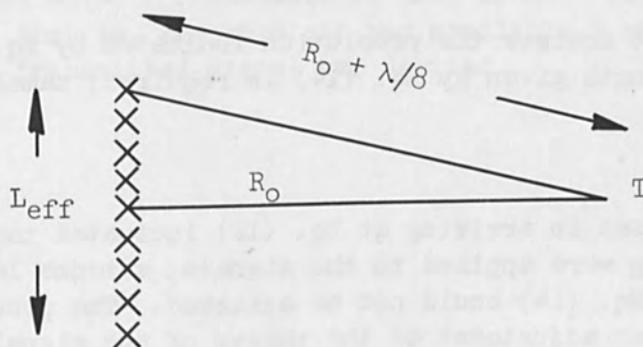


Figure 17

Geometry for Unfocused Synthetic Antenna

$$\beta = \frac{\lambda}{2L_{\text{eff}}} \text{ radians.} \quad (13)$$

Multiplying this beamwidth by range results in the resolution given by Eq. (7).

As shown by Eq. (7) the transverse linear resolution is independent of the antenna aperture size. For the unfocused case fineness of resolution is increased by the use of shorter wavelengths, but in comparison with the conventional case, the improvement in fineness of resolution varies as the square root of λ rather than directly as λ .

It will also be observed that the resolution as a function of range deteriorates as the square root of range rather than directly as range for the conventional case. A plot of Eq. (7) is given in Figure 16.

F. The Focused Case

An expression for the resolution achievable in the focused case was given in Cutrona, et al.^{3,4} This expression has been given as Eq. (8). It is significant to note that the azimuth resolution achievable for this case depends only upon the synthetic antenna aperture and that in contrast to the conventional case, fine resolution is achievable by using small rather than large antennas.

Also significant for this case is the fact that the achievable resolution for a given antenna size is independent both of the range and of the wavelength used. A graph of Eq. (8) is also shown on Figure 16. Inasmuch as the resolution achievable in this case is a constant, this curve is a straight line parallel to the range axis.

In order to achieve the resolution indicated by Eq. (8), a synthetic antenna length given by Eq. (14) is required; namely,

$$L_{\text{eff}} = \frac{\lambda R}{D} \quad (14)$$

The considerations used in arriving at Eq. (12) indicated that unless additional processing were applied to the signals, antenna lengths such as those implied by Eq. (14) could not be achieved. The processing required is that of an adjustment of the phases of the signals received at each point of the synthetic antenna which makes these signals co-phase for a given target. If this is done, the restrictions which limited the maximum antenna length to that given by Eq. (12) are no longer pertinent and the new limitation on the length of synthetic antenna achievable becomes simply the linear width of the radiated beam at the range of the target. It will be noted that Eq. (14) gives an expression

for the maximum length in this case.

In some cases a resolution coarser than $D/2$ is sufficient. Then, a fraction α of the maximum focused synthetic antenna length can be used. Eq. (14) becomes

$$L_{\text{eff}} = \frac{\alpha \lambda R}{D} \quad (14a)$$

and the achievable resolution is

$$\text{Resolution}_{\text{foc}} = \frac{D}{2\alpha} \quad (8a)$$

For those situations for which the synthetic antenna length given by Eq. (14a) is less than or equal to the synthetic antenna length for the unfocused case as given by Eq. (12) only a limited improvement in resolution is achievable for the focused case. However, if a resolution finer than that given by Eq. (7) is desired, then focusing must be used, and in this sense the use of focusing, in effect, removes the restriction on synthetic antenna length which would otherwise apply.

G. The Radar Signal

A common configuration for a ground-mapping radar of the synthetic antenna type is shown in Figure 18. An aircraft carries an antenna which illuminates a ground-swath parallel to the flight path; the radar beam is oriented in azimuth roughly normal to the direction of flight; at range R , the azimuth linear beamwidth βR is much larger than the desired azimuth resolution at that range. Finally, the radar is "coherent," that is, the receiver has available a reference signal from which the transmitted signal was derived.

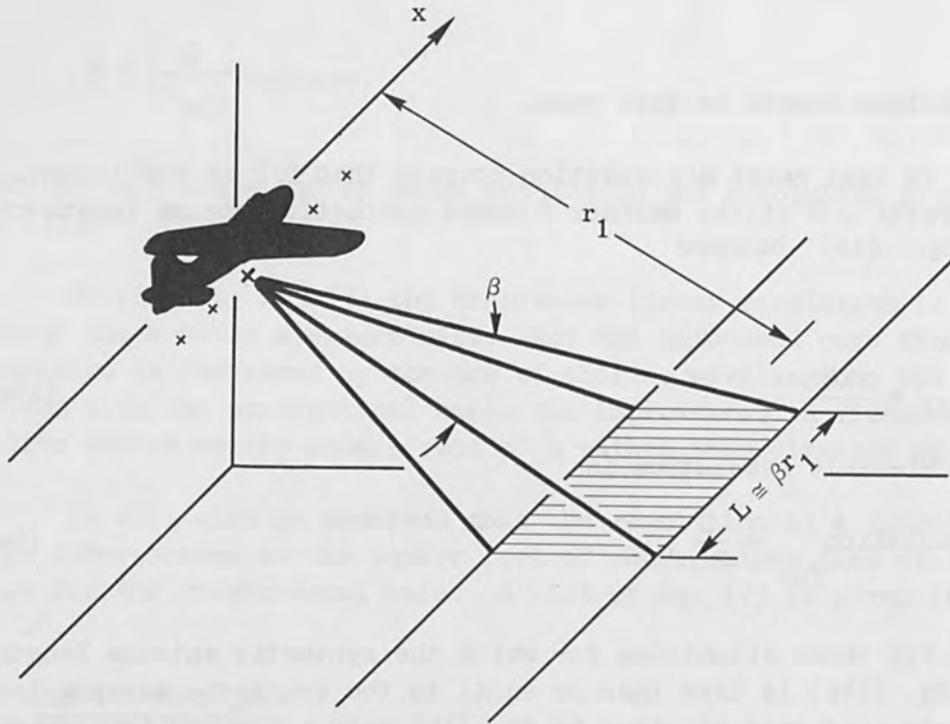


Figure 18

Sidelooking Radar Configuration

Synthetic-antenna radars derive range information through pulsing, and derive fine azimuth resolution by processing doppler-shifted radar returns which lie in a spectral band which is adequately sampled by the pulse rate of the radar. For our purposes, we may neglect this intermediate sampling process and consider only the reconstructed azimuth histories which are easily derived from the samples. We will assume that the entire radar receiver and processor behave as a linear system; we can then investigate the response of the radar to a single "point" target,* and by superposition extend the results to apply to realistic reflective complexes.

To find the form of the signal returned to the radar antenna, consider a point target located at a position x_1 along the ground track, at a range r_1 from the aircraft flight path (Fig. 19). Let a sinusoid of radian frequency ω be transmitted. The signal returned to the

* "Point" targets are defined to be those targets which have physical dimensions that are small compared with the resolution capability of the radar system.

aircraft then has the form

$$s_1(t) = \sigma_1 e^{j[\omega(t - \frac{2r}{c})]} \quad (15)$$

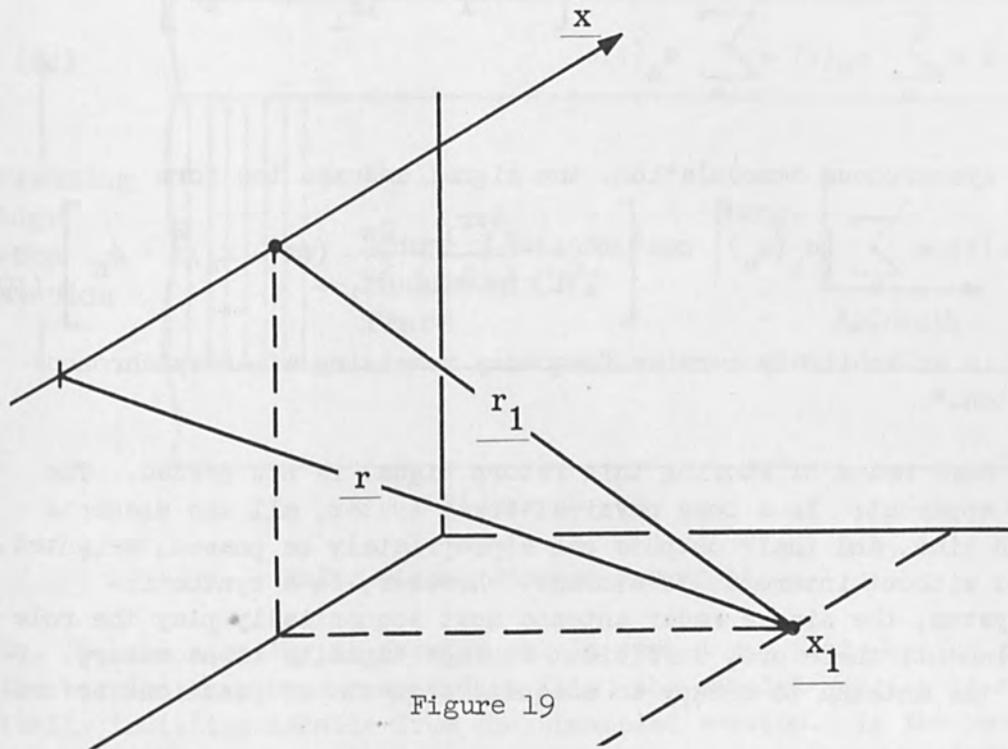


Figure 19
System Geometry

where r is the range from aircraft to target, c is the velocity of light, and σ_1 is the complex amplitude of the signal as received by the radar. Into σ_1 we have absorbed such factors as the radiated power level, the target reflectivity and phase shift, and the inverse fourth-power law. By use of the approximation

$$r = \sqrt{r_1^2 + (x - x_1)^2} \approx r_1 + \frac{(x - x_1)^2}{2r_1} \quad (16)$$

Eq. 15 can readily be written as

$$s_1(t) = \sigma_1(t) e^{j\left[\omega t - \frac{4\pi r_1}{\lambda} - \frac{2\pi}{\lambda} \frac{(x - x_1)^2}{r_1}\right]} \quad (17)$$

The motion of the aircraft links the temporal behavior of the signal with the spatial coordinates of the target, through the relation

$$x = vt$$

where we have chosen the origins of both systems so that $t = 0$ when $x = 0$. If the terrain at range r_1 is regarded as a collection of numerous point scatterers, the return signal is of the form

$$s = \sum_n s_n(t) = \sum_n \sigma_n(t) e^{j \left[\omega t - \frac{4\pi r_1}{\lambda} - \frac{2\pi}{\lambda r_1} (vt - x_n)^2 \right]} \quad (18)$$

and after synchronous demodulation, the signal assumes the form

$$s(t) = \sum_n \left| \sigma_n(x_n) \right| \cos \left[\omega'_c t - \frac{4\pi r_1}{\lambda} - \frac{2\pi}{\lambda r_1} (vt - x_n)^2 + \phi_n \right] \quad (19)$$

where ω'_c is an arbitrary carrier frequency remaining after synchronous demodulation.*

Some means of storing this return signal is now needed. The reason is apparent: In a long physical-array system, all the elements coexist in time, and their outputs can appropriately be phased, weighted, and summed without intermediate storage. However, in a synthetic-antenna system, the single radar antenna must sequentially play the role of each element; therefore, sufficient storage capacity is necessary to permit the antenna to occupy an adequate sequence of positions before summing.

The returning radar signals will have a bandwidth determined by the range-resolution capability built into the radar system. The photographic process permits recording of wideband signals with greater ease than is possible with certain other recording systems; consequently, it is well suited for use with synthetic-antenna radars, provided one can tolerate some of the film-processing inconveniences. Photographic recording also permits high-density storage, another quality which is desirable in fine-resolution mapping systems.

The radar receiver output is a sequence of reflected range pulses; these are used to intensity-modulate a cathode-ray tube, the electron beam of which is swept in synchronism with the returning pulses. Successive range traces are recorded side by side, producing a two-dimensional format in which the dimension across the film represents range, and the dimension along the film corresponds to the along-track

*Equations 17, 18, and 19 above omit amplitude weighting which is inserted by the azimuth cross section of the radiated beam; this weighting is not essential to the present discussion.

dimension. This format is shown in Fig. 20. The signals in the along-track dimension are the sampled versions of $s(t)$ as given by

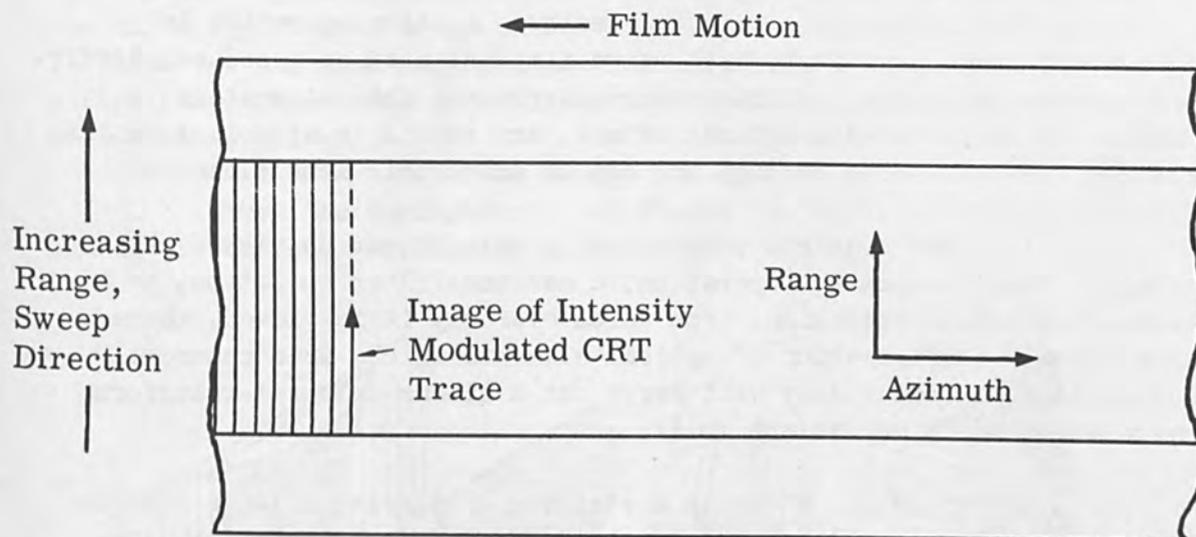


Figure 20

Radar Signal Storage Format

Eq. 19; the sampling is carried out in such a way that, by the time the samples have been recorded on film, the sampled version is essentially indistinguishable from the unsampled version. In the recording process, the time variable is converted to a space variable defined in terms of distance along the film. The recording process itself introduces a pair of scale factors, one in azimuth and one in range, into the data record; however, for our purposes, all relationships are handled in "radar space," and the presence of these scale factors is neglected. The radar signal is also placed on a bias level s_b prior to recording; thus the recorded signal is now in a form which is a spatial analog to that given by Eq. 19:

$$s_b + s(x) = s_b + \sum_n |\sigma_n(x)| \cos \left[\omega_x x - \frac{4\pi r_1}{\lambda} - \frac{2\pi}{\lambda r_1} (x-x_n)^2 + \phi_n \right] \quad (20)$$

H. Reasons for Processing Optically

There are several methods for carrying out the data-processing operation which is required to generate a fine-resolution radar image from the raw radar data. Coherent optical processing is one method which is particularly well suited to this task for the following reasons:

- (a) The video data are generally stored on photographic film

because of its large storage capacity; therefore, the video data record is available for direct insertion into an optical channel.

(b) Because the weighted vector-summation operation in azimuth is range dependent, each range interval must be processed differently; this creates a multichannel requirement. Optical systems, because of their two-dimensional nature, are readily adaptable to multichannel configurations through the use of anamorphic lens elements.

(c) The required processing in azimuth can be viewed as a weighted vector-summation operation, a matched-filter operation, or a cross-correlation operation. For which ever way it is viewed, there is a certain configuration of optical elements which, when coherently illuminated, automatically will carry out a linear integral-transform operation that is equivalent to it.

(d) Finally, there is a striking similarity between the coherent radar system which acquires the data, and the coherently illuminated processor which converts the data to a radar image. The coherent optical system, therefore, fits its application in a very natural way, and it tends readily to perform precisely those operations which are required of it. Hence, the optical system is both simple and elegant. This analogy will be developed later in this paper.

I. Optical Radar Processor

Let attention be given to the configuration given as Figure 21. In this case, an illumination system to the left of plane P_1 is not shown. However, a means for coherently illuminating this signal history with a plane wave of spatially coherent light is required. The processor is required to perform the operation of synthetic antenna generation and focusing.

Figure 2 and Eq. (3) show that the combination of the cylindrical lens and the spherical lens can be used to perform multi-channel integration. However, before performing this integration it is necessary to re-adjust the phases on the signal history film to achieve focusing. The signal histories at a given range have a quadrature phase progression. This phase progression, as a function of along-track position, is different at different ranges. It turns out that a conical lens can be used to make the phase corrections necessary for focusing. This conical lens is shown adjacent to the signal history film. Thus, signals after having passed through both the signal history film and the conical lens will now have phases adjusted so that focusing can take place simultaneously at all ranges.

The cylindrical lens together with the spherical lens then

performs a multiplicity of integrations, one for each range. Thus, one range sweep of fine resolution radar data is found in the slit in plane P_2 . As the signal history in plane P_1 is translated through the optical system, a sequence of range sweeps of fine resolution radar data occurs in plane P_2 . As a second unexposed film is made to pass the slit P_2 , a fine resolution radar map will be formed after photographic development of this film.

Thus, the configuration of Figure 21 is essentially an optical analog of the radar world.

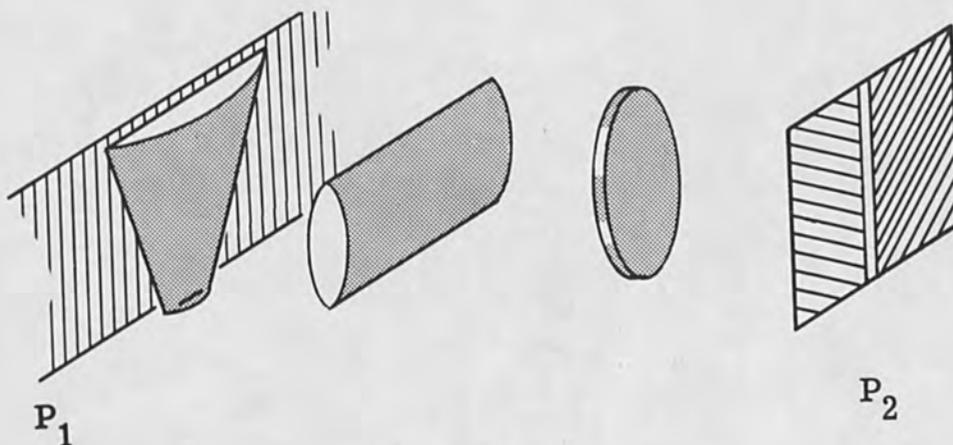


Figure 21a

Optical Processor Using a Conical Lens

As an example of the improvement in radar resolution, Figures 21b through 21d are shown. Figure 21b shows the output of a conventional radar in which the beamwidth radiated provides the resolution. Figures 21c and 21d show the types of images achievable using the synthetic antenna approach. The latter pictures were generated using the AN/UPD-1 radar system.

J. Communications

Another application of optics is to communications. In this case the optical techniques are used for terminal signal operations rather than for transmission from one point to another. Thus, operations such as coding and decoding are considered. Some methods of performing these operations using optical techniques are discussed in this section.

K. General Features of a Communications System

The essential components of a communications system are

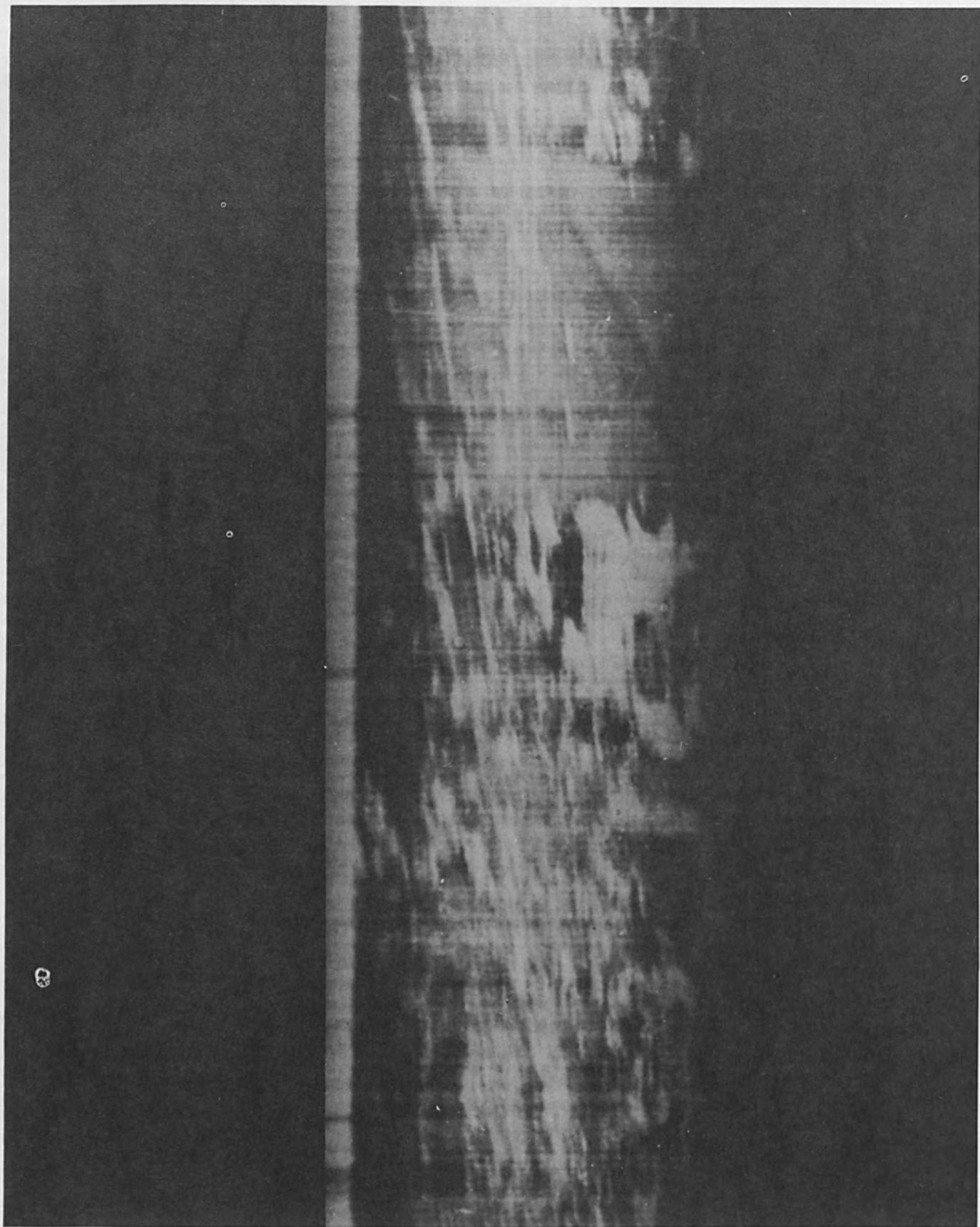


Figure 21b.

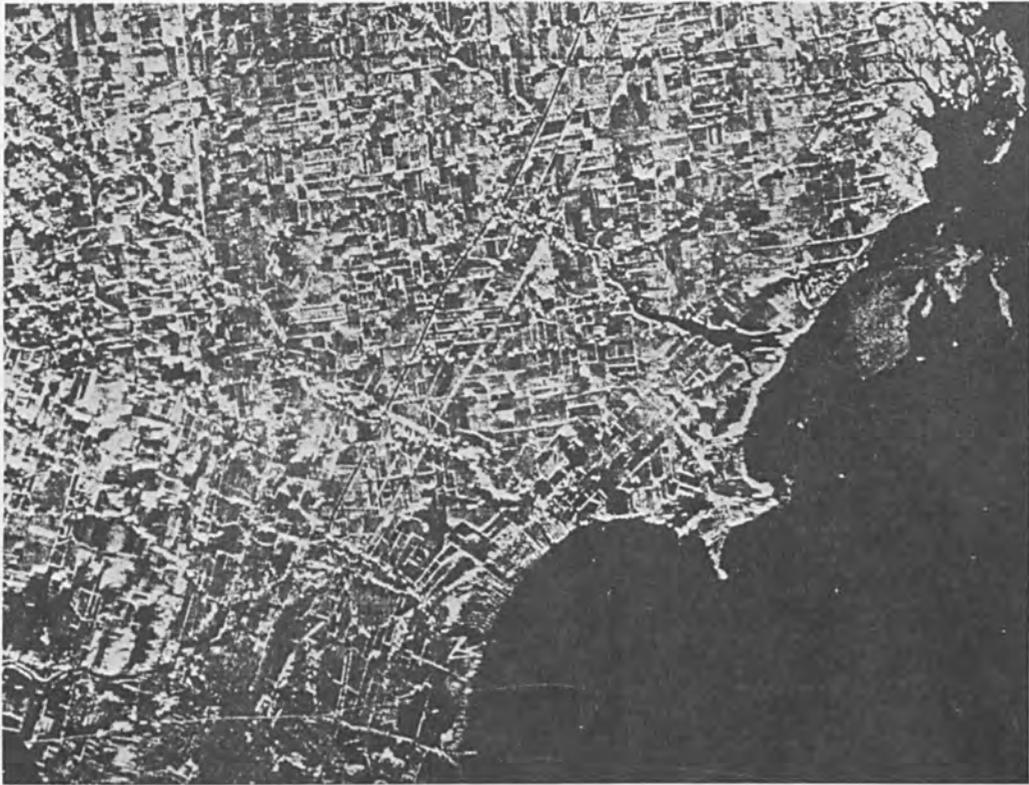


Figure 21c.

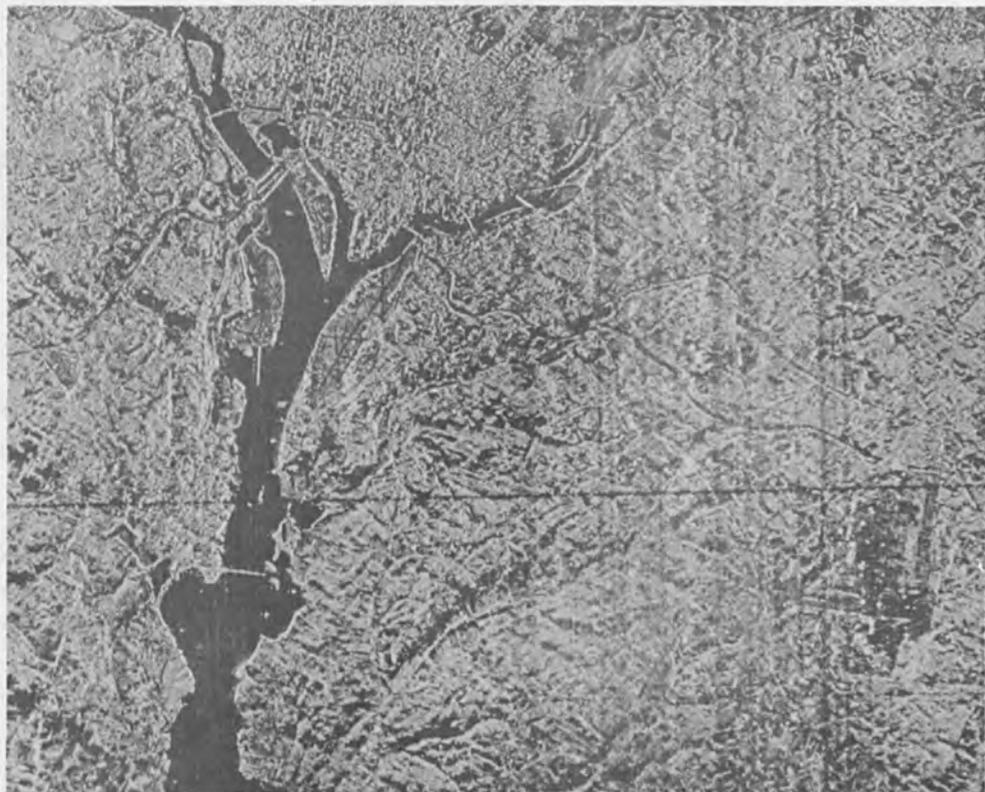


Figure 21d.

shown in Figure 22. In such a system the source generates the information which is eventually delivered to the destination. The source may be any of a number of items. It may be oral as in voice communication; it may be a picture as in TV transmission; it may be a series of digits if digital data is transmitted; it may be an analog voltage.

The output of the source is sent to a coder. It is the object of the coder to perform such operations on the output of the source as will adapt these signals to the transmission channel. Often a transducer of some type is included in the coder such as a microphone in the case of oral sounds, a TV camera for scenes, so that the information is converted into electrical form. Additional operations are also performed. The most common of these is to impress the information onto some electromagnetic carrier. Many types of modulation exist, such as amplitude modulation, frequency modulation, and a number of pulse modulations.

These operations are primarily for the purpose of adapting the signal to the channel. However, in more sophisticated communications systems additional coding is employed for a variety of purposes. For example, such additional coding may be used for the purpose of signal-to-noise improvement, error correction, etc.

The output of the coder is fed to the communications channel. Two general types of channels are employed, in one case a wire system connects the transmitter to the receiver; in the other case (radio transmission) an electromagnetic wave is launched by way of a transmitting antenna and a receiving antenna associated with the receiver abstracts from this electromagnetic energy. A wide range of frequencies is available for either wire transmission or radio transmission. Recently, this spectrum has been extended to include visual frequencies so that modulation of light sources such as lasers may now be considered.

In either wire transmission or radio transmission, noise is added to the signals. Often, this is additive noise. However, multiplicative noise and multi-path transmission are operations (generally undesirable) which may also be performed by signals as they pass through the channel.

The signals from the channel become the input to a receiver. It is the function of the receiver to detect the signal and to convert it into a form suitable for acceptance by the destination. If coding for signal-to-noise improvement or error correction has also been included in the coder, the operations necessary to achieve these improvements are also performed by the decoder.

Not shown in Fig. 22 but sometimes used and necessary are a

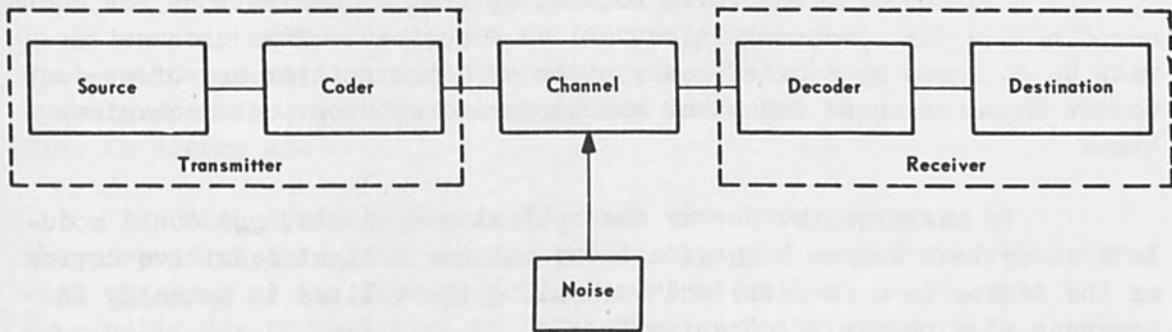


Figure 22. Essential Components of A Communications System.

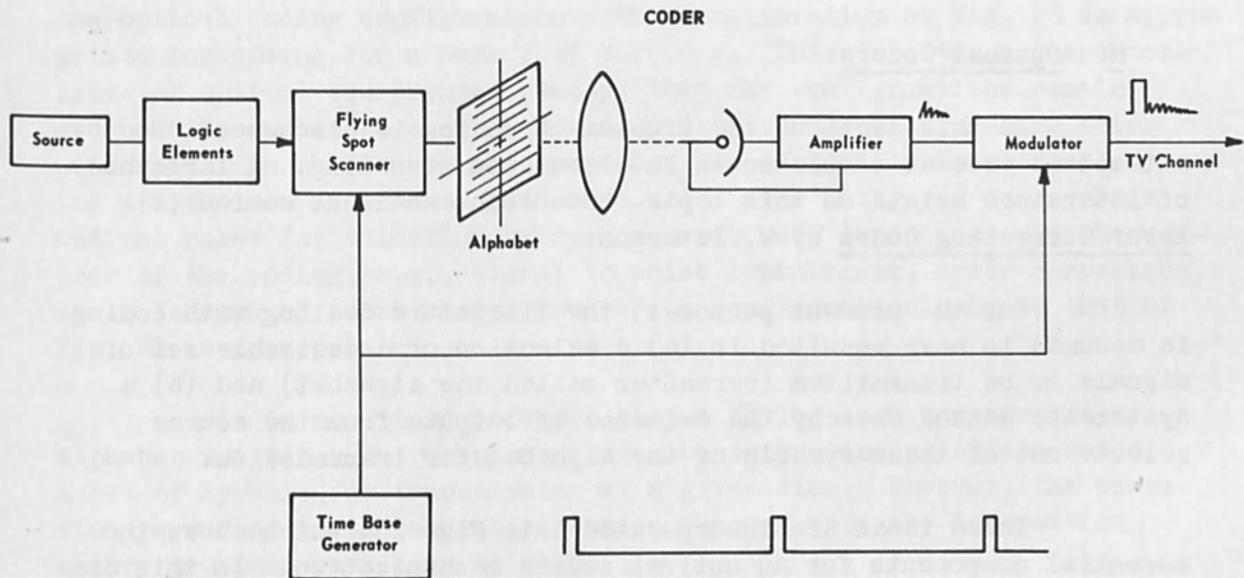


Figure 23. Optical Coding.

number of administrative equipments. These equipments perform such functions as synchronization of transmitter and receiver, automatic testing, automatic switching to a standby unit in case of failure, etc. Of these administrative functions, only that of synchronization will be discussed further in this report.

In the sections which follow, optical mechanizations for coding, decoding and for synchronization will be described. This discussion will be followed by a brief description of a transmitter and of a receiver in which these functions are performed with optical mechanizations.

To maximize the use of the optical components, one could modulate an optical source such as a laser and use a light sensitive device as the sensor in a receiver and work along these lines is actually in progress at a number of organizations.

L. Optical Systems

In this section, optical mechanizations for coding, decoding and synchronization will be discussed. The coder and some of its functions were described previously. The role of the decoder and of the synchronization circuits were also discussed above. In the paragraphs which follow, a more detailed discussion of these operations and of means available for their mechanization are discussed.

M. Optical Coders

In this section, the problem of coding is discussed. No attempt is made at completeness in discussion of coding. A large body of literature exists on this topic. One such excellent source is: Error Correcting Codes by W. Peterson.

For the present purposes, the literature dealing with coding is assumed to have resulted in (a) a selection of a desirable set of signals to be transmitted (hereafter called the alphabet) and (b) a systematic method whereby the sequence of outputs from the source selects one of these symbols of the alphabet for transmission.

These ideas are incorporated into Fig. 23, which shows the essential components for an optical coding mechanization. In this diagram, the source is shown feeding an equipment labeled "logic elements". The output of this device controls the flying spot scanner (or equivalent). Depending upon the output of the logic elements, the flying spot scanner will select and readout one of the symbols of the alphabet shown in Fig. 23. As this flying spot scanner scans the specified line in this alphabet, a photosensitive device receives the output and converts

the light to electrical form. The role of information theory and of coding theory is to arrive at an appropriate alphabet (set of waveforms) and a logic for selecting the proper symbol in response to the output of the source.

For the present purposes, it is assumed that the alphabet consists of a number of one-dimensional waveforms written along parallel lines onto a transparency. As the output of a source varies, the flying spot scanner reads out in accordance with this logic a sequence of such symbols so that a sequence of waveforms forms the input to the amplifier in Figure 23.

Also shown in Figure 23 is a time base generator. It is the purpose of this generator to synchronize the sampling of the source with the scan of the flying spot scanner, as well as to generate a sync pulse for transmission in cases when such a synchronization pulse is used. The configuration of Fig. 23 assumes that synchronization information is transmitted along with the signal information. Thus, an output from the time base generator is fed to the modulator together with the amplifier output. These are summed and put onto an appropriate carrier and fed to the channel.

It will be noted that except for the selection of an alphabet and for the logic of selecting a specified symbol of the alphabet, in response to outputs from the source, there is a general invariance to the optical coding configuration. The configuration of Fig. 23 is appropriate for coding for a number of purposes. This is often a characteristic of optical equipments; namely, that the configurations remain essentially unchanged even when widely different operations are being performed. In this case, the operation is essentially that of selecting a symbol in accordance with some rule. The nature of the symbols and the rules for selection of symbols will vary depending upon the purpose of the coding, e.g., signal to noise improvement, error correction, etc. In each of these cases, however, a configuration such as that of Fig. 23 is appropriate.

It is not to be inferred that a flying spot scanner and an alphabet written on a transparency are the only means to select one of a set of symbols for transmission at a given time. However, the operation performed is the essential one in a coder once the information theory problems have been solved.

N. Optical Decoders

In this section, a discussion of the decoding operation is given. In general, the decoder performs operations inverse to those of the coder. Specifically, it is the function of the decoder to recognize

which symbol of the alphabet was received and then employ a logic element to generate output appropriate for sending to the destination. A number of optical mechanizations for decoding exist. In what follows in this section, it is assumed that the coder has a finite number of symbols in its alphabet and that, therefore, the decoder must recognize only this finite set of symbols. The information regarding this set of symbols is written on the reference function in Figure 24.

The configuration shown in Fig. 24 is essentially a multi-channel cross correlator in which the received signal is cross correlated simultaneously with each of the symbols of the alphabet. The output will have a maximum value when the received signal is correlated against the symbol which has the identical waveform. Thus, an equipment which recognizes the maximum output signal determines which symbol was transmitted and this, in turn, can be used to convert this symbol into the form appropriate for sending to the destination.

The signal from the channel in Fig. 24 is fed to an ultrasonic light modulator. This ultrasonic light modulator is illuminated coherently by a light source S through collimating lens L_1 .

The signal from the channel launches a compressional wave down the ultrasonic light cell. This causes a phase modulation of the light transmitted through the ultrasonic light modulator. Lenses L_2 and L_3 image the ultrasonic light cell upon the reference functions. Thus, the received symbol is imaged simultaneously upon each of the reference functions in plane P_2 . Looking to the left through the reference function in plane P_2 , one sees the product of the received symbol by each of the reference functions.

The combination of cylindrical lens L_4 with lens L_5 performs a multi-channel spectrum analysis of the signals emerging from plane P_2 . If $s(x)$ represents the symbol in the ultrasonic light modulator in plane P_1 and $r_n(x)$ represents each of the symbols in plane P_2 , then a typical signal in plane P_2 is the product $s(x-vt)r(x)$, where the displacement variable "vt" accounts for the signal motion through the ultrasonic cell. The distribution of light in plane P_3 is given by Eq. (21)

$$E_o(\omega, n, t) = \int s(x-vt) r_n(x) e^{-j\omega x} dx. \quad (21)$$

This result is well known.^{5,6,7,8}

If only the light in a central slit in plane P_3 is observed, this corresponds to setting $\omega = 0$. In this case, one obtains Eq. (22).

$$E_o(0, n, t) = \int s(x-vt) r_n(x) dx \quad (22)$$

Equation (22) is seen to be the simultaneous cross correlation of $S(t)$ with each of the $r_n(t)$ which form the alphabet.

In writing both Eqs. (21) and (22) synchronization has been assumed.

An alternate mechanization for the decoder is given in Figure 25. Superficially, the equipment in Fig. 25 resembles that in Figure 24. However, in this case there is a single lens between planes P_1 and P_2 and the lens L_2 is a cylindrical lens and lenses L_2 and L_3 cause the one-dimensional spectrum of the signal in the ultrasonic light modulator to be incident upon plane P_2 . Moreover, plane P_2 contains not the reference functions but their one-dimensional transforms. Thus, looking toward the source through plane P_2 , one has the product given by

$$S(\omega) e^{-j \frac{\omega x}{v}} R_n(\omega). \quad (23)$$

$S(\omega)$ represents the Fourier transform of $s(x)$, $R_n(\omega)$ represents the Fourier transform of $r_n(x)$ and the exponential factor is associated with $S(\omega)$ as the Fourier transform of $s(x-vt)$. In Eq. (23) and Eq. (24)

$$\int s(x-vt) e^{-j\omega x} dx = S(\omega) e^{-j\omega vt} \quad (24)$$

x is a parameter indicating the location of the signal in the ultrasonic light modulator and v is the speed of propagation of this signal in the ultrasonic cell. Eq. (24) is of course the Fourier translation theorem.

The distribution of light in plane P_3 is again the multi-channel spectrum analysis of the signals in plane P_2 , that is, the Fourier transform of Eq. (23). The distribution of light in plane P_3 is given by Eq. (25).

$$\begin{aligned} e(x,n) &= \int S(\omega) e^{-j\omega vt} R_n(\omega) e^{-j\omega x} d\omega \\ &= s(x-vt-\tau) r_n(\tau) d\tau = \phi_{sr_n}(x-vt) \end{aligned} \quad (25)$$

In writing the second form of Eq. (25), use has been made of the translation theorem as well as the fact that the Fourier transform of a product is a convolution of the respective transforms.

It will be noted that the second form for Eq. (25) is also

a simultaneous cross correlation of the signal with each of the reference functions but that in this case synchronization is not necessary, the cross correlation function being translated along the t axis by the quantity vt . In this case the entire plane of P_3 must be searched for the maximum signal, or a slit may be used to monitor one "x" position as before. Searching an area rather than a line may be a disadvantage of Fig. 25 with respect to the equipment in Fig. 24. However, the configuration of Fig. 25 is advantageous for synchronization purposes inasmuch as a measurement of the location of this peak can be used to determine time to synchronization.

As was the case in Fig. 24, a maximum signal detection device as well as a logical device to convert received waveforms into an appropriate form for the source is also required.

The comments made with respect to Fig. 23 concerning the invariance of the configuration for the decoder apply to Figs. 24 and 25.

A change in coding simply requires a change in the signals placed into planes P_2 of these signals and of course the change in the signal received. The components otherwise are identical.

0. Optical Synchronization

In the above discussion, synchronization has been assumed. There is no real synchronization problem in the coder. However, in the decoder, particularly that given in Fig. 24, a reading of the correlation function at the proper time presupposed synchronization of the signal at each of the reference functions. Consequently, it is desirable to include means for recognizing synchronization so that readout of the decoder occurs only during such synchronization.

The configuration of Fig. 25, although described as one for performing simultaneous cross correlations of the received signal with the alphabet, was shown to give an output from which the location of the signal in the ultrasonic light cell could be determined. Thus, by using a recognizable code, one can determine the location of the synchronization signals in the ultrasonic delay line and predict the time that information signals will be in the correct positions. Thus, if a signal representing a synchronization signal is interwoven with message symbols, a configuration such as that of Fig. 25 can be used to determine the instant of synchronization for the message symbols and the message symbol determined by the use of equipments having configurations of that in Figs. 24 and 25. Thus, optical synchronization can be achieved.

There are a number of codes which are appropriate for

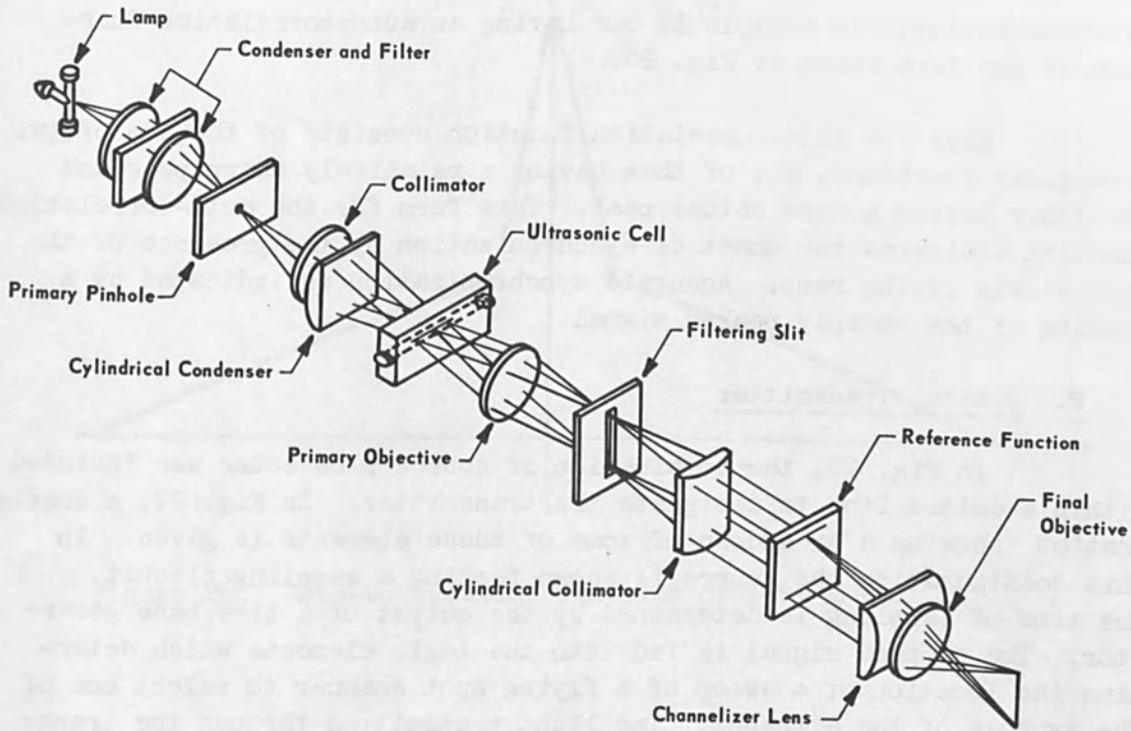


Figure 24. Optical Decoding (Cross-Correlation Configuration).

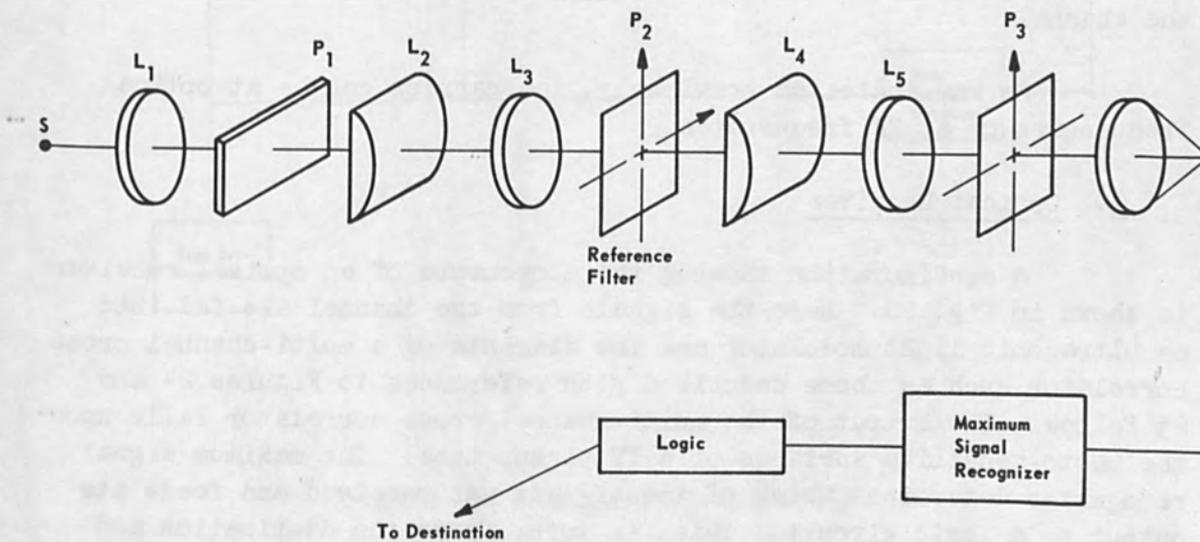


Figure 25. Optical Decoding (Matched Filter Configuration).

synchronization. An example is one having an auto-correlation function of the form shown by Fig. 26.

Here the auto-correlation function consists of the sum of two triangular functions, one of them having a relatively sharp peak and the other having a more obtuse peak. This form for the auto-correlation function indicates the onset of synchronization by the presence of the more slowly rising ramp. Accurate synchronization is indicated by a peaking of the sharply peaked signal.

P. Optical Transmitter

In Fig. 22, the combination of source plus coder was included within a dotted line to designate the transmitter. In Fig. 27, a configuration showing a breakdown of some of these elements is given. In this configuration the source is shown feeding a sampling circuit. The time of sampling is determined by the output of a time base generator. The sampled signal is fed into the logic elements which determine the location of a sweep of a flying spot scanner to select one of the symbols of the alphabet. The light transmitted through the transparency upon which this alphabet is written is converted to electrical form by a photoelectric sensitive device and amplified. The output of the amplifier is fed into a modulator. The output from a time base generator is fed to a synchronization signal generator. This synchronization signal is chosen to have desirable characteristics as a synchronization code. The output of this synchronization signal generator is also fed to the modulator. The modulator interweaves the sync signals with the message symbols, puts them onto a carrier, and feeds them to the channel.

As was indicated previously, the carrier can be at optical frequencies or at RF frequencies.

Q. Optical Receiver

A configuration showing the components of an optical receiver is shown in Fig. 28. Here the signals from the channel are fed into an ultrasonic light modulator and the elements of a multi-channel cross correlator such as those described with references to Figures 24 and 25 follow. The output of the multi-channel cross correlator falls upon the photo-sensitive surfaces of a TV pickup tube. The maximum signal recognizer determines which of the signals was received and feeds its output to a logic circuit. This, in turn, feeds the destination and regenerates the information from the source. For purposes of synchronization a network recognizing the occurrence of a sync pulse is shown. The output of this device feeds timing signals to the logic elements so that they examine the outputs from the receiver only at times of

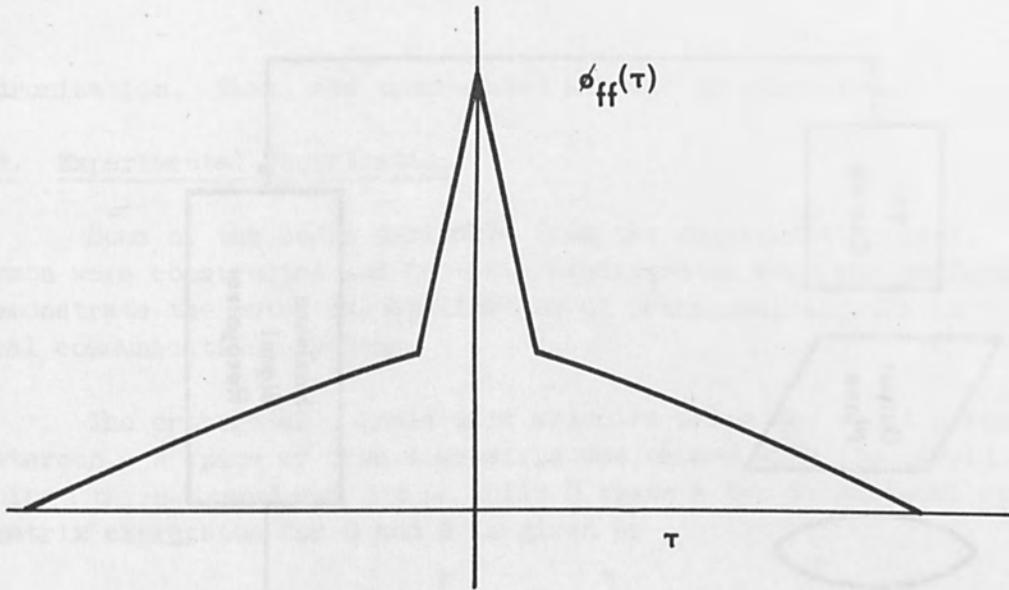


Figure 26. Auto-Correlation Function of Sync Code.

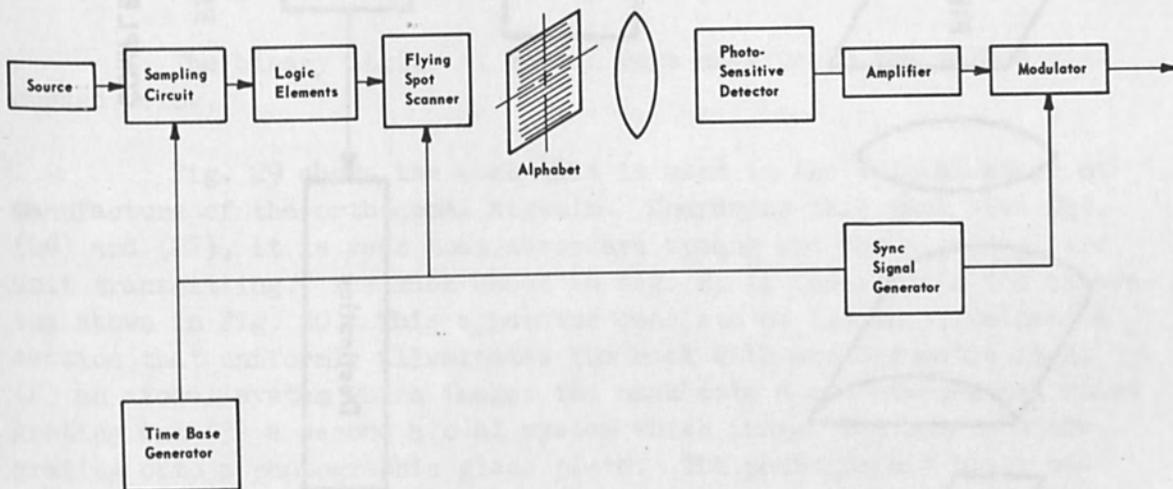


Figure 27. Complete Transmitter.

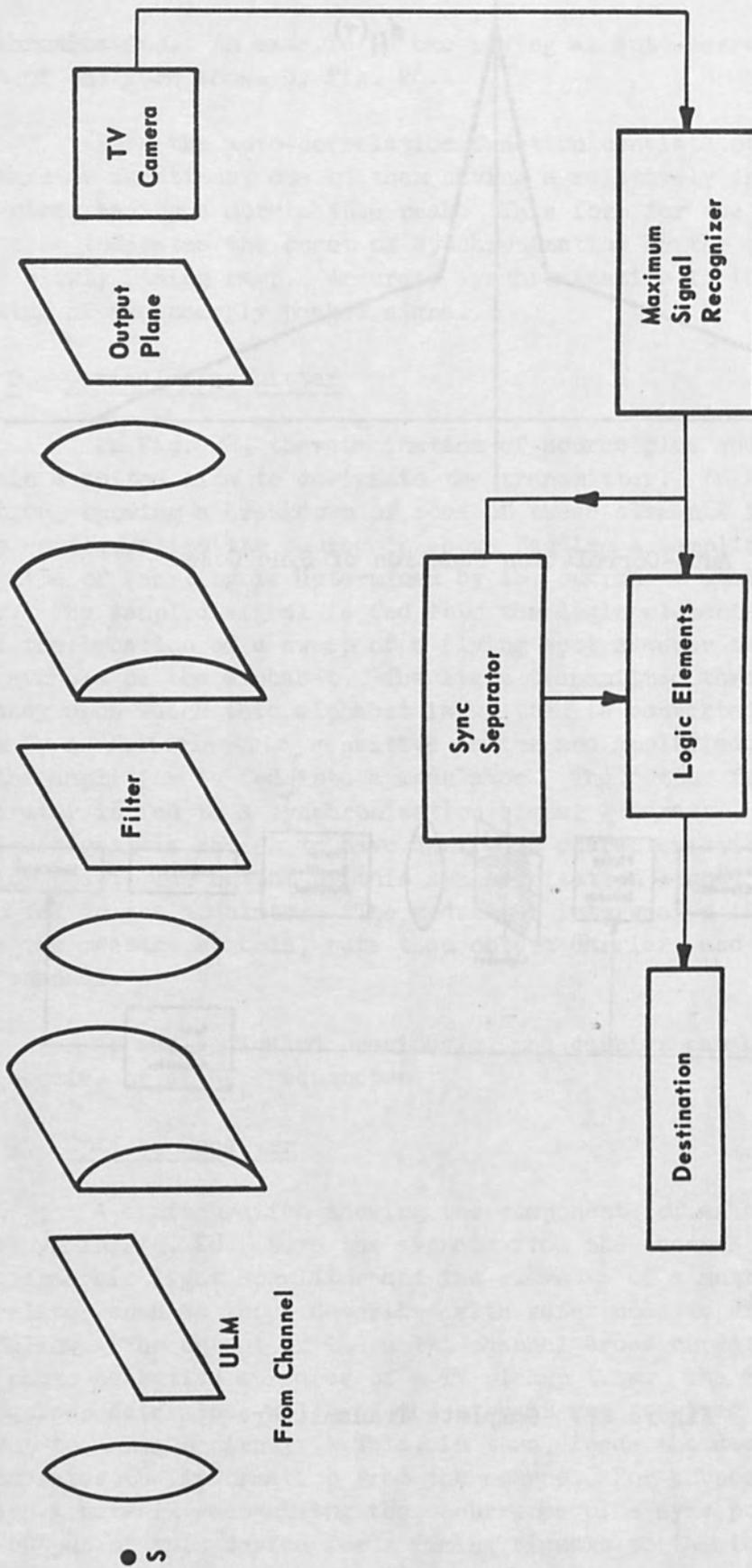


Figure 28.

COMPLETE RECEIVER

synchronization. Thus, the transmitted message is recovered.

R. Experimental Verification

Some of the codes derivable from the consideration by W. Peterson were constructed and tested. Experimental work was performed to demonstrate the potential application of orthogonal signals in optical communications systems.

The orthogonal signals were selected using the rules given by Peterson. A space of five dimensions was chosen with the matrix G spanning a three-dimensional space, while H spans a two dimensional space. The matrix expression for G and H is given by

$$G = [I_3P] = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (26)$$

$$H = [-P^T I_2] = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \end{bmatrix} \quad (27)$$

It can be shown by direct multiplication that the product of $GH^T = HG^T = 0$. Zeros can be synthesized optically by blanks or opaque areas.

The binary digits +1 and -1 were made using the method discussed below.

Fig. 29 shows the mask that is used in the initial stage of manufacture of the orthogonal signals. Comparing this mask with Eqs. (26) and (27), it is seen that zeros are opaque and the digits +1 are unit transmitting. The mask shown in Fig. 29 is inserted in the apparatus shown in Fig. 30. This apparatus consists of (1) an illumination section that uniformly illuminates the mask with monochromatic light (2) an afocal system which images the mask onto a one-dimensional ruled grating and (3) a second afocal system which images the mask and the grating onto a photographic glass plate. The photographic plate has an optical tolerance of $\lambda/4$ deviation from plane over the entire aperture. The emulsion is Kodak 649-F.

The grating may be displaced transverse to the optical axis of the apparatus by the precision micrometer.

The procedure used to expose the photographic glass plate is listed below.

1. Cover the "negative" or -1 apertures so that only the +1

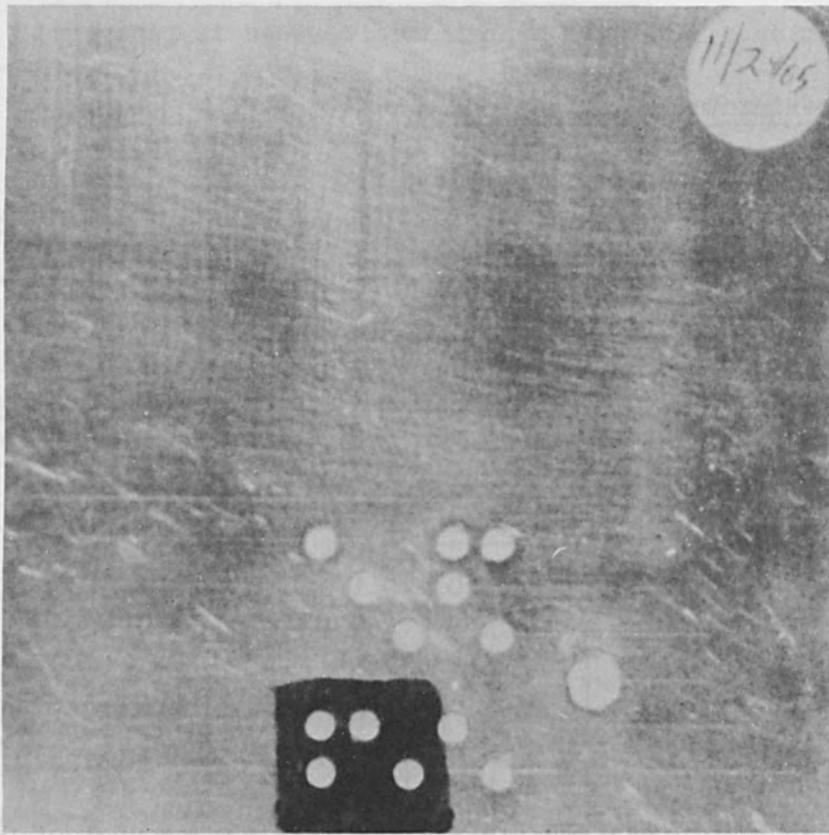


Figure 29. Mask for Making Optical Orthogonal Codes.

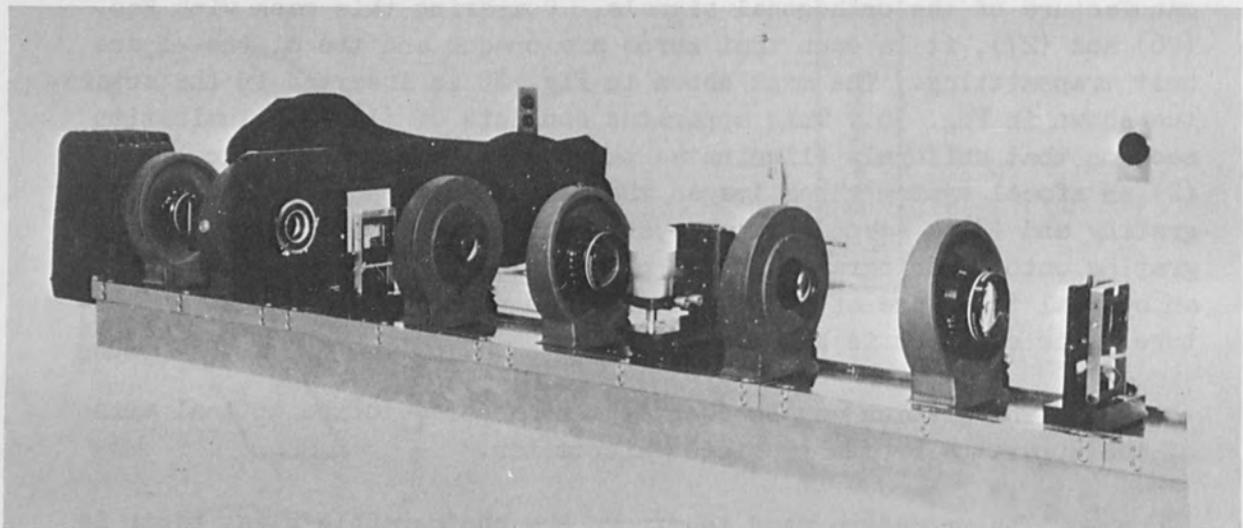


Figure 30. Equipment for Making Optical Orthogonal Codes.

apertures are imaged onto the grating. Expose the photographic plate.

2. Cover the "positive" or +1 apertures so that only the -1 apertures are imaged onto the grating. Use the precision micrometer to displace the ruled grating $1/2$ cycle. Thus the opaque portions of the grating now occupy the space previously occupied by the transparent portions of the grating. Expose the photographic plate. The developed photographic plate is shown in Figure 31.

The result of the above procedure is to record +1 in the form of plane waves as shown below. Neglecting higher order harmonic terms

Record	Plane Waves
1	$e^{i0}, e^{ikx}, e^{-ikx}$
-1	$e^{i0}, e^{ikx+i\pi}, e^{-ikx+i\pi}$

where $k = 2\pi v$

$v \equiv$ spatial frequency of grating line, pairs/mm

$x \equiv$ coordinate axis transverse to the optical axis.

Experimental apparatus was assembled to demonstrate

- (1) the capability to record binary phase
- (2) the capability to correctly multiply two orthogonal signals.

The apparatus is shown in Figure 32. The laser beam is divided into two beams; one preceding to the left through a condensing lens, pinhole and collimating lens, the other to the right through a neutral density filter, condensing lens, pinhole and collimating lens.

The collimated beam preceding to the left is reflected by two mirrors onto the first signal (matrix G) which is in a liquid film gate to minimize phase errors. The succeeding optics Fourier transforms spatially filter and image the filtered matrix onto the second signal (matrix H) also in a liquid film gate. The multiplication of matrixes G and H are Fourier transforms spatially filtered and transformed onto the polaroid film holder.

Sketch 33 shows the detail of the spatial filters and the

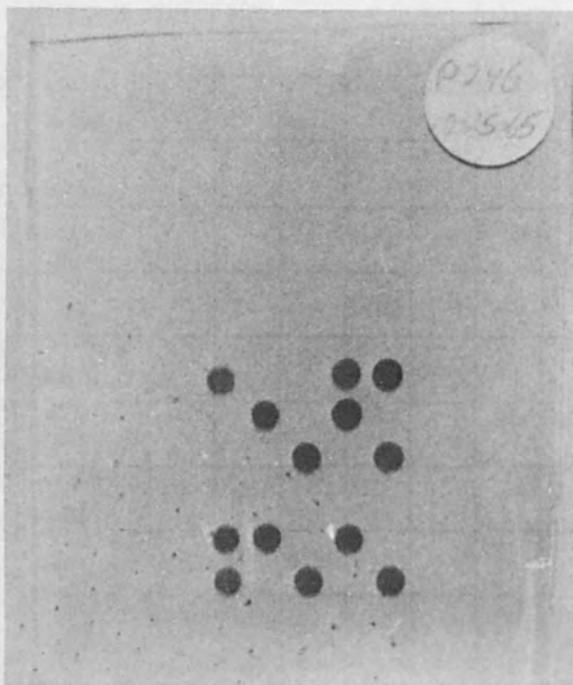


Figure 31. Optical Orthogonal Codes.

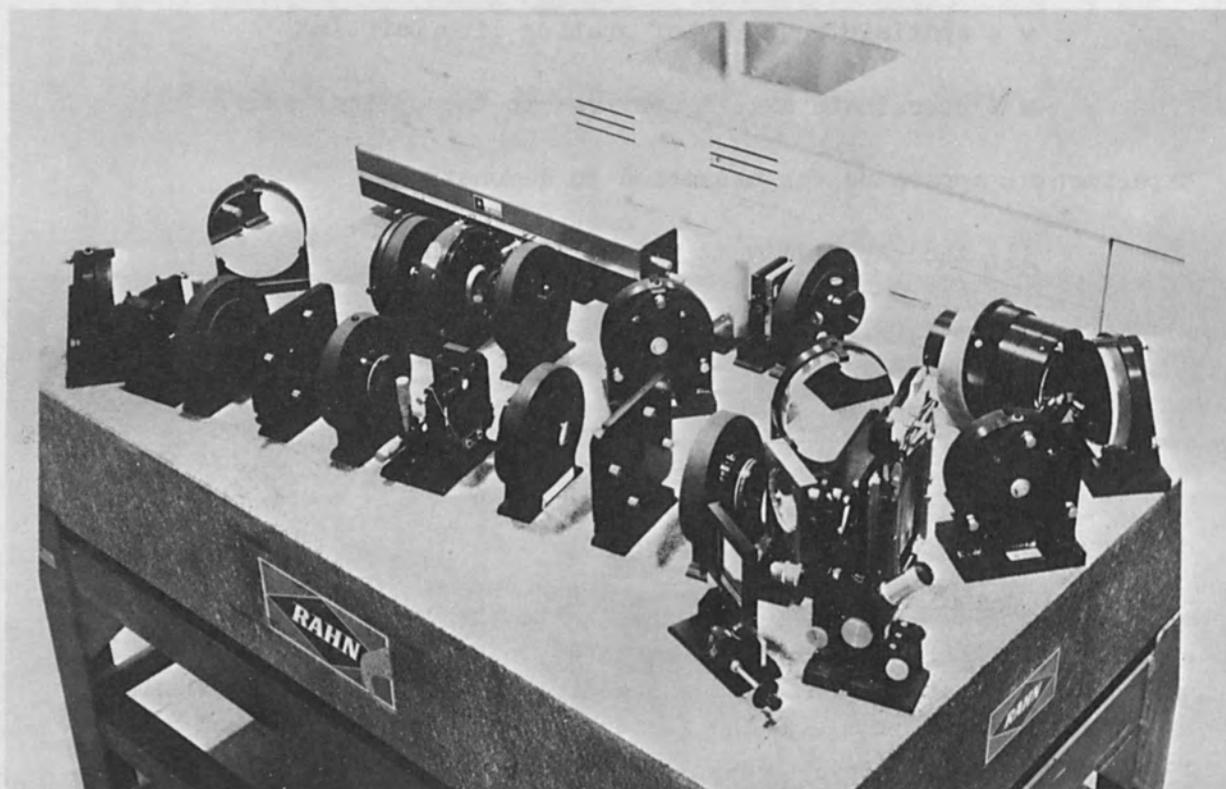


Figure 32. Orthogonal Code Processor.

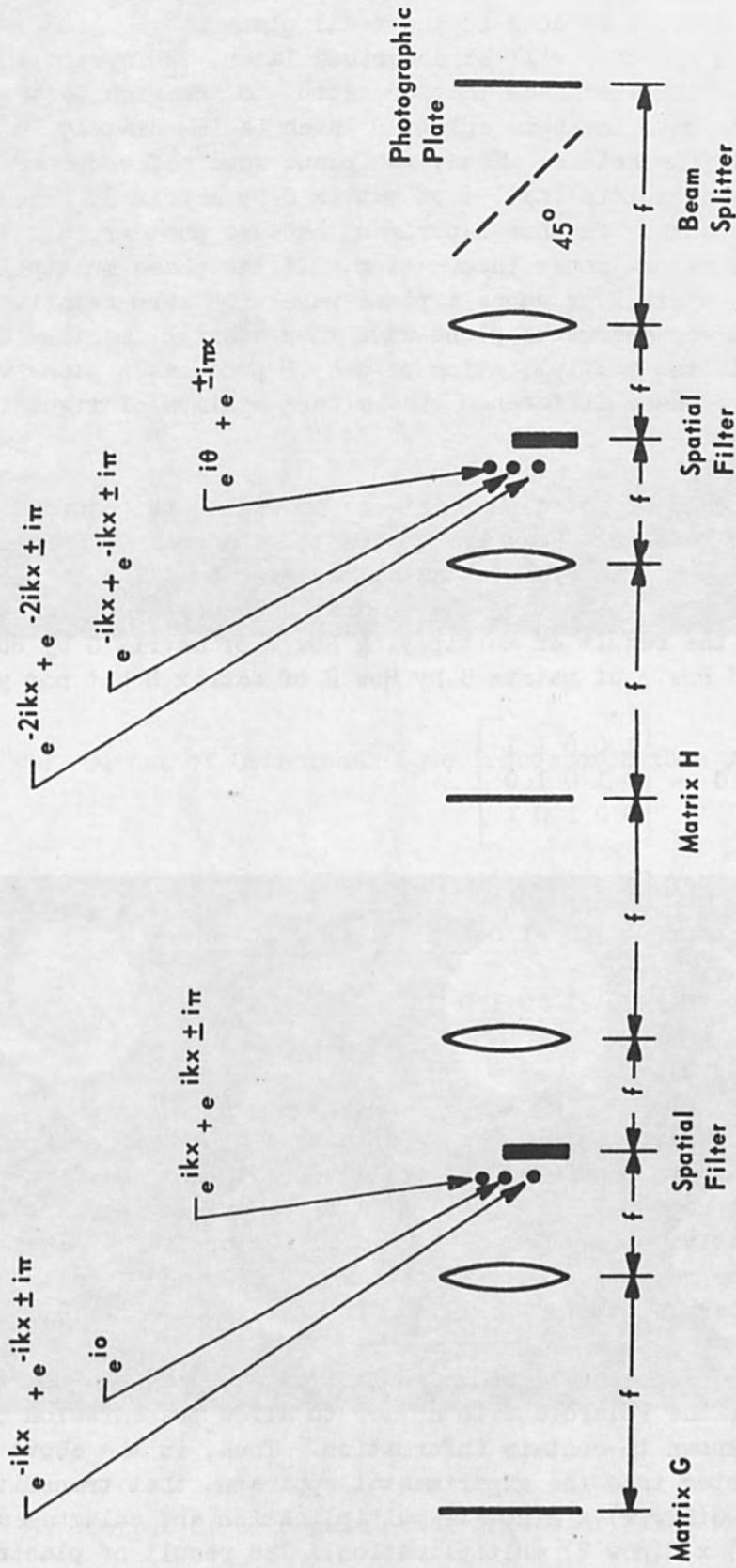


Figure 33.

OPTICAL SCHEMATIC OF ORTHOGONAL CODE PROCESSOR

plane waves that are filtered. The Fourier transform of the various plane waves is denoted by dots in the focal plane of the lens. The function of the beam splitter will be described later. Referring again to Figure 32, the collimated beam that preceded to the right is reflected by these mirrors onto the beam splitter which is immediately in front of the polaroid film holder. Thus, the plane wave reference beam is added to the phase multiplication of matrix G by matrix H. The reference beam is necessary in this experiment because photographic film cannot directly record phase information. If the phase multiplication of matrix G by matrix H produces a plane wave with zero relative phase difference with the reference plane wave then a bright maximum of light is recorded. If the multiplication of G by H produces a plane wave with $\pi(180^\circ)$ relative phase difference then a dark minimum of light is recorded.

It should be noted that extreme precision is required so that the signals and reference beam are collimated and coaxial. If the reference beam is tilted from the axial position by $\lambda/2$ over the matrix aperture the recorded output of the apparatus becomes meaningless. Shown below is the result of multiplying Row 2 of matrix G by Row 1 of matrix H and Row 3 of matrix G by Row 2 of matrix H but not yet summing.

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \end{bmatrix}$$

$$G(\text{Row } 2) \times H(\text{Row } 1) = 0 \quad -1 \quad 0 \quad 1 \quad 0$$

$$G(\text{Row } 3) \times H(\text{Row } 2) = 0 \quad 0 \quad -1 \quad 0 \quad 1$$

Figure 34 shows the output of the experimental apparatus as a result of performing the above mentioned multiplications. In the top row a dark spot appears in column 2 indicative of a -1; a bright spot appears in column 4 indicative of a +1. The same analysis applies to the lower Row which is the result of multiplying matrix G(Row 3) by matrix H(Row 2). Column 3 is a dark spot (-1); column 5 is a bright spot (+1).

If the signals are known a priori, a mask may be inserted in the position of the Polaroid film holder to allow transmission of only those regions known to contain information. Thus, in the above example a mask is inserted into the experimental apparatus that transmits column 2 and 4 of the $G(\text{Row } 2) \times H(\text{Row } 1)$ multiplication and columns 3 and 5 of the $G(\text{Row } 3) \times H(\text{Row } 2)$ multiplication. The result of placing this mask in the position of the image of the multiplication, i.e., position

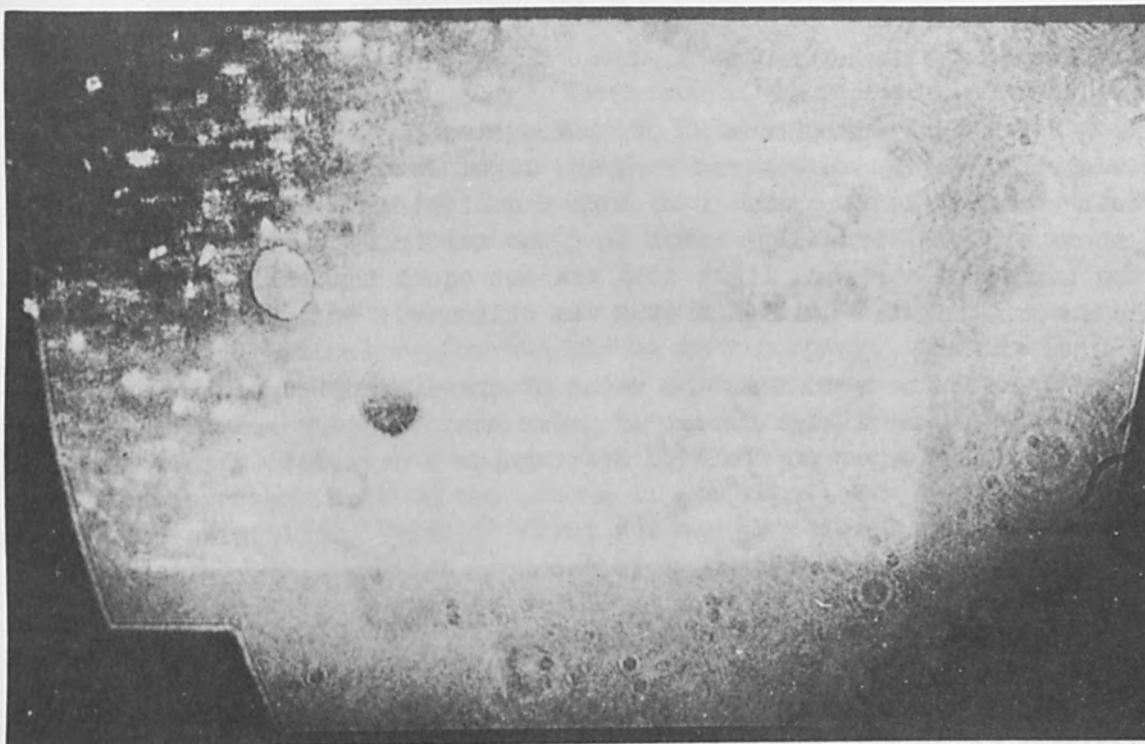


Figure 34. Output of Orthogonal Code Processor Before Masking.

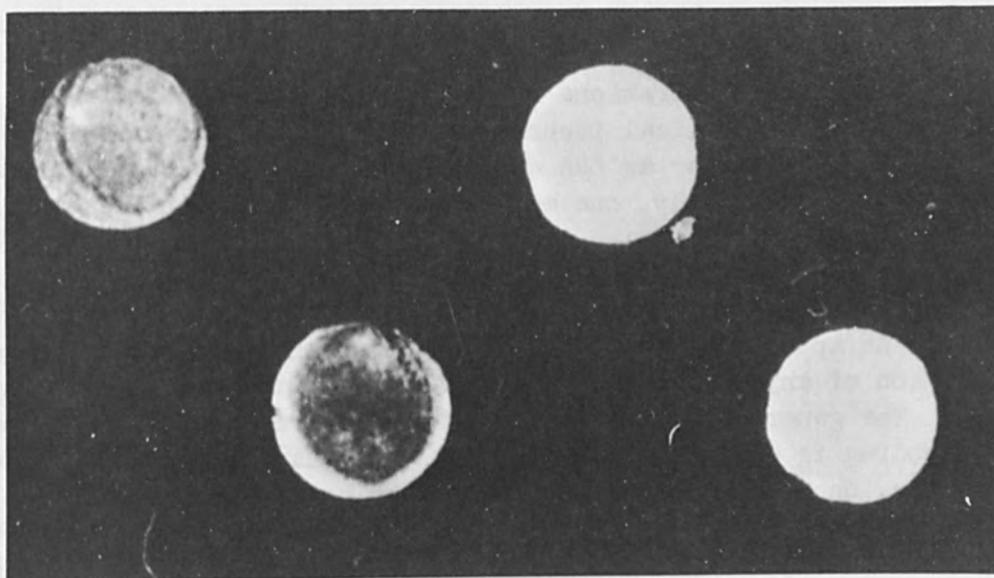


Figure 35. Output of Orthogonal Code Processor After Masking.

of the Polaroid film holder, is shown in Figure 35.

A suggested refinement of this system is to incorporate a logic circuit which can discern dark and light dots. If the logic circuit counted the same number of dark and light dots in a row (as in the above example) it would reason that the multiplication was orthogonal. If the number of dark and light dots was not equal the logic circuit would reason that the multiplication was colinear.

It can be seen that the value of these error correcting orthogonal codes is that a high amount of noise must be added to the system before it is no longer possible to distinguish a dark dot (-1) from a light dot (+1).

III. CONCLUSIONS

The discussion above has considered a number of optical techniques in the design of communications equipment. The operations described were those of coding, decoding and synchronization by optical means. In addition, reference was made to the use of an optical carrier for the transmission of the signals from one point to another. The merits or disadvantages of optical mechanization as compared to non-optical mechanizations were not discussed. It is appropriate to make such a comparison at this point.

Information theory considerations indicate that the result of specified operations on signals achieve certain results. Such considerations indicate that the results are independent of the mechanization provided identical operations are performed. On this basis, the use of optics over non-optical techniques results in no advantage or disadvantage per se insofar as the characteristics of the output signal are concerned. Consequently, one must look for other advantages or disadvantages.

Consider, for example, the complexity of codes which can be mechanized. An appropriate exercise in information theory must precede the conception of any appropriate code. Again, neither system has an advantage. The generation of a complicated waveform by non-optical means and its decoding is often a difficult task to mechanize. In the optical case, however, once the desired waveforms have been designed, one can draw and photograph them; one then has the essential ingredients of both the coder and of the decoder. Furthermore, if a code is to be changed periodically, the invariance of the equipments referred to in the text is a great advantage inasmuch as the changing of the code represents a change in a transparency in the transmitter and a transparency in the receiver. For non-optical change of codes, it may sometimes be necessary to replace large items of equipment.

Due to the ease of optical mask generation, one can mechanize more complicated codes optically. The non-optical mechanization would also result in a very complex equipment. In some cases, one could get finer performance by optical means than by electronic means, not because of any inherent advantage of the optics, but rather because of a self-imposed limitation on the complexity of a non-optical mechanization.

Finally, the size, cost and reliability of equipments should be considered. Optical equipments can be very compact. Equipments have been built in which 50 channels per mm have been recorded. This is not an upper limit but represents present day comfortable technology. Thus, a 27 mm wide film can accommodate 1350 channels. A similar length in the orthogonal direction can accommodate a signal having a time bandwidth product of 1350. With 5" film, one can process 5000 simultaneous channels with time bandwidth products of 5000. These equipments are largely passive. Their size is compact and their flexibility is very great. It is believed that these advantages suffice to make probable the incorporation of some of these techniques in future communications systems.

REFERENCES

1. A. L. Ingals, "Optical Simulation of Microwave Antennas," presented at the 1964 PTGAP International Symposium, New York, September 21-24, 1964 and published in the Program and Digest, pp. 203-208.
2. L. J. Cutrona, E. N. Leith, L. J. Porcello, and W. E. Vivian, "On the Application of Modern Optical Techniques to Radar Data Processing," presented at the 9th Symposium of the AGARD Avionics Panel on Opto-Electronic Components and Devices, September 1965, Paris, France.
3. L. J. Cutrona, W. E. Vivian, E. N. Leith, and G. O. Hall, "A High Resolution Radar Combat-Surveillance System," IRE Transactions on Military Electronics, April 1961, pp 127-131.
4. L. J. Cutrona and G. O. Hall, "A Comparison of Techniques for Achieving Fine Azimuth Resolution," IRE Transactions on Military Electronics, Vol. MIL-6, No. 2, April 1962, pp 119-121.
5. L. J. Cutrona, E. N. Leith, C. J. Palermo, and L. J. Porcello, "Optical Data Processing and Filtering Systems," IRE Transactions on Information Theory, June 1960, pp 386-400.
6. Conductron Technical Documentary Report No. D-5210-70-T80-110, "Coherent Light Investigation, Vol. I," March 1964, prepared under Prime Contract AF33(615)-1014.
7. L. J. Cutrona, "Optical Computing Techniques," IEEE Spectrum, October 1964.
8. L. J. Cutrona, "Recent Developments in Coherent Optical Technology," 1964.
9. Emmett N. Leith, and Juris Upatnieks, "Wavefront Reconstruction with Continuous Tone Objects," J. of the Optical Society of America, Vol. 53, No. 12, December 1963, pp 1377-1381.
10. Bernard Friedman, Principles and Techniques of Applied Mathematics, John Wiley and Sons, Inc., New York and London, 1956.
11. L. A. Zadeh, "A General Theory of Linear Signal Transmission Systems," J. of Franklin Institute, 253, January-June 1952, pp 293-312.
12. M. Born and E. Wolf, Principles of Optics, Pergamon Press, New York, 1959.

13. P. Elias, D. Grey and D. Robinson, "Fourier Treatment of Optical Processes," J. of the Optical Society of America, Vol. 42, February 1952, pp 127-134.
14. T. P. Cheatham, Jr., and A. Kohlenberg, "Analysis and Synthesis of Optical Processes," Boston Univ. Physics Res. Labs, Boston, Mass., Technical Note 84, Part I, March 1952.
15. P. Elias, "Optics and Communication Theory," J. of the Optical Society of America, Vol. 43, pp 229-232, April 1953.
16. J. Rhodes, "Analysis and Synthesis of Optical Images," American J. of Physics, Vol. 21, January 1953, pp 337-343.
17. T. P. Cheatham, Jr. and A. Kohlenberg, "Optical Filters - Their Equivalence to and Differences from Electrical Networks," 1954 IRE National Convention Record, pp 6-12.
18. E. O'Neill, "The Analysis and Synthesis of Linear Coherent and Incoherent Optical Systems, Boston Univ. Physics Res. Labs, Technical Note 122, September 1955.
19. E. O'Neill, "Spatial Filtering in Optics," IRE Transactions on Information Theory, Vol. IT-2, June 1956, pp 56-65.
20. E. O'Neill, "Selected Topics in Optics and Communication Theory, Boston University Physics Research Laboratories, Boston, Mass., Technical Note 133, October 1957.
21. L. J. Cutrona, E. N. Leith, and L. J. Porcello, "Filtering Operations Using Coherent Optics," Proceedings of the National Electronics Conference, Vol. XV, October 1959.
22. L. J. Cutrona, E. N. Leith, and L. J. Porcello, "Data Processing by Optical Techniques," 3rd National Convention on Military Electronics Conference Proceedings, Washington, D. C., June 1959.
23. L. J. Cutrona, E. N. Leith and L. J. Porcello, "Coherent Optical Data Processing," 1959 IRE WESCON Convention Record, Part 4, pp 141-153, and IRE Transactions on Automatic Control, Vol. AC-4, No. 2, pp 137-149, November 1959.
24. B. A. Vander Lugt, "Signal Detection for Complex Spatial Filtering," IEEE Trans. on Information Theory, Vol. IT-10, No. 2, April 1964, pp. 139-145.
25. P. M. Woodward, Probability and Information Theory, with Applications to Radar, Pergamon Press, New York, 1960.

13. H. H. Gans and J. H. Goldstein, *Journal of Polymer Science*, **14**, 455 (1954).

14. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

15. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

16. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

17. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

18. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

19. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

20. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

21. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

22. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

23. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

24. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

25. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

26. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

27. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

28. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

29. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

30. T. P. Greenham, Jr. and A. K. Kricheldorf, *Journal of Polymer Science*, **14**, 455 (1954).

AIRBORNE INVESTIGATIONS OF CLEAR AIR
TURBULENCE WITH LASER RADARS*

by

P. A. Franken, J. A. Jenney, D. M. Rank
Physics Department
The University of Michigan
Ann Arbor, Michigan

ABSTRACT

Over 1300 experiments have been performed with a laser radar mounted in a light twin-engine airplane. Ancillary equipment included an accelerometer, thermistor, data recording equipment, and a forward-looking camera for recording the local meteorology. The studies were performed in mountain waves near Denver, in the severe storm activity of Oklahoma, and the local Detroit area. A description of the equipment will be given, together with examples of the data.

I. INTRODUCTION

This program was initiated to explore the possibility that characteristic optical radar echoes might actually be correlated with clear-air turbulence. A light twin engine airplane was equipped with a laser radar and ancillary equipment for monitoring acceleration, temperature variations, and relevant meteorological data. The design of this equipment and the development of the flight program were predicated on theoretical considerations of optical scattering from particulate matter.

A priori considerations, as well as actual experimental data, strongly support the recognition that severe turbulence of the type under investigation arises at the interface of air streams characterized by different temperatures and, quite likely, by different particulate content. The hypothesis underlying our work was that the resultant aerosol distributions in the turbulent region are characterized by some distinct changes in their optical radar responses. We had in mind, particularly, that there might be changes either in the number or size of the particles right in the region characterizing the turbulent interface between two different air streams.

This program was supported jointly by the Office of Naval Research, the United States Weather Bureau, and the Bureau of Naval

*This work was sponsored by the Office of Naval Research under Contract NONR 1224(51)

Weapons. The optical radar and ancillary equipment were built and installed by Lear Siegler Laser Systems Center, in an Army L23-D (a light twin engine Beech Bonanza) operated by the Willow Run Laboratories of the University of Michigan.

Data was gathered by flying into severe thunderstorm areas near Norman, Oklahoma, and mountain wave regions in the Denver, Colorado vicinity, as well as around the Detroit area. The results of these experiments indicate that a fine grain structure does often exist in atmospheric aerosols, and the various implications of this are discussed in Section 5.

A more detailed description of the equipment and these experiments, including composite photographs of the data from over 100 of the experiments, is available in a report¹ by the authors that has been prepared at the University of Michigan.

II. THEORY

A. Pulsed Laser Radar Sensitivity

The expected capability of a pulsed laser radar can be determined from quite general considerations of geometry, targets and noise sources. However, a few specific assumptions will be made about systems in order to develop equations directly appropriate to this program. A schematic presentation of the radar system and target is shown in Figure 2.1. The type of target of interest to the present work is thought of as some layer or discontinuity of particulate matter, at a range R , whose lateral extent is greater than the spread of the laser beam.

While it is quite possible to analyze an optical radar in terms of signal-to-noise power ratios, as with conventional microwave radars, it must be recognized that an optical receiver of the photomultiplier type is specifically sensitive to the individual photons in the light beam, rather than to the power per se. In the limit of strong signals, of course, the discussion can be perfectly well treated with traditional concepts of signal power, equivalent noise power, etc. However, in the very weak signal case the fact that the receiver is literally sensitive to individual photons suggests the utility of describing light in terms of the number of photons involved rather than the energy.

The effects of Johnson noise and dark current for a modern photomultiplier tube are generally negligible for optical radar systems. Consequently the dominant noise phenomena that must be considered are statistical variations in the total number of photons arriving at the photocathode in the minimum response time of the circuitry (the "resolution time"), and the current noise due to the statistical emission of

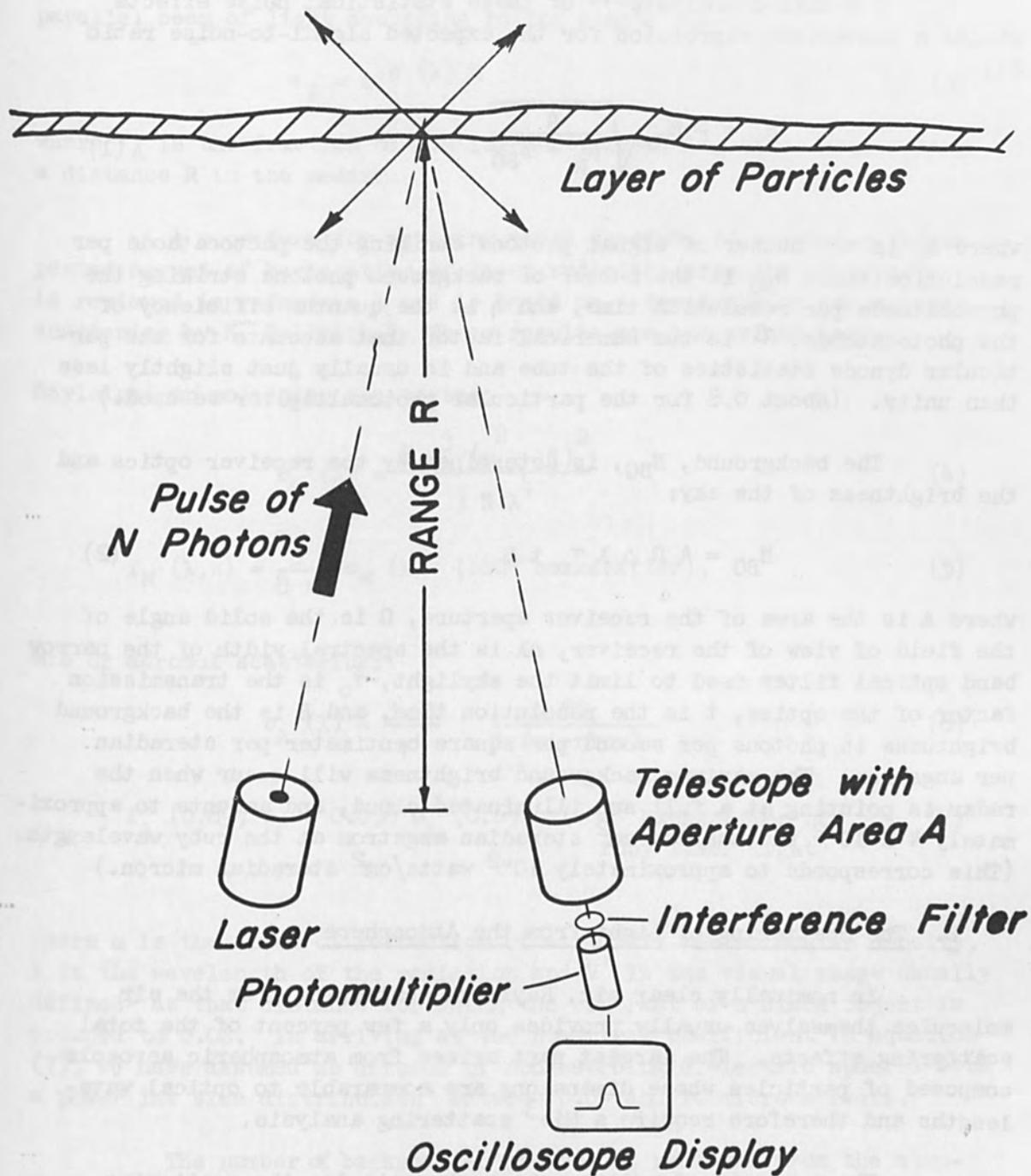


Figure 2.1. A Schematic Representation of the Detection of Particulate Matter in the Atmosphere with a Laser Radar.

electrons by the photocathode and the dynodes.

A detailed analysis^{1,2} of these statistical noise effects yields a convenient expression for the expected signal-to-noise ratio S' :

$$S' = \beta N_R \sqrt{\frac{\eta}{N_R + N_{BG}}} \quad (1)$$

where N_R is the number of signal photons striking the photocathode per resolution time, N_{BG} is the number of background photons striking the photocathode per resolution time, and η is the quantum efficiency of the photocathode. β is the numerical factor that accounts for the particular dynode statistics of the tube and is usually just slightly less than unity. (About 0.8 for the particular photomultiplier we used.)

The background, N_{BG} , is determined by the receiver optics and the brightness of the sky:

$$N_{BG} = A \Omega \Delta \lambda \tau_o t B \quad (2)$$

where A is the area of the receiver aperture, Ω is the solid angle of the field of view of the receiver, $\Delta \lambda$ is the spectral width of the narrow band optical filter used to limit the skylight, τ_o is the transmission factor of the optics, t is the resolution time, and B is the background brightness in photons per second per square centimeter per steradian per angstrom. The maximum background brightness will occur when the radar is pointing at a full sun illuminated cloud, and amounts to approximately 4×10^{12} photons/sec cm^2 steradian angstrom at the ruby wavelength. (This corresponds to approximately 10^{-2} watts/ cm^2 steradian micron.)

B. The Scattering of Light from the Atmosphere

In nominally clear air, Rayleigh scattering from the air molecules themselves usually provides only a few percent of the total scattering effects. The largest part arises from atmospheric aerosols composed of particles whose dimensions are comparable to optical wavelengths and therefore require a Mie³ scattering analysis.

There are two functions which characterize the scattering process for the present considerations. The first is usually called the scattering function, $f(\lambda, \phi)$, and is the fraction of radiation scattered per unit path length per unit solid angle in a given direction ϕ . The second, often designated as the scattering coefficient $\sigma(\lambda)$, is obtained by integrating the scattering function over a sphere

and is the total attenuation per unit path length due to the scattering process. This scattering coefficient describes the attenuation of a parallel beam of light according to the simple law:

$$\tau_A = e^{-\sigma(\lambda) R} \quad (3)$$

where τ_A is the fraction of the intensity remaining after penetrating a distance R in the medium.

A procedure for applying these concepts to computing the expected amount of backscatter of laser radiation from the atmosphere is reviewed in reference 1 and is based on a treatment of atmospheric scattering by K. Bullrich.⁴ These results are summarized here:

Rayleigh or molecular scattering:

$$\sigma_M(\lambda) = \frac{8 \pi^3 (m^2 - 1)^2}{3 N \lambda^4} \quad (4)$$

$$f_M(\lambda, \pi) = \frac{3}{8 \pi} \sigma_M(\lambda) \quad (180^\circ \text{ backscatter}). \quad (5)$$

Mie or aerosol scattering:

$$\sigma_A(\lambda) = \frac{3.91}{V'} \frac{0.55}{\lambda \text{ (microns)}} \quad (6)$$

$$f_A(0.69, \pi) = 0.038 \sigma_A(0.69) \quad (180^\circ \text{ backscatter of ruby laser light}), \quad (7)$$

where m is the index of refraction of air, N is the molecular density, λ is the wavelength of the radiation and V' is the visual range usually defined⁵ as that distance for which the contrast of a black object is reduced to 0.02. In arriving at the numerical coefficient in equation (7), we have assumed an aerosol of nonabsorbing dielectric spheres with a power law size distribution⁴ between 0.04 and 10 microns radii.

The number of backscatter photons N_R received from the atmosphere at a range R per resolution time t can be computed from the expression:

$$N_R = N_T f \frac{A}{R^2} \frac{ct}{2} \ell \quad (8)$$

where N_T is the total number of transmitted photons (and the transmission

time is assumed less than t), A is the receiver aperture area, c is the velocity of light, and l is the loss factor equal to the product of the receiver optics transmission τ_o and the square of the atmospheric transmission τ_A . The symbol f is either $f_M(\lambda)$ or $f_A(\lambda)$ or the sum of the two, depending on whether one wishes to calculate molecular, aerosol or total scattering. The signal amplitude at the receiver output is then calculated in the usual manner from the characteristics of the photomultiplier tube and the receiver electronics.

III. SYSTEM DESIGN AND PERFORMANCE

The principal components of the system used for this program were a laser radar, an oscillograph that recorded both the vertical acceleration of the airplane and fluctuations in air temperature, and a forward camera that photographed the cloud conditions in the area probed by the laser beam. A block diagram of the system is shown in Figure 3.1.

A basic laser radar and most of the ancillary equipment were built by Lear Siegler, Inc. and installed by them into the passenger compartment of an Army L23-D (Twin Beech Bonanza) operated by the University of Michigan Willow Run Laboratories. The oscillograph (flight recorder), accelerometer and bridge network were obtained on loan from the Navy, and were modified as necessary.

A. Laser Transmitter

The transmitter was a ruby laser, Q-switched with a liquid saturable filter composed of vanadium phthalocyanine dye dissolved in nitrobenzene. It was capable of producing plane polarized pulses of 10^{17} to 3×10^{17} photons in approximately 30 nanoseconds (roughly one to three megawatts peak power). The beam divergence was such that approximately 90 percent of the energy was contained in a full angle of about $1/3$ of a degree. Air cooling of the ruby rod and flashlamps permitted the laser to be fired three times per minute.

B. The Receiver

The receiver consisted of a Questar telescope coupled to an EMI 9558A photomultiplier tube via a 10 \AA bandpass interference filter and a camera shutter that was automatically opened just before the laser was fired. The effective aperture area of the telescope is 5.8×10^{-2} square feet (corresponding to an effective diameter of $3 \frac{1}{4}$ inches) and its full angular field of view is one degree. The photomultiplier is a factory selected tube with a stated quantum efficiency of about 4 percent at 6943 \AA .

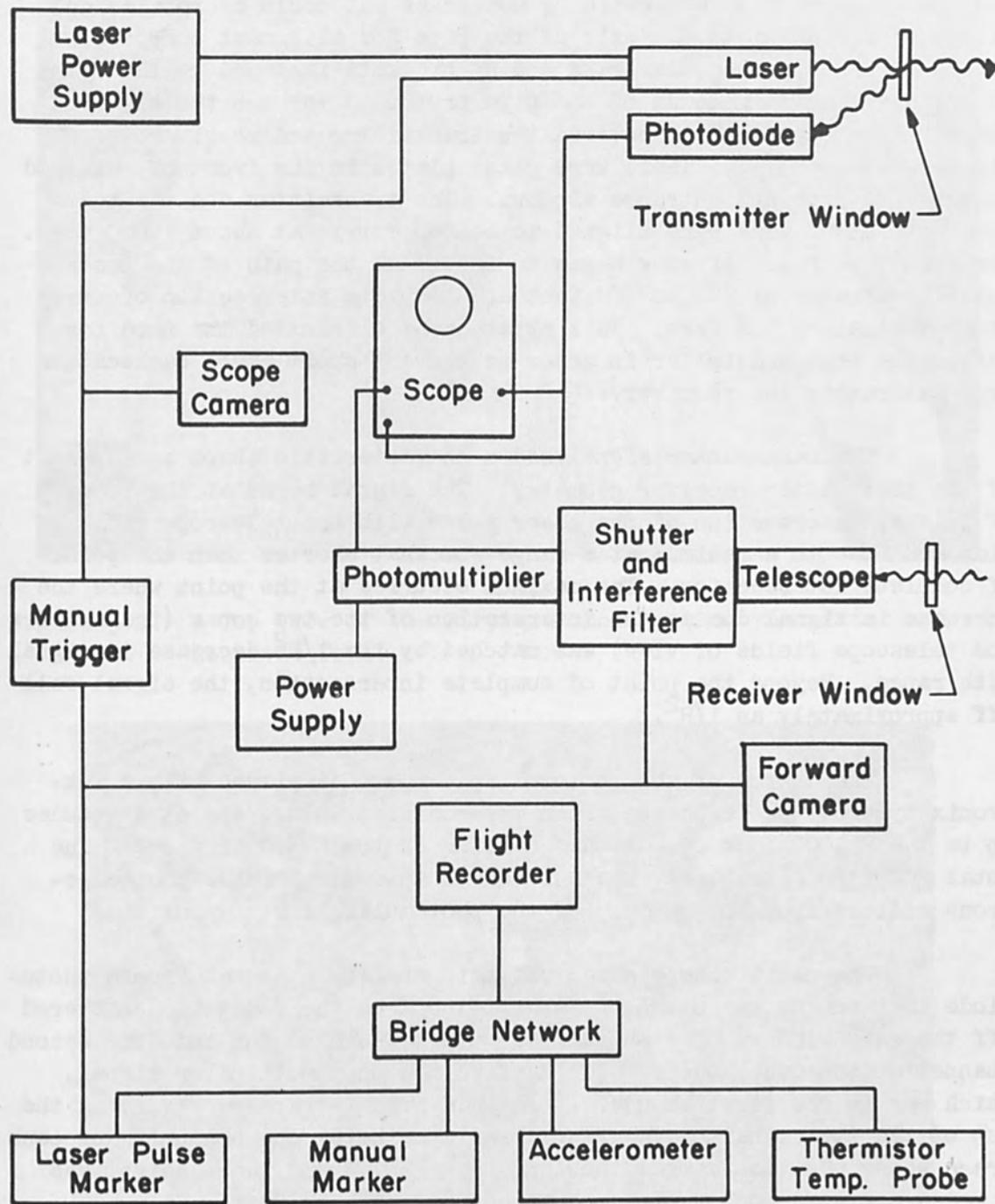


Figure 3.1. A Block Diagram of the Integrated Airborne System.

The transmitter and receiver were mounted on opposite ends of a 3 inch diameter aluminum pipe. The optic axes of the transmitter and the receiver were separated by 60 inches but could be rotated and tilted with respect to the axis of the pipe for alignment purposes. The pipe was supported near each end by brackets fastened to the frame of the airplane. The ends of the pipe protruded through the sides of the plane into pods which covered the transmitter and receiver (see Figures 3.2 and 3.3). There were glass plates in the front of each pod to serve as exit and entrance windows. The transmitter and the telescope fields of view were aligned to be concentric at about 1,000 feet. The telescope field of view began to intersect the path of the laser pulse at a range of 200 to 300 feet and complete intersection occurred at approximately 500 feet. This arrangement eliminated the need for gating the photomultiplier in order to prevent short range backscatter from saturating the receiver.

The backscatter signal had a characteristic shape as a result of the transmitter-receiver geometry. The signal began at the point of initial intersection of the laser pulse with the telescope field of view and rose to a maximum at a range somewhat shorter than the point of complete intersection. The maximum occurred at the point where the increase in signal due to the intersection of the two cones (transmitter and telescope fields of view) was matched by the $1/R^2$ decrease in signal with range. Beyond the point of complete intersection, the signal fell off approximately as $1/R^2$.

The output of the photomultiplier was displayed with a Tektronix type 647 oscilloscope which was modified to operate at altitudes up to the 22,000 foot operational ceiling of the L23-D airplane. The total receiver sensitivity was sufficient to detect single photoelectrons emitted from the cathode of the photomultiplier.

The oscilloscope sweep was triggered by a signal from a photodiode that sensed the outgoing laser pulse from the radiation scattered off the exit window. The photodiode output was also put into the second channel of the oscilloscope and added to the photomultiplier signal, which was in the first channel, forming a signal trace representing the sum of the two signals. The result was a spike at the beginning of each trace whose position marked the zero range point and whose height was a measure of the number of transmitted photons.

C. Ancillary Apparatus

A fully automatic camera photographed the visible meteorological conditions in the direction of each laser pulse. An optical beam oscillograph recorded accelerometer and temperature probe outputs and a tape recorder was used to keep track of the experimental conditions for each laser shot.

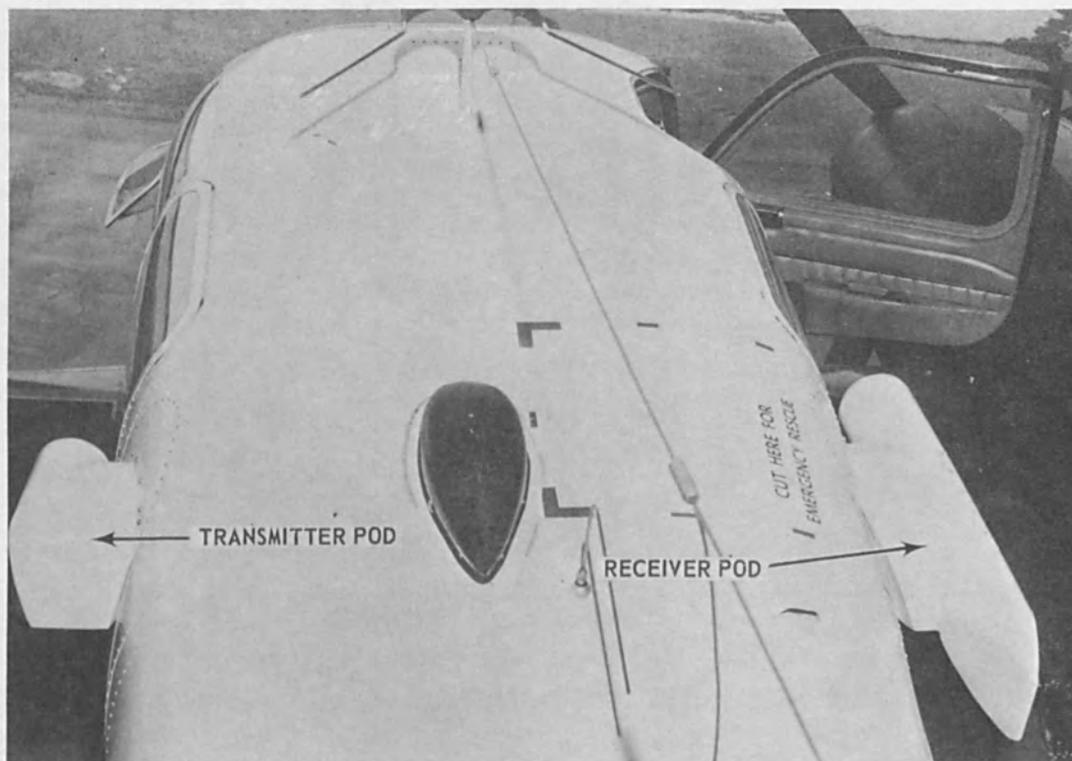


Figure 3.2. A Top View of the Airplane showing the transmitter and Receiver Pods.

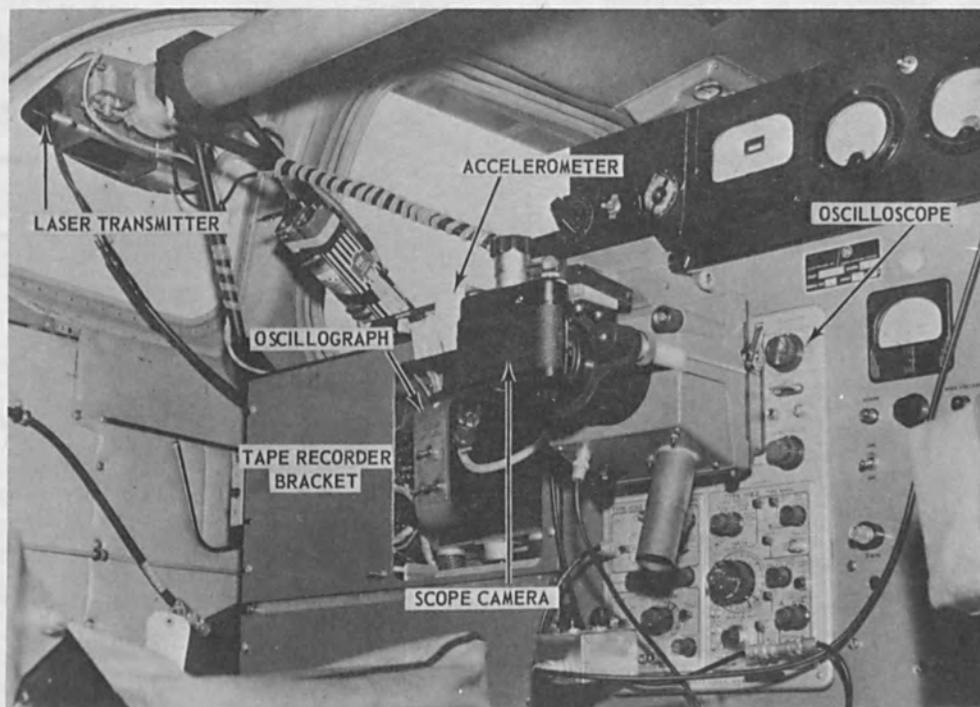


Figure 3.3. The Data Display and Recording Area in the Rear of the Airplane. The Tape Recorder and Film Magazine for the Oscillograph were not in Place in this Photograph.

D. Calibration of the Transmitter and Receiver

The performance of the optical radar system was established by measurements of the laser output and the receiver sensitivity. The photodiode monitoring the laser output was calibrated with a Lear Siegler Laser Systems Center calorimeter accurate to \pm three percent. The receiver sensitivity was determined by reflecting the laser beam from a diffuse screen of known reflectivity and comparing the return signal to the number of transmitted photons.

E. Experimental Measurements of Aerosol Backscatter

A number of backscatter experiments were performed on July 29, 1965, including the variation of backscatter with altitude, the low level horizontal variations and a comparison of rural and urban areas. The rural area was between Ann Arbor and Toledo and the urban area was Detroit. Three average values of the total scattering function (aerosol and molecular) were obtained from these measurements and are plotted, along with the theoretically predicted total scattering function, against visual range in Figure 3.4.

F. Airborne Performance Checks of the Integrated System

The optical alignment of the radar transmitter and receiver were frequently checked by ranging on the ground while diving the airplane at about a 45° angle. Checks on the timing system of the oscillograph and our ability to fly through the area probed by the laser pulse were provided by ranging on smoke from a ground fire and the exhaust wakes of departing jet aircraft. Upon receiving an optical radar echo from such targets we would continue our flight in a straight line through the target area and record the temperature fluctuations and accelerations produced by the wake or smoke column.

IV. FLIGHT PROGRAM

We explored three sources of turbulence during the flight program. These were: 1) clear air turbulence adjacent to heavy thunderstorms in Oklahoma; 2) low level orographic or terrain turbulence in the Ann Arbor area; 3) mountain wave activity in the Denver area. Clear air turbulence associated with the jet stream was not explored because of the limited altitude capability of our aircraft.

The data from these flights which exhibited turbulence, distinct anomalies in the backscatter curve, or both, were combined into composite pictures. The accelerometer and temperature probe records

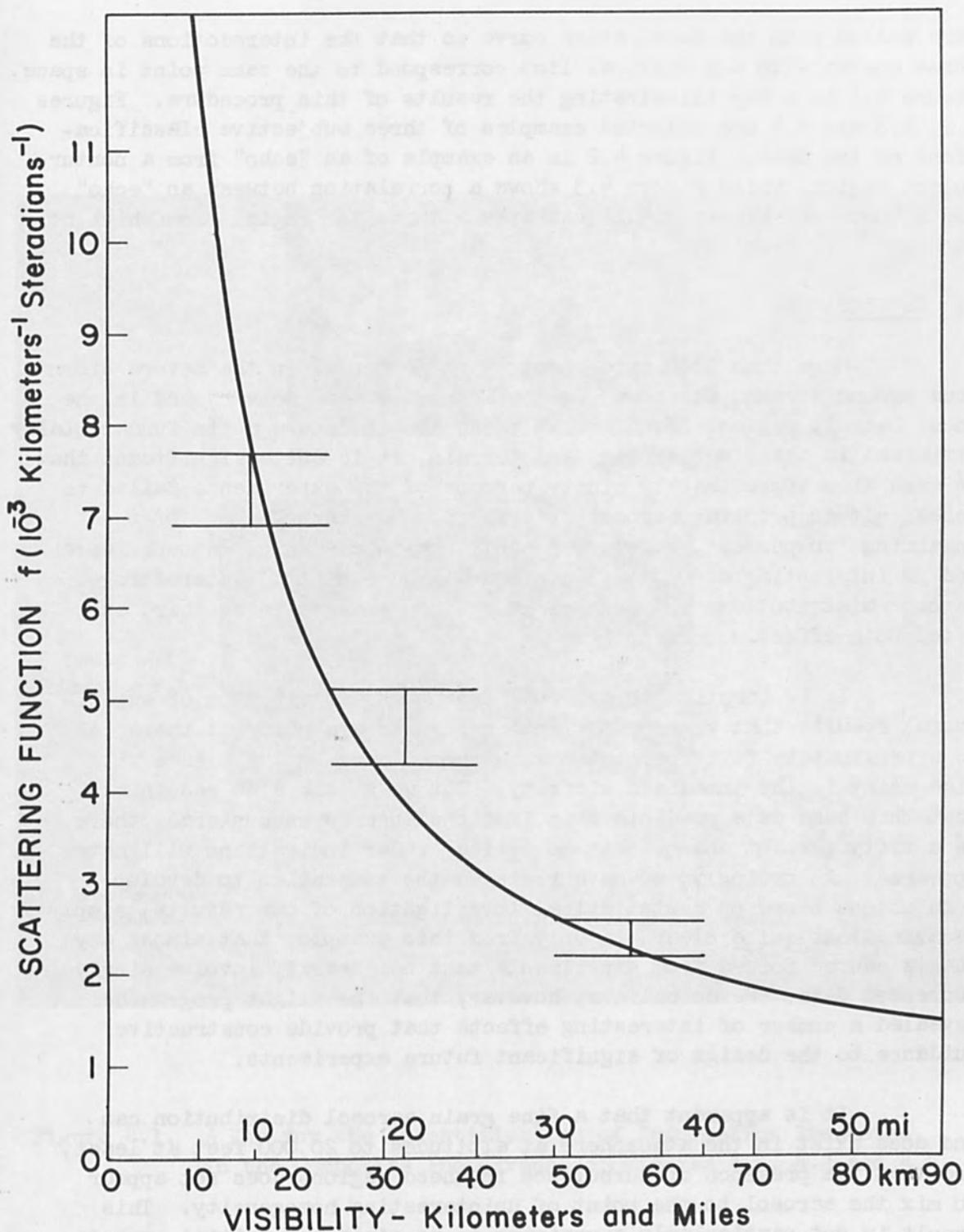


Figure 3.4. A Theoretical Plot of the Sum of Aerosol and Molecular Scattering Functions Versus Visibility for Ruby Laser Radiation and 180° Backscatter. Three Experimentally Determined Points are also Shown.

were scaled with the backscatter curve so that the intersections of the three curves with any vertical line correspond to the same point in space. Figure 4.1 is a key illustrating the results of this procedure. Figures 4.2, 4.3 and 4.4 are selected examples of three subjective classifications of the data. Figure 4.2 is an example of an "echo" from a nonturbulent region, while Figure 4.3 shows a correlation between an "echo" and a bump, and Figure 4.4 illustrates a turbulent region from which no "echoes" were received.

V. CONCLUSIONS

More than 1300 experiments were performed in the severe storm area around Norman, Oklahoma, in the mountains near Denver, and in the local Detroit region. Even though these locations are quite fundamentally different in their meteorology and terrain, it is quite significant that in each area approximately ninety percent of the experiments failed to reveal either peculiar aerosol distributions or turbulence. Of the remaining ten percent, however, roughly one-third exhibited turbulence and no interesting aerosol effects, one-third exhibited interesting aerosol distributions but no turbulence, and the remaining third exhibited both effect concurrently.

It is tempting to conclude from this distribution of experimental results that when optical radar signals are observed there is an approximately fifty percent chance that clear air turbulence will also exist in the immediate vicinity. But we should also recognize that this same data predicts that if turbulence is encountered, there is a fifty percent chance that no optical radar indications will have appeared. Accordingly, we have resisted the temptation to develop conclusions based on a statistical investigation of our results, simply because it is quite clear, if only from this example, that almost any thesis can be forced from experiments that necessarily involve singly connected data. We do believe, however, that the flight program has revealed a number of interesting effects that provide constructive guidance to the design of significant future experiments.

It is apparent that a fine grain aerosol distribution can and does exist in the atmosphere at altitudes to 20,000 feet at least, and that the presence of turbulence in these regions does not appear to mix the aerosol to the point of uninteresting homogeneity. This result is not particularly surprising to us since our initial expectation was that turbulent interfaces might indeed be characterized by detectable discontinuities in the aerosol. Unfortunately, however, the mere presence of aerosol discontinuities does not provide a unique signature for turbulent activity because the aerosol signals can apparently be associated with quite friendly atmospheres that simply contain sheets or striae of particulate matter.

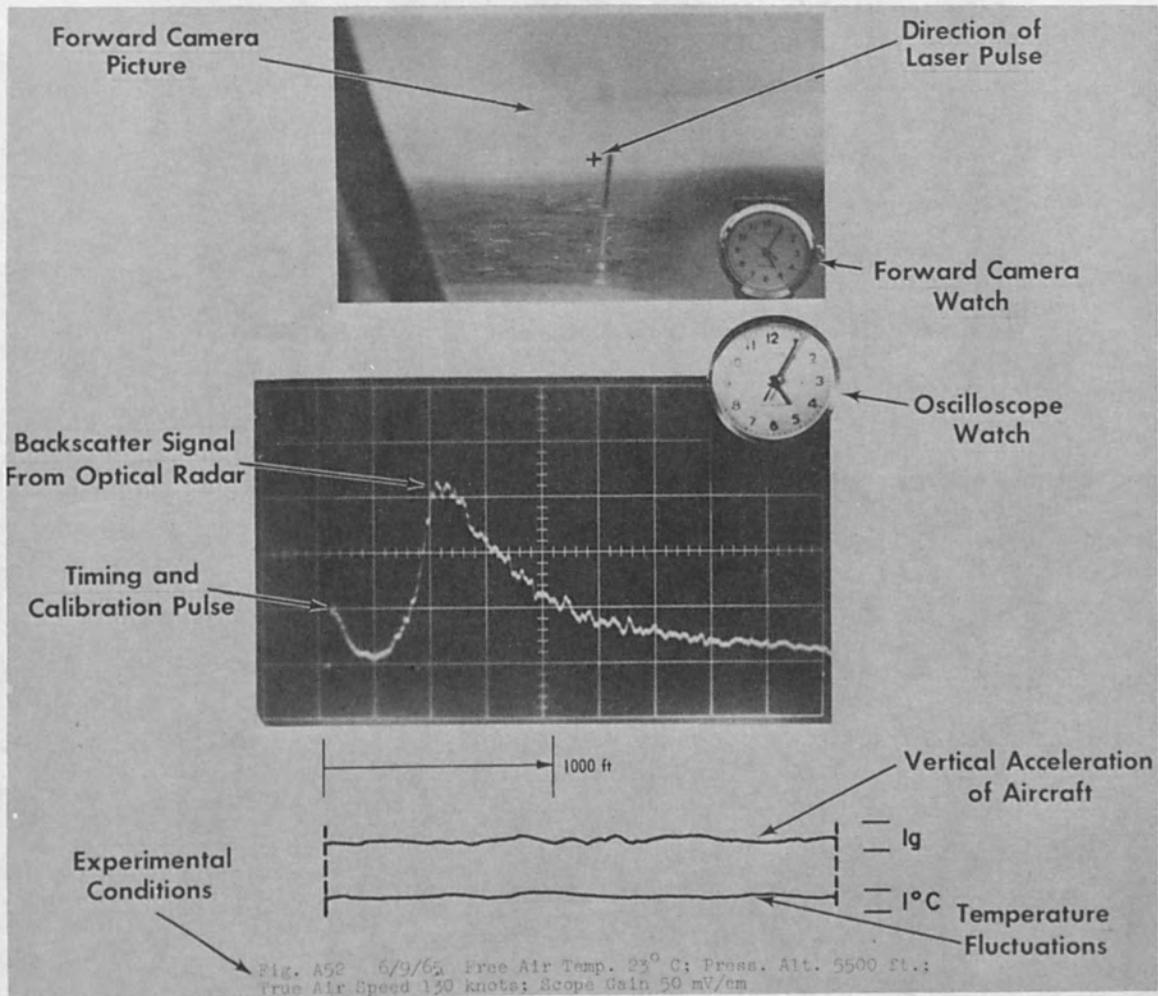
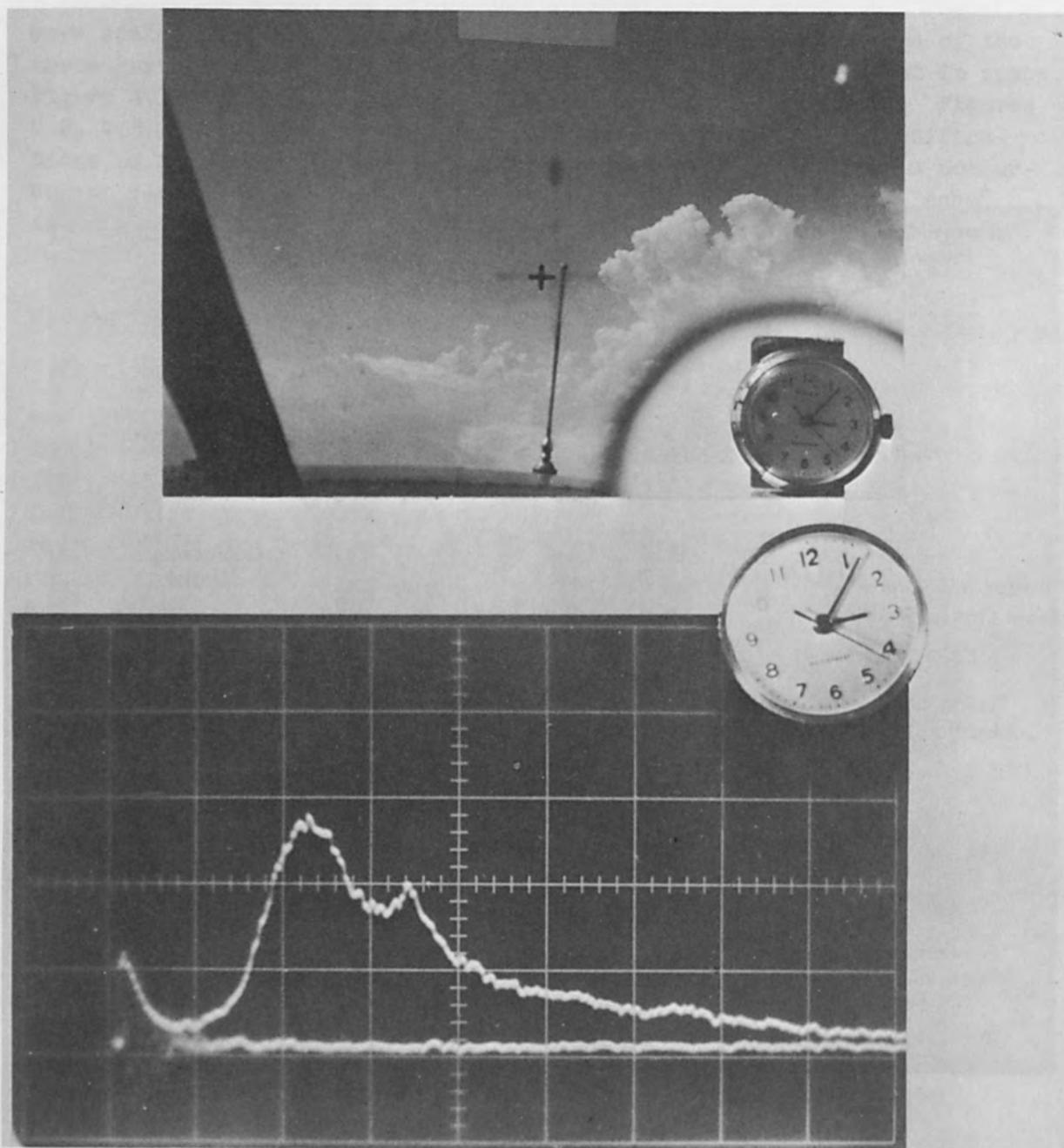


Figure 4.1. A Key for the Identification of Information Presented in the Composite Photographs of Figures 4.2, 4.3 and 4.4.



1000 ft

lg
1°C

Figure 4.2. 6/8/65 Near a Thunderstorm in Oklahoma; Free Air Temp. 10°C; Press. Alt. 15,800 ft.; True Air Speed 136 knots; Scope Gain 50 mV/cm.

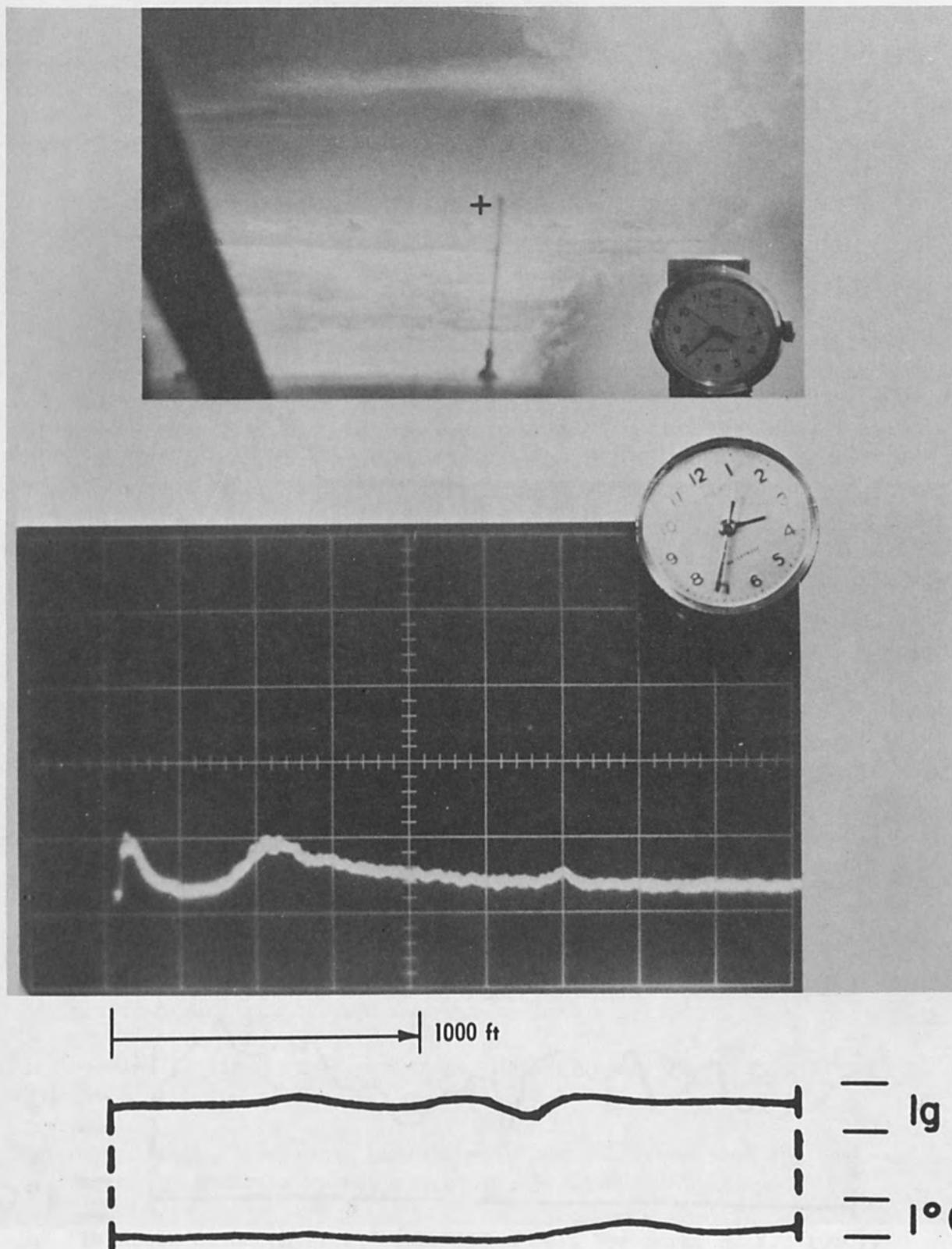


Figure 4.3. 6/10/65 Near an Oklahoma Thunderstorm; Free Air Temp. 0°C;
Press. Alt. 15,800 ft.; True Air Speed 163 knots; Scope Gain 200,
200 mV/cm.

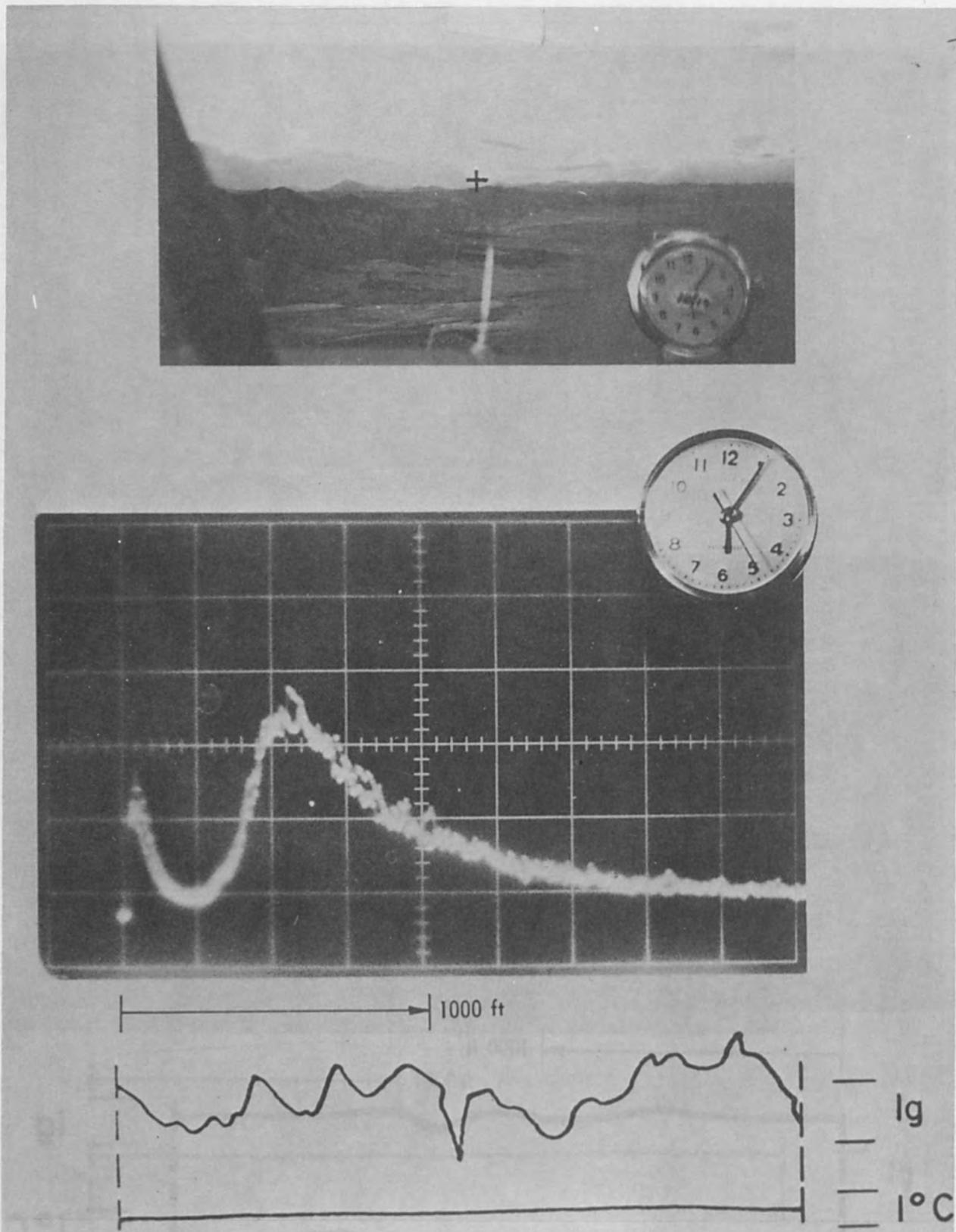


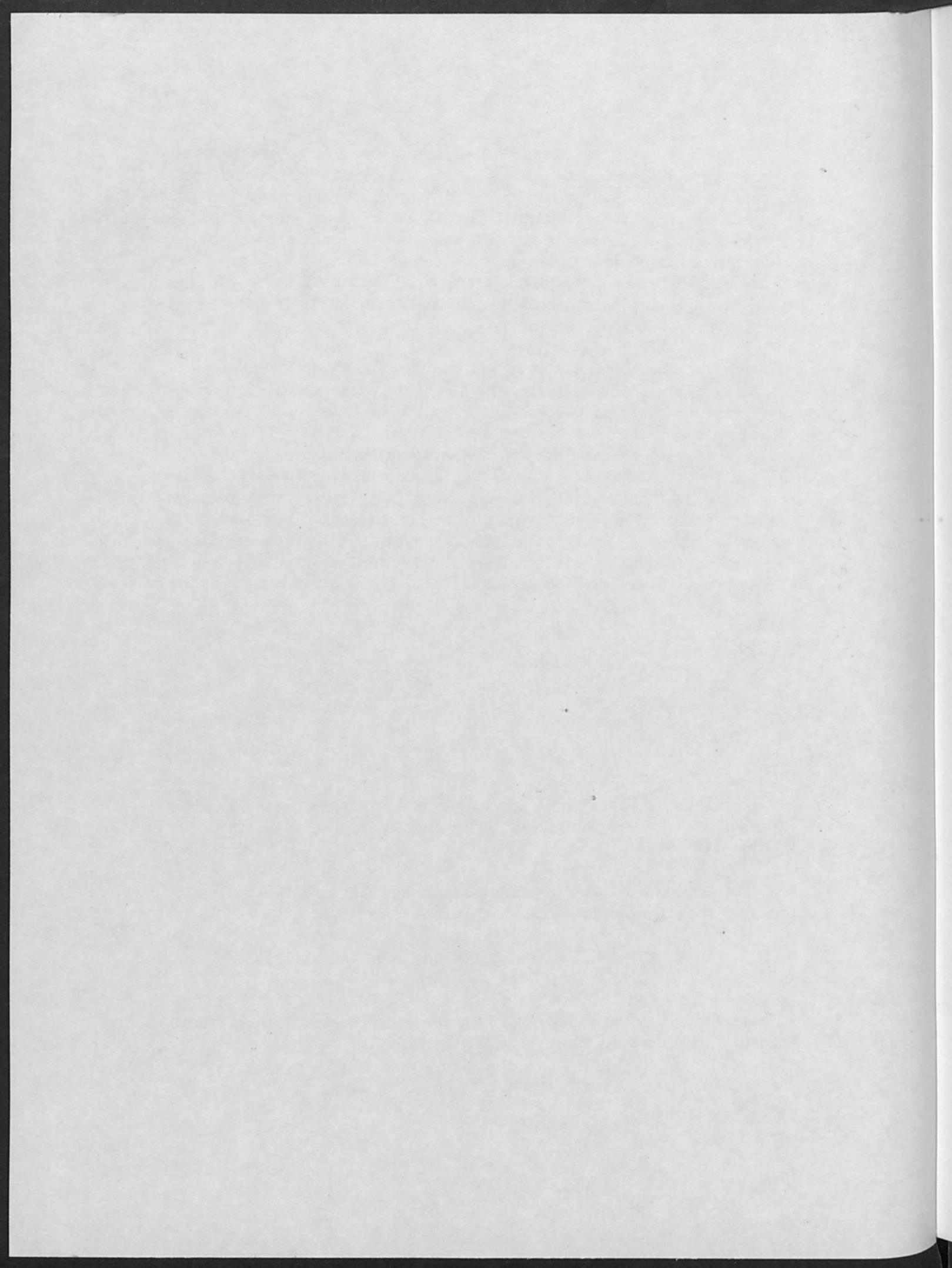
Figure 4.4. 10/19/65 Orographic Turbulence in the Boulder, Colorado Area; Free Air Temp. 7°C; Press. Alt. 8000 ft.; True Air Speed 155 Knots; Scope Gain 50 mV/cm.

The critical question now is whether there are characteristic differences in the distribution of the particulate matter in turbulent versus quiet atmospheres. We have in mind the possibility that turbulent regions could provide a bunching or "blotchiness" of particulate matter that might have merely a layered structure in calm atmospheres. This flight program could not distinguish these different possible distributions simply because the "A-scope" type of presentation of our optical radar provided only a one-dimensional exploration of the atmosphere forward of the aircraft.

These considerations have led us to initiating the development of a PPI (Plan Position Indicating) optical radar which will provide information about the aerosol distribution in "clear air" regions that we can still probe with aircraft. For example, should it be that regions of severe turbulence are characterized by a clumping of the particulate matter into several hundred-foot sized cells, then a PPI optical radar provides just the two-dimensional map of the forward area required for identification of such reticulation in the aerosol. The new equipment is presently being designed and it is hoped that definitive data can be achieved in 1966 at altitudes to 30,000 feet, and in jet stream as well as mountain wave and severe storm associated clear air turbulence.

REFERENCES

1. P. A. Franken, J. A. Jenney, and D. M. Rank, "Airborne Investigations of Clear Air Turbulence with Optical Radar," University of Michigan (December 1965).
2. See, for example, J. Sharpe, "Photoelectric Cells and Photomultipliers," Electronics Technology, June and July 1961.
3. G. Mie, "Beitrag zur Optik truber Medien," Ann. Phys. 25, 377 (1908).
4. K. Bullrich, "Scattered Radiation in the Atmosphere and the Natural Aerosol," Advances in Geophysics 10, 99 (1964).
5. "Handbook of Geophysics," (MacMillan Co., New York, N. Y., 1960), Chap. 8, p. 12.



ZAPPING PAPER

by

Jon H. Myer
Hughes Aircraft Company
Newport Beach, California

ABSTRACT

A novel and simple method for monitoring the temporal and spatial distribution of pulsed laser energy by means of a multilayer coated paper is described.

I. PROPERTIES OF ZAPPING PAPER

A multi-layer coated paper is located in the focused or unfocused laser beam to show the thermal effects of this beam on the various layers. In the present zapping paper the first layer has a sensitivity of approximately 1/10 of a Joule per square millimeter at 6943 Å wavelength, when applied with millisecond pulses, and a sensitivity of approximately one Joule per square millimeter through the second layer. Approximately ten Joules per square millimeter will pierce the paper.

At slower rates and with tungsten spectral distribution the sensitivity of the first layer is only approximately one Joule per square millimeter.

To obtain a time record this paper is wrapped around a drum driven by a synchronous motor and moving at known peripheral speed. Time records of laser pulses are easily obtained in this manner. Both stationary and time records are shown in Figure 1.

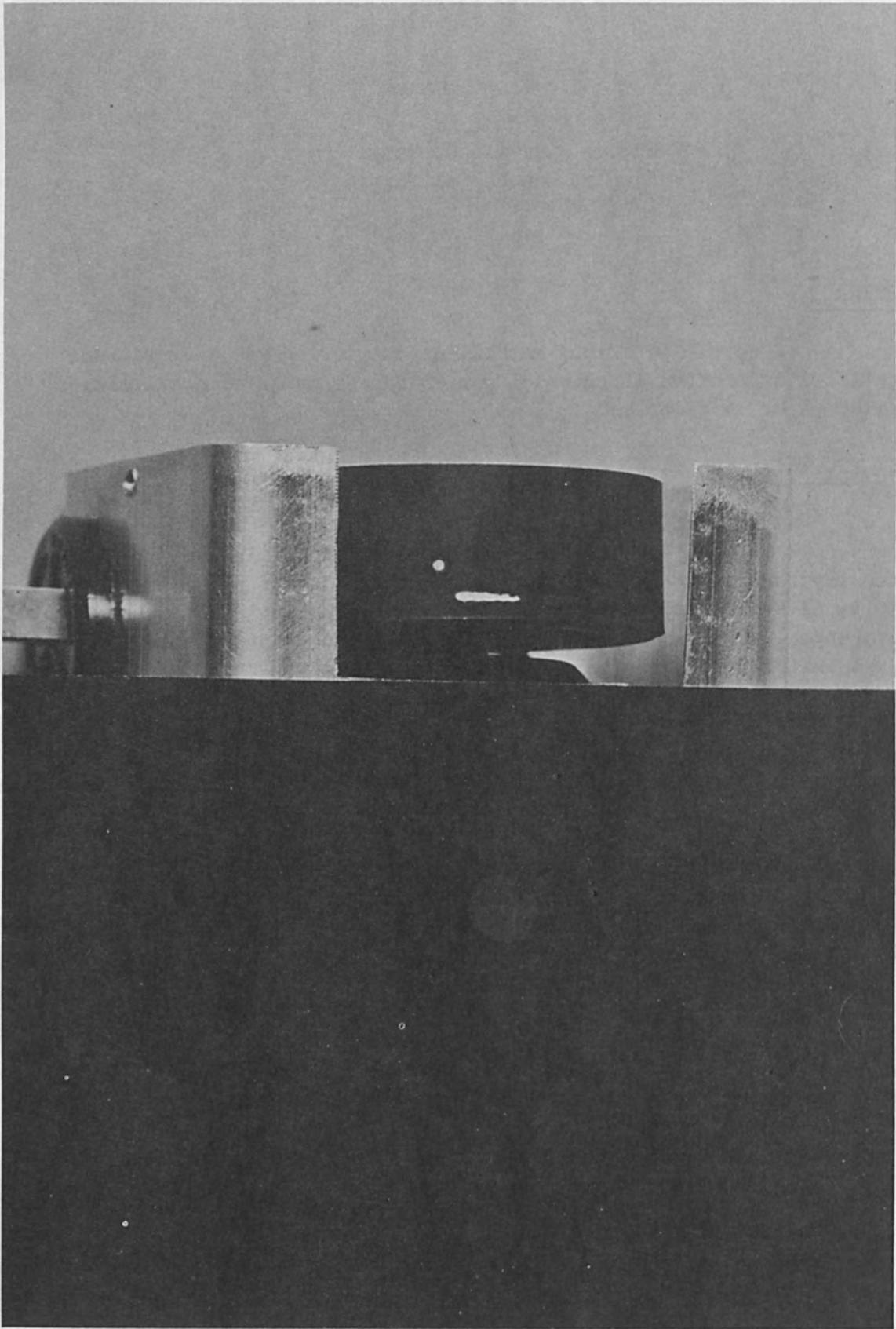


Figure 1.

THE APPLICATION OF LASERS
IN THERMOPHYSICAL PROPERTIES MEASUREMENTS *

by

M. M. Nakata
Atomics International
Canoga Park, California

ABSTRACT

The application of a ruby laser as the pulsed energy source for the flash technique of measuring the thermophysical properties of materials is discussed. The advantages and limitations of its application, as well as other related problems, are treated, and the results of measurements performed are presented.

I. INTRODUCTION

As a result of rapid advances in the application of new materials at high temperatures, notably in the space and nuclear energy fields, there has been a renewed interest in recent years in the development of new techniques of measuring the thermophysical properties of materials. In many instances, the classical methods of measuring thermal conductivity and thermal diffusivity are inadequate for high temperature applications, due to sample size limitations, length of time required for measurements, and difficulties in modifying the apparatus for high temperature measurements. Many of the materials currently being investigated are not obtainable in large sizes, due either to scarcity of raw material or to difficulties in fabricating large homogeneous bodies. Furthermore, many of the materials are toxic or present radiological hazards and are best handled in small quantities. Also, at elevated temperatures the kinetics of physicochemical processes increase appreciably over those at moderate temperatures. Transient techniques are thus preferred over steady-state techniques at high temperatures.

One of the new techniques recently reported in the literature for measuring the thermophysical properties of materials is the flash method, developed by Parker et al. (1). In the flash method, a thin sample, usually in the form of a disc, is irradiated on the front face with a high intensity, short duration pulse of light. The resulting temperature transient on the rear face of the sample is then related to the thermal diffusivity. When the heat pulse is of infinitesimal duration, the diffusivity, α , is obtained from the transient response

*This work was supported in part by the National Aeronautics and Space Administration and the U. S. Atomic Energy Commission under Contract AT-(11-1)-GEN-8.

of the rear face temperature by

$$\alpha = a^2/\pi^2 t_c = 1.37 a^2/\pi^2 t_{1/2} \quad (1)$$

In (1), $t_{1/2}$ is the time required for the rear face temperature to reach half-maximum, a is the sample thickness, and t_c is the characteristic rise time of the rear face temperature. It is not necessary to know the amount of energy absorbed in the front surface in order to determine the thermal diffusivity. Only t (time) and a (length) are needed, both of which can be measured with high accuracy. In theory, heat capacity can also be determined using this technique, since

$$DC = Q/aT_M, \quad (2)$$

where D and C are the density and heat capacity of the sample, respectively, and T_M is the maximum temperature rise on the back face of the sample. However, in this case, the energy absorbed by the sample, Q , must be known. Thermal conductivity is given by the relation

$$K = \alpha CD, \quad (3)$$

so that, in essence, the flash technique is capable of generating thermal diffusivity, thermal conductivity, and heat capacity data in the same apparatus.

Parker *et al.* employed a helical xenon flash lamp as the source of pulsed energy. The flash lamp, however, is restricted in its application by the fact that it must be located close to the sample in order to transfer a sufficient amount of energy to the sample. At elevated temperatures, the heat from a furnace prevents the placing of the flash lamp near the sample. The laser, on the other hand, is particularly suitable for this application in that a highly energetic well-collimated beam of light energy is available, and can be operated at a distance from the furnace. Other pulsed-energy sources, such as electron beam bombardment and arc imaging techniques can be used. However, in terms of simplicity of design and adaptability to various experimental conditions, the laser appears to be superior. For example, electron beam bombardment techniques require the source to be located within the system so that the system is more complicated. Furthermore, where measurements need be done in gaseous atmospheres, electron beam bombardment is not practical.

In spite of the many favorable characteristics of the ruby laser, several difficulties are encountered in applying the laser to the flash method. Although the difficulties have not been completely

resolved, sufficient progress has been made so that reliable thermal diffusivity measurements can be made. Heat capacity measurements using the flash technique are more difficult and have not been successfully demonstrated as yet. It is the subject of this paper to discuss some of the problems involved in the development of the flash technique and to report the progress made to date.

II. THERMAL DIFFUSIVITY MEASUREMENTS

A. Apparatus

A schematic diagram of the flash thermal diffusivity apparatus is shown in Figure 1. The sample, which is nominally 1/4 inch diameter, is supported within a tube furnace, using a sample holder which makes minimum thermal contact with the sample. The thickness of the sample is dictated by the properties of the material and the boundary requirements of the method, a typical thickness being about 0.1 inch. Diffusivity measurements are usually made under vacuum, but the furnace chamber can also be filled with inert gases or hydrogen. The laser head is a Raytheon LH3 model, which uses a 3/8 inch diameter by 6-5/8 inch long ruby rod. The laser output is rated at 15-20 joules at the input power of 2000 joules, when operated at the temperature of liquid nitrogen. The wavelength of the laser beam is 6934 Å and the pulse width is 1.5 msec. Between room temperature and about 1300°C, thermocouples are used to measure the steady-state temperature, as well as to follow the transient on the back face of the sample. The thermocouple signal is amplified using a Dana amplifier (Model 2200), and with an appropriate biasing network, displayed as one beam on a Tektronix 502 dual-beam oscilloscope. By means of a knife switch, the thermocouple signal can be switched to a Rubicon potentiometer (Model 2745), to determine the ambient temperature of the sample. A small portion of the laser beam is reflected off a prism onto a lead sulfide cell and the output of the cell is displayed by the second beam of the oscilloscope, for the purpose of accurately establishing the zero time. The oscilloscope traces are permanently recorded on Polaroid Type 47 films. Figure 2 is a photograph of the apparatus.

For diffusivity measurements above 1300°C, a second apparatus has been constructed in which the axis of the tube furnace is vertical, instead of on the horizontal plane. A front surface mirror placed above the furnace is used to deflect the laser beam onto the sample. Temperature measurements for the apparatus are made using a photoelectric pyrometer. The sample holder is also designed to prevent the laser beam from passing through to the sensitive pyrometer and saturating, or even damaging, the photoelement.

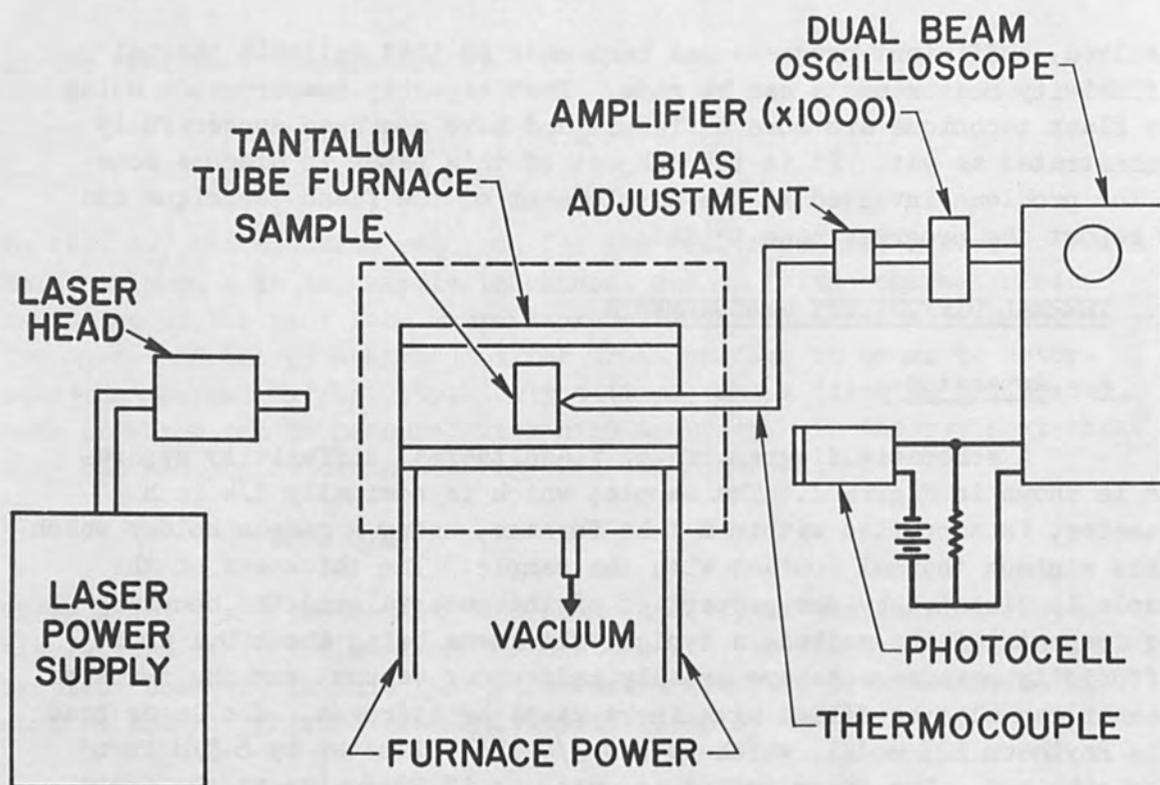


Figure 1. Schematic Diagram of Flash Diffusivity Apparatus.

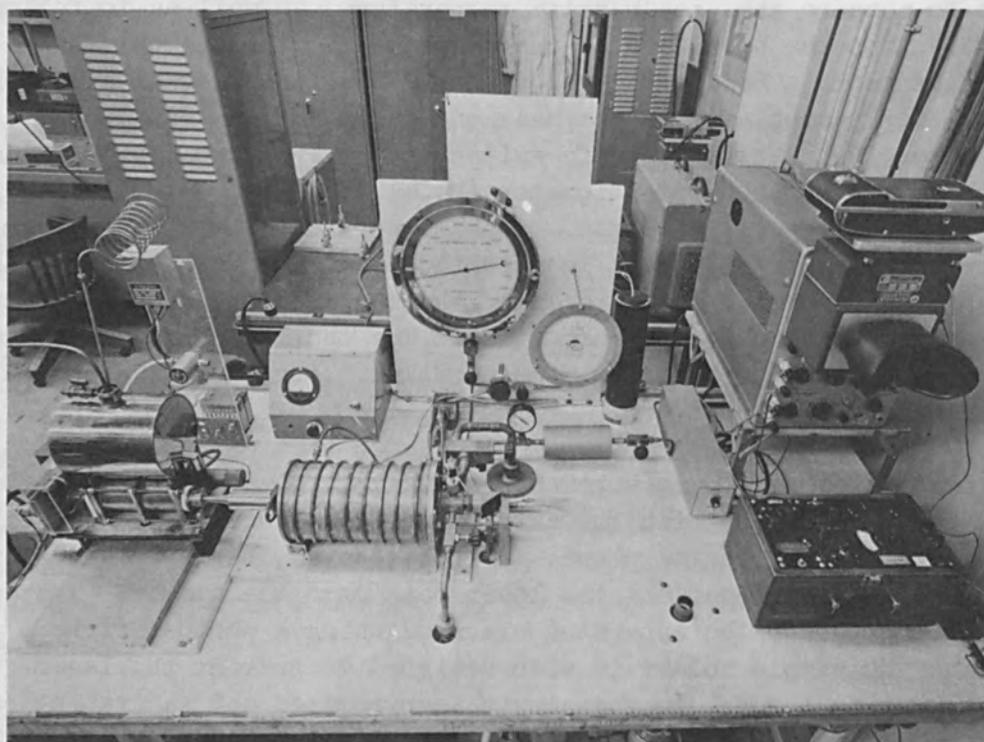


Figure 2. Photograph of Flash Diffusivity Apparatus.

B. Method

Accurate thermal diffusivity measurements using the flash technique require that certain boundary conditions be met. These are 1) the faces of the sample must be parallel, 2) the sample must be thermally isolated, i.e., losses due to conduction, convection and radiation must be negligible, 3) the pulse time must be small compared to the characteristic time of the sample, which is roughly the time required for the heat to travel through the sample, and 4) the pulse of energy on the front face must be uniform. The first requirement can be easily met by careful machining of the sample. Thermal conduction losses can be minimized by reducing the contact area between the sample and sample holder. This is simple when thermocouples are used for temperature measurements, but when a radiation pyrometer is used, the pyrometer must be completely shielded from the laser beam. About the only way to protect the pyrometer is to design the sample holder in such a way that no light can pass between the sample and sample holder, i.e., the sample and its holder must be in intimate contact. The problem of thermally isolating the sample then becomes serious. The sample is therefore supported on a 3/16 inch diameter thin-walled refractory metal tubing, the end of which has been ground to a knife-edge. A fine groove is machined into the back face of the sample to fit the knife-edge of the tubing, and the sample is held in place by gravity. The use of optical filters to cut off the laser beam from the pyrometer does not appear to be feasible because the laser beam consists not only of the characteristic wavelength of the ruby laser, but also a wide spectrum of wavelengths due to the fluorescent discharge of the ruby and perhaps also to a light-pipe effect of the rod in transmitting the light of the xenon flash lamp. Since measurements are usually made in vacuum, convection losses are relatively unimportant. For materials that dissociate at elevated temperatures, e.g., the hydrides, measurements must be made with an overpressure of gas, so that convection losses can become significant. However, by using thin samples or by reducing the temperature rise of the sample, the effect of convection can be minimized. Radiation losses become significant at high temperatures, but techniques have been developed (2,3) to correct for this effect.

In some applications of the flash technique, the duration of the energy pulse is comparable to the characteristic thermal diffusion time, t_c , of the sample. When the pulse time, τ , is not infinitesimal compared to t_c , the rise time of the rear face will be retarded, and $t_{1/2}$ will be greater than $1.37 t_c$. This is called the "finite pulse-time effect" (2). This situation arises when the sample thickness is limited, due to unavailability of larger samples, or when the thickness is reduced to minimize heat losses. Q-spoiling devices may be used to shorten the pulse time, but not without losing a large fraction of the laser energy in the process. Furthermore, the cost of the instrumentation can be quite high.

A technique has been developed (2,4), therefore, in which a correction for the finite pulse-time effect can be determined graphically from measured quantities (viz., τ and $t_1/2$) to obtain the true diffusivity values. The accuracy of this technique was tested experimentally using Armco iron, which is the generally accepted thermal conductivity standard (5). Figure 3 illustrates the results obtained. The solid curve represents the accepted values. The open points represent the data obtained by the flash technique using no corrections for the finite pulse-time effect, i.e., using Eq. (1), and the solid points are obtained using the corrected thermal response curves (2,4). The different types of data points denote samples of different thicknesses.

In the operation of the laser, the experience has been that the spatial and temporal distribution of a ruby laser was quite unstable. Factors such as the condition of the crystal reflective coating, condition of the laser cavity, and temperature of the ruby, contribute to variations in the laser output. In view of the requirement in the flash technique, that the front face of the sample be subjected to a uniform pulse, it is important, therefore, that the laser beam be monitored and, if possible, modified to give the uniformity desired.

An image converter camera was employed to analyze the laser output, which clearly showed the short-time instability of the pulse. It was difficult, however, to estimate the integrated spatial distribution of a single pulse from the data obtained.

Two other approaches have been taken, therefore, to seek a solution to the problem. First, a qualitative method has been used to check the spatial distribution of the laser beam. By clipping together several layers of "Labelon" thermal copy paper (a heat-sensitive paper used in Thermo-fax copying machines) and pulsing them with the laser, an indication of the intensity profile of the pulse could be obtained from the number of layers penetrated and the size, location and shape of the hot spots on each layer. The characteristics of the laser beam were analyzed as functions of distance (laser to target) and input power. In general, it was observed that the intensity was maximum at the center of the beam with a nearly linear attenuation in going from the center to the edge. The magnitude of this effect was proportional to input power and inversely proportional to distance. By operating the laser at approximately 90% of its rated input power and placing the sample approximately 12 inches from the laser rod, a more nearly uniform heat pattern was obtained. Diffusivity measurements of Armco iron samples made with these settings have agreed with accepted values to within 2%, indicating that the laser beam had been made sufficiently uniform. The spatial distribution of the laser beam, however, can change from time to time, so that it is still necessary to monitor the pulses.

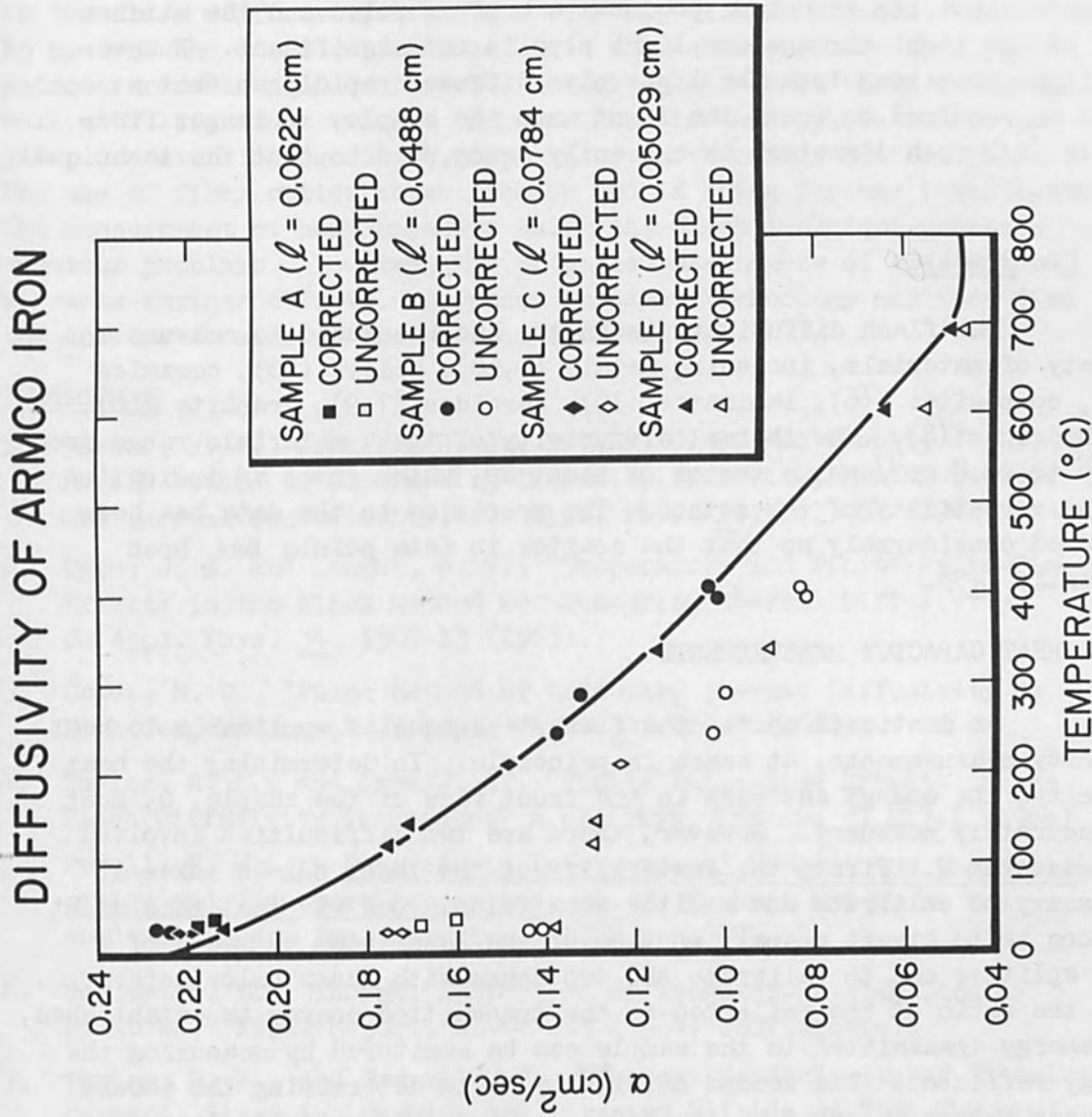


Figure 3. Thermal Diffusivity of ARMCO Iron.

The second approach was to modify the laser beam using fiber optics. By orienting the fibers of a fiber bundle in a completely random fashion so that the light entering the bundle over any portion of the surface at one end is distributed uniformly over the entire surface area at the other end, a uniform pulse can be obtained, regardless of the shape of the incident beam. A 1/4 inch diameter by 12 inch long light pipe was obtained for this purpose. Tests conducted thus far have indicated that the technique produces a uniform pulse and the attenuation of the light through the light pipe is not significant. However, the light emanating from the light pipe diffuses rapidly so that a lens would be required to focus the light onto the sample. A larger fiber bundle (1/2 inch diameter) is currently being used to test the technique further.

C. Results

The flash diffusivity technique has been used to measure a variety of materials, including metals (4,10), alloys (10), ceramics (10), composites (6), laminates (10), carbides (7,9), graphite (10), and hydrides (8). The thermal diffusivity of these materials range from ~ 0.04 to ~ 0.8 cm^2/sec , a factor of about 20, which gives an indication of the versatility of the method. The precision in the data has been improved considerably so that the scatter in data points has been typically $< \pm 5\%$.

III. HEAT CAPACITY MEASUREMENTS

As mentioned above, the flash technique is applicable to heat capacity measurements, at least in principle. In determining the heat capacity, the energy absorbed in the front face of the sample, Q , must be accurately measured. However, there are two difficulties involved in measuring Q . First, the instability of the laser pulses makes it necessary to calibrate and monitor each pulse. One way that this might be done is to divert a small portion of the laser beam by means of a beam splitter and to calibrate the two beams with laser calorimeters. Once the ratio of the reflected to the transmitted energy is established, the energy transmitted to the sample can be monitored by measuring the energy reflected. The second difficulty is in determining the amount of laser energy absorbed by the sample. Since absorptivity is a function of the sample material, surface conditions, and temperature, it would be extremely difficult to determine accurately. A comparative method has therefore been considered for which the absorptivity need not be known. In this method, a sample holder of mass m_1 made of material of which the heat capacity, C_1 , is well known is first pulsed and the rise in temperature, T_1 , measured. The sample of mass m_2 is then placed in the holder and the holder is again pulsed to obtain the temperature rise, T_2 . If the pulsed energy is the same in each case,

the heat capacity of the sample, C_2 , can be determined from

$$Q = C_1 m_1 T_1 = (C_1 m_1 + C_2 m_2) T_2 . \quad (4)$$

IV. CONCLUSIONS

The ruby laser has been used as the source of pulsed energy in the flash diffusivity technique with appreciable success. Solutions to most of the difficulties related to the technique have been found and measurements made on numerous materials have been shown to agree well with literature values. Improvements in the application of the laser are still being sought, especially in obtaining uniform pulses. The use of fiber optics shows promise and is being further investigated. The measurement of heat capacity using the flash technique presents numerous problems. However, the potential usefulness of this method warrants further efforts. Advances in laser technology may soon find the answers to some of the present difficulties.

REFERENCES

1. Parker, W. J., Jenkins, R. J., Butler, C. P. and Abbott, J. L., "Flash Method of Determining Thermal Diffusivity, Heat Capacity, and Thermal Conductivity," *J. Appl. Phys.* 32, 1679-84 (1961).
2. Cape, J. A. and Lehman, G. W., "Temperature and Finite Pulse-Time Effects in the Flash Method for Measuring Thermal Diffusivity," *J. Appl. Phys.* 34, 1909-13 (1963).
3. Cowan, R. D., "Pulse Method of Measuring Thermal Diffusivity at High Temperatures," *J. Appl. Phys.* 34, 926-7 (1963).
4. Taylor, R. E., and Cape J. A., "Finite Pulse-Time Effects in the Flash Diffusivity Technique," *Appl. Phys. Letters* 5, 212-3 (1964)
5. Powell, R. W., in Progress in International Research on Thermodynamics and Transport Properties, ed. J. F. Masi and D. H. Tsai, Academic Press, Inc., New York, 1962, pp. 454-73.
6. Taylor, R. E., "Thermal Properties of Tungsten-Uranium Dioxide Mixtures," Final Report, AI-64-153 (1964) Confidential.
7. Taylor, R. E., and Morreale, J., "Thermal Conductivity of Titanium Carbide, Zirconium Carbide and Titanium Nitride at High Temperatures," *J. Am. Ceram. Soc.* 47, 69-73 (1964).
8. Taylor, R. E., and Ambrose, C. J., "Thermophysical Properties of SNAP Fuels," NAA-SR-9782 (1964) Confidential.
9. Carniglia, S.C., "Single Crystal and Dense Polycrystal Uranium Carbide: Thermal, Mechanical, and Chemical Properties," Paper presented at the International Conference on Carbides in Nuclear Energy, Harwell, England. November 1963.
10. Unpublished data.

VANADIUM CHARGE COMPENSATOR SITE IN LASER CaWO_4

by

C. Kikuchi and N. Mahootian*
Department of Nuclear Engineering
The University of Michigan
Ann Arbor, Michigan

and

W. Viehmann and R. T. Farrar
Harry Diamond Laboratories
Washington, D. C.

ABSTRACT

This paper is intended to give certain preliminary results of the program to investigate the role of charge compensators in laser CaWO_4 . Vanadium, which lies above niobium in the periodic table, is introduced as the charge compensator and micro-probe. Electron spin resonance measurements show that vanadium occupies W- and/or Ca-sites depending upon the relative concentration of vanadium to rare-earth. These and the preliminary optical results are discussed.

I. INTRODUCTION

The present investigation was undertaken to see what information about CaWO_4 and the included rare-earth laser active centers can be obtained by the EPR (electron paramagnetic resonance) studies of vanadium in such crystals. Following the discovery of laser action in CaWO_4 doped with a variety of paramagnetic impurities¹, a beginning in the systematic study of these materials has been made in several laboratories. Among these are the programs carried out by the workers²⁻⁴ at the Bell Telephone Laboratory, where the unambiguous evidence for the importance of charge compensation was obtained, and the results by Kedzie and his associates and others⁵⁻⁷ pointing to the possible multiple sites occupied by the paramagnetic ions, and hence of the laser active centers.

In our effort to obtain further information on the questions raised in the references quoted above, a preliminary study of CaWO_4 ;Nd, doped with Na or Nb was carried out. Unfortunately, the EPR spectrum is extremely complex, dependent to a certain extent upon the Nd concentration.

*Supported in part by contract with Harry Diamond Laboratories.

Consequently it was felt that the task of unraveling the complex spectrum would be difficult, so that our attention was turned to vanadium, which we felt would provide less ambiguous results. The merits for using vanadium stem from (1) the nuclear spin $7/2$, which gives the readily recognizable paramagnetic resonance signature of 8 HFS (hyperfine structure) lines, (2) the valence state can be readily changed by such radiations as x-rays and gamma-rays, (3) the EPR properties of vanadium are sensitive to the crystalline environment, and (4) of importance to the present specific problem is the fact that vanadium lies immediately above Nb in the Group V-B column of the periodic table. Thus it seems to us that vanadium may be particularly suited for the study of the diadochy of impurities in CaWO_4 .

The results obtained to date indicate that vanadium can occupy both W- and Ca-sites, and that there are several crystal imperfection and impurity sensitive EPR spectra. To date four types of spectra have been observed, of which one results from neutron irradiation. In this paper, we shall limit ourselves primarily to the experimental results. The theoretical interpretation and the additional information obtained by neutron irradiation will be reported later.

We wish to point out that the study reported here is a part of our general program using vanadium to probe the crystalline properties of solids.⁸⁻¹¹ To date we have investigated V^{2+} , V^{3+} , and V^{4+} in $\alpha\text{-Al}_2\text{O}_3$, V^{++} and VO^{2+} in Tutton salt, and V^{++} in SnO_2 , TiO_2 , and GeO_2 . A very brief account of V in CaWO_4 was given earlier.¹²

II. EXPERIMENTAL PROCEDURE

The crystals were grown by the Czochralski technique in an apparatus similar to that described by Nassau and Broyer.¹³ Optical grade calcium tungstate, rare-earth oxides of 99.9% purity, and reagent grade vanadium pentoxide were used in the preparation of the melt. The melt was contained in an irridium crucible. The melt temperature was sensed by a jacketed platinum-platinum/rhodium thermocouple immersed in the melt and was stabilized to about $+2.5^\circ\text{C}$ by the automatic adjustment of the RF generator power output. The use of a slightly conical alumina chimney above the crucible reduced markedly the temperature fluctuations on the melt surface and resulted in improved uniformity and optical quality of the crystals. The growth interface was still of the desirable conical shape. Pulling was performed in air with typical growth speeds of 0.25 to 0.30 inch per hour and turning rates ranging from 60 to 120 rpm.

The crystal axis determination and alignment were carried out by x-rays. To produce the paramagnetic centers, the samples were

x-irradiated at 50 kvp, 35 ma for about 15 minutes. The EPR spectra for such samples ($\text{CaWO}_4:\text{Tb},\text{V}$) were studied at 77° and 4.2°K using the Varian X-band EPR spectrometer.

III. CRYSTAL STRUCTURE AND SITE SYMMETRY

The crystal structure of calcium tungstate (Scheelite) is characterized by space group C_{4h} , or the tetragonal I $4_1/a$, with four molecules in the unit cell. The tetragonal body-centered unit cell has the dimensions

$$a = b = 5.243 \pm .002 \text{ \AA}, c = 11.376 \pm .003 \text{ \AA}$$

The site symmetry of both Ca and W is S_4 . The structure of CaWO_4 may be viewed as composed of WO_4 anions ionically bonded to the Ca^{++} cations. The metallic sites are found on planes separated by $c/4$ and perpendicular to the c -axis, as shown in Figure 1. The metal ions on the planes at $c/4$ and $3c/4$ are related to those on the ab - and $c/2$ planes by glides. The nearest metal ions lie in planes separated by $c/4$ and the direction joining the two ions, Ca-Ca , W-W , and W-Ca , makes an angle of $42^\circ 40'$ with the c -axis.

Each tungsten ion in CaWO_4 is bonded to four oxygens, forming a slightly distorted tetrahedron as shown in Figure 2a. The W-O distance is about 1.78 \AA , and the base and the height of the squashed tetrahedron are 2.11 and 1.96 \AA , respectively.^{14,15} The diagonal of this tetrahedron makes an angle of $31^\circ 54'$ with the a -axis. Figure 2b gives the projection of these tetrahedra on the ab -plane.

The calcium atoms, on the other hand, are surrounded by eight oxygens which form two distorted tetrahedra, as shown in Figure 3. The orientations of these tetrahedra relative to the crystal axes are shown in Figure 4.

IV. CHEMICAL ANALYSIS

It was found that all samples contained Mn^{2+} , producing EPR spectra with an intensity of about $1/4$ of that of the vanadium spectra. In one sample, Nd^{3+} was observed, although this rare-earth was not intentionally added. Its presence is due probably to Nd adhering to the iridium crucible, or evaporating off the walls of the crystal growing equipment, thus showing the extreme sensitivity of the equipment to paramagnetic ions. The vanadium concentration in these samples were determined for us using arc discharge by the Physics Instrumentation Laboratory, Department of Industrial Health, of the University of Michigan. The results are tabulated in Table I. The last column in Table I shows the rare-earth concentration estimated from their distribution coefficients in calcium tungstate. These values give $k' = 0.12$

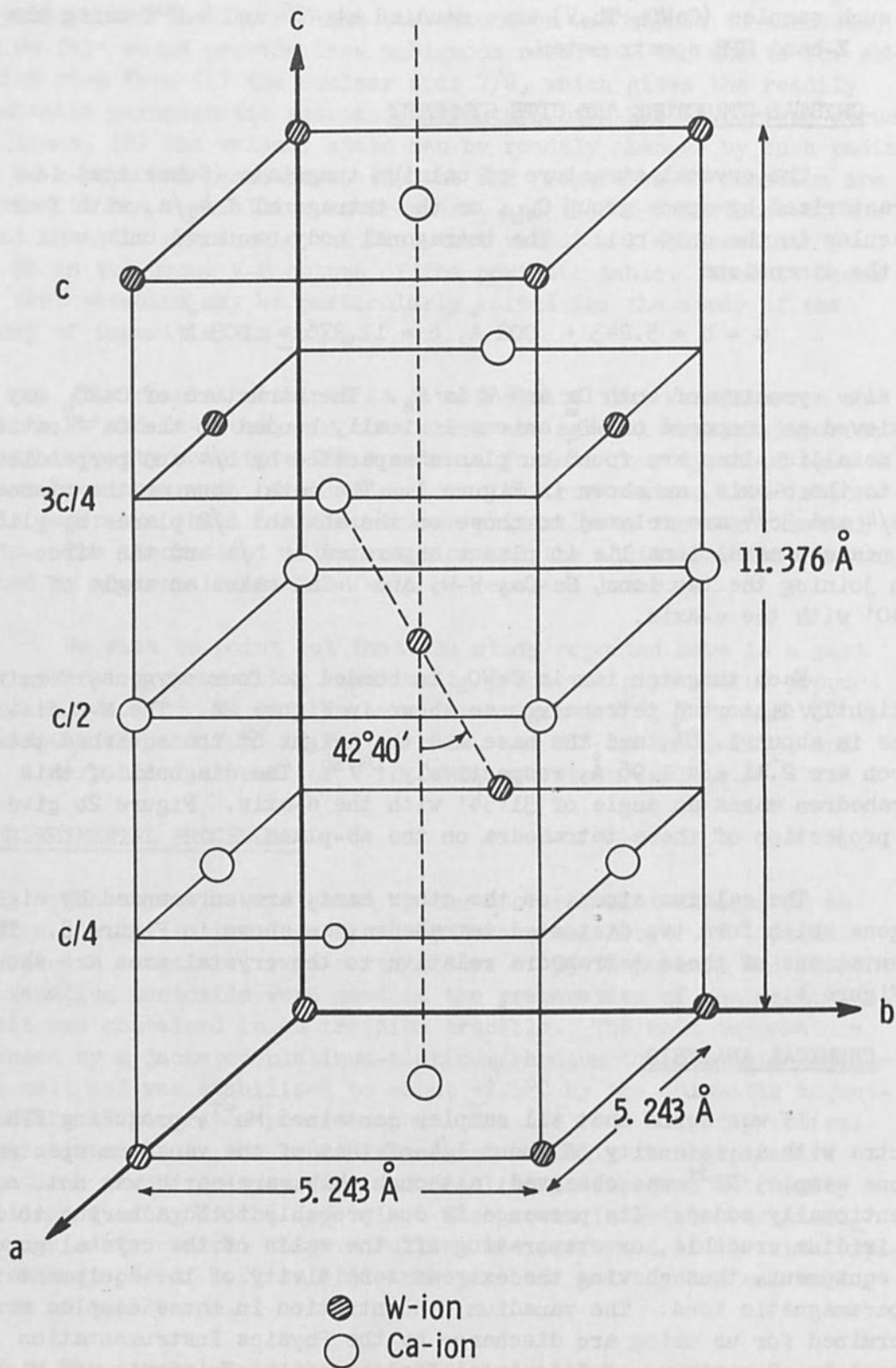


Figure 1. Metal Ions in CaWO_4 Unit Cell.

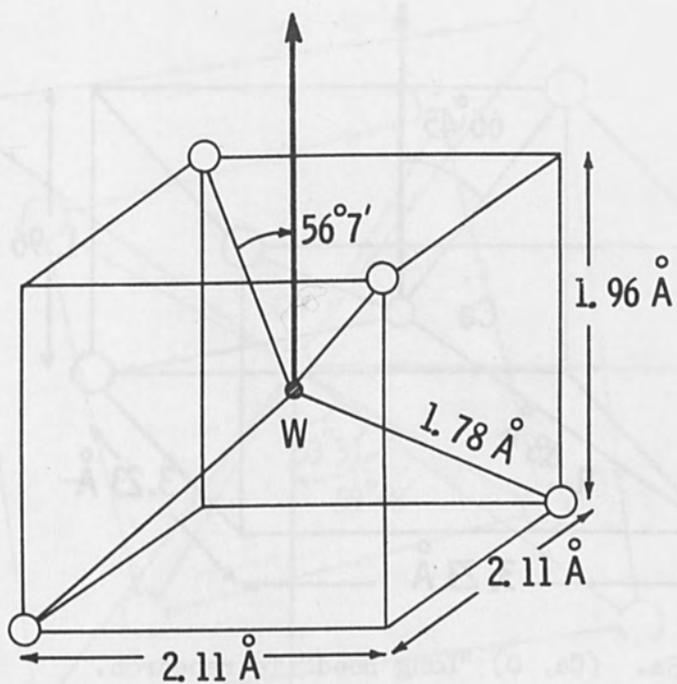


Figure 2a. $(\text{WO}_4)^{2-}$ Tetrahedron. W at (0.0.0).

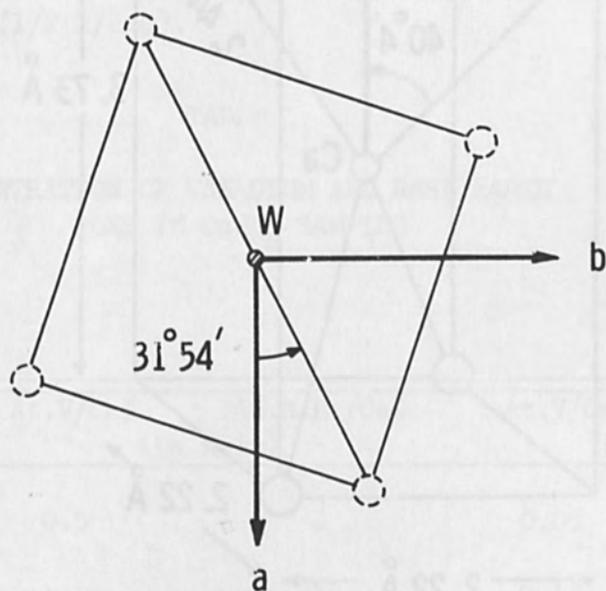


Figure 2b. Projection of the $(\text{WO}_4)^{2-}$ Tetrahedron on the ab-plane.

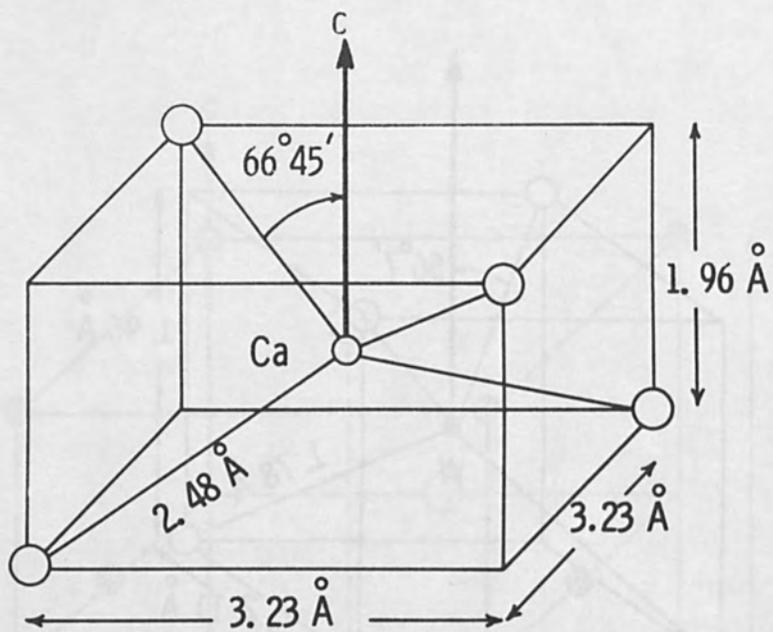


Figure 3a. (Ca, O) "Long Bond" Tetrahedron.

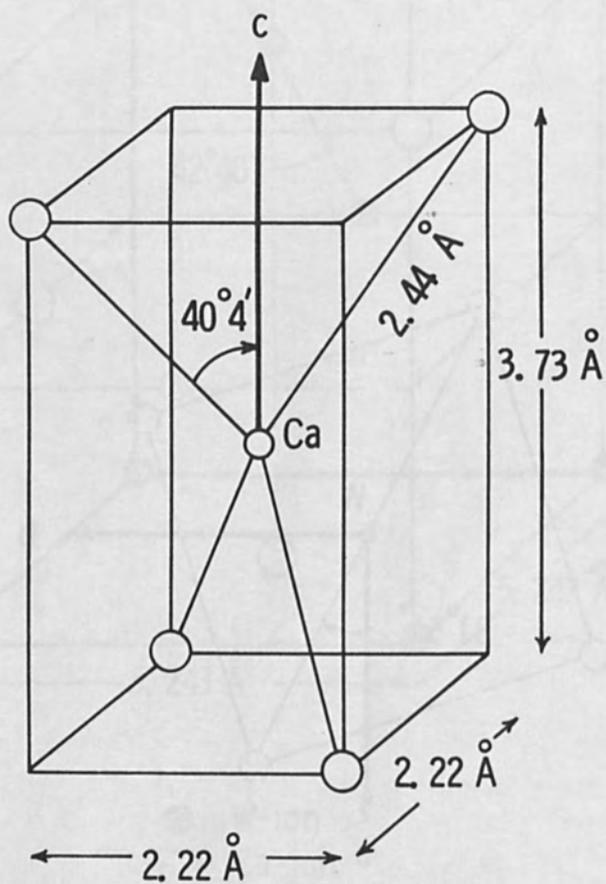


Figure 3b. (Ca, O) "Short Bond" Tetrahedron.

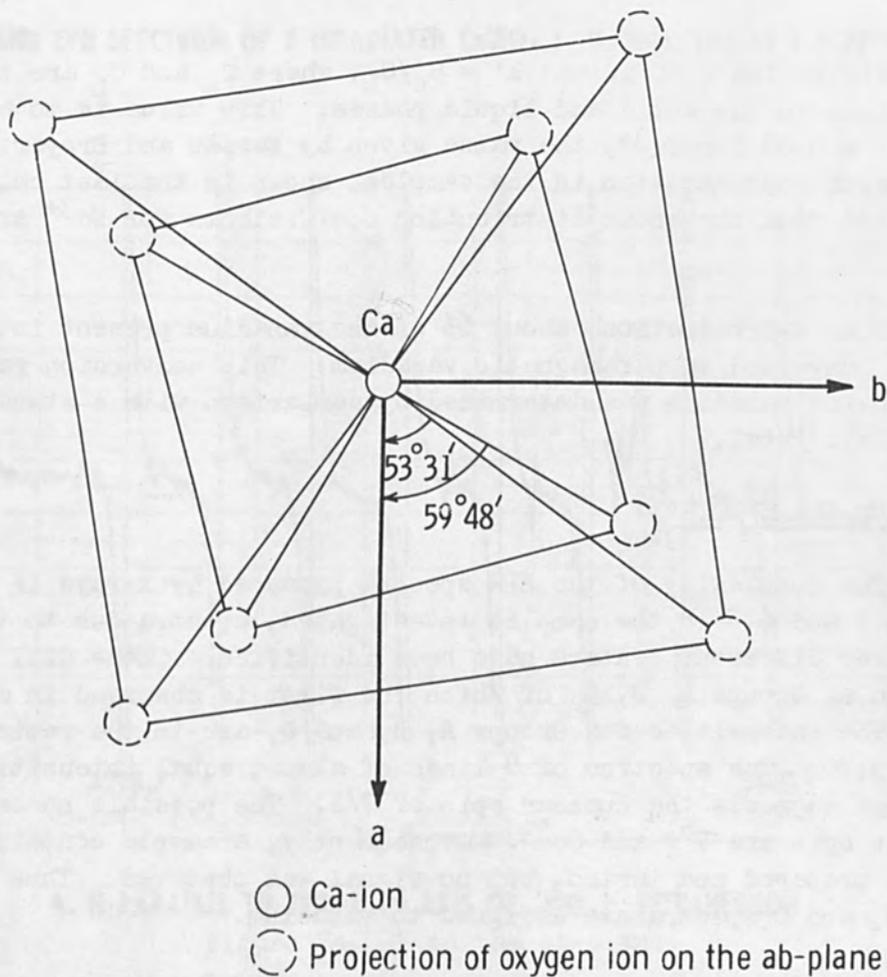


Figure 4. Projection of the (Ca,O) Double-tetrahedron on the ab-plane. Ca at $(\frac{1}{2} \frac{1}{2} 0)$.

TABLE I

CONCENTRATION OF VANADIUM AND RARE EARTH IONS IN CaWO_4 SAMPLES

Sample	At.V/Ca%	At.R.E./Ca%	At.V/Ca%	At.R.E./Ca%
		(In Melt)	(In Crystal)	
$\text{CaWO}_4:\text{V}$	0.5	0	0.05	0
$\text{CaWO}_4:\text{V, Tb}$	1	1	0.13	0.3
$\text{CaWO}_4:\text{V, Nd}$	1	1	0.13	0.2

for the distribution coefficient $k' = C_s/C_L$, where C_s and C_L are the concentrations in the solid and liquid phases. This value is to be compared to $k' = 0.08$ for Cr^{3+} , the value given by Nassau and Broyer.¹³ The rare-earth concentration in the samples, shown in the last column, were computed from the known distribution coefficients for Nd^{3+} and Tb^{3+} .

Upon x-irradiation, about 6% of the vanadium present in the crystal is converted to paramagnetic vanadium. This conversion ratio to paramagnetic vanadium was determined by comparison with a standard $CuSO_4$ single crystal.

V. VANADIUM EPR SPECTRA

The complexity of the EPR spectra produced by x-rays is shown in Figures 5 and 6. In the samples investigated, spectra due to vanadium in three different centers have been identified. These will be referred to as Groups A, B, C, of which the first is observed in all samples. The intensities for Groups A, B, and C, are in the ratio of about 100:5:70. The spectrum of 8 lines of almost equal intensities and spacings suggests the nuclear spin of $7/2$. The possible nuclei having this spin are V^{51} and Co^{59} . Consequently, a sample containing cobalt was prepared and tested, but no signal was observed. Thus the Group A, B, and C spectra are assigned to vanadium.

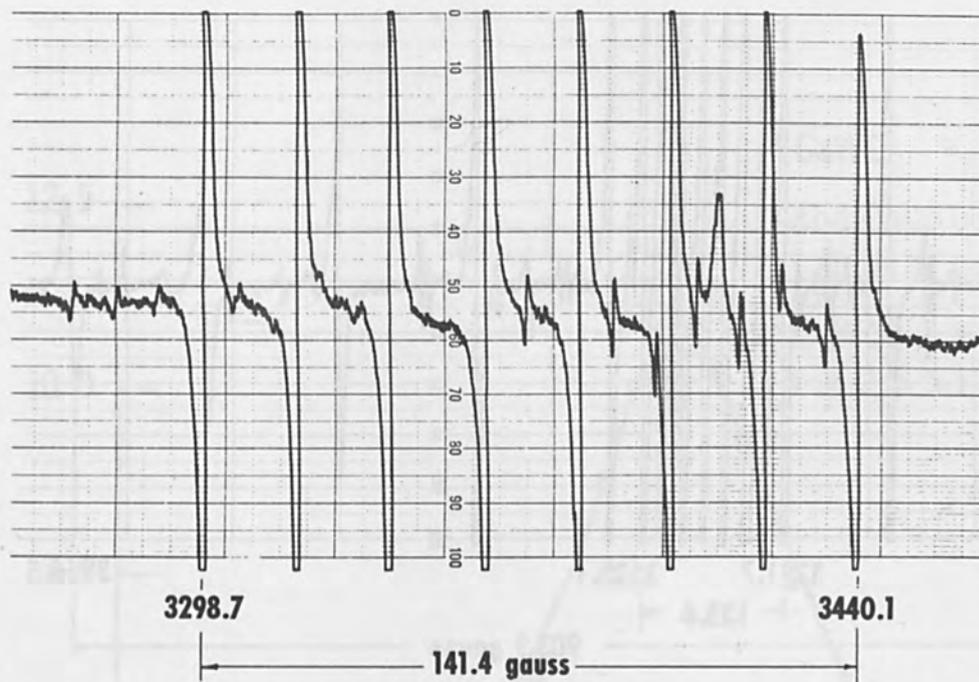
At $77^\circ K$, the Group A spectrum consists of 8 lines separated by about 20 gauss, isotropic g-value, and positive Δg . At $4.2^\circ K$, the spectrum becomes more complex, as will be discussed later. The Group B spectrum has been seen in the $CaWO_4:V$, Tb and $CaWO_4:V$, Nd samples. It is characterized by a hyperfine structure of about 10 gauss and also an anisotropic g-value with positive Δg . Its intensity is only about $1/20$ of Group A.

In contrast, for the Group C spectrum, the hyperfine structure is large, varying from about 176 to about 36 gauss, and the Δg is negative. This spectrum has been observed in samples containing vanadium only, or excess of vanadium over neodymium.

VI. GROUP A

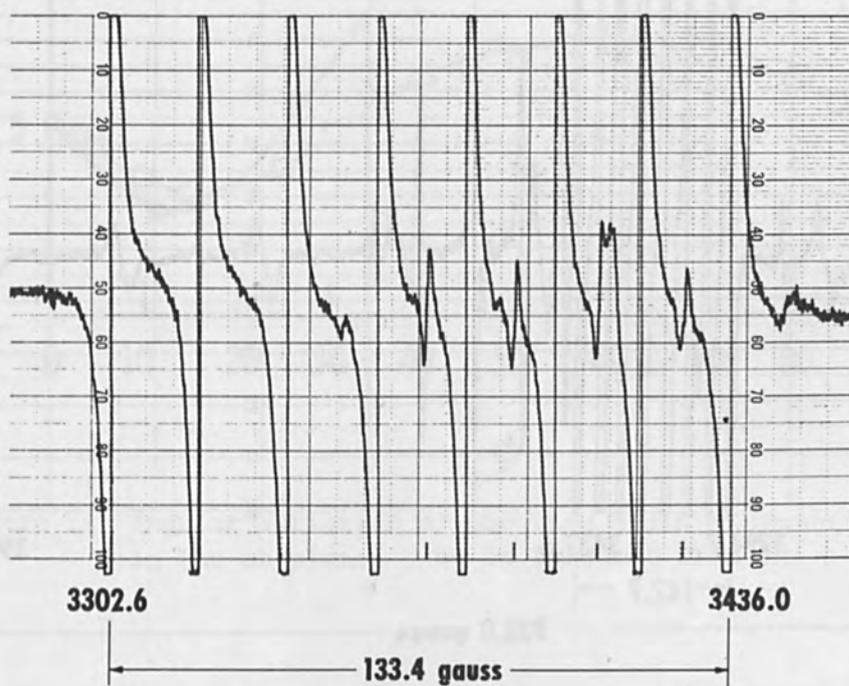
This is the strongest and the sharpest of the three groups and has been seen in all samples at $77^\circ K$. The g-value has an isotropic value of $2.0245 + .0005$, and the A-value is slightly anisotropic, with $A_{||} = (19.1 + .2)$ gauss = $(18.0 + .2) \times 10^{-4}$ cm⁻¹ along the c-axis, and $A_{\perp} = 20.3 + .2$ gauss = $(19.0 + .2) \times 10^{-4}$ cm⁻¹ in the ab-plane. The linewidth is angle dependent varying from about 2.5 gauss along the c-axis to about 1.6 gauss along the a-axis. Figure 7 gives the

X-BAND EPR SPECTRUM OF X-IRRADIATED $\text{CaWO}_4 : \text{V}, \text{Tb}(1, 1\%)$ AT $T = 77^\circ\text{K}$



A. H PARALLEL TO TWO-FOLD AXIS OF $[\text{WO}_4]$ TETRAHEDRON

Figure 5a. H in the ab-plane.



B. H PARALLEL TO c-AXIS

Figure 5b. H Parallel to the c-axis.

X-BAND EPR SPECTRUM OF IRRADIATED CaWO_4 : V (.5%) AT $T = 77^\circ\text{K}$

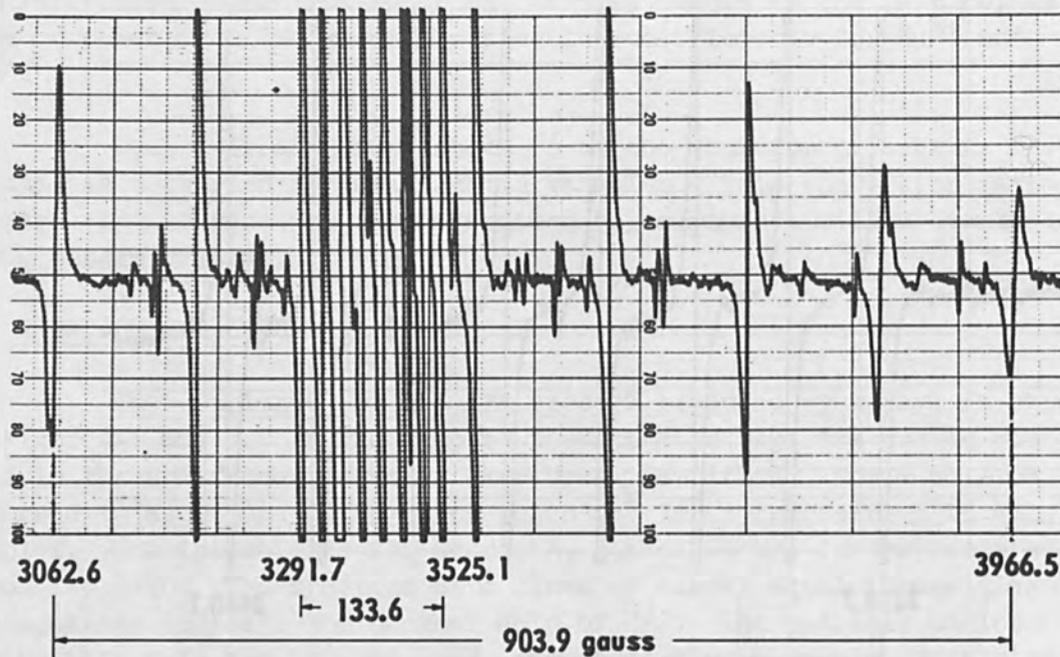


Figure 6a. H Parallel to the c-axis.

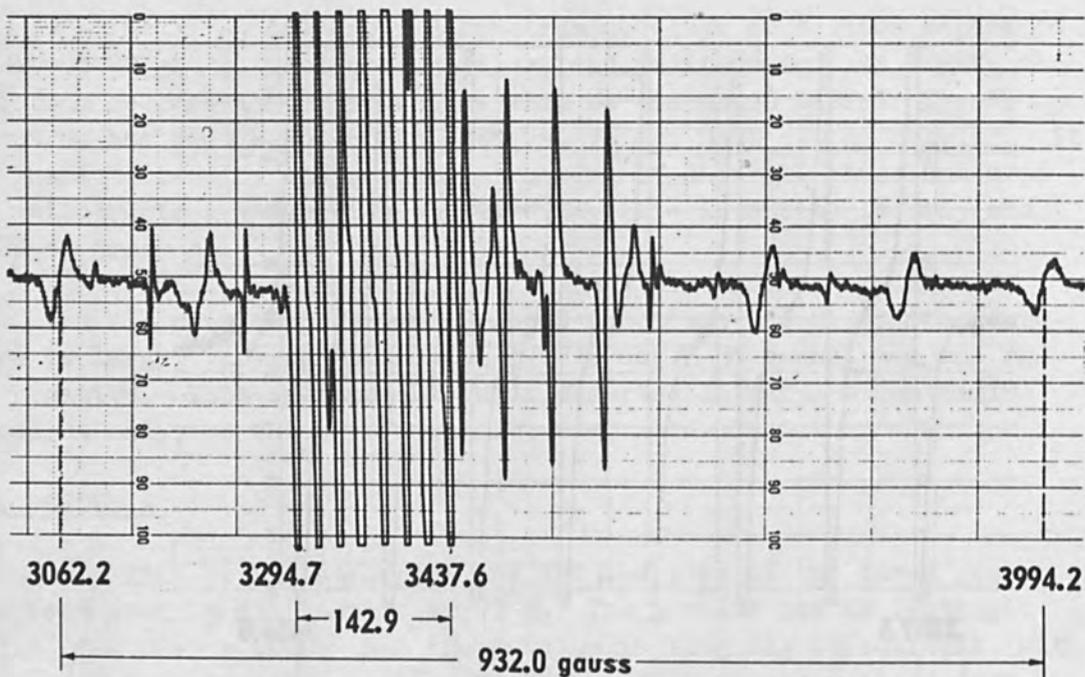


Figure 6b. H Parallel to the a-axis.

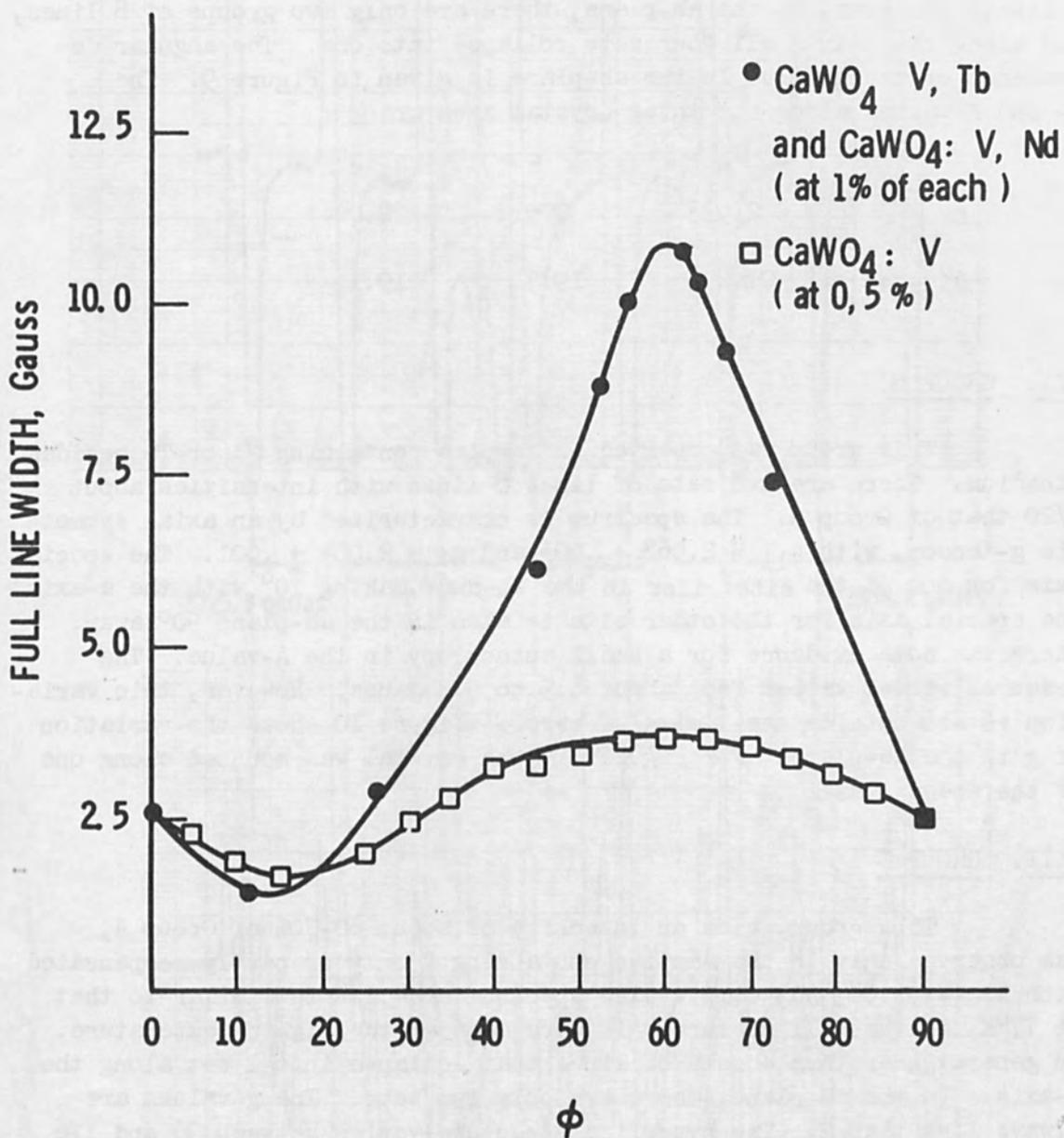


Figure 7. Angular Variation of the Line Width of Group A Lines at 77°K in the ab-plane. ϕ is Measured from the a-axis.

angular dependence. Note that the linewidth angular dependence is large for samples containing the rare-earths Nd and Tb, in addition to vanadium.

Figure 8 provides a comparison of the Group A spectra at 77°K and 4.2°K. In general, each hyperfine structure component splits into 4 lines. However, in the ab-plane, there are only two groups of 8 lines, and along the c-axis all four sets collapse into one. The angular dependence of the g-value in the ab-plane is given in Figure 9. The g- and A-values along the three crystal axes are

	a	b	c
g	2.033	2.017	2.024
A(gauss)	21.2	19.7	19.1

VII. GROUP B

This group was observed in samples containing Nd or Tb besides vanadium. There are two sets of lines 8 lines with intensities about 1/20 that of Group A. The spectrum is characterized by an axial symmetric g-tensor, with $g_{11} = 2.068 \pm .001$ and $g_1 = 2.004 \pm .001$. The special axis for one of the sites lies in the ab-plane, making 10° with the a-axis. The special axis for the other site is also in the ab-plane 90° away. There was some evidence for a small anisotropy in the A-value. The measured values varied from about 8.9 to 9.3 gauss. However, this variation is not outside the limits of error. Figure 10 shows the variation of g in the ab-plane. For Figure 11, the crystal was mounted along one of the special axes.

VIII. GROUP C

This group, with an intensity of about 60-70% of Group A, was observed only in the samples containing V only or partly compensated with Nd (V, 1.0%; Nd, 0.1%). The spectrum at 4.2°K is similar to that at 77°K, so that all measurements were made at the higher temperature. In general there are 4 sets of lines that collapse into 1 set along the c-axis. In the ab-plane, there are only two sets. The g-values are always less than 2. The hyperfine structure varies between 36 and 176 gauss. Four non-equivalent sites with the principal axes are shown in Table II. In the table, θ is the angle that each principal axis makes with the c-axis and ϕ is the angle of the projection of each principal axis on the ab-plane from the a-axis. It is to be noted that the axis associated with the smallest g-value (g_z) lies in the crystal planes ac and bc, making an angle of about 46° with the c-axis.

For the A-tensor, only the z-axis coincides with the corresponding

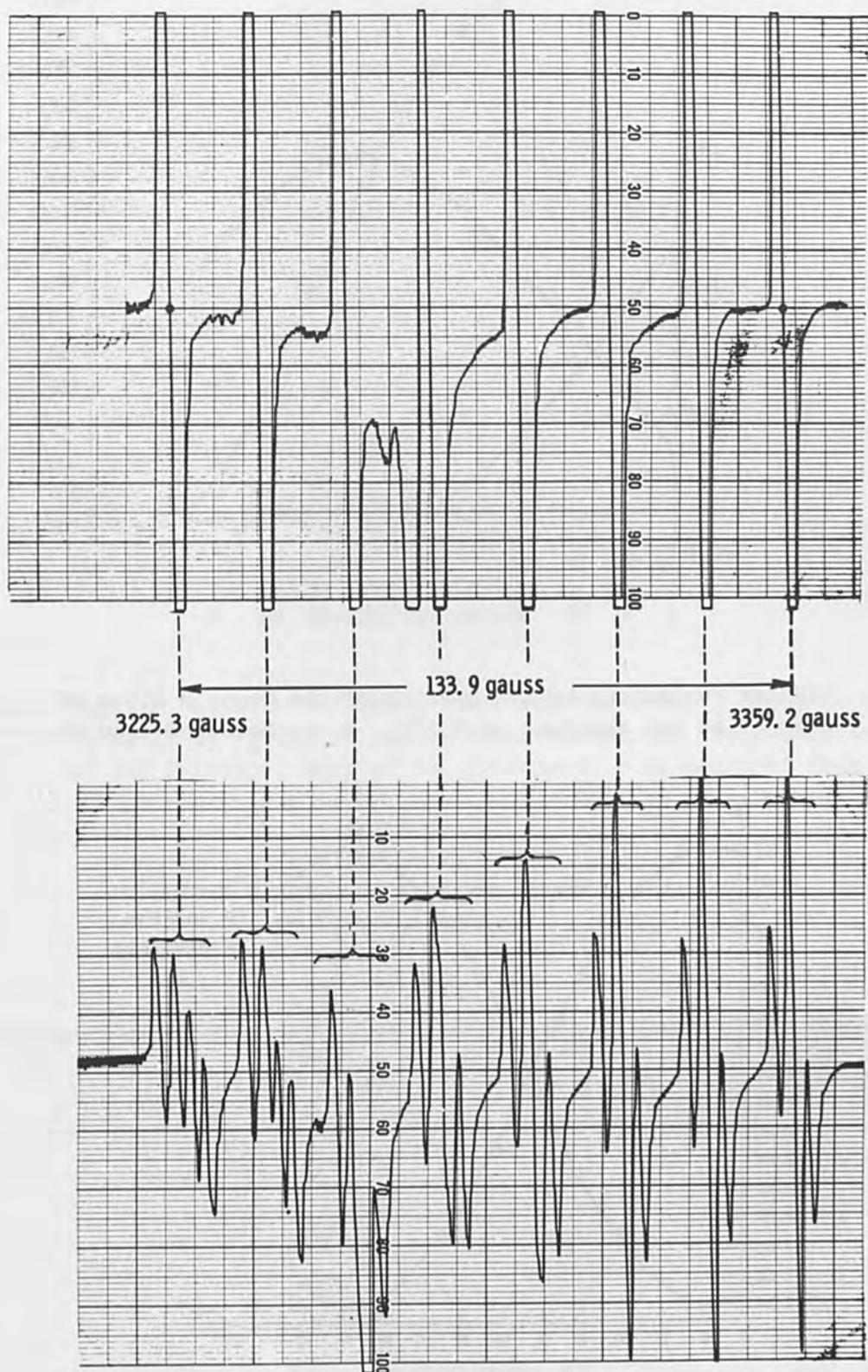


Figure 8. Group A Lines at 77°K (Top) and at 4.2°K (Bottom).

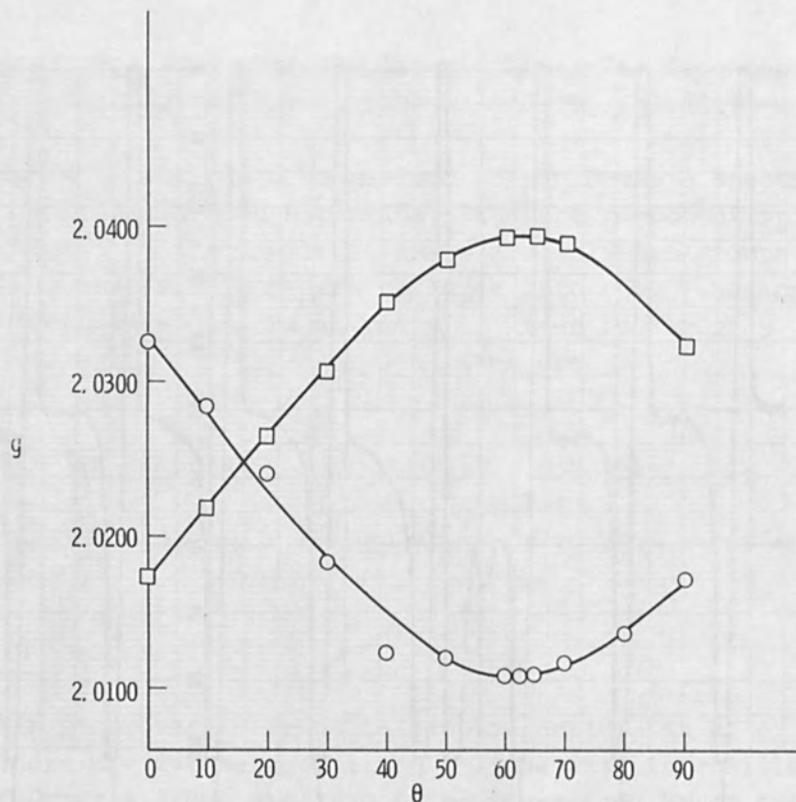


Figure 9. Angular Variations of the g-value of the Group A Lines of CaWO₄:V in the ab-plane at 4.2°K. ϕ is Measured from the a-axis.

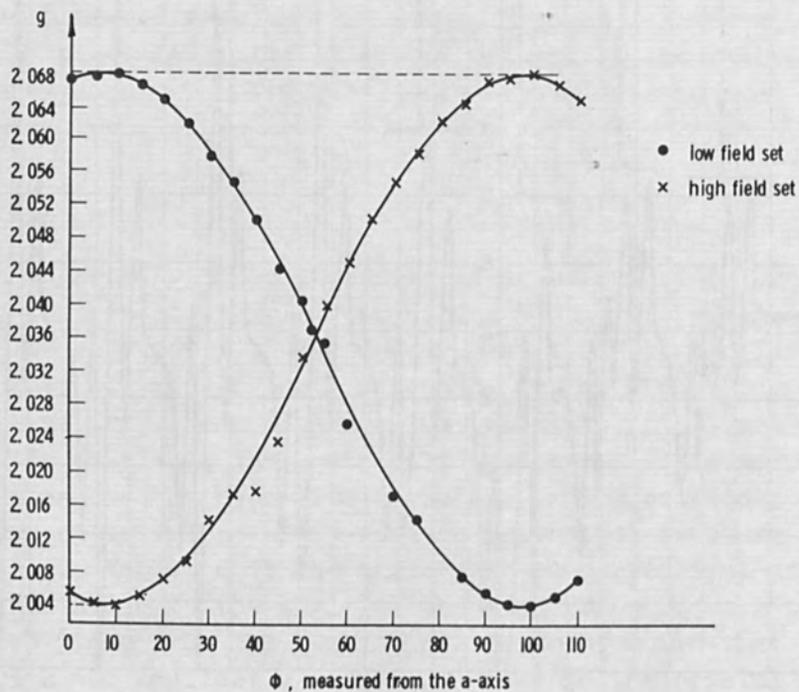


Figure 10. Angular Variation of the g-value for the Group B Lines of CaWO₄:V, Tb in the ab-plane, $T = 77^\circ\text{K}$. ϕ is Measured from the a-axis.

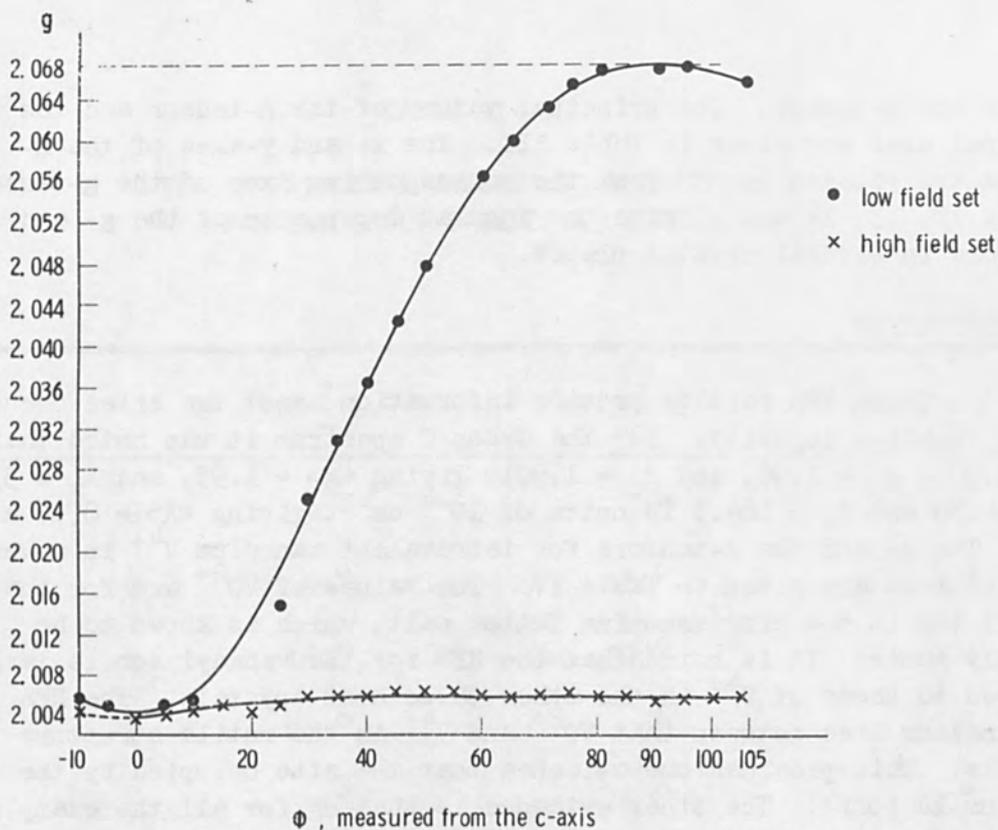


Figure 11. Angular Variation of the g-value for the Group B Lines of $\text{CaWO}_4:\text{V}$, Tb at 77°K . \underline{H} Varies in the Plane Containing the c-axis and one of the Principal Axes of the g-tensor. θ is measured from the c-axis.

TABLE II

PRINCIPAL VALUES AND AXES FOR GROUP C LINES
AT 77°K (g-TENSOR)

	g_x $1.985 \pm .005$	g_y $1.96 \pm .005$	g_z $1.9010 \pm .0005$
θ	66°	53°	46°
ϕ_1	115°	224°	0°
ϕ_2	205°	314°	90°
ϕ_3	295°	44°	180°
ϕ_4	25°	134°	370°

one for the g-tensor. The principal values of the A-tensor and the principal axes are given in Table III. The x- and y-axes of the A-tensors are rotated by 33° from the corresponding axes of the g-tensor. Figures 12, 13, 14 and 15 give the angular dependence of the g- and A-tensors in several crystal planes.

IX. DISCUSSION

These EPR results provide information about the sites occupied by the vanadium impurity. For the Group C spectrum it was noted that $g_x = 1.985$, $g_y = 1.96$, and $g_z = 1.9010$ giving $\langle g \rangle = 1.95$, and $A_x = 33.7$, $A_y = 64.5$, and $A_z = 164.5$ in units of 10^{-4} cm^{-1} , giving $\langle A \rangle = 87.6 \times 10^{-4} \text{ cm}^{-1}$. The g- and the A-tensors for tetravalent vanadium V^{4+} in several host lattices are given in Table IV. The values of VO^{2+} are for the vanadyl ion in the zinc ammonium Tutton salt, which is known to be strongly ionic. It is noted that the HFS for the vanadyl ion is large compared to those of V^{4+} in the other three host crystals. The HFS for vanadium lies between that VO^{2+} and V^{4+} in the rutile structure crystals. This provides one evidence that the site occupied by the vanadium is ionic. The other evidence is that Δg for all the examples cited is negative. In general this is to be expected for ions occupying ionic sites, for which case the crystal field calculations apply. Thus we conclude that the Group C spectrum is due to vanadium occupying Ca-sites.

For the Group A spectrum, on the other hand, measurements show that Δg is positive and the HFS is small. To account for the positive Δg , Karavelas and Kikuchi have made preliminary molecular orbital calculations for the orthovanadate complex situated in the $CaWO_4$ crystal. The results indicate that the unpaired electron has the symmetry of the Γ_2 irreducible representation of the S_4 group, and the g-values are $g = 2.0268$, $g = 2.0231$ and $\langle g \rangle = 2.024$, after making an electrostatic correction stemming from the transfer of an electron from the peripheral ligand orbital to the central metal orbital. These calculated results are in excellent agreement with the experimental value of $g = 2.0245$. Thus Group A spectrum is assigned to vanadium in W-site. Additional evidence is provided by the facts that the orthovanadate ion VO_4^{3-} exists chemically and that the Group A has been observed in all samples containing vanadium.

Since x-irradiation is required for the production of both the Group A and C spectra, it was suggested to us¹⁶ that the mechanism might be one of electron transfer, following the splitting of the non-bonding P_z -orbital pair on the oxygen. Consequently a series of measurements investigating the growth and annealing of Group A and C lines were carried out at 23, 68, and 100°C . At 100°C , for example, the Group A spectrum grows more slowly but anneals out more quickly than the Group

TABLE III

A-TENSOR PRINCIPAL VALUES AND DIRECTIONS
FOR THE GROUP C LINES AT 77°K

	A_x	A_y	A_z^*
	33.7 ± 1.5	64.5 ± 1	164.5 ± 1
θ	90°	44°	46°
φ_1	90°	180°	0°
φ_2	180°	270°	90°
φ_3	270°	0°	180°
φ_4	0°	90°	270°

*In units of 10^{-4} cm^{-1}

TABLE IV

V^{4+} g- AND A-TENSORS IN SEVERAL CRYSTALS

Crystal	g_x	g_y	g_z	$\langle g \rangle$	A_x	A_y	A_z	$\langle A \rangle^*$
TiO ₂	1.915	1.912	1.956	1.928	31	43	142	72
SnO ₂	1.939	1.903	1.943	1.928	21	44	144	70
GeO ₂	1.921	1.921	1.963	1.935	36.7	37.5	134.4	69.5
VO ²⁺	1.9813	1.9801	1.9331	1.9648	71.2	72.4	182.8	108.8

*In units of 10^{-4} cm^{-1} .

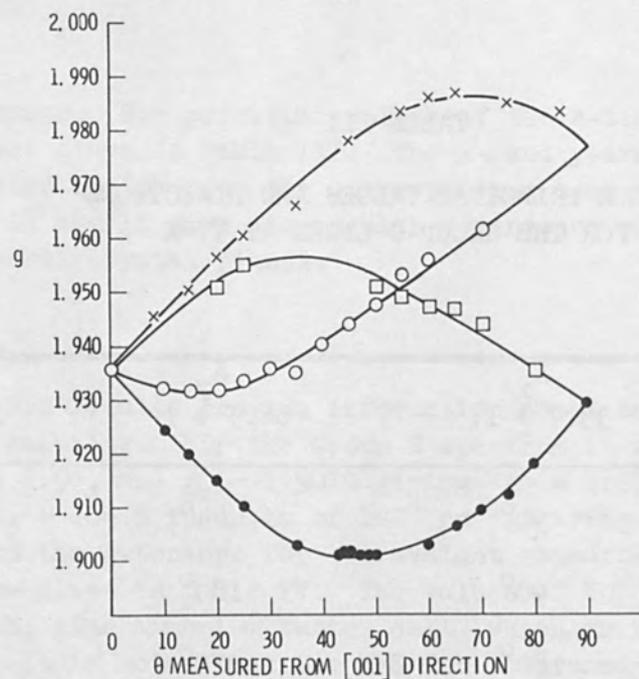


Figure 12. Angular Variation of the g-value for the Group C Lines in $\text{CaWO}_4:\text{V}$ in the ac-plane at 77°K . θ Is Measured from the c-axis.

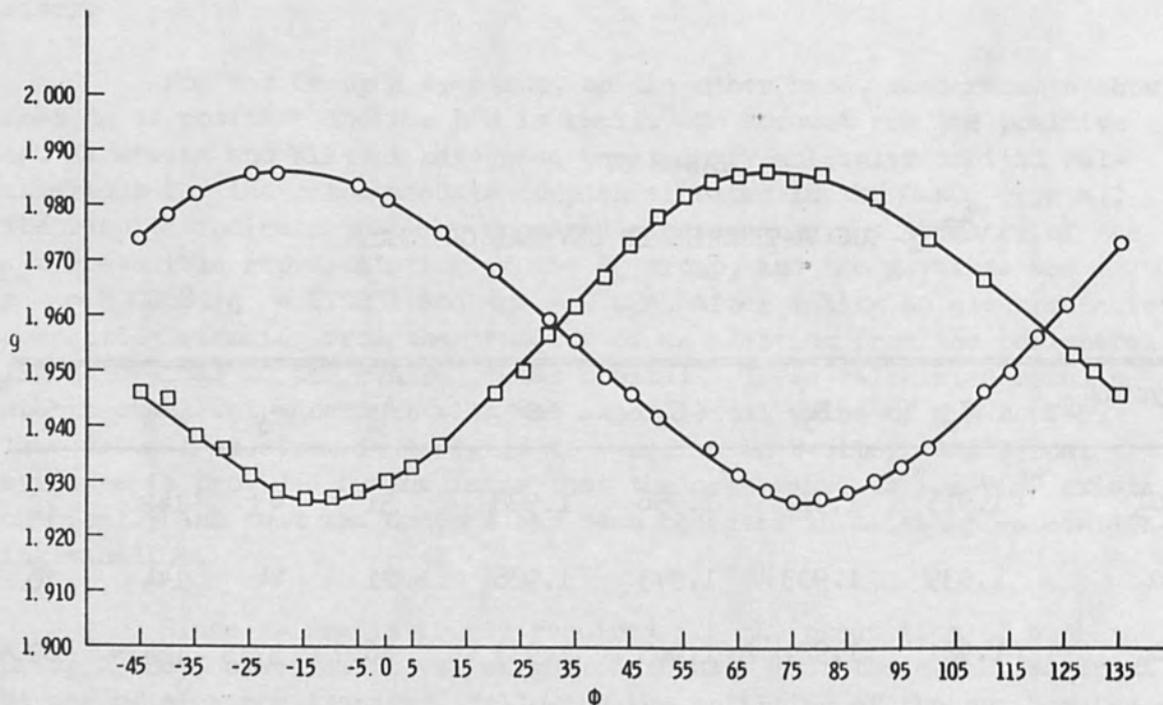


Figure 13. Angular Variation of the g-value for the Group C Lines of $\text{CaWO}_4:\text{V}$ in the ab-plane. ϕ is Measured from the a-axis.

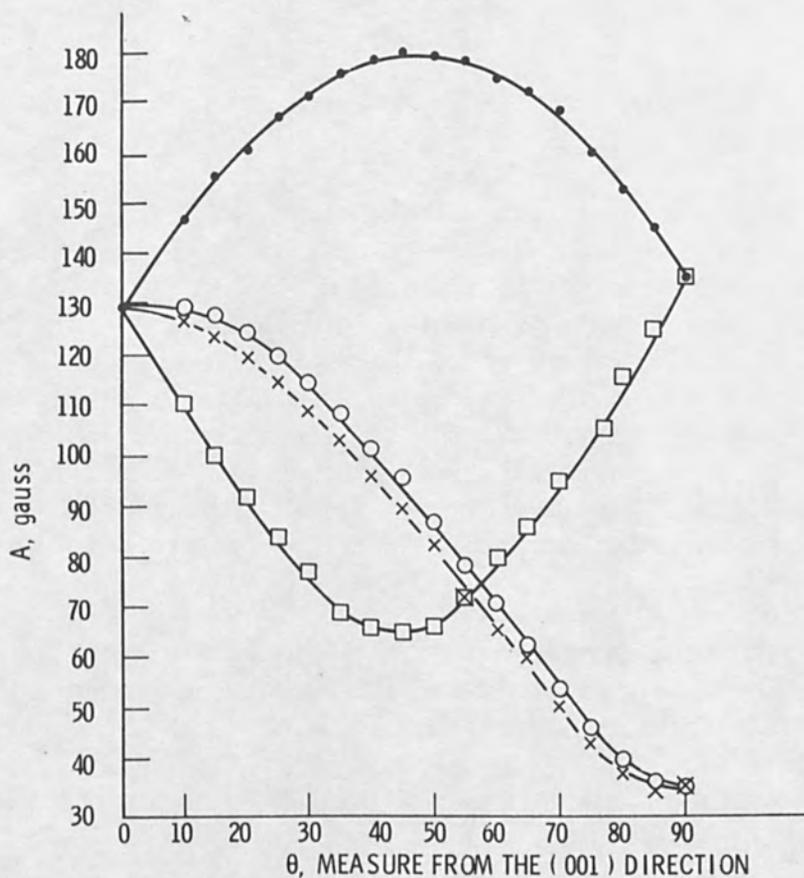


Figure 14. Angular Variation of the HFS for the Group C Lines of $\text{CaWO}_4:\text{v}$ in the ac-plane at 77°K . θ is Measured from the c-axis.

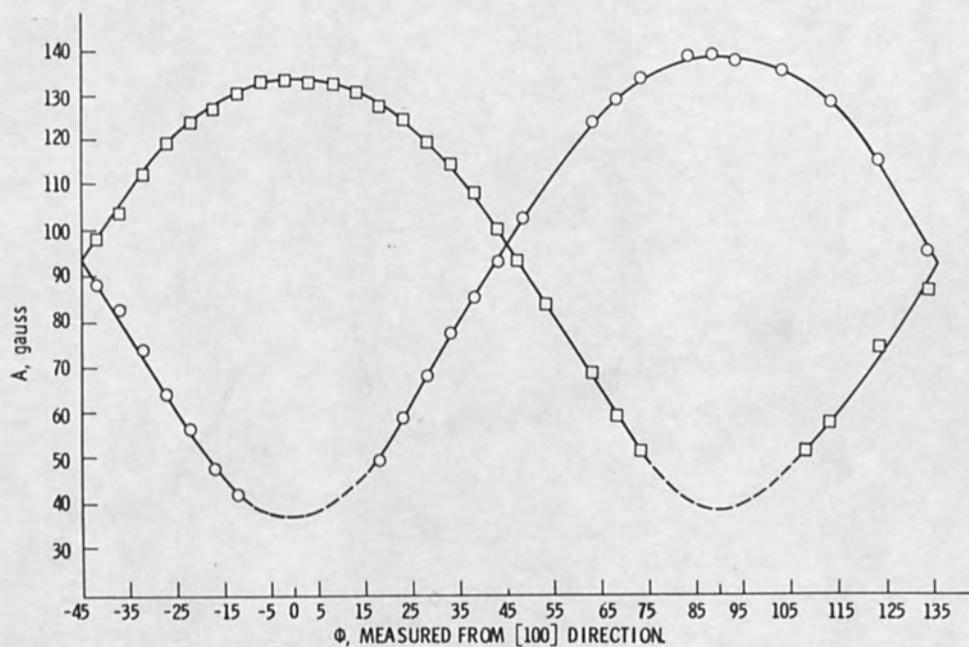


Figure 15. Angular Variation of the HFS for the Group C Lines of $\text{CaWO}_4:\text{v}$ at 77°K in the ab-plane. ϕ is Measured from the a-axis.

C spectrum. Thus it appears that the relation between the Group A and C spectra is not a simple one.

Preliminary investigations show that $\text{CaWO}_4:\text{Nd}$ samples compensated with Nb or V have a pair of sharp optical absorption lines on the long wave length side of the transition to the ${}^4\text{F}_{3/2}$ level. As Figure 16 shows, the extra pair occur at 8785.5 Å ($11,382.4 \text{ cm}^{-1}$) and 8783.3 Å ($11,385.3 \text{ cm}^{-1}$) compared to the sharp component of ${}^4\text{F}_{3/2}$ at 8766.8 Å ($11,406.8 \text{ cm}^{-1}$), for the sample containing vanadium. For Nb-compensated sample (Figure 17) the two lines are at 8782.6 ($11,386.0 \text{ cm}^{-1}$) and 8778.6 Å ($11,391.3 \text{ cm}^{-1}$). Similar effect has been observed in $\text{CaMO}_4:\text{Nd}$, Nb. (See Figure 18.) C. Morrison** has tentatively suggested that the extra pair stems from charge compensator occupying the two nearby sites.

We wish to acknowledge the continued interest and stimulating discussions with Clyde Morrison of the Harry Diamond Laboratories, and Dr. Sophocles Karavelas of the Nuclear Research Center Democritus, Athens, Greece.

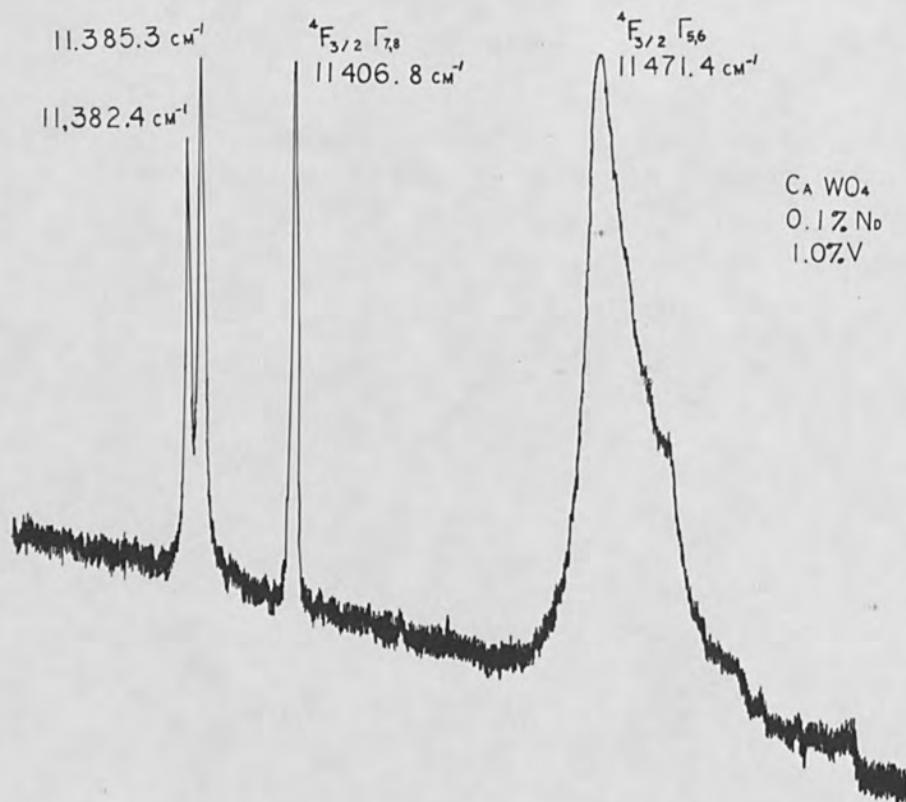


Figure 16. Absorption Spectrum of CaWO_4 Containing Nd and V.

** Private Communication.

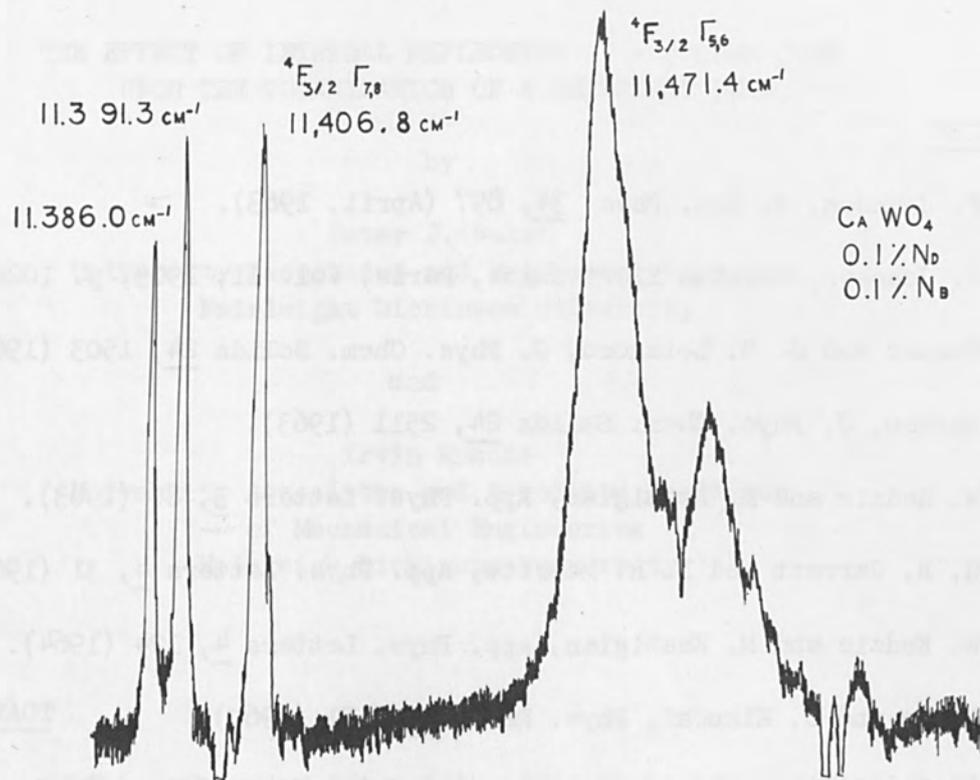


Figure 17. Absorption Spectrum CaWO₄ Containing Nd and Nb.

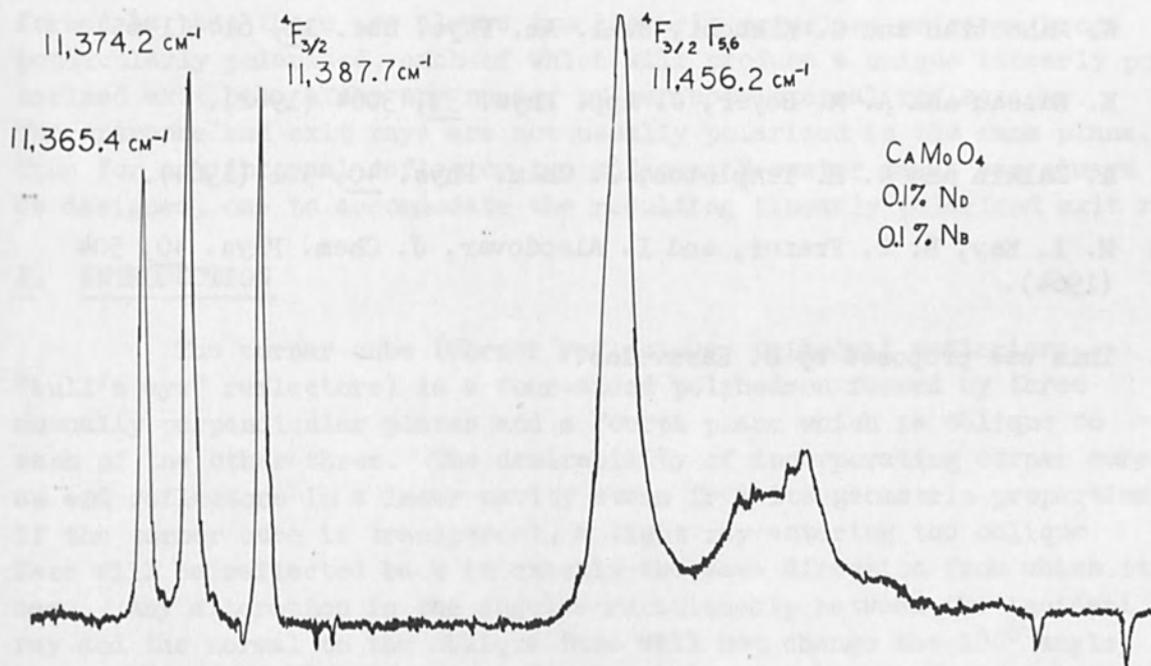


Figure 18. Absorption Spectrum of CaMoO₄ Containing Both Nd and Nb.

REFERENCES

1. L. F. Johnson, J. App. Phys. 34, 897 (April, 1963).
2. L. F. Johnson, Quantum Electronics, Paris, Vol. II, 1963, p. 1021.
3. K. Nassau and G. M. Loiacono, J. Phys. Chem. Solids 24, 1503 (1963).
4. K. Nassau, J. Phys. Chem. Solids 24, 2511 (1963).
5. R. W. Kedzie and M. Kestigian, App. Phys. Letters 3, 86 (1963).
6. C. G. B. Garrett and F. R. Merritt, App. Phys. Letters 4, 31 (1964).
7. R. W. Kedzie and M. Kestigian, App. Phys. Letters 4, 124 (1964).
8. J. Lambe and C. Kikuchi, Phys. Rev. 118, 71 (1960).
9. R. H. Borcherts and C. Kikuchi, J. Chem. Phys. 40, 2270 (1964).
10. C. Kikuchi, I. Chen, W. H. From, and P. B. Dorain, J. Chem. Phys. 42, 181 (1965).
11. S. Karavelas and C. Kikuchi, Unpublished
12. N. Mahootian and C. Kikuchi, Bull. Am. Phys. Soc. 11, 614 (1965).
13. K. Nassau and A. M. Boyer, J. App. Phys. 33, 3064 (1962).
14. A. Zalkin and D. H. Templeton, J. Chem. Phys. 40, 501 (1964).
15. M. I. Kay, B. C. Frazer, and I. Almodovar, J. Chem. Phys. 40, 504 (1964).
16. This was proposed by S. Karavelas.

THE EFFECT OF INTERNAL REFLECTION IN A CORNER CUBE
UPON THE POLARIZATION OF A REFLECTED BEAM

by

Peter J. Walsh
University Associates and Professor of Physics
Fairleigh Dickinson University

and

Irvin Krause
University Associates and Associate Professor
of Mechanical Engineering
Fairleigh Dickinson University

ABSTRACT

The problem of determining the change of polarization state of a ray due to multiple internal reflections within a many-sided reflectors is discussed. Particular attention is paid to the corner cube which because of its retro direction properties acts as a self-aligning element in a laser cavity. A matrix method and a graphical method, are presented for dealing with the problem. The graphical method is based on the use of the Poincaré sphere. A very important conclusion formed is that there are always two linearly polarized entrance beams perpendicularly polarized, each of which will produce a unique linearly polarized exit beam after any number of perfect internal reflections. The entrance and exit rays are not usually polarized in the same plane. Thus for any internal reflector two adjacent Brewster faces can always be designed, one to accommodate the resulting linearly polarized exit ray.

I. INTRODUCTION

The corner cube (corner reflectors, trihedral reflectors, "bull's eye" reflectors) is a four-sided polyhedron formed by three mutually perpendicular planes and a fourth plane which is oblique to each of the other three. The desirability of incorporating corner cubes as end reflectors in a laser cavity stems from its geometric properties. If the corner cube is transparent, a light ray entering the oblique face will be reflected back in exactly the same direction from which it came. Any alteration in the angular relationship between the incident ray and the normal to the oblique face will not change the 180° angle between the incident ray and the reflected ray. Hence, for laser

applications, the deleterious effect of reflector misalignments may be minimized. If the corner cube reflector is solid, it has the additional advantage that its reflections are internal. This results in low reflection losses.

In order to insure low transmission losses of a linear polarized beam, the front of a solid corner cube should be a Brewster face which would limit the reflection off that face. An important difficulty in the use of the corner cube arises from the fact that an incident ray must undergo several internal reflections before it emerges from the oblique face. Unfortunately, the internal reflections in the corner cube will usually result in elliptical polarization of light which was originally plane polarized. Thus the effectiveness of the Brewster face in transmitting the emerging beam is hampered. If a direction of linear polarization is to be maintained, then reflection from the three mutually orthogonal surfaces must result in an emerging ray which has the same state (plane) of polarization. Maintenance of the same polarization direction through successive internal reflections leads to a unique design problem encountered in the use of cube corner reflectors.

It is the purpose of this paper to present an analytical method and its graphical counterpart for determining a final state of polarization of an initially polarized light ray which has had multiple reflections from surfaces. The analytical method utilizes transformation matrices which operate on the incident ray vector. The graphical analogue is based upon the mapping technique of the Poincaré sphere¹ which can be used as long as the magnitude of the amplitude of the wave remains constant throughout its reflections. Although a specific application is made to the aforementioned reflector, the method is generally applicable to successive internal reflection in other geometrical systems.

II. MATHEMATICAL DESCRIPTION OF THE PROBLEM

The problem undertaken here is the description of the state of polarization produced by multiple complete internal reflections². We are particularly interested in the effect on linear polarization of three complete internal reflections in a corner cube. If the same direction of linear polarization can be maintained, then a single Brewster plane can be cut on the oblique entrance face and extremely low loss may be assured for a suitably polarized beam entering and exiting the corner cube. The techniques presented are obviously applicable to a wide variety of problems which are related to the one delineated above.

The geometry of the corner cube is illustrated in Figure 1. We are concerned with the change in polarization between points A and B due to complete internal reflection at the orthogonal surfaces in the sequence 1, 2, and 3.

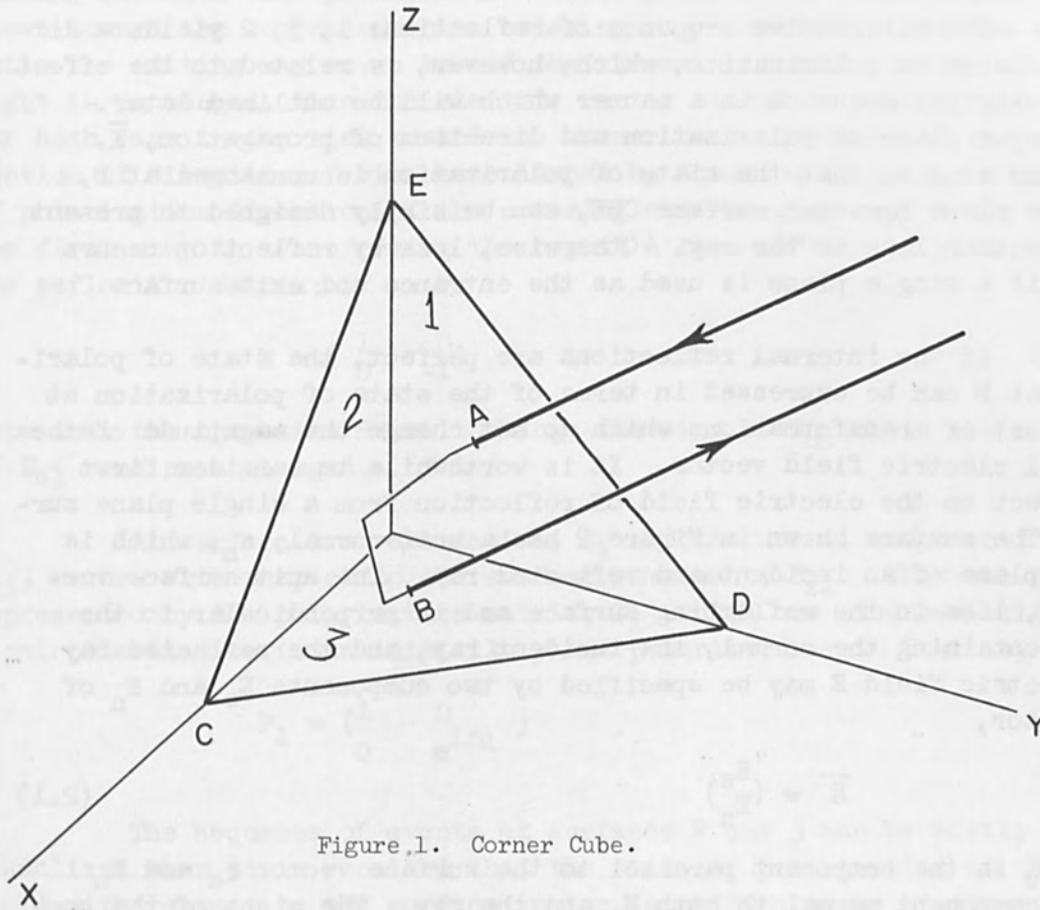


Figure 1. Corner Cube.

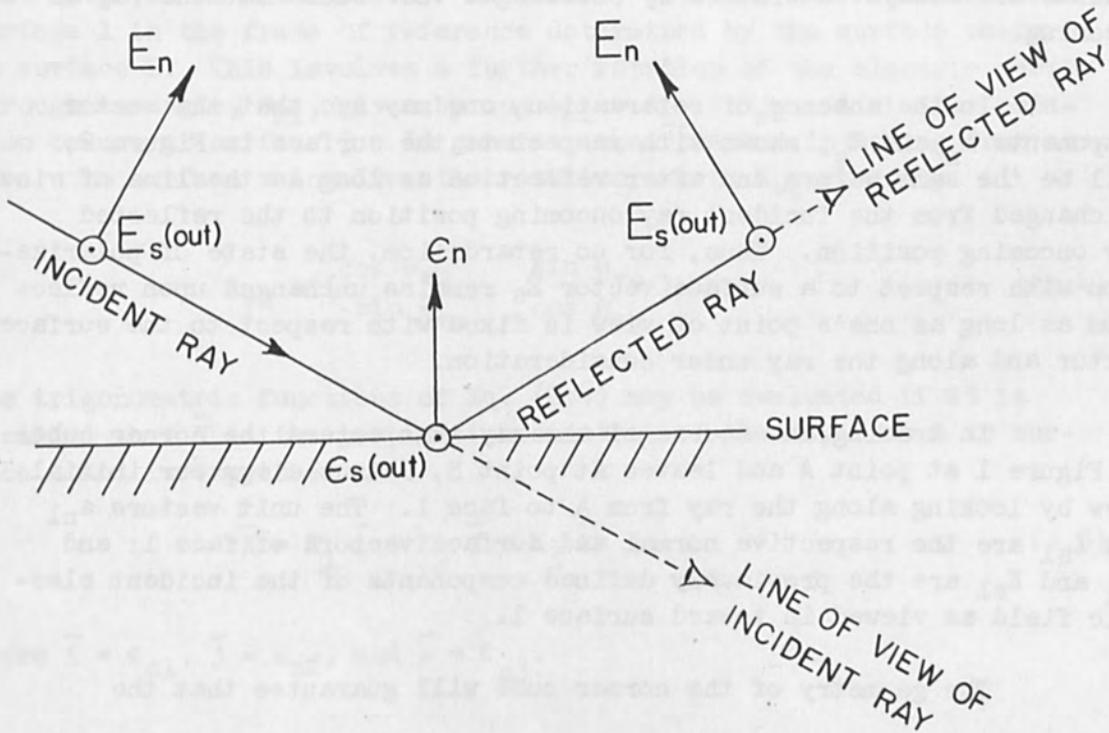


Figure 2. Geometry of A Reflected Polarized Ray.

The alternative sequence of reflection: 1, 3, 2 yields a different change in polarization, which, however, is related to the effect of the original sequence in a manner which will be outlined later. If a proper plane of polarization and direction of propagation, \bar{K} , are found at A so that the state of polarization is unchanged at B, then the plane Brewster surface CDE, can be simply designed to present no reflection loss to the ray. Otherwise, loss by reflection occurs at CDE if a single plane is used as the entrance and exit surface.

If the internal reflections are perfect, the state of polarization at B can be expressed in terms of the state of polarization at A by a set of transformations which do not change the magnitude of the original electric field vector. It is worthwhile to consider first the effect on the electric field of reflection from a single plane surface. The surface shown in Figure 2 has a unit normal, ϵ_n , which is in the plane of an incident and reflected ray. The unit surface vector, ϵ_s , lies in the reflecting surface and is perpendicular to the plane containing the normal, the incident ray, and the reflected ray. The electric field E may be specified by two components E_s and E_n of the vector,

$$\bar{E} = \begin{pmatrix} E_s \\ E_n \end{pmatrix} \quad (2.1)$$

where E_s is the component parallel to the surface vector ϵ_s and E_n is the component normal to both E_s and the ray. The signs of the components are always determined by choosing a view such that the ray is oncoming.

In the absence of retardation, one may see that the vector components E_n and E_s , shown with respect to the surface in Figure 2, will be the same before and after reflection as long as the line of view is changed from the incident ray oncoming position to the reflected ray oncoming position. Thus, for no retardation, the state of polarization with respect to a surface vector E_s remains unchanged upon reflection as long as one's point of view is fixed with respect to the surface vector and along the ray under consideration.

In tracing the course of the ray that enters the corner cube of Figure 1 at point A and leaves at point B, let us align our initial view by looking along the ray from A to face 1. The unit vectors ϵ_{n1} and ϵ_{s1} are the respective normal and surface vectors of face 1; and E_{n1} and E_{s1} are the previously defined components of the incident electric field as viewed in toward surface 1.

The geometry of the corner cube will guarantee that the

exiting ray at B will parallel the entering ray at A. However, if one views toward the corner cube the entering ray moves away from the viewer while the exiting ray moves toward the viewer. If polarization of the light is unchanged, then the electric vector should be exactly the same for both rays in the view toward the corner cube. Because of the convention of direction of view established by Equation (2.1) the identity of polarization direction of the entering and emerging rays means that the first step in transforming the ray at A is a reflection given by the reflection matrix:

$$R = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \quad (2.2)$$

Note that the operation $R\bar{E}$ will change the sign of E_{n1} while the sign of E_{s1} remains unchanged.

At surface 1 a retardation, Δ_1 of the normal component, E_{n1} , occurs with respect to the surface component, E_{s1} . This can be represented by the polarization matrix, P_1 , which now produces elliptical polarization in an initially plane polarized beam. That is,

$$P_1 = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\Delta_1} \end{pmatrix} . \quad (2.3)$$

The sequence of events at surfaces 2 and 3 can be easily visualized when it is recalled that the state of polarization of the field with respect to the surface vector is unaffected by a reflection that involves no retardation. As before, view the ray reflected from surface 1 in the frame of reference determined by the surface vector in surface 2. This involves a further rotation of the electric vector through an angle, θ_{21} , given by $\cos \theta_{21} = \epsilon_{s1} \cdot \epsilon_{s2}$. A similar rotation occurs at surface 3. In general, the rotation in moving from surface u to surface v involves the rotation matrix R_{vu} given by

$$R_{vu} = \begin{pmatrix} \cos \theta_{vu} & \sin \theta_{vu} \\ -\sin \theta_{vu} & \cos \theta_{vu} \end{pmatrix} . \quad (2.4)$$

The trigonometric functions of Eq. (2.4) may be evaluated if it is recalled that \bar{K} is the propagation constant of the ray from A to surface 1 and is expressed as

$$\bar{K} = K_1 \bar{i} + K_2 \bar{j} + K_3 \bar{k} \quad (2.5)$$

where $\bar{i} = \epsilon_{n1}$, $\bar{j} = \epsilon_{n2}$, and $\bar{k} = \epsilon_{n3}$.

The orthogonality conditions between \bar{K} and the unit surface vectors and between the unit normal and the unit surface vector demand that,

$$\epsilon_{su} \cdot \bar{K} = 0 \text{ and } \epsilon_{su} \cdot \epsilon_{nu} = 0$$

From Equation (2.5) and the above orthogonality requirement, one finds that

$$\cos \theta_{vu} = \frac{K_u K_v}{\sqrt{K_u^2 + K_v^2} \sqrt{K_v^2 + K_w^2}} \quad (2.6)$$

The complete transformation matrix T, is now given by

$$T = R_{13} P_3 R_{32} P_2 R_{21} P_1 R \quad (2.7)$$

The final rotation matrix, R_{13} , is needed to compare our final view upon reflection from surface 3 with the initial view when the ray is incident to surface 1. In general, the final state of polarization of the light from the corner cube is given by $T\bar{E}$ which can obviously be generalized to other internal reflectors besides a cube corner. If the special requirement is imposed that initially plane polarized light, \bar{E} , be plane polarized in the same plane after the three internal reflections, then

$$T \begin{pmatrix} E_{s1} \\ E_{n1} \end{pmatrix} = e^{i\beta} \begin{pmatrix} E_{s1} \\ E_{n1} \end{pmatrix} \quad (2.8)$$

The quantity β represents a phase shift common to both components of \bar{E} . This phase shift will add to the total phase shift within a laser. Since the laser will automatically choose modes such that the round trip phase shift is an integral number of 2π radians, this phase shift need not concern us here.

The retardation for complete internal reflection is not completely known experimentally.³ In the absence of experimental data theoretical formulas are available. The retardation, Δ , produced by complete internal reflection within a material of relative index, n , is given by³

$$\tan \frac{\Delta}{2} = \sqrt{1 - \left(\frac{1}{n \sin \gamma}\right)^2} \cot \gamma \quad (2.9)$$

γ is the angle between the ray vector and the normal to the surface and may be expressed as:

$$\cos \gamma_u = - \frac{K_u}{\sqrt{K_1^2 + K_2^2 + K_3^2}} \quad (2.10)$$

Figure 3 shows a graph of Δ vs γ for $n = 1.50$.

A simple solution of Equation (2.8) exists when the retardations are all zero. In this special case, the polarization matrices are all unit matrices. The product of the three rotation matrices corresponds to a single rotation by the angle; $\theta_{21} + \theta_{32} + \theta_{13}$. Let $E_{s1} = E \cos \theta_p$ and $E_{n1} = E \sin \theta_p$ where θ_p is the initial polarization angle with respect to ϵ_{s1} . Then Equation (2.8) yields

$$\begin{aligned} \cos (\theta_{21} + \theta_{32} + \theta_{13} + \theta_p) &= e^{i \beta} \cos \theta_p \\ -\sin (\theta_{21} + \theta_{32} + \theta_{13} + \theta_p) &= e^{i \beta} \sin \theta_p \end{aligned}$$

Two simple solutions exist:

$$\begin{aligned} \beta &= 0, \theta_p = \frac{\theta_{21} + \theta_{32} + \theta_{13}}{2} \\ \beta &= \pi, \theta_p = \frac{\theta_{21} + \theta_{32} + \theta_{13}}{2} + \frac{\pi}{2} \end{aligned} \quad (2.11)$$

These simple solutions are extremely useful as a starting point in practical calculations, especially in using the graphical techniques discussed in the next section.

The problem of finding the final state of polarization produced by reflections of a given incident ray from surfaces in the order: 1, 3, 2 is equivalent to that of reflections of a ray of different propagation constant from surfaces in the order: 1, 2, 3. If K_x, K_y, K_z are the components of a ray suffering reflections from surfaces in the order: 1, 2, 3, then the final state of polarization is identical to that produced by the same incident electric field with propagation components K_x, K_z, K_y and undergoing reflections sequential from surfaces: 1, 3, 2. This can be seen by direct calculation of the angles θ and ϕ for both cases and follows from the fact that for the sequence: 1, 3, 2 planes XY and XZ exchange the roles that they had for the sequence: 1, 2, 3.

III. GRAPHICAL TECHNIQUE. GENERAL

The mathematical operations describing the effect on polarized light of successive reflections from surfaces that produce retardations can be described in an elegant geometrical manner by the use of the Poincaré sphere.^{1,4}

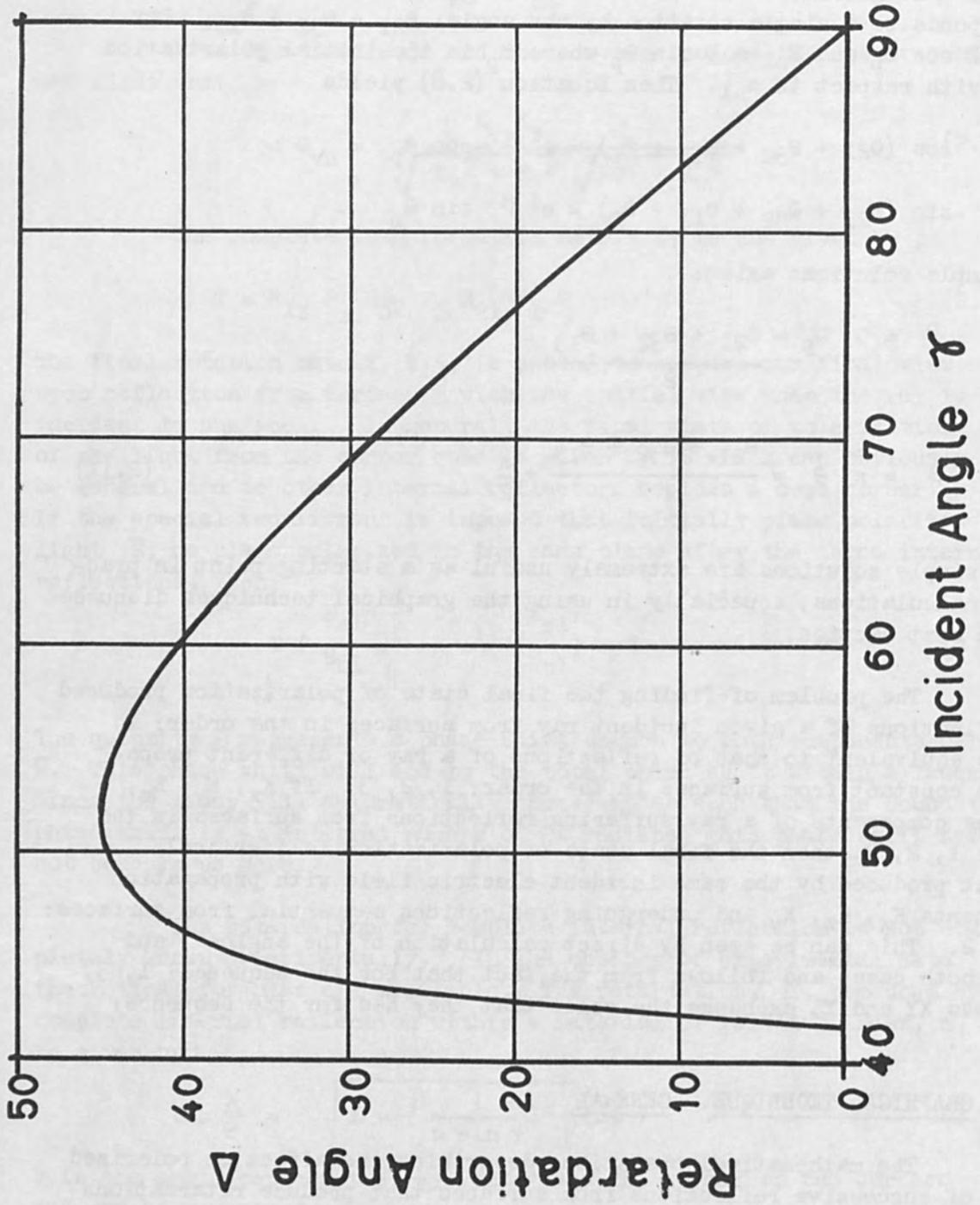


Figure 3. Graph of Δ vs. γ for $n = 1.50$

Any possible state of polarization of a monochromatic, constant intensity light ray is represented by a single point on the surface of the Poincaré sphere. One such sphere, whose center is at 0, is shown in the OXYZ coordinate system in Figure 4. The equator, which is the intersection of the spherical surface with the equatorial plane OXY, represents the locus of all states of linear polarization. When the plane of polarization, represented by the electric vector, is parallel to the unit surface vector of a reflecting surface, the plane of polarization is said to be horizontal. Note that a polarization angle of 180° from the horizontal results in exactly the same plane of polarization as the horizontal. Thus in a full progression westward around the equator, the plane of polarization of the corresponding linear polarized light is taken to rotate uniformly counterclockwise through 180° degrees with respect to the horizontal. If the horizontal is taken as the OX axis then the intercepts of the OY axis on the equator represent linear polarization angles of 45° and 135° with respect to the horizontal.

Various states of elliptically polarized light are represented by points that lie off the equator. The north and south poles represent, respectively, left hand and right hand circular polarized light. Latitudes are loci of constant ellipticity. In general the upper half of the sphere represents left hand polarized light whose major axis is oriented at an angle α which is half the longitudinal angle, and whose ellipticity is determined by the latitude, 2ω . Specifically, the ellipticity, is given by

$$b/a = \tan |\omega|$$

The advantage of the Poincaré sphere is that the operations of reflection, axis rotation, and retardation can be simply represented upon a single sphere, when no loss in amplitude is involved. Reflection of a state of polarization, \bar{E} , through the vertical axis means that the point, \bar{E} , on the sphere of Fig. 4 is transferred to the point, $R\bar{E}$, at the same latitude but to the corresponding negative longitude. Thus a westward rotation about the polar axis by $(360 - 4\alpha)$ degrees occurs. A rotation of the axis of polarization is accomplished by a westward rotation about the polar axis by twice the rotation angle and is represented by point $R_{UV}\bar{E}$ in the figure. Finally, retardation is accomplished by a rotation about the OX axis by twice the retardation angle. This is shown as point $P\bar{E}$ on the figure.

In the reflector that uses only completely internal reflection from any number of surfaces, the operations that must be performed on the entrance ray to find the resulting polarization of the exit ray are a combination of change of view or reflection, rotations of view, and retardations as described above. If no energy loss is involved,

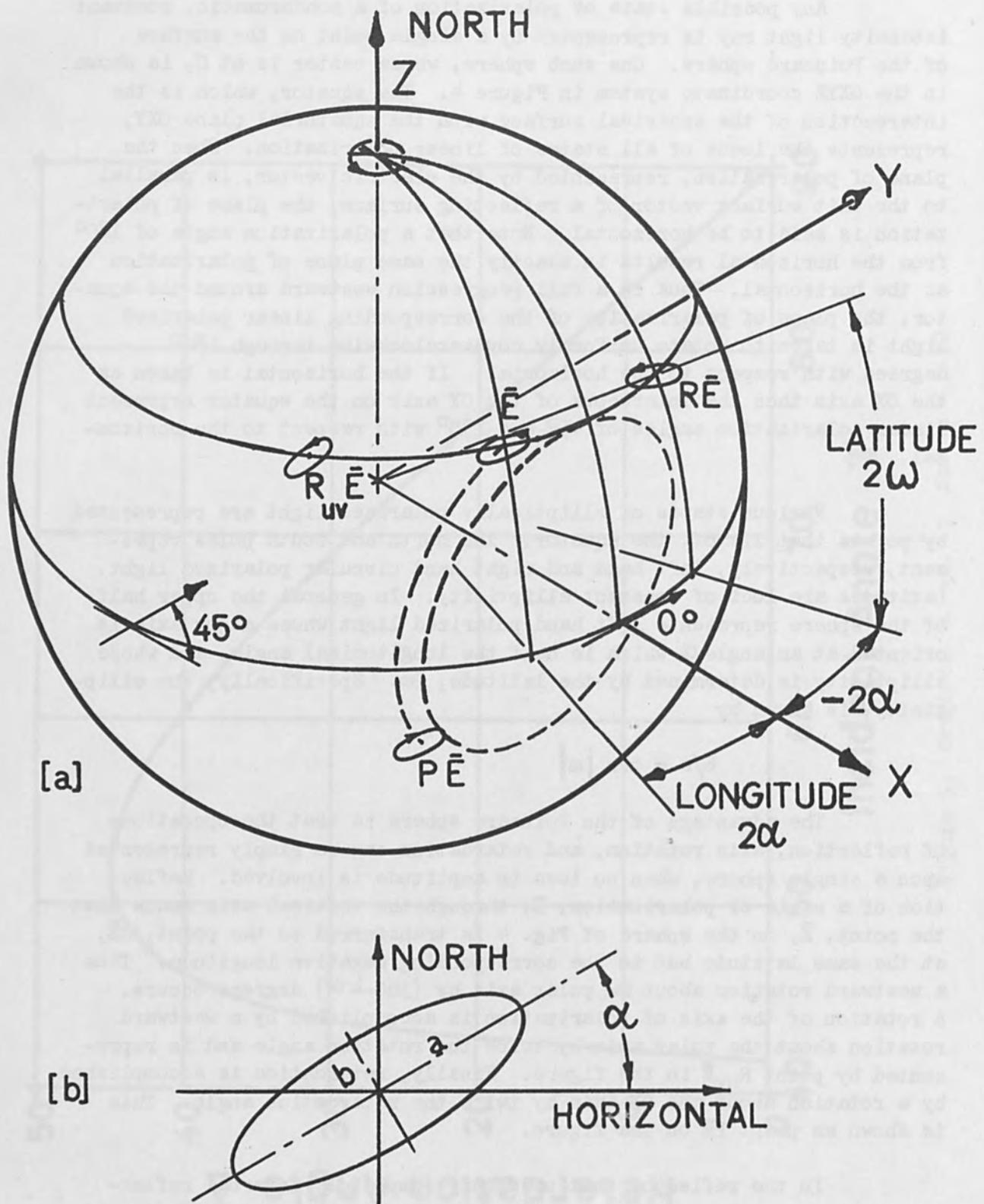


Figure 4. [a] Illustrating Operations on the Poincaré Sphere.
[b] Ellipse Geometry and Polarization.

all of these operations in the internal reflector correspond to rigid body rotations of an initial point on a single Poincaré sphere. If we wish to find the state of exit polarization for arbitrary entrance polarization states the whole Poincaré sphere may be put through a specific set of rigid body rotations, the final position of the sphere then being measured relative to its original position. This implies, among other things that there is a one to one correspondence of exit polarization states to entrance states in any completely internal reflector. Some conclusions immediately follow from the representation of the operations in the internal reflector as rigid body rotations.

A. There are always two linear polarized entrance rays, each of which will produce a unique linear polarized exit ray after any number of perfect internal reflections. The two entrance rays are polarized at right angles to one another and the exit rays will also be polarized at right angles to one another. The entrance and exit linear polarizations are not usually the same.

The above conclusion is illustrated in Fig. 5 and follows from the fact that the locus of all linear polarized entrance rays is the initial equator, a great circle. Although the rigid body rotations of the equatorial great circle change the orientation of the great circle, it must still intersect its initial position in at least two points. The intersection points represent states where initially plane polarized light yields plane polarized light in the exit ray. The intersection points are always separated by a diameter of the sphere and hence lie 180° apart, and representing rays that are orthogonally polarized.

The very important practical implication of this conclusion is that for any internal reflector, unique Brewster surfaces may be designed, one for the linear polarized entrance ray and the second for the linear polarized exit ray, thus permitting use of corner cubes as an extremely low loss reflector for linear polarized light.

B. For a given internal reflector, any desired exit state of polarization can be obtained. However, there is only one entrance state that will yield this exit state. This follows directly from the one to one correspondence of the exit polarization states to the entrance states.

IV. GRAPHICAL TECHNIQUE. CORNER CUBE

The use of the Poincaré sphere will be applied to the case of a corner cube in this section. We will be particularly interested in the change of polarization produced in linear polarized light and to the loss of energy by subsequent reflection of part of the exit beam from the entrance face when a single entrance face is used on the cube.

onto the north pole. In the equatorial view, any retardation or polarization operation, P, appears as a rotation on a circle centered at the horizontal polarization position. By inspection of Fig. 4 or Fig. 6 it can be seen that in the polar view of the operation of retardation P, is represented by motion on a straight line parallel to the OY axis, while on the equatorial view both the reflector and rotation operations R and R_{UV} are represented by motion on a straight line parallel to the equator.

Since both the equatorial and polar views are views of the same sphere, these two views can be superimposed on the same polar graph paper and the required operations carried on in sequence. As in any orthographic projection procedure, exchange between the two views is accomplished by dropping a straight line from one view and finding the intersection with the corresponding straight or circular line of the companion view. This is seen more clearly in the illustration remembering that the polar view represents a 90° rotation of the equatorial view about the OY axis.

Fig. 7 illustrates the operation of the graphical technique for a particular problem. (See Table 1)

It is worthwhile to note that if two points representing the final exit polarizations are known for two given plane polarized entrance polarizations, then the complete set of final polarizations states can be found for all plane polarized initial states. This simply requires passing a great circle through the two points on the Poincaré sphere, preserving the original spacing of polarization points on the great circle. On the aforementioned projections this requires passing an ellipse through the two points with the major axis of the ellipse represented by a diameter of the projected circle. Incidentally, all intermediate positions of the great circle must obviously also appear as ellipses on the projection plots.

V. CONCLUSIONS

We have presented analytical and graphical methods for dealing with the change of polarization produced by complete internal reflectors. Although we have dealt more extensively with the graphical techniques, the analytical results expressed by the matrix formula of Eq. 2.7 can be used for digital computation directly.

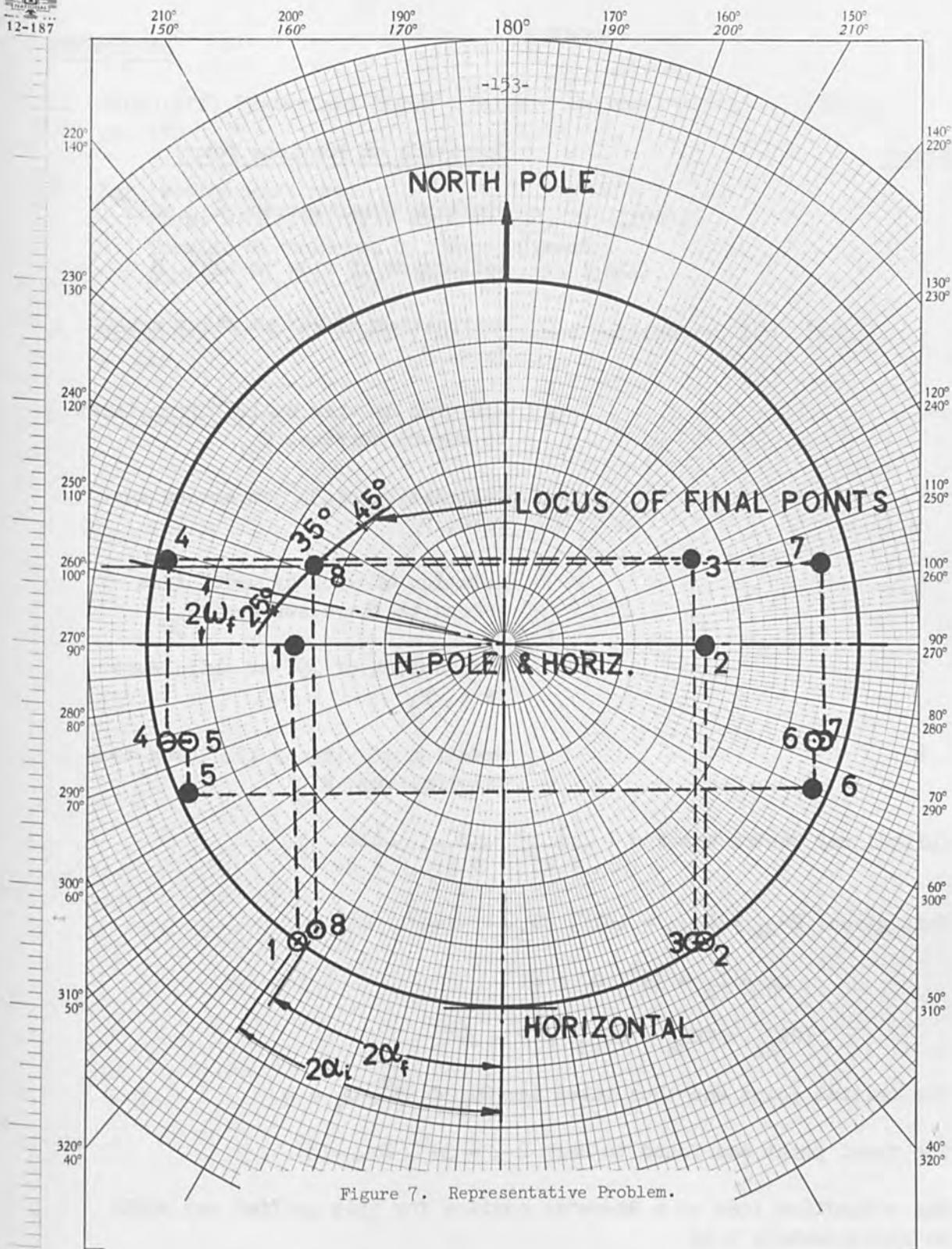


Figure 7. Representative Problem.

TABLE 1

<u>Number</u>	<u>Operation on Previous Point</u>
1	Starting polarization, $2_i = 35^\circ$
2	Reflection of view to -35° , R
3	Retardation of 24° at first surface, P_1
4	Rotation of view by 107° from first to second surface, R_{21}
5	Retardation of 39° at second surface, P_2
6	Rotation of view by -146° from second to third surface, R_{32}
7	Retardation of 39° at third surface, P_3
8	Rotation of view by 107° back to original ray, R_{13}

Initial ray vector was $k = \left(\frac{1}{4.25}, \frac{1}{4.25}, \frac{1.5}{4.25} \right)$

This gives: $2\theta_{21} = 2\theta_{13} = 107^\circ$, $2\theta_{32} = -146^\circ$

$$\Delta_2 = \Delta_3 = 39^\circ, \Delta_1 = 24^\circ$$

$$(2\theta_{21} + 2\theta_{32} + 2\theta_{13}) / 2 = 34^\circ$$

The initial point was thus taken as: $2\omega_i = 35^\circ$, $\omega_i = 0$

The final point was found to be: $2\omega_f = 34^\circ$, $2\omega_f = 13^\circ$

The reflection loss at a Brewster surface for this initial ray would be approximately 0.2%.

REFERENCES

1. Shurcliff, "Polarized Light", Harvard University Press, (1962), pp. 120.
2. For related work see:
Peck, E. R., J. Opt. Soc. Am. 52, 253 (1963).
Toraldo di Francia, G., Nuovo Cimento.
Bobroff, D. L., Appl. Opt. 3, 1485 (1964).
3. Ditchburn, R. W.; "Light", Interscience Publishers, Inc. (1963); p. 542.
4. Ditchburn, R. W.; Op. Cit; pp. 481-482.

LASER MACHINING STUDY

by

Warren V. Trammell
GM Defense Research Laboratories
Santa Barbara, California

ABSTRACT

This report presents the results of an experimental program to define the current status of pulsed lasers as machine tools for material removal. The laser machining systems used for the tests were sufficiently flexible so that an extensive study could be made of the effects of energy, power, pulse duration, beam divergence and focusing of the beam on machining capability. The machining tests were limited to the following materials: Rene-41, Inconel 713, L-605, Titanium 6-4, Aluminum Oxide, Hastelloy X, C-2 Carbide, Stellite 19, and D-6 Steel.

The smallest diameter hole obtained was .0035 inch. The largest area vaporized and ejected was .104 inch in diameter. The deepest hole drilled with a single pulse was .183 inch and with multiple pulses was .402 inch. The depth of multiple pulse holes was limited by metal resolidification to a depth-diameter ratio of 5-10 depending on the work material. Grooves up to .080 inch wide and .170 inch deep were formed by overlapping single pulses. All drilled holes had resolidified surface layers approximately .001 inch thick in which microstructural changes had occurred.

This project was sponsored by Allison Division, General Motors Corporation.

I. INTRODUCTION

The design engineer in his search for better materials for use in aerospace systems creates almost insurmountable problems for the tool designer who must provide tools to work these materials. As new materials are discovered for tools, the design engineer immediately specifies them for use in his product. A system capable of machining any and all materials, past, present and future, would be immensely helpful. This study was made to determine if the current state-of-the-art laser might be such a machine tool. The basic objectives of this program were to accomplish the following:

1. Investigate various techniques of hole drilling using a laser (power densities, energy levels, pulse length, spot size, etc.)

2. Analyze drilled holes for accuracy of shape, diameter, depth, location, and repeatability.
3. Investigate results in various materials.

In order to accomplish these objectives the results of making the following changes in machining techniques were examined:

1. Vary laser cavity length.
2. Vary focal length of condensing lenses.
3. Vary location of focal plane with respect to surface of work piece.
4. Vary energy while maintaining constant power density.
5. Vary power density while maintaining constant energy.
6. Drill in air, nitrogen, and helium atmospheres and in a vacuum.
7. Drill at various work piece temperatures.
8. Drill using single and multiple pulses.

It became evident quite early in the program that preparation of the samples for evaluation after drilling would be difficult. In the first place, the materials were very hard. Moreover, in the case of small-diameter holes, any attempt to cut the sample open for visual observation would result in loss of the entire hole. To overcome this difficulty, a scheme was devised which proved very satisfactory. The procedure used is to make the work piece out of two butting pieces whose adjoining surfaces have been lapped. The work piece therefore contains a tight seam. The laser is positioned so that its beam will be bisected by the seam (Figure 1). Due to the azimuthal symmetry of the heat diffusion, the split in the sample has no effect on the drilling capability. It now becomes possible to open the seam after each shot to see how the drilling is progressing and then to restore it for the next shot. Furthermore, after the final shot, the piece may be simply separated and is immediately available for inspection.

II. LASER MACHINING SYSTEM

Three basic laser machining systems were used in conducting this program. All three systems consisted of an air-cooled laser head, a charging supply, an energy storage bank, a micropositioning unit to hold and



Figure 1. Titanium Sample, Showing In-the-Seam Drilling Technique.

align the work piece, and the required instrumentation to permit monitoring the flash lamp voltage, current and output during discharge and the ruby output energy and pulse shape. Figure No. 2 is a photograph of the GM DRL Mark IV laser system. This system is capable of producing an output of 164 joules in a 3-millisecond pulse. The energy output may be varied from 164 joules to as low as 2 joules. The range of laser pulse lengths obtainable is from one-half millisecond to six milliseconds. The Mark IV laser uses a ruby rod that is $5/8$ " in diameter and 7 inches long. The rod may be pumped simultaneously or consecutively by four linear flash lamps. The reflecting cavity is formed by four symmetrically intersecting circular cylinders finished by silver-plating and buffing. The regeneration cavity has a rear mirror mounted on a slide, permitting adjustment of the length of the cavity from $11-1/2$ " to $33-1/2$ ". The output beam divergence is varied by varying this cavity length.

The 50-joule system uses a $3/8$ " diameter by $6-1/2$ " ruby rod pumped by one linear flash lamp. The one-joule system uses a $1/4$ " diameter by 3" ruby rod pumped by one linear flash lamp. The majority of the machining was accomplished with the large laser. The one-joule system was used to define the smallest diameter holes. The 50-joule system was used to fill in the range between the other two.

III. SMALL-DIAMETER HOLES

The laser machining conducted to define the smallest diameter, d , that could be drilled was accomplished with the one-joule laser. The focal length of the condensing lens used ranged from one inch to $1/4$ inch. Shorter focal length lenses were not practical due to spherical aberration effects and the inability to interpose glass shields to protect the lenses from the spattered molten metal. The laser cavity length was varied from $11-1/4$ " to $23-1/4$ " in order to vary the beam divergence.

The smallest d was obtained, as anticipated, by using a $1/4$ " focal length lens focused on the surface of the material, with the longest possible laser cavity length. The resulting hole ($d = .0035$ ") is shown in Figure No. 3 (Sample No. 202). The apparent plugging of the hole visible in the photograph is due to two facts: the hole is at a slight angle to the lapped joint, and the hole is necked down at that point. The hole when viewed along its axis is clear. The holes in Samples Nos. 23 and 37 were obtained with short cavity lengths. The ability to drill small diameter holes at an angle is demonstrated by Sample No. 27.

If the work-stock thickness is increased, it finally becomes impossible to pierce the sample. For a given diameter there is a depth at which the holes start to plug. The vaporized material is not ejected completely from the hole; resolidification occurs before it can get completely free of the sample. This results in the formation of a cap over the entrance to the hole or of a plug somewhere along the length of the

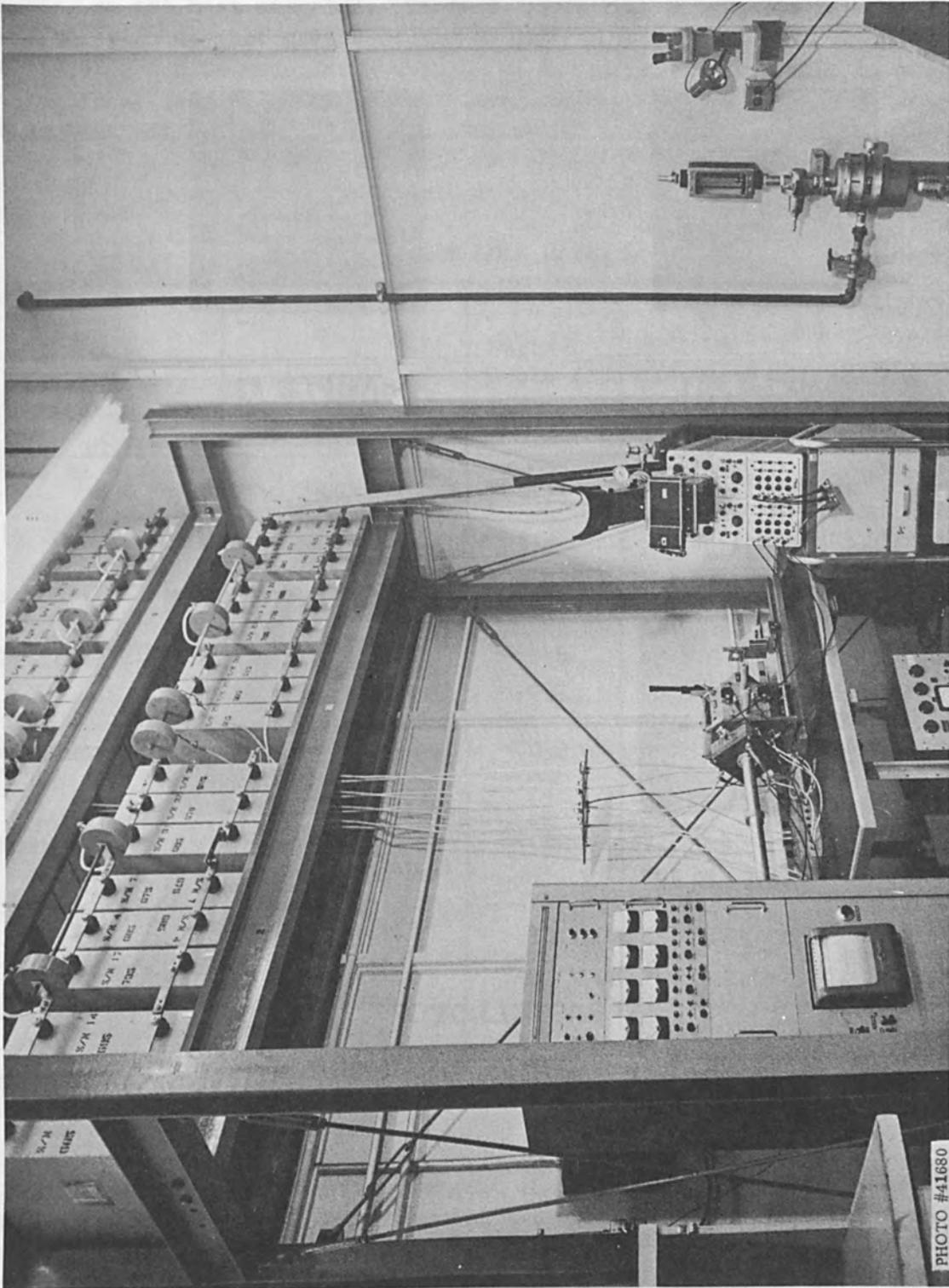
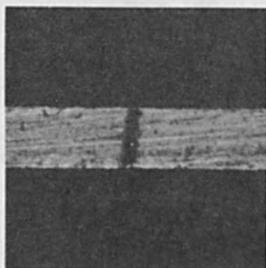
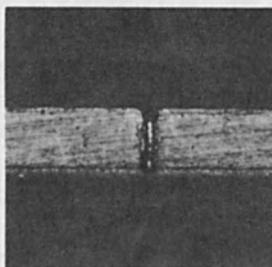


Figure 2. 150-joule Laser System.

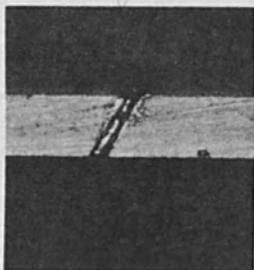
PHOTO #41680



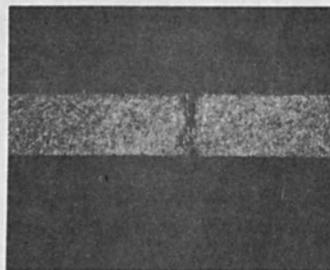
SAMPLE #37



SAMPLE #23



SAMPLE #27



SAMPLE #202

HASTELLOY X
MAGNIFICATION 10X

Figure 3. Small-Diameter Holes Through 0.031-Inch Material.

hole. This phenomenon limits the ratio ℓ/d . At the test energy level of one joule, values of the maximum ℓ/d ranged from about 5 to 10 depending on the test material. Table 1 summarizes the values found. Figure No. 4 shows photographs of the samples listed in this table. The hole diameter can be made large enough to keep ℓ/d within limits by using a longer focal length lens and/or by locating the focal plane of the condensing lens off the surface of the sample.

Table 1

SMALL-DIAMETER HOLE DATA

Material	Diameter (d) - Inches -	Maximum Depth (ℓ) Inches	Maximum Slenderness Ratio (ℓ/d)	Sample Number
Rene 41	.009	.045	5	201
Inconel 713	.007	.038	5.4	6a
L-605	.007	.039	5.6	173
Titanium 6-4	.008	.048	6.0	136d
	.009	.061	6.8	136b
Aluminum Oxide	.008	.032	4.0	110b
	.011	.042	3.8	110a
Hastelloy X	.009	.082	9.1	10
Carbide C-2	.012	.026	2.2	129
	.011	.065	5.9	118
Carbon	.008	.046	5.8	121

IV. LARGE-DIAMETER HOLES

In order to define the largest diameter holes available, the output beam of the Mark IV laser system was spread over as large an area as possible while maintaining sufficient power density to vaporize the various materials. Two techniques were used to do this. One technique consisted of using lenses of long focal lengths and maintaining the focal plane of each lens on the surface of the material. The other method used a single-focal-length lens and varied the location of the focal plane with respect to the surface of the material.



SAMPLE #201
RENE 41



SAMPLE #6 a
INCONEL 713



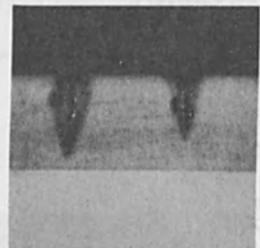
SAMPLE #173
L-605



SAMPLE # 136d
TITANIUM 6-4



SAMPLE # 136b
TITANIUM 6-4



SAMPLE #110
ALUMINUM OXIDE



SAMPLE #10
HASTELLOY X



SAMPLE #118
CARBIDE C-2



SAMPLE #121
CARBON

MAGNIFICATION 10X

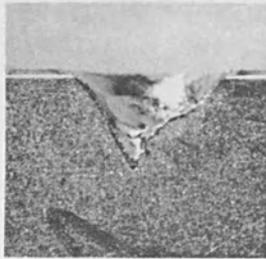
Figure 4. Small-Diameter Holes in Thick Material.

The maximum hole diameter obtained with either system was .108 inch. It was drilled in a sample of L-605 steel (No. 90, Figure 5) using a three-inch focal-length lens focused on the surface. Lenses with focal lengths over three inches resulted in molten material rather than vaporization for most of the materials. The largest-diameter hole obtained with a fixed-focal-length lens and variable-focal-plane location was .100 inch. It was drilled in a sample of Hastelloy X (No. 11, Figure 5). The maximum-diameter holes obtained in other materials are shown in Table 2. The samples of D-6 steel and Carbide (No. 5 and No. 107) were drilled with a 1-1/4" focal-length lens focused on the surface. Samples Nos. 5 and 107 do not necessarily represent the largest holes obtainable in the two materials, as a change in the program emphasis did not permit completion of the tests on those materials.

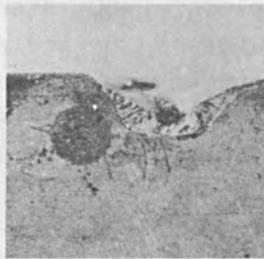
Table 2

LARGE-DIAMETER HOLE DATA

Material	Energy Joules	Diameter Inches	Depth Inches	Sample Number
Rene 41	60	.088	.049	98
Inconel 713	65	.096	.033	63
L-605	60	.108	.035	90
Titanium 6-4	65	.104	.044	41
Aluminum Oxide	65	.104	.050	124
Hastelloy X	142	.100	.048	11j
Stellite 19	60	.102	.028	84
D-6	114	.041	.085	5
Carbide C-2	125	.040	.075	107



SAMPLE #98
RENE 41



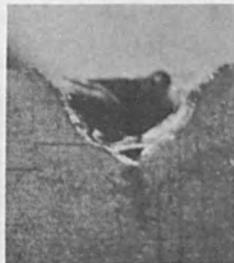
SAMPLE #63
INCONEL 713



SAMPLE #90
L-605



SAMPLE #124
ALUMINUM OXIDE



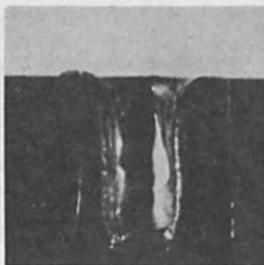
SAMPLE #11j
HASTELLOY X.



SAMPLE #84
STELLITE 19



SAMPLE #41
TITANIUM 6-4



SAMPLE #5
D-6



SAMPLE #107
CARBIDE C-2

MAGNIFICATION 10X

Figure 5. Large-Diameter Holes.

V. DEPTH CAPABILITIES

An investigation was made to define the maximum depth to which a hole about .050 inch in diameter could be drilled. The Mark IV laser system was used for this test. Lens focal length, beam divergence, pulse energy, and pulse length were varied.

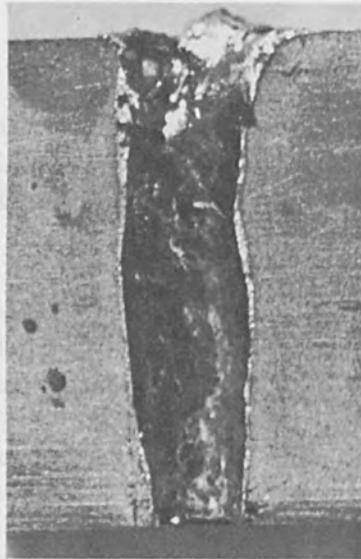
The deepest hole obtained with a single pulse was drilled in Sample No. 11a, Figure 6. This hole is .045 inch in diameter and .183 inch deep. It was drilled with a single pulse of 142 joules.

Drilling was also done by the use of multiple pulses in order to obtain still deeper holes of about .050 inch diameter. For the first pulse, the beam was focused on the surface of the work. For each succeeding pulse, the focus was advanced to the bottom of the hole formed by the previous shot. This technique did permit increased penetration up to some limiting value; after a certain number of pulses, however, the hole depth would not increase. Material appeared to melt but then resolidify along the sides of the hole as it was being ejected. The next pulse, fired into a constricted hole, melted but did not eject the constriction. This continued pulse after pulse without significant deepening of the hole. The following typical capability was found for multiple-pulse drilling. The results are reported starting with the use of low-energy pulses and going on up.

1. A .173-inch-deep hole was machined in a 1/4 inch piece of Titanium 6-4 (No. 59) using six pulses of 8 joules each. Additional pulses did not materially increase the hole depth. The slenderness ratio achieved was 4-1/2.
2. A hole was machined in a 1/4 inch piece of Stellite 19 (No. 85) using three pulses of 13 joules each. By way of comparison, a single pulse of 113 joules was fired at the same sample. The size and shape of the two holes are nearly identical (Figure 18).
3. When the pulse energy exceeded 50 joules, .250-inch-thick samples of all materials were easily pierced.
4. When the energy was increased to about 95 joules per pulse, in an attempt to pierce .500-inch stock, the results were mostly unsuccessful. Thus, for Titanium 6-4 (Sample No. 46) no significant deepening occurred after the seventh pulse of a series of twenty. For Inconel 713 (Sample No. 67), nine shots were used with the same limiting value of about 7 for depth-to-diameter ratio.



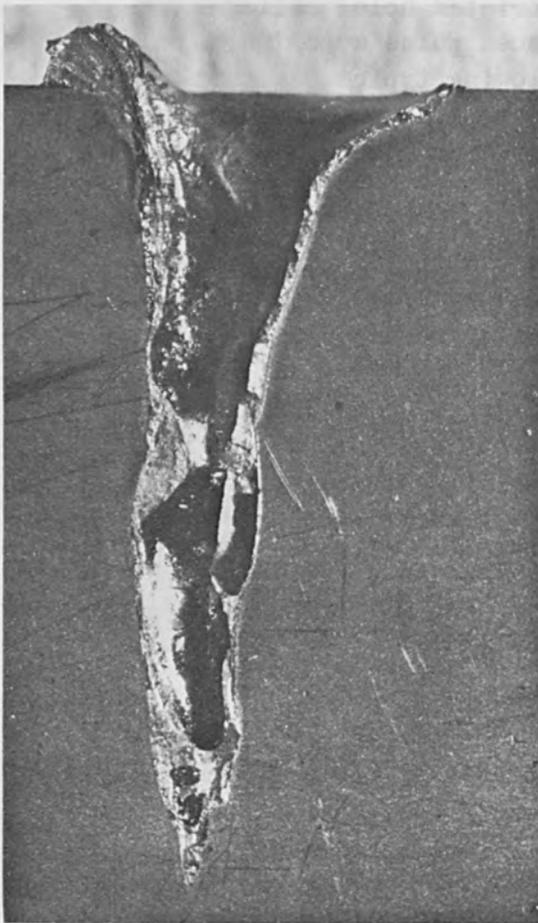
SAMPLE #126 b
ALUMINUM OXIDE



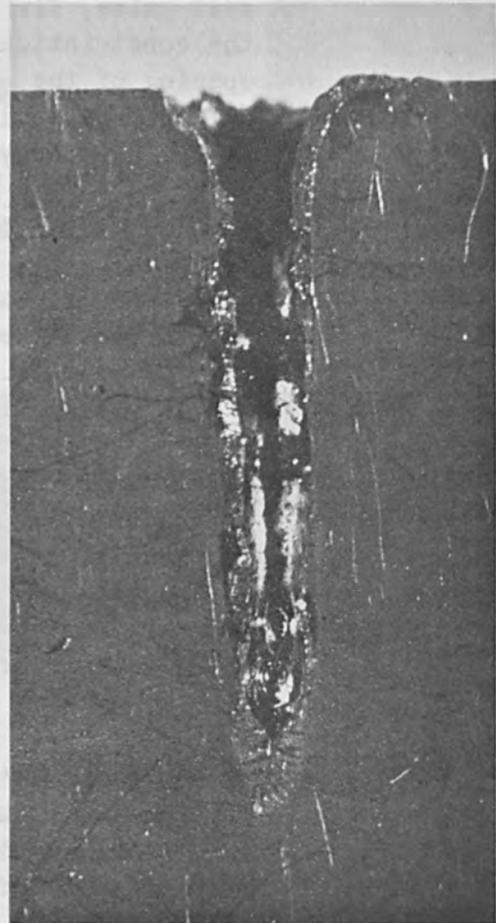
SAMPLE #99
RENE 41



SAMPLE #11 a
HASTELLOY X



SAMPLE #46
TITANIUM 6-4



SAMPLE #67
INCONEL 713

MAGNIFICATION 10X

Figure 6. Maximum Depth Holes.

Typical results of the single- and multiple-pulse techniques are shown in Table 3 and in Figure 6. It is interesting to note that multiple-pulse drilling for depth does not increase the initial hole diameter.

Table 3

MAXIMUM DEPTH CAPABILITY DATA

Material	Diameter Inches	Depth Inches	Sample Number
Rene 41	.086	.158	101
	.035	.172	191a
	.089	.250 (3 Pulses)	99
Inconel 713	.047	.156	71e
	.042	.187 (2 pulses)	72e
	.051	.377 (9 pulses)	67
L-605	.023	.142	91d
	.033	.133	163a
	.056	.130	167
	.039	.143	171
	.054	.219 (2 pulses)	165d
Titanium 6-4	.036	.169	58d
	.025	.181	58a
	.050	.249 (2 pulses)	57a
	.059	.402 (20 pulses)	46
Aluminum Oxide	.061	.250 (7 pulses)	126b
Hastelloy X	.045	.183	11a
	.075	.100	34
	.050	.250 (3 pulses)	36
Stellite 19	.053	.098	85b
	.037	.116	87a
D-6	.041	.085	5
	.036	.129	95d

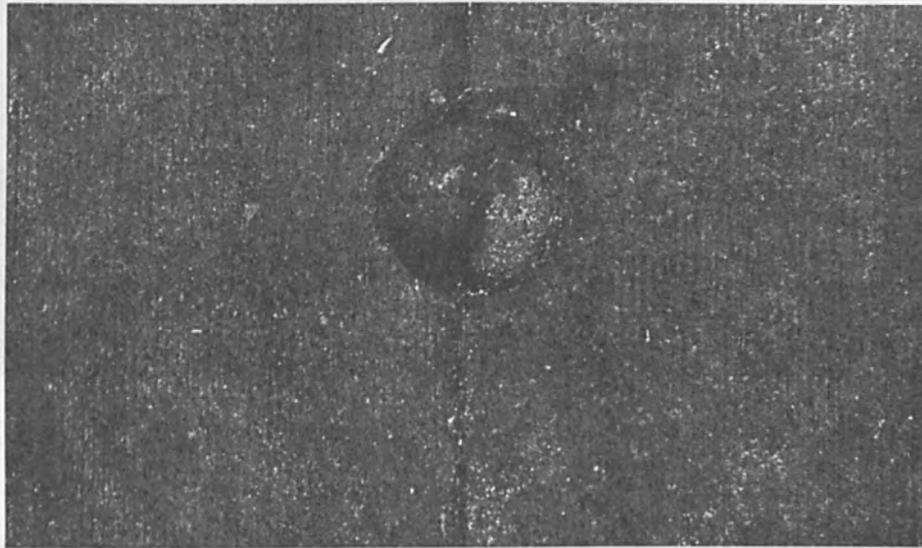
VI. SPOT REPETITION AND UNIFORMITY

The following procedure was used to determine the variation in location of the drilled hole from firing to firing. First, the laser head was securely clamped to an optical bench rail. The fixture holding the work, which was positioned and securely clamped to the rail, was designed with a 90° corner stop to assure positive location of the test sample. A rectangular sample was drilled once and then rotated 90°. With four such operations, a hole was obtained near each corner. The distances from the edges of the sample to the center of the hole were then measured for all four holes. The variation in hole diameters (.003 inch) and in hole shape set a limit on the accuracy of the measurements. However, the variation in hole location was definitely less than the variation in hole diameters. For this test, drilling was not "in the seam".

The laser must be bore-sighted to assure that the center of the machined hole will be at the proper location. This was done after every re-assembly of the laser head following replacement of parts. The bore-sighting operation was accomplished as follows: A test sample was installed in the holding fixture and a single low-energy pulse was fired to drill a small hole in the sample. The optical axis of the alignment telescope was then adjusted to bring the cross-hairs onto the center of the machined hole. Locking the scope in this position completed the bore-sighting operation.

The precision with which a machined hole may be centered on a sample depends primarily on the skill of the operator. All three technicians employed on this program found it quite easy to locate the drilled hole so that it would be bisected when drilling in the seam. A photograph of a typical sample is shown in Figure 7.

Five different techniques were used to drill four holes each in samples of Rene-41, Inconel 713 and L-605. In addition, one sample of Hastelloy X, Titanium 6-4, and Stellite 19 were drilled with four holes per sample. The test conditions used and the sample number are shown in Table 4, and the maximum variation in diameter and depth for each sample is shown in Table 5. The Hastelloy X and Stellite 19 samples were machined using a short laser cavity and a short-focal-length lens. The Titanium 6-4 sample was machined with the one-joule laser system. In general, as can be seen in Figure 8, the uniformity of the four holes in any one sample is very good.



SAMPLE #93
L-605

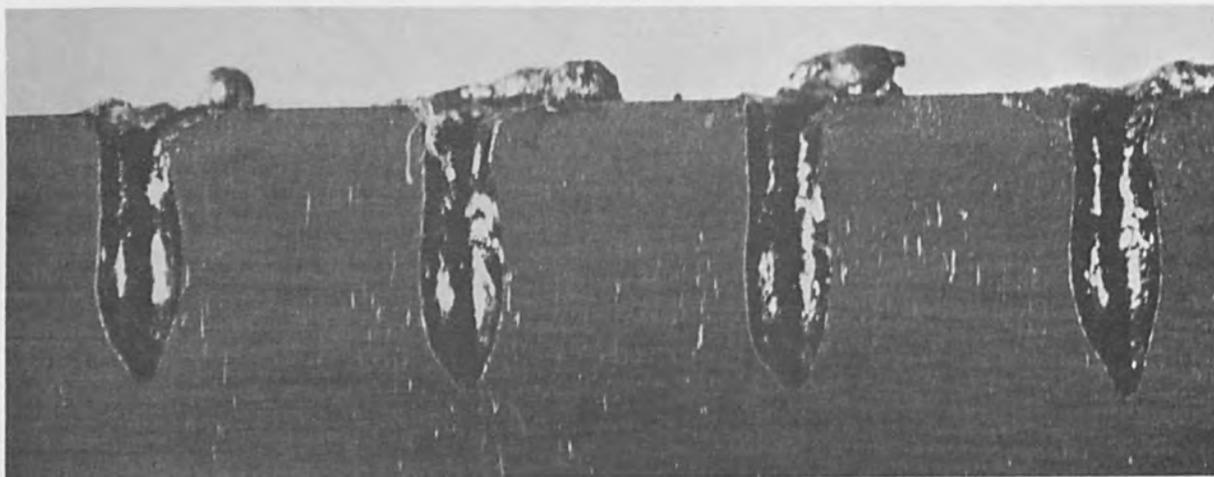
MAGNIFICATION 10X

Figure 7. Typical Sample as Machined.

Table 4

CONDITIONS FOR UNIFORMITY TEST

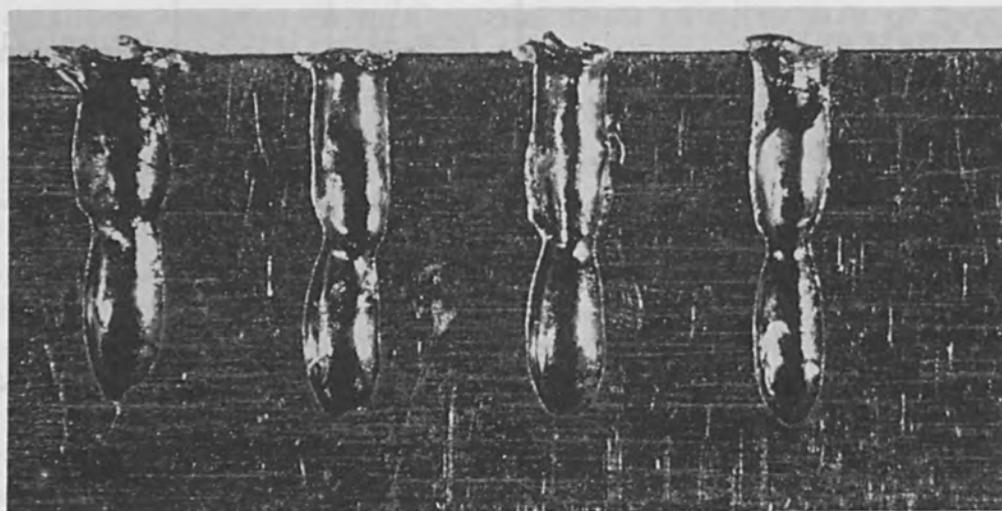
Laser Configuration	Rene 41	L-605	Inconel 713
Short cavity, long-focal-length lens, 75 joules, 3-millisecond pulse	193	92	74
Long cavity, short-focal-length lens, 64 joules, 3-millisecond pulse, two pulses per hole	194	165	72
Long cavity, short-focal-length lens, 50 joules, 3 millisecond pulse	100	91	75
Long cavity, short-focal-length lens, 45 joules, 4.1-millisecond pulse	191	163	77
Long cavity, short-focal-length lens, 8 joules, 0.65-millisecond pulse	192	164	175



SAMPLE #100 RENE 41 LONG CAVITY, SHORT FOCAL LENGTH LENS
52 JOULE PULSE



SAMPLE #193 RENE 41 SHORT CAVITY, LONG FOCAL LENGTH LENS
75 JOULE PULSE



SAMPLE #194 RENE 41 LONG CAVITY, SHORT FOCAL LENGTH LENS
TWO 64 JOULE PULSES
MAGNIFICATION 10X

Table 5

DIAMETER AND DEPTH VARIATIONS

Sample No.	Material	Maximum Variation		Remarks
		Diameter	Depth	
100	Rene 41	.002	.015	
191	Rene 41	.003	.018	
192	Rene 41	.007	.009	
193	Rene 41	.003	.006	
194	Rene 41	.012	.017	Two pulses.
91	L-605	.010	.028	Lamp failed to fire.
92	L-605	.006	.028	
163	L-605	.004	.017	
164	L-605	.004	.011	
165	L-605	.012	.040	Two pulses.
72	Inconel 713	.005	.029	Two pulses.
74	Inconel 713	.006	.009	
75	Inconel 713	.005	.039	Lamp failed to fire.
77	Inconel 713	.001	.007	
175	Inconel 713	.005	.006	
136	Titanium 6-4	.002	.013	Drilled with small laser.
4105	Hastelloy X	.009	.016	
87	Stellite 19	.002	.011	

VII. TREPANNING TESTS

All of the trepanning tests were conducted using the Mark IV laser system. A short laser cavity and a short-focal-length lens were used. Grooves were successfully machined in all materials tried, which included Inconel 713, Titanium 6-4, Carbide C-2, and Aluminum Oxide. The distance the laser head was translated between firings was varied from 1/4 of a hole diameter to a full hole diameter. The best grooves were obtained when the translation was one half of the initial hole diameter. Grooves approximately .050 inch wide, .180 inch deep and .250 inch long were machined in three of the four materials. The deepest groove obtained in Carbide was .086 inch. Photographs of two of the samples are shown in Figure 9.

VIII. HEAT ZONE TESTS

The effects of pre-heat, of surrounding atmosphere, and of heating rates were examined in an effort to reduce the heat-affected zone by drilling the samples as follows:

1. At near-melting temperature
2. At 1000° C and 500° C
3. In vacuum
4. In nitrogen at atmospheric pressure
5. In helium at atmospheric pressure
6. At constant power density with variable pulse length
7. With a single high-energy pulse and with multiple low-energy pulses.

The tests concentrated primarily on Rene-41, L-605 and Inconel 713. None of the variations tried gave appreciable improvement of the heat-affected zone in the vicinity of the hole.

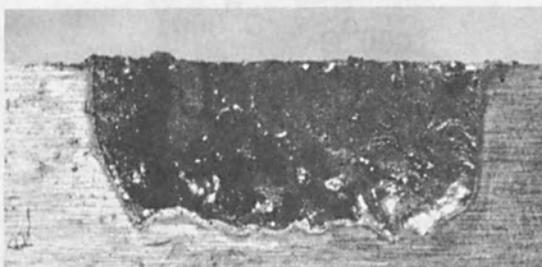
The resulting holes for techniques Nos. 1-5 above may be seen in Figures 10, 11, and 12.

IX. TECHNIQUES

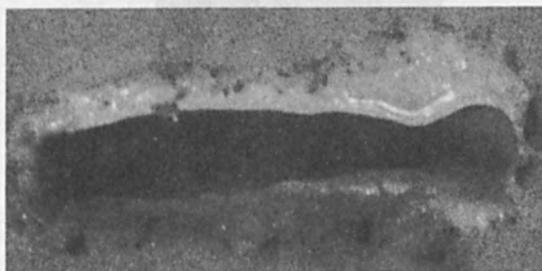
This section is devoted to a discussion of the results obtained when various laser machining techniques were applied to the same material. One of the first laser machining methods to be investigated was the effect of changing the focal length of the condensing



SAMPLE #117
CARBIDE C-2
MACHINING POSITION



SAMPLE #117
CARBIDE C-2
CROSS SECTION



SAMPLE #159
ALUMINUM OXIDE

MAGNIFICATION 10X

Figure 9. Trepanning Test Samples.



SAMPLE #101
1400°C



SAMPLE #195
1000°C



SAMPLE #196
500°C



SAMPLE #197
VACUUM



SAMPLE #198
NITROGEN
ATMOSPHERE



SAMPLE #199
HELIUM
ATMOSPHERE

RENE 41
MAGNIFICATION 10X

Figure 10. Rene 41 Heat Zone Samples.



SAMPLE #93
1400°C



SAMPLE #166
1000°C



SAMPLE #167
500°C



SAMPLE #168
VACUUM



SAMPLE #169
NITROGEN
ATMOSPHERE



SAMPLE #170
HELIUM
ATMOSPHERE

L-605

MAGNIFICATION 10X

Figure 11. L-605 Heat Zone Samples.



SAMPLE #76
1325°C



SAMPLE #176
1000°C



SAMPLE #177
500°C



SAMPLE #178
VACUUM



SAMPLE #179
NITROGEN
ATMOSPHERE



SAMPLE #180
HELIUM
ATMOSPHERE

INCONEL 713
MAGNIFICATION 10X

Figure 12. Inconel 713 Heat Zone Samples.

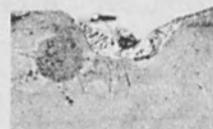
lens used to concentrate the beam onto the surface of the material. For this test, a one-inch lens was compared with a three-inch lens by drilling holes in Rene-41, L-605 and Inconel 713. The results are shown pictorially in Figure 13. The three-inch lens formed shallow craters whose depths were $1/3$ to $1/2$ the diameter of the hole. The one-inch lens formed holes with parallel sides whose depths were 2 to 3 times the diameter of the hole.

Using a 2-inch focal-length condensing lens, ten holes were drilled in a sample of Hastelloy X. The purpose of this test was to determine the effect of defocusing the beam (varying the position of the focal plane of the lens with respect to the surface of the material). The first hole was machined with the beam focused on the surface. For each succeeding firing, the surface of the sample was moved .025 inch farther away from the focal plane of the lens. The results are shown in Figure 14. As the beam was defocused, the machined hole grew larger and shallower.

The next series of firings was conducted to define the effects of maintaining a constant power density while varying the total energy. Samples of Rene-41, L-605, and Inconel 713 were used for this investigation. The laser cavity length, focal length of lens, and location of focal plane were maintained constant. A long laser pulse length was obtained by sequential firings of the four flash lamps. The resultant energy output was a 26-joule pulse 6 milliseconds in length. A short laser pulse ($1/2$ millisecond) was obtained by using only one third of the capacitor bank and firing all four lamps simultaneously. The applied voltage was varied until a total energy output of 2 joules was obtained, which resulted in the same power (4 kilowatts) and therefore the same power density as the long pulse. The results are shown in Figure 15. The longer pulse drilled a hole which was not only deeper but considerably larger in diameter as well. The ratio of the laser energy in the two pulses was 13 to 1; however, the ratio of the volume of material removed was nearly 50 to 1.

The 50-joule laser machining system was used to perform the next series of tests. Five holes were drilled in one sample of Inconel 713 using one laser pulse per hole. Each pulse contained the same energy (15 joules); however, the length of each pulse was different ($1/2$, 1, 2, 3, and 4 milliseconds). This was accomplished by varying the charged voltage to maintain a constant laser energy output. The resulting holes are shown in Figure 16 and are approximately equivalent. The results of this test and the previous one indicate that the diameter and depth of a laser machined hole is primarily dependent on total energy rather than power density or pulse length.

The next test consisted of firing more than one laser pulse



SAMPLE #98
RENE 41
3" FOCAL LENGTH LENS

SAMPLE #90
L-605
3" FOCAL LENGTH LENS

SAMPLE #63
INCONEL 713
3" FOCAL LENGTH LENS



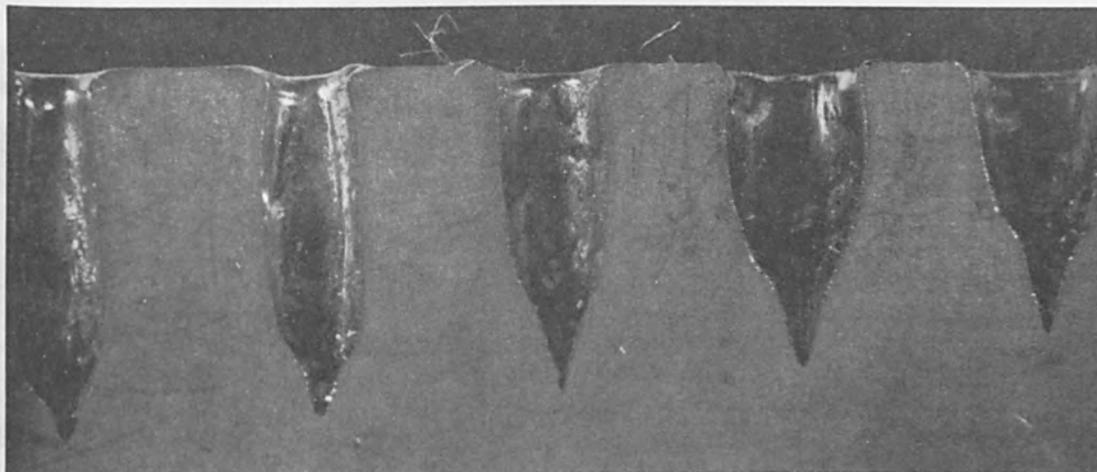
SAMPLE #100
RENE 41
1" FOCAL LENGTH LENS

SAMPLE #91
L-605
1" FOCAL LENGTH LENS

SAMPLE #71
INCONEL 713
1" FOCAL LENGTH LENS

MAGNIFICATION 10X

Figure 13. Lens Focal Length Effects.



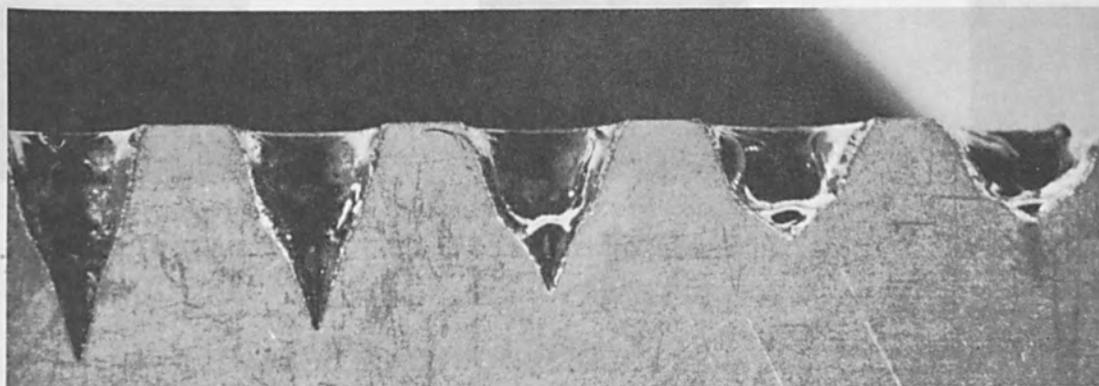
FOCAL PLANE
ON SURFACE

.025" OFF
SURFACE

.050 OFF
SURFACE

.075" OFF
SURFACE

.100" OFF
SURFACE



.125" OFF
SURFACE

.150" OFF
SURFACE

.175" OFF
SURFACE

.200" OFF
SURFACE

.250" OFF
SURFACE

SAMPLE #11 HASTELLOY X
MAGNIFICATION 10X

Figure 14. Lens Focal Plane Location Effects.



SAMPLE #200
RENE 41



SAMPLE #171
L-605



SAMPLE #181
INCONEL 713

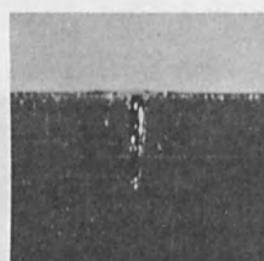
26 JOULE PULSE 6 MILLISECONDS IN LENGTH



SAMPLE #201
RENE 41



SAMPLE #172
L-605

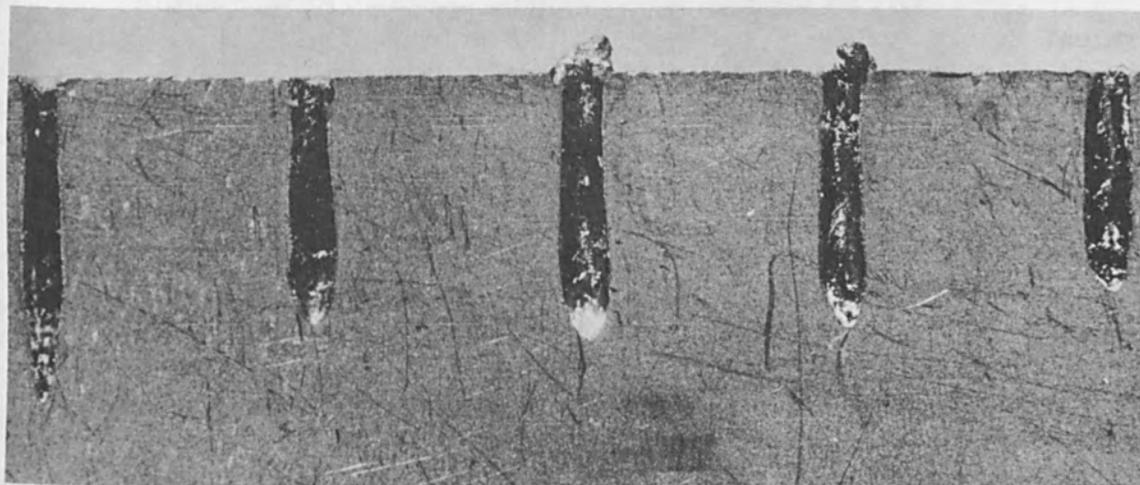


SAMPLE #182
INCONEL 713

2 JOULE PULSE 0.5 MILLISECONDS IN LENGTH

MAGNIFICATION 10X
POWER = 4 KILOWATTS

Figure 15. Constant Power Density-Variable Pulse Length.



4

3

2

1

1/2

LASER PULSE LENGTH IN MILLISECONDS

SAMPLE #61 TITANIUM 6-4

MAGNIFICATION 10X

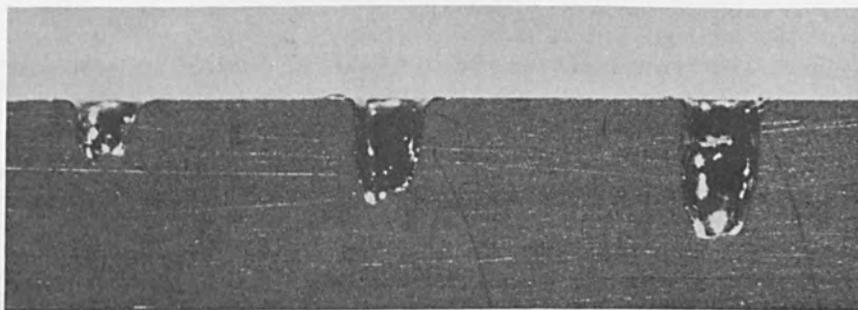
Figure 16. Constant Energy-Variable Pulse Length.

to form a single hole. Three holes were machined in a sample of Stellite 19. The first hole was formed with a single low-energy pulse. The second and third holes were formed by two and three similar laser pulses respectively. In each case, the drilling laser pulse was focused on the bottom of the hole formed by the previous pulse. The results are shown in Figure 17. The diameters of the holes are equal and the ratio of the depths are approximately 1, 2 and 3, which are equivalent to the number of pulses used to form the holes.

A sample of Titanium 6-4 (Figure 17) was drilled with 3, 4, 5, 6 and 7 pulses of similar energy using the same technique. Once again the diameters of the holes are equal and the depths are approximately a ratio corresponding to the number of shots fired. It should be noted that this statement about the depth does not apply to the hole that was drilled with seven pulses. Upon closer examination of the hole drilled with six pulses, a quantity of resolidified material is observed in the entrance region of the hole. A considerable portion of the energy in pulse No. 7 is therefore consumed in vaporizing this material, leaving very little to increase the depth of the hole. The entrance to the last hole is half plugged, which would make pulse No. 8 very ineffective. For this set of conditions, six is the limiting number of useful pulses that may be used to drill the hole. The next test, Figure 18, was a comparison of multiple pulsing of low-energy vs a single high-energy pulse. Two holes were drilled in a sample of Stellite 19. One was formed with three 13-joule pulses while the other was drilled with a single 113-joule pulse. Both were drilled using a short-focal-length lens focused on the surface (bottom of the hole for the multiple-pulsed hole) and a short laser cavity. The resulting holes are nearly equal in diameter, depth, and shape.

X. ACKNOWLEDGEMENTS

The author wishes to express his gratitude to I. Hodes for his helpful suggestion on this project and to M. Green, H. Tucker and J. Oshiro for their help in conducting the tests.



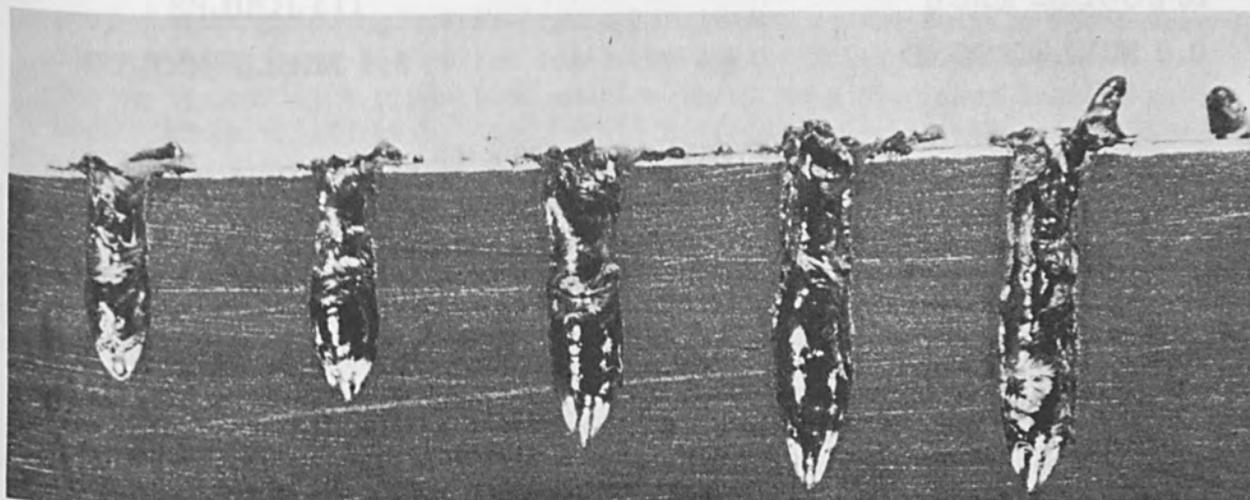
ONE PULSE

TWO PULSES

THREE PULSES

EACH LASER PULSE 13 JOULES 0.8 MILLISECONDS IN LENGTH

SAMPLE #86 STELLITE 19



3 PULSES

4

5

6

7 PULSES

EACH LASER PULSE 8 JOULES 0.65 MILLISECONDS IN LENGTH

SAMPLE #59 TITANIUM 6-4

MAGNIFICATION 10X

Figure 17. Multiple Pulse Test Samples.



3 PULSES
13 JOULES EACH
0.8 MILLISECOND



1 PULSE
113 JOULES
3.1 MILLISECONDS

SAMPLE #85 STELLITE 19

MAGNIFICATION 10X

Figure 18. Multiple Pulse Versus Single Pulse.

THE LASER
NOW A PRODUCTION TOOL

by

J. P. Epperson and R. W. Dyer
Western Electric Company, Inc.,
Princeton, New Jersey

and

J. C. Grzywa
Western Electric Company, Inc.,
Buffalo, New York

ABSTRACT

The laser, one of the great scientific discoveries of our era, has modernized one of Western Electric's most tedious mechanical processes, diamond drilling. Wire-drawing engineers at the Buffalo Plant worked closely with laser and optics engineers at the Engineering Research Center to develop a production machine which uses the laser beam to re-work diamond wire-drawing dies. This article describes the laser system which made diamond piercing possible and the unique closed-circuit television viewing system which made it practical. Operator safety is of paramount importance and was considered carefully in the design of all components of the production machine and in its installation.

I. INTRODUCTION

The lifeline of telephone communication has been and continues to be the copper wire. Although research and development efforts have been successful in devising other means for transmitting voices and pictures, copper wire remains the chief carrier of communications.

In addition to a very large demand for standard gauges of copper wire, advances in component miniturization have made it necessary to produce increasing quantities of fine and ultrafine gauge wire in sizes down in 42 AWG¹ (0.0025 inch diameter).

II. WIRE DRAWING

Copper rod, 5/16 inch in diameter, is drawn through dies with progressively smaller openings until it is reduced to the specified diameter. The wire drawing process is illustrated in Figure 1.

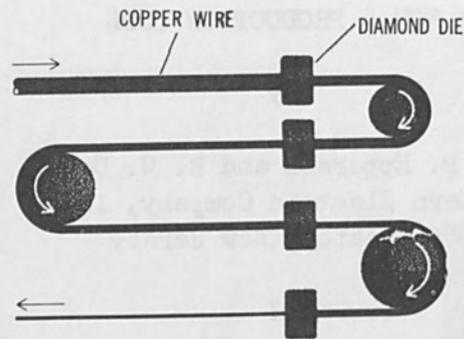


Figure 1. Wire Drawing Takes Place in Incremental Steps as Illustrated Schematically Here. In Each Step, the Wire Diameter is Reduced to the Next Smaller Gauge Size.

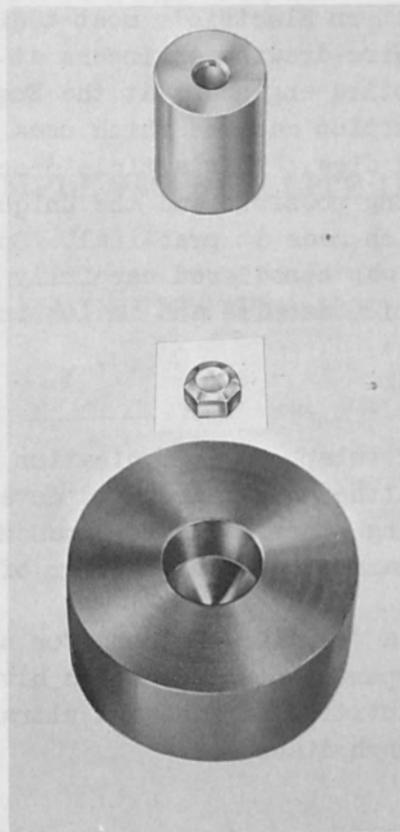
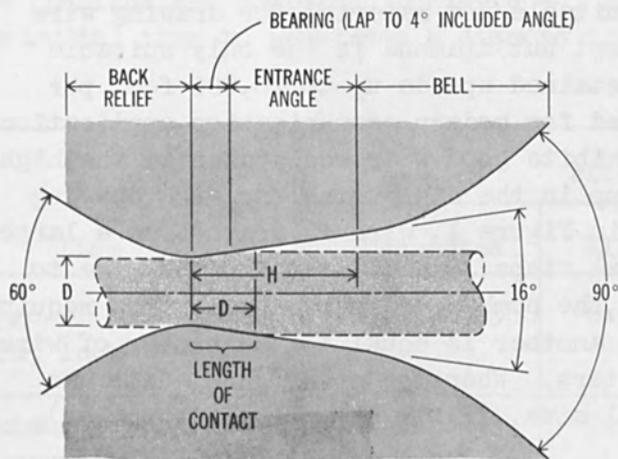


Figure 2.



(a)



(b)



(c)

Figure 3. Important Features of Diamond Die Shape as Detailed in Tool Specification 27,029 are Shown in the Sketch at Left. A Cross Section of a Diamond Die Having these Features is Pictured at Right.

Tungsten carbide is a suitable die material for drawing wire down to 15 AWG (0.057 inch diameter) but diamond is the only suitable substance for drawing wire, at sustained speeds up to 10,000 feet per minute, to the fine gauges required for modern communication applications. Diamond is suitable because it exhibits good wear resistance at the high temperatures and pressures built up in the die during the wire drawing process. As shown schematically in Figure 1, wire is drawn from a large size to a small size in incremental steps. Each die draws the wire to the next smaller gauge size; hence, the number of incremental steps required to draw wire from one diameter to another is equal to the number of wire-gauge sizes between the two diameters. When the opening in a die has worn too large for one incremental size, it can be resized (reworked) to the next larger size.

III. DIAMOND DIE FABRICATION

A. Mounting the Diamond

Making a die blank is the first step in fabricating a diamond wire drawing die. A diamond die stone of predetermined carat weight is centrally placed within a casing made of monel or stainless steel, after which a mounting plug of the same material is inserted as shown in Figure 2. The mounting plug and the casing are then heated and compressed. Under pressure, the plug metal flows around the diamond holding it securely and providing the mechanical support to protect the brittle die stone. A tapered recess is now machined into the die blank from both sides exposing the uneven faces of the diamond.

In normal preparation the uneven faces must be flattened before the hole can be started. The die blank is firmly seated on a rotating table and is centered under a rotating and reciprocating steel pin. Olive oil containing diamond particles is charged into the tapered recess, and the rotating pin reciprocates forcing diamond particles to impinge on the uneven surface of the die stone. The peaks of the uneven surface are abraded away yielding a relatively flat surface which is suitable for starting the drilling action.

B. Drilling the Diamond

Drilling the initial hole is accomplished by the same basic methods described above except that a tapered pin is substituted for the blunt pin. Recently high-voltage drilling, electrolytic drilling, and chemical drilling have been used to assist mechanical preparation² of some specialized dies, but these techniques are not suited to the bulk of die preparation. Subsequent to this "roughing" operation, the hole is lapped to achieve the desired shape and surface finish, as shown in Figure 3. The characteristics of the different methods of preparing

die openings are compared in Figure 4. As can be seen in this comparison, the initial time to penetrate a diamond is extensive.

Item	Piercing Technique		
	Mechanical	High-voltage	Electrolytic
Range of hole diameters, inch	0.0022-0.0560	0.002-0.004	0.002-0.004
Maximum depth of penetration, inch	unlimited	0.025	0.008*
Time to penetrate to maximum depth	24 hours (for 20 point stone)	20-25 minutes	2 hours
Normal application	general	initial preparation of dies for intermediate gauge sizes	Final preparation of dies for ultra fine gauge sizes
Disadvantages	Time consuming	limited penetration depth and danger of structural damage	extremely limited penetration depth

*Holes must be first mechanically pierced to within 0.008 inch of breakthrough.

Figure 4
Comparison of Techniques Used to Pierce Diamonds

However, Western Electric currently uses diamond dies already mounted and with the initial size hole already finished. Since the diamond dies soon wear, Western Electric's interest is mainly in reprocessing the die to the next larger gauge size. The resizing operation is identical to the initial preparation of the hole (i.e., roughing followed by lapping) except that typical drilling times are about 12 hours.

IV. LASER FEASIBILITY

Because of the vast effort expended annually in resizing diamond dies to produce drawn copper wire, the engineers at Buffalo are constantly seeking better and faster methods of preparing diamond dies for production. The Engineering Research Center in Princeton, New Jersey, was contacted to explore the possibility of developing the laser as a tool to pierce³ diamonds. It was envisioned that such a laser tool could be used to rough out the original hole in a diamond die or to remove material around the periphery of a worn die. To perform feasibility studies, diamond chips, mounted and unmounted die stones, and worn diamond dies were submitted to the Engineering Research Center.

A. Laboratory Apparatus

The laser at the Engineering Research Center pictured in Figure 5 was used for feasibility studies. It consists of a laser head containing a xenon flash lamp and a synthetic ruby rod, a power supply for the flash lamp, a triocular microscope for viewing the workpiece and focusing the laser beam, and an XYZ micropositioning stage to accurately position the diamond under the focused laser beam.⁴

B. Initial Experiments

The first attempts at piercing holes in diamonds were made in diamond chips mounted in either putty or solder. A 10-joule beam of 0.5 millisecond duration⁵ was focused to a 5-mil spot⁶ on a flat surface of the chip. An explosive sound is produced by the rapid vaporization of the material, and a plume of diamond vapor is generated, see Figure 6. Initial results were discouraging because most of the chips were fractured as shown in Figure 7. Previous studies had proved that most materials could be pierced efficiently with a beam pulse-duration of 0.5 to 1.0 millisecond.⁷ However, since there was no specific data on diamond piercing, shorter pulse durations were again considered. A Q-switched laser⁸ with a beam pulse duration of approximately 0.1 microsecond was used. Previous investigations in drilling with such a laser had shown a tendency to shatter glass.⁹ Experiments at the Engineering Research Center using the Q-switched laser revealed that the diamond could be pierced without structural damage. However, the penetration per pulse of the Q-switched laser was less than that obtained with the non-Q-switched laser. In addition, the lens which focuses the Q-switched laser beam is more likely to be damaged.¹⁰ Thus Q-switching was determined unsuitable for diamond piercing.

Besides the short-duration pulses of the Q-switched laser, pulse lengths of 4 and 6 milliseconds were also used. The longer pulse duration eliminated the diamond fracturing problem, but it introduced

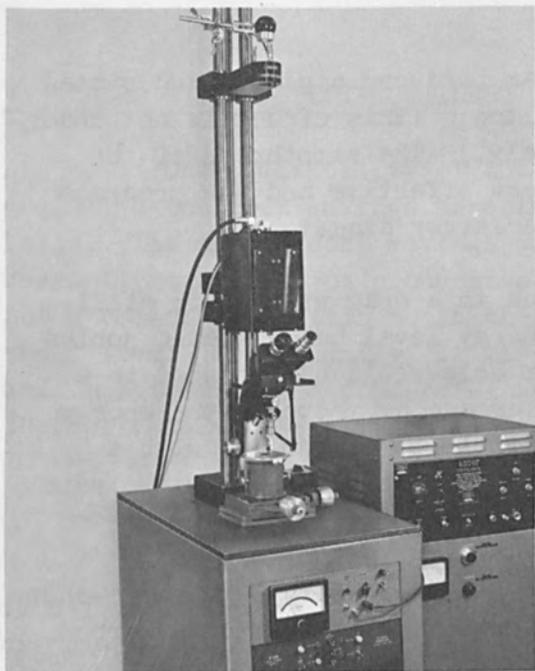


Figure 5.

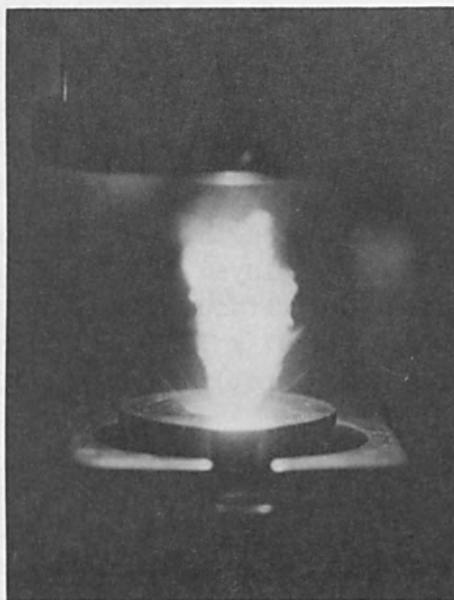


Figure 6.

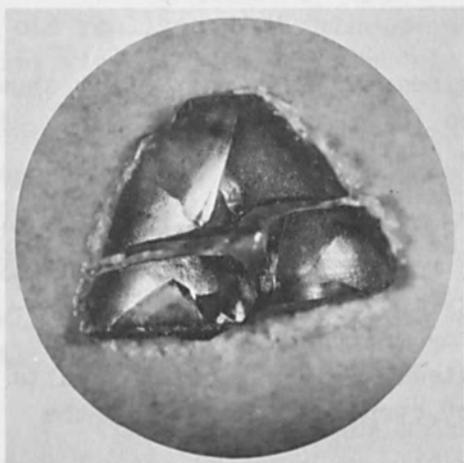


Figure 7.

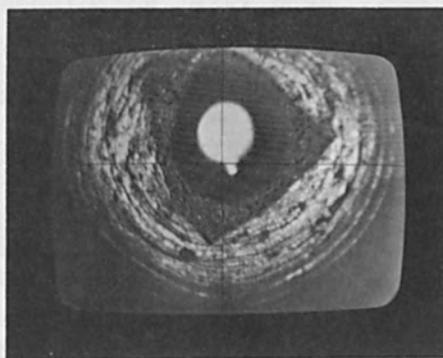


Figure 8.

another problem. The longer-duration beam produced a glazed and pitted surface, almost as if the surface had melted. (This effect is not understood, since actual melt is highly unlikely.) The smoothness of the glazed surface made the lapping action less effective and the presence of the pits made it necessary to lap to a larger diameter.

The next approach was to go back to a beam with a 0.5 milli-second pulse-length, but to reduce the energy level below the 10 joules previously employed. It was found that a hole could be produced in a diamond chip without apparent structural damage by reducing the energy density to 2.37×10^4 joules per square centimeter (3 joules with a 5-mil diameter spot) and repeatedly applying the laser energy to the same spot.

A die was pierced by this relatively low-energy, multiple-shot technique and sent to the Buffalo Plant for analysis. Residual carbon produced by laser piercing was easily removed by ultrasonic cleaning, there was no structural damage to the diamond, and the finish in the tapered hole was ideal for lapping. Thus, it was shown to be technically possible to prepare holes rapidly in diamond dies with the laser technique.

C. Enlarging Existing Holes

Making the initial hole in a diamond die is but a small part of the overall work done on the die during its useful life. Because diameter reductions in wire drawing are incremental, it is practical to enlarge and relap a die when the opening becomes worn beyond a given incremental step.

Experimental results at the Engineering Research Center showed that the cone of laser light could be used to enlarge the hole in a die by the low-energy, multiple-shot approach described above.

Production dies with 33 gauge (0.007 inch) holes were selected for most of these experiments. They were roughed-out to approximately 0.014 inch with 22 shots a 4-joules energy and 0.5 millisecond pulse-length. Figure 8 shows a diamond die being enlarged. Some of these dies were then machined-lapped and inspected. The inspection results were favorable and showed no indication of structural damage to the diamond.

As final proof of feasibility, five dies which had been prepared by the above method were placed on production machines with results which were indistinguishable from mechanically drilled dies.

V. LASER SYSTEM SPECIFICATIONS

Once feasibility of the process had been demonstrated, Buffalo evaluated the cost savings aspects of applying the laser on a production basis. The annual cost savings was determined to be sufficient to warrant procurement of production laser equipment. Since the laser is not a common production tool, the required machine was not an off-the-shelf item. Consequently, it was necessary to write specifications for and to determine the commercial availability of individual components to fabricate a laser system having the desired features.

A. Laser Equipment

Some of the features desired in the laser are listed as specifications in Figure 9. Additional features to ensure operator safety are discussed later.

B. Optical Equipment

It was originally expected that the laser supplier would provide the optical system for simultaneously focusing the laser and viewing the diamond. None of the proposed viewing systems, however, met Western Electric's requirements for operator safety, flexibility, and reliability. Operator protection is of paramount importance, particularly since the cumulative effects of exposure to laser radiation are unknown.

The difficulty of providing such protection can be appreciated when one realizes that the operator must have a clearly focused, high quality view of the very area which is exposed to the laser beam. She also must be able to accurately and quickly position the diamond with respect to the point at which the laser beam is focused. In addition, the operator must be able to determine that the laser beam is perfectly focused at the diamond surface. Therefore, a unique viewing system was designed at the Engineering Research Center, with design work so paced that the viewing system would be built by the laser manufacturer and installed on the laser drilling machine before it was delivered to the Buffalo Plant.

The optical system which was designed at the Engineering Research Center is based on a Closed Circuit television (CCTV) camera and monitor. CCTV is absolutely safe and promises to be more trouble free than other systems which must employ shutters and/or filters to block both the light reflected from the workpiece and that emitted by the fiery plume (which may contain ultraviolet or even weak X-ray radiation).

While performing the laser feasibility studies, certain optical

requirements were determined. The objective lens that focuses the laser beam also must be used to focus the viewing beam. In addition, the objective lens must be at least 1 inch above the diamond surface, otherwise vapor from the plume can cloud or pit the lens. The lens must be slightly larger in diameter than the 3/8-inch laser beam, and the lens must focus the laser beam to a spot approximately 0.005¹¹ inch in diameter.

Attribute	Specification
Energy output from optics	0-10 joules, continuously variable
Beam divergence at 10 joules output	5 milliradians
Pulse duration at 10 joules output	0.5 to 1.0 millisecond
Cooling rod and xenon flash lamp	Closed cycle deionized water
Maximum repetition rate	1 pulse/second
Modes of operation	1. remote footswitch 2. manual pushbutton 3. automatic a. 1 pulse/sec. b. 1 pulse/5 sec. c. 1 pulse/10 sec.
Flashlamp lifetime guarantee	30,000 shots at the 10 joule, 10 watt level.

Figure 9

Design specifications for laser to pierce diamond wire drawing dies

The optical system is illustrated in Figure 10. The laser beam travels downward through a dichroic filter¹² designed to pass laser light efficiently. The paths of three separate optical systems (laser beam, viewing beam and illumination beam) pass through a single objective lens.

One of the significant features of this design is that the same objective lens is used for both viewing the work and for focusing the laser. This double use allows the laser beam's focus to be confirmed by the visual clarity of the image. Dual usage also insures that the operator is viewing the area which will be affected by the laser.

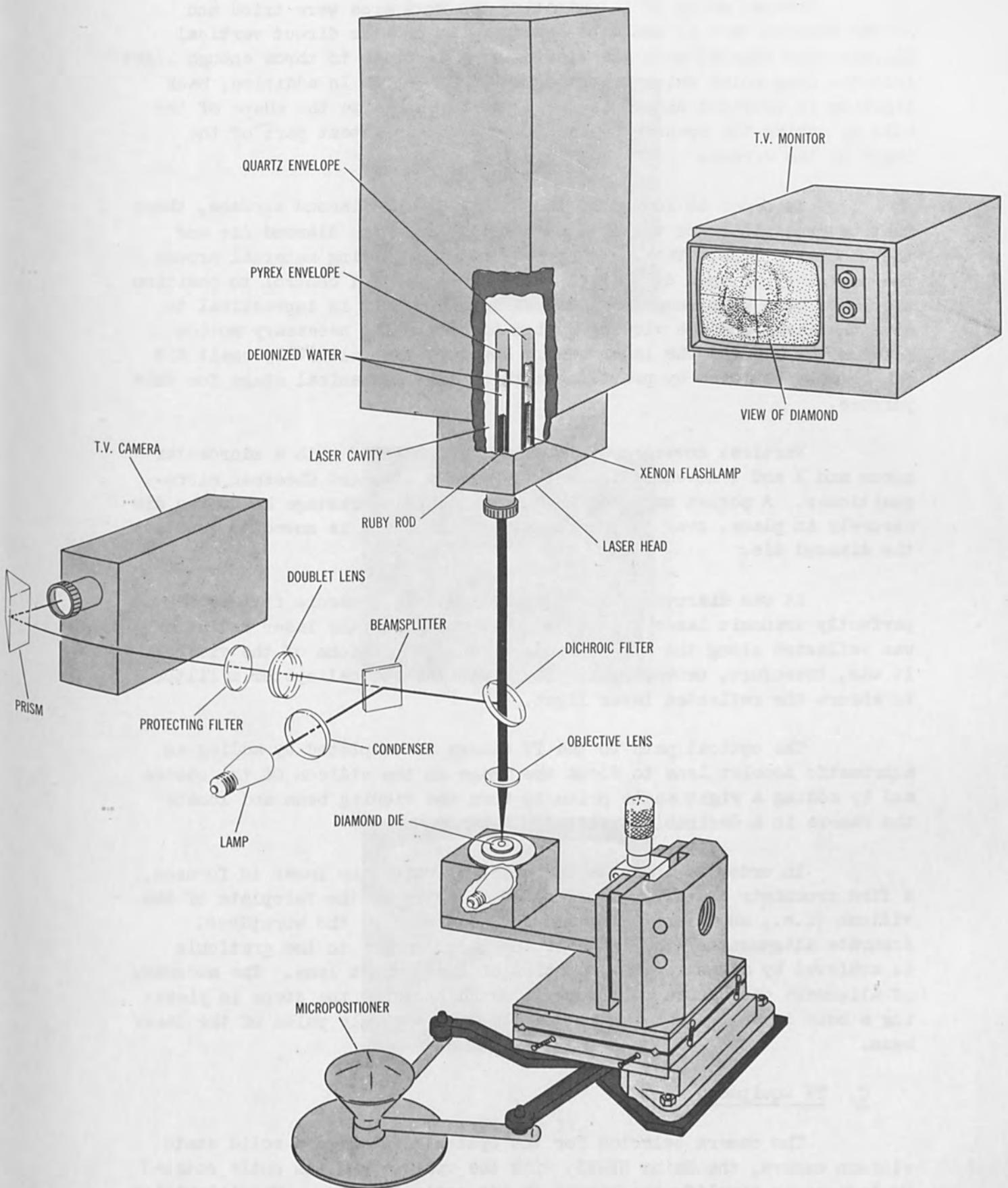


Figure 10. Schematic Diagram of Overall Laser System.

Several means of illuminating the work area were tried and it was decided that it would be necessary to provide direct vertical illumination coaxial with the viewing beam in order to throw enough light into the deep holes which are made in the diamond. In addition, back lighting is provided under the workpiece to emphasize the shape of the hole by making the opening in the diamond the brightest part of the image on the screen.

In order to focus the laser beam at the diamond surface, there must be provisions for vertical movement between the diamond die and the objective lens. The movement required in removing material around the periphery of worn dies also necessitates X and Y control to position the laser beam with respect to the die opening. It is impractical to move the laser and the viewing system to obtain the necessary motion between the die and the laser beam. However, the relatively small die can readily be moved by providing a three axis mechanical stage for this purpose.

Vertical movement of the stage is achieved with a micrometer screw and X and Y movement is achieved with a standard Chessman micro-positioner. A pocket machined in the mechanical workstage holds the die securely in place, over the backlight, as the stage is moved to position the diamond die.

It was discovered quite early that the dichroic filter did not perfectly transmit laser light, but that enough of the laser radiation was reflected along the viewing path to damage portions of the vidicon. It was, therefore, necessary to place into the optical system a filter to absorb the reflected laser light.

The optical path to the TV camera is completed by adding an achromatic doublet lens to focus the image on the vidicon of the camera and by adding a right angle prism to turn the viewing beam and locate the camera in a desirable position.

In order to indicate the point at which the laser is focused, a fine crosshair graticule is cemented directly on the faceplate of the vidicon (i.e., superimposed on the optical image of the workpiece. Accurate alignment of the image of the laser crater to the graticule is achieved by adjusting the position of the doublet lens. The accuracy of alignment can be seen in Figure 11 which pictures the steps in piercing a hole in an ordinary sewing needle with a single pulse of the laser beam.

C. TV Equipment

The camera selected for the optical system is a solid state vidicon camera, the Shiba HV-13, with the vidicon and its coils rotated 90 degrees to simplify the layout of the optical system. The television

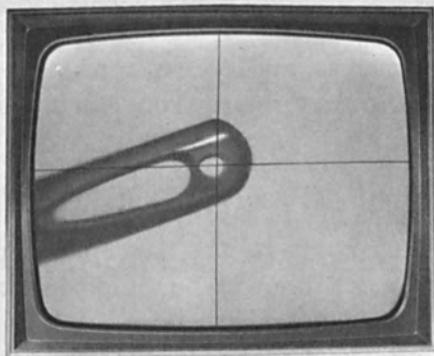
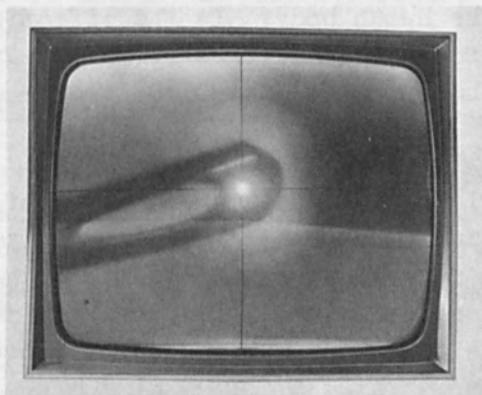
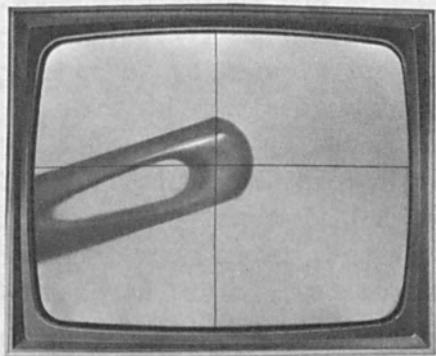


Figure 11.

receiver which is used as a monitor is a standard Zenith 12-inch portable modified to bypass the RF and tuning circuitry. The portable monitor can be moved as freely as a home TV set, and the monitor can be easily located for convenient viewing by an operator. Moreover, the picture can be viewed simultaneously by several people, other than the operator, while the laser actually proceeds to drill the diamond. This particular combination of camera and monitor, as discovered in prior studies of CCTV, provides a high-quality picture roughly comparable to that received on a home TV set. Even more important, this television system provides the high-quality picture at about \$600, or 1/3 the cost of a conventional closed-circuit television system.

VI. OPERATOR SAFETY

Introduction of the laser into a production environment brought with it new and unanswered questions regarding operator safety. Is laser radiation harmful to the human body? Is the effect of laser radiation cumulative? How dangerous are the fumes resulting from the laser action on toxic materials? What minimum safeguards are necessary to protect an operator?

At the Buffalo plant, a team, comprised of members from the engineering, medical, and industrial hygiene organizations, was organized to answer these questions. The team made a thorough search of laser literature and personally contacted leading laser experimenters. As a result of the team effort, the following minimum safety recommendations¹³ were made for industrial laser installations:

1. The laser head and the workstage shall be completely shielded by a light tight enclosure. All access ports to the enclosure shall be fitted with electrical interlocks to prevent an accidental laser discharge while the ports are open.

2. The laser and its enclosure shall be housed in a light tight room, so that experimental operation can be safely performed with the light shields removed.

3. The laser room shall be provided with air conditioning and positive ventilation.¹⁴ To relieve the load on air conditioning equipment, power supplies and pumps should be located outside the room.

4. The laser enclosure shall be ventilated to draw off any gases or fumes which might be generated by the laser's operation. Baffles shall be used in the exhaust ducts to ensure that no light can be reflected from the enclosure.

5. Means to visually confirm the well being of the operator

in the laser room shall be provided. Voice communication without opening the laser room door shall also be provided.

6. A high level of illumination shall be provided within the laser room, so that the iris diaphragms of the operator's eyes are normally contracted.

7. The walls of the laser room shall be painted with a light colored very flat paint. Laser light that accidentally hits the walls will be diffused.

To meet the safety requirements recommended by the team, a new light tight, air conditioned, well ventilated, well lighted room was constructed. A second CCTV system and an intercom were installed to provide a means of supervising the well being of the operator, who was carefully educated in the safe use of the laser. Figure 12 illustrates how an operator can be viewed from outside the room.

VII. SIMPLICITY OF OPERATION

From her position at the drilling machine, Figure 13, the operator must occasionally adjust the electrical controls grouped to her right. In addition, she must control the location of the die with respect to the laser beam during the drilling operation. Viewed at eyelevel, crosshairs, on a 60 X image of the diamond, mark the exact spot at which the laser beam will strike. Thus, the operator quickly and accurately can relocate the diamond between pulses of the laser beam.

VIII. PERFORMANCE OF THE LASER MACHINE

A laser piercing machine built to Western Electric specifications was constructed by Raytheon (following competitive bidding) and delivered to the Buffalo Plant in October 1965. Two weeks after delivery, the machine was ready to start processing diamonds.

Rather than use diamond die stones to gain experience in operating the new piercing machine, less expensive diamond macles were used. (A macle is thinner than a die stone and triangular in shape.) Piercing macles is somewhat more difficult than piercing die stones, because the macles are not supported by a die casing and are, therefore, more subject to cleavage. After the operating parameters were established, 50 macles were pierced with a high yield.

Experience gained from successfully piercing macles led to resizing worn production dies. The resizing was achieved with only minor adjustments of the operating parameters required to drill macles. The worn dies were resized quickly.



Figure 12.



Figure 13.

The final step in proving-in the laser machine was piercing the initial hole in a mounted diamond die stone. Test stones representing the common sizes 20 points, 55 points, and 100 points were selected. (The thickness of a stone is related to its weight and 100 points are equal to 1 carat.) As can be seen from the data in Figure 14, it takes less than 10 minutes to pierce a 20 point die stone, which is far less than the 24 hours required to mechanically drill a stone of the same weight.

Item	Nominal Stone Weight		
	20 point	55 point	100 point
Approximate dimensions, inch (Diam. x thick.)	1/8 x 0.050	3/16 x 0.080	1/4 x 0.150
Range of gauges used for	42 - 30	29 - 20	19 - 15
Energy (Joules)	10	10	10
Pulses, front side (@ 5 sec. repetition rate)	75-100	150-175	250-275
Pulses, back side	3	10	10
Time, minutes	6-9	12.5-14.6	20.9-23
Objective lens focal length, mm	33	33	33

Figure 14

This table summarizes the production results for initially drilling virgin die stones. The 33 mm focal length of the objective lens focuses the laser beam to a spot 0.010 inch in diameter, resulting in an energy density of 1.84×10^4 joules/cm².

IX. CONCLUSIONS

Laser experiments with many materials led ultimately to production processing of diamond wire-drawing dies with the laser. This step is but the first of many that undoubtedly will be taken in harnessing the laser beam. The team responsible for this development was comprised of specialists in wire drawing from the Buffalo Plant and specialists

in lasers and optics from the Engineering Research Center. The joint efforts of these groups were essential in achieving final success. The wire-drawing group was aware of the need for faster diamond piercing and of the possibility that laser piercing could meet that need. The laser and optics groups translated the need into a prototype machine and later into a production laser machine.

FOOTNOTES

1. American Wire Gauge.
2. Chauncey G. Peters, Walter B. Emerson, Karl F. Nefflen, Forest K. Harris, Irvin L. Cooter, "Electrical Methods for Diamond Die Production," U.S. Department of Commerce RR1787, Volume 38, May, 1947.
3. Laser piercing denotes either producing the initial hole in a diamond or enlarging the existing hole. In laser jargon, the piercing process is called drilling.
4. For a more detailed description of laser apparatus and theory of operation see Donald S. Young, "The Laser as an Industrial Tool," The Western Electric Engineer, VIII, 4 (October 1964) pp. 2-10.
5. The 10-joule, 0.5 millisecond laser beam previously had been used to drill metal, ceramic, and plastic.
6. This setting gives an energy density of 7.90×10^4 joules per square centimeter.
7. Clyde M. Adams, Jr. and Glenn A. Hardway, "Fundamentals of Laser Beam Machining and Drilling," IEEE Transactions, IGA 1, No. 2 March/April 1965.
8. Arthur L. Schawlow, "Advances in Optical Masers," Scientific American, 209 No. 1, July 1963, pp. 34-35.
9. M. Hercher, "Laser Induced Damage in Transparent Media," Presented at the spring meeting of the Optical Society of America, April 1964.
10. Surface crazed, cracked, shattered, and internally clouded.
11. The spot diameter of a focused laser beam is determined by two parameters: the focal length of the lens and the beam divergence, θ , of the laser. $S = f\theta$.
 - θ is beam divergence in milliradians.
 - f is lens focal length in inches.
 - S is spot size in mils.
12. A dichroic (two-color) filter is a mirror which transmits selected colors of light and reflects others.
13. Safety studies were continued and a company guide "Guidelines for the Safe Use of Lasers," which contains the answers to the questions posed above, is being published by the General Medical Director's Organization at headquarters.
14. The laser-room-ventilation requirement has not been verified.

1. The first part of the report deals with the general principles of the method of moments. It is shown that the method of moments is a special case of the maximum likelihood method. The method of moments is based on the assumption that the population parameters are known and the sample moments are used to estimate them. The method of moments is a simple and easy to use method for estimating the parameters of a distribution. It is particularly useful for distributions that are not normal. The method of moments is also useful for distributions that are skewed or have heavy tails. The method of moments is a good alternative to the maximum likelihood method when the maximum likelihood method is difficult to apply.

2. The second part of the report deals with the application of the method of moments to the estimation of the parameters of a normal distribution. It is shown that the method of moments estimates of the mean and variance of a normal distribution are the same as the maximum likelihood estimates. This is not true for all distributions. For example, for a skewed distribution, the method of moments estimates of the mean and variance are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

3. The third part of the report deals with the application of the method of moments to the estimation of the parameters of a binomial distribution. It is shown that the method of moments estimates of the probability of success and the number of trials of a binomial distribution are the same as the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

4. The fourth part of the report deals with the application of the method of moments to the estimation of the parameters of a Poisson distribution. It is shown that the method of moments estimates of the mean and variance of a Poisson distribution are the same as the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

5. The fifth part of the report deals with the application of the method of moments to the estimation of the parameters of a gamma distribution. It is shown that the method of moments estimates of the shape and scale parameters of a gamma distribution are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

6. The sixth part of the report deals with the application of the method of moments to the estimation of the parameters of a beta distribution. It is shown that the method of moments estimates of the two parameters of a beta distribution are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

7. The seventh part of the report deals with the application of the method of moments to the estimation of the parameters of a chi-squared distribution. It is shown that the method of moments estimates of the degrees of freedom and the variance of a chi-squared distribution are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

8. The eighth part of the report deals with the application of the method of moments to the estimation of the parameters of an F-distribution. It is shown that the method of moments estimates of the two parameters of an F-distribution are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

9. The ninth part of the report deals with the application of the method of moments to the estimation of the parameters of a t-distribution. It is shown that the method of moments estimates of the mean and variance of a t-distribution are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

10. The tenth part of the report deals with the application of the method of moments to the estimation of the parameters of a normal distribution with unknown mean and variance. It is shown that the method of moments estimates of the mean and variance of a normal distribution are the same as the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

11. The eleventh part of the report deals with the application of the method of moments to the estimation of the parameters of a normal distribution with unknown mean and unknown variance. It is shown that the method of moments estimates of the mean and variance of a normal distribution are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

12. The twelfth part of the report deals with the application of the method of moments to the estimation of the parameters of a normal distribution with unknown mean and unknown variance. It is shown that the method of moments estimates of the mean and variance of a normal distribution are different from the maximum likelihood estimates. This is because the method of moments estimates are based on the first two moments of the distribution, while the maximum likelihood estimates are based on the entire distribution.

DESIGN OF A PRODUCTION-WORTHY LASER MICROWELDER

by

Jon H. Myer
Hughes Aircraft Company
Newport Beach, California

ABSTRACT

The evolutionary steps in the design and construction of a production-worthy, safe, and economical laser microwelder are described. A novel, modular, liquid-cooled, cylindrical ruby laser head as well as a novel method of adjustment of the location of the welding focal plane are described.

I. INTRODUCTION

Early work in laser welding at Hughes Aircraft Company by Budenhagen, Engquist, and Martin indicated the need for a production-worthy, safe, and economical laser microwelder. Various methods of aiming, focusing, and weldment placement were analyzed and it was found that an optimal system comprises a horizontal laser head with a dichroic beam splitter mirror performing the dual function of deflecting the beam vertically and providing a coaxial aiming light path. Due to the immaturity of the laser welding process the design incorporated modular elements permitting later additions and modifications.

II. SYSTEM DESCRIPTION

The evolution of the present machine came about as follows: Our first laser welder employed the principle of Figure 1. Alignment of the weldment was carried out by viewing through the ruby. A solenoid activated shutter protected the operator. This system had the disadvantage that workpiece location was low in relation to the eye and workpiece placement was unnecessarily cumbersome. Another disadvantage was the hazard of viewing through the laser, particularly since the magnetic shutter jammed every once in a while. Last but not least the visual image attenuation caused by the red ruby and its reflecting coatings interfered with weldment alignment and observation.

Our next design, illustrated in Figure 2, placed the weldment vertically, the observer looking down into a suitably placed beam splitter. The inconvenience of vertical weldment placement, the hazard of

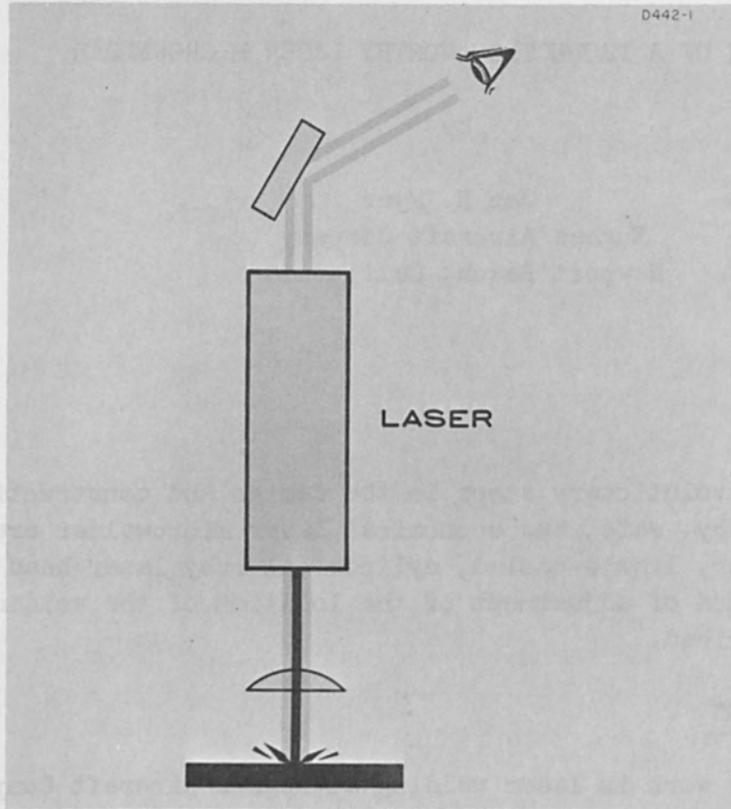


Figure 1.

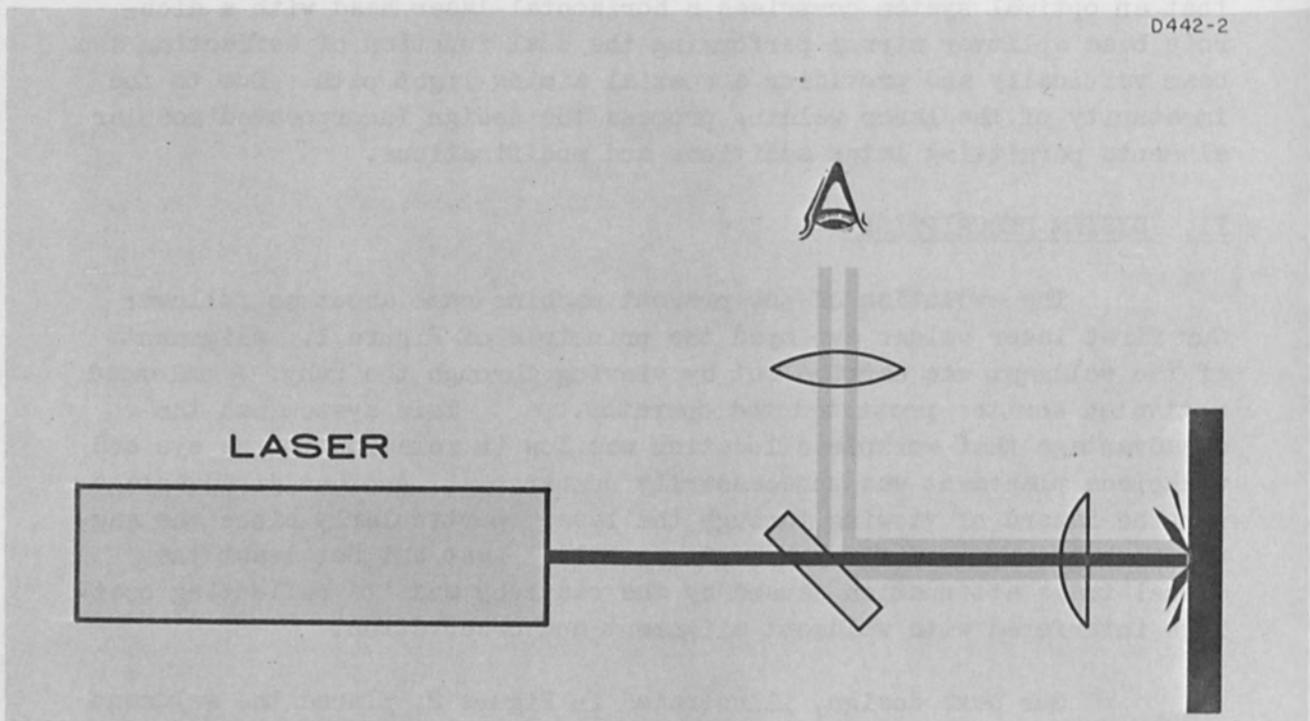


Figure 2.

a horizontally emerging beam requiring permanently placed backstops, and the losses at both surfaces and the bulk of the beam splitter made this a less than ideal arrangement.

Figure 3 shows an early method of removing the disadvantages of vertical weldment placement. However, the problems of excess losses on the beam splitter still remained.

Figure 4 shows the arrangement presently selected. Here only a single reflection loss is encountered at the first surface of the beam splitter, laser radiation is effectively filtered from the observing beam due to the color selective action of the beam splitter, and standard microscope components can be applied in a production-worthy configuration.

To permit the adjustment of spot size without introducing optical elements into the laser beam light path, the method of Figure 5 was chosen. It consists of defocusing the viewing path which is shown in the right half of the lenses while maintaining the location of the focal plane of the laser beam. As the left figure shows with both the welder and the viewing focal plane coplanar, the highest energy density will be obtained at the surface of the weldment. As shown in the center illustration, a converging viewing path will cause the elevation of the viewing focal plane in relation to the laser focal plane effectively submerging the focal spot of the laser when the surface of the weldment is in focus. As shown in the right illustration, a divergent viewing beam will force the operator to lower the weldment to retain a sharp image surface effectively elevating the focal plane of the laser above the surface of the weldment.

If we add the zap control to the features of Figure 4 and Figure 5 we have the Model 5500 microwelder, illustrated in Figure 6. Here the light path from a standard microscope eyepiece head (American Optical Microstar) which is designed to be focused at infinity is passed through a relay lens, a reticle and a spot size control lens and dichroic mirror toward the objective lens and the weldment, giving a true view of the product. All the necessary optical components with their unavoidable absorption and reflection losses are in the observation light path. The laser beam encounters only the slight loss of reflection at the dichroic mirror.

The welder head assembly is shown in Figure 7. The zap control activates all interlocks and initiates the firing sequence. This control includes the shutter and the necessary switches to provide one-hand control for laser pulse initiation, and provides a safe transition from viewing-aiming through charging and firing. As long as the shutter is open the capacitor bank is shorted to ground. The firing cycle is

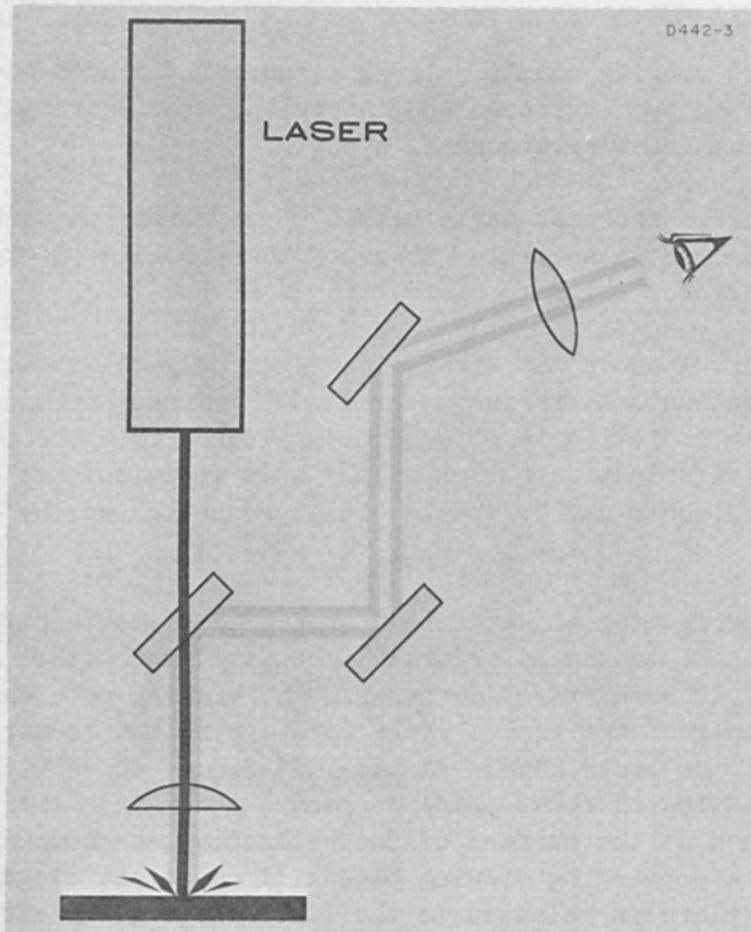


Figure 3.

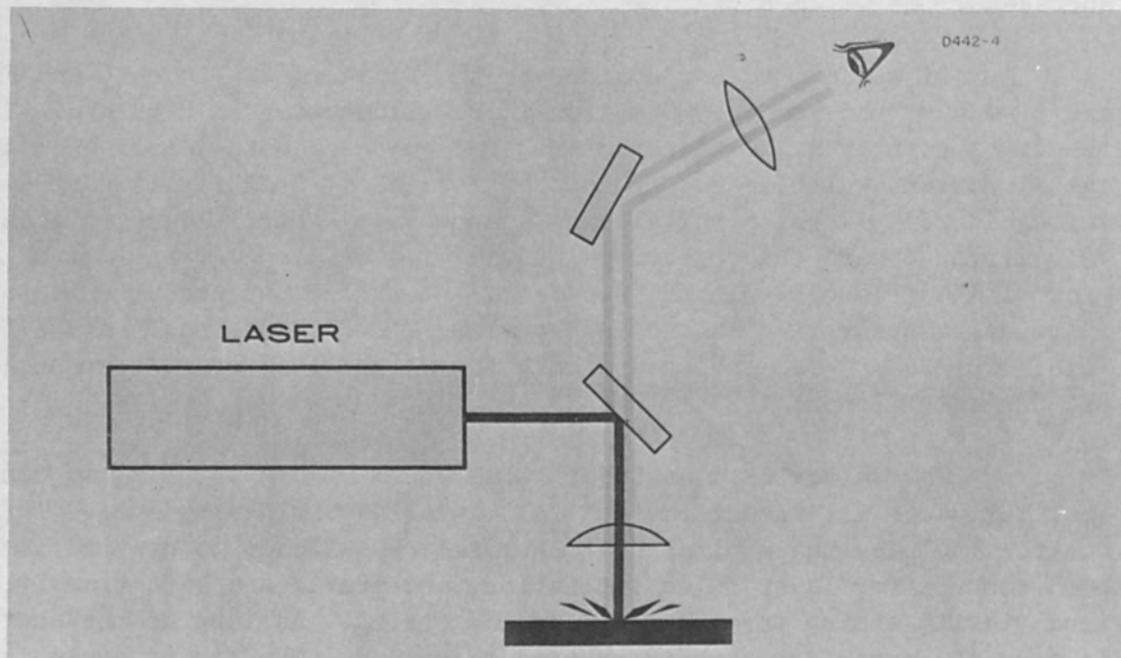


Figure 4.

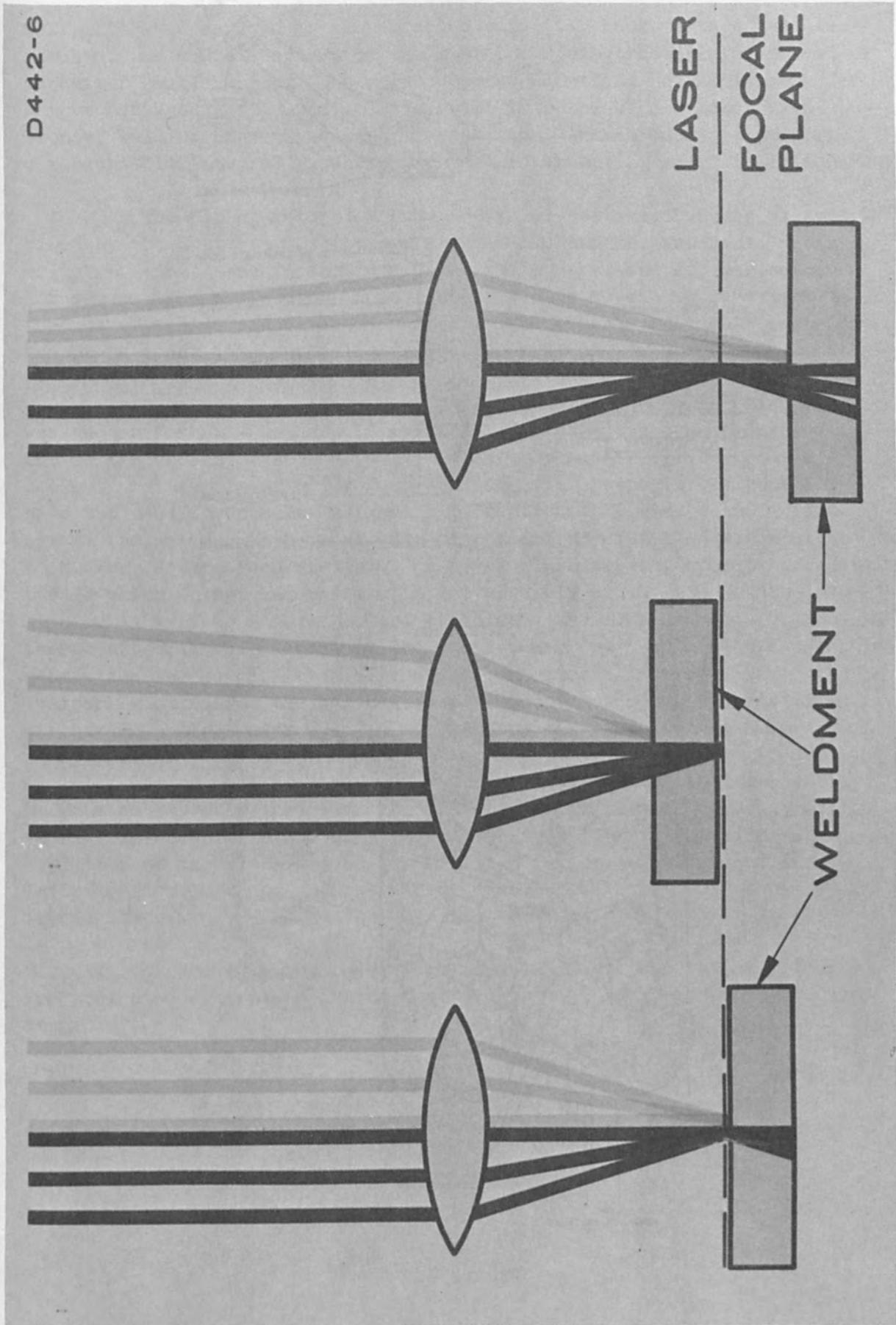


Figure 5.

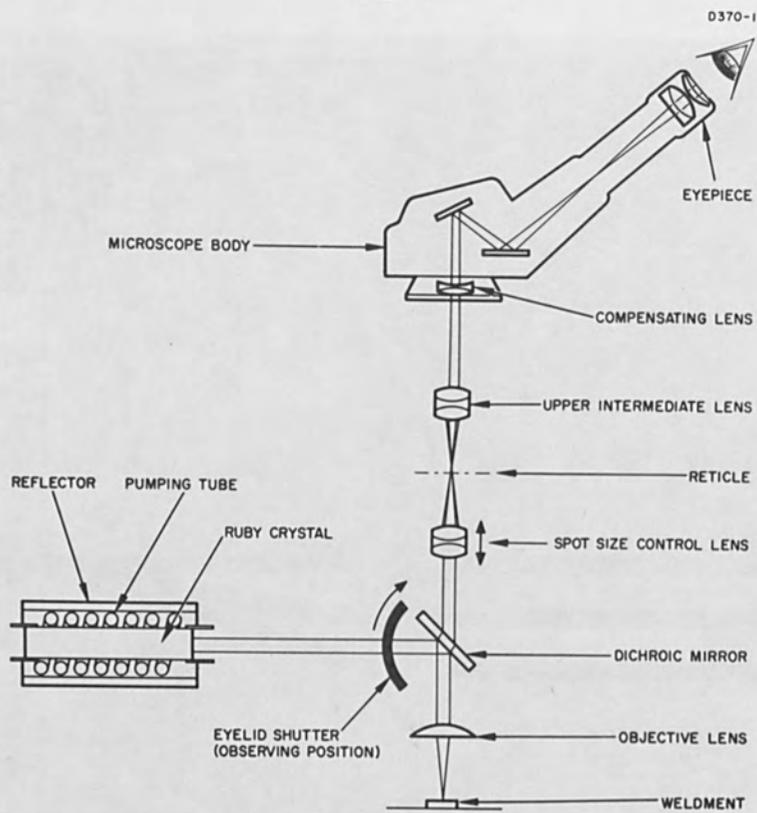


Figure 6.

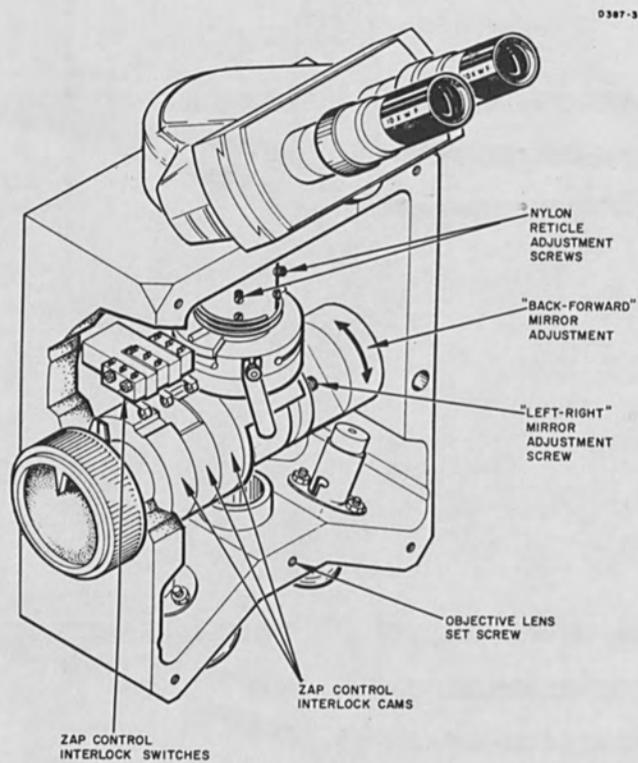


Figure 7.

self-initiated as soon as the selected capacitor bank voltage has been reached. An audio alert warns the operator during the brief interval of capacitor bank charging. To reduce operator fatigue powerful lamps for bright background illumination are built into the welder head. A selection of welding objectives of different focal length and incorporating a replaceable integral flashback protector can be installed in this head.

The firing rate of a human operator-activated welder is determined by the dexterity of the operator in placing the weldment. This criterion limits the firing rate to no faster than one zap per second nor slower than a zap every four seconds. Such a welding rate permits the use of a high power vacuum switch in the discharge circuit reducing the parts count and increasing the reliability of the power supply.

A cylindrical laser cavity with a helical pumping flashlamp was chosen for this welder. Significant refinement has been carried out on the design of such a cavity. The ruby has integrally coated dichroic reflectors of 98% and 65% reflectivity respectively to eliminate the need for complex alignment procedures. Figures 8 and 9 illustrate the design which is capable of containing distilled water at up to 20 lbs./square inch pressure. A novel flashlamp incorporating electrodes which do not interfere with the assembly of the ruby or the reflector as well as a novel method of holding the ruby are combined in this configuration. Access to the laser beam from both sides is possible permitting the use of integral coatings or external mirrors. A 15 KV triggering pulse for the helical lamp is coupled through the distilled deionized cooling water which envelops both the ruby and the lamp.

Aging data at maximum power settings on both the power supply and the cavity show that with the present design thermal capacity of the refrigeration unit is the limiting factor on firing rate, while gas depletion of the flashlamp is the limiting factor on laser head life. Investigations are now underway to determine feasibility of using re-filled lamps.

Figure 10 shows the welder in operation. Conveniently located are spot size control, focusing control, voltage level control, and zap control.

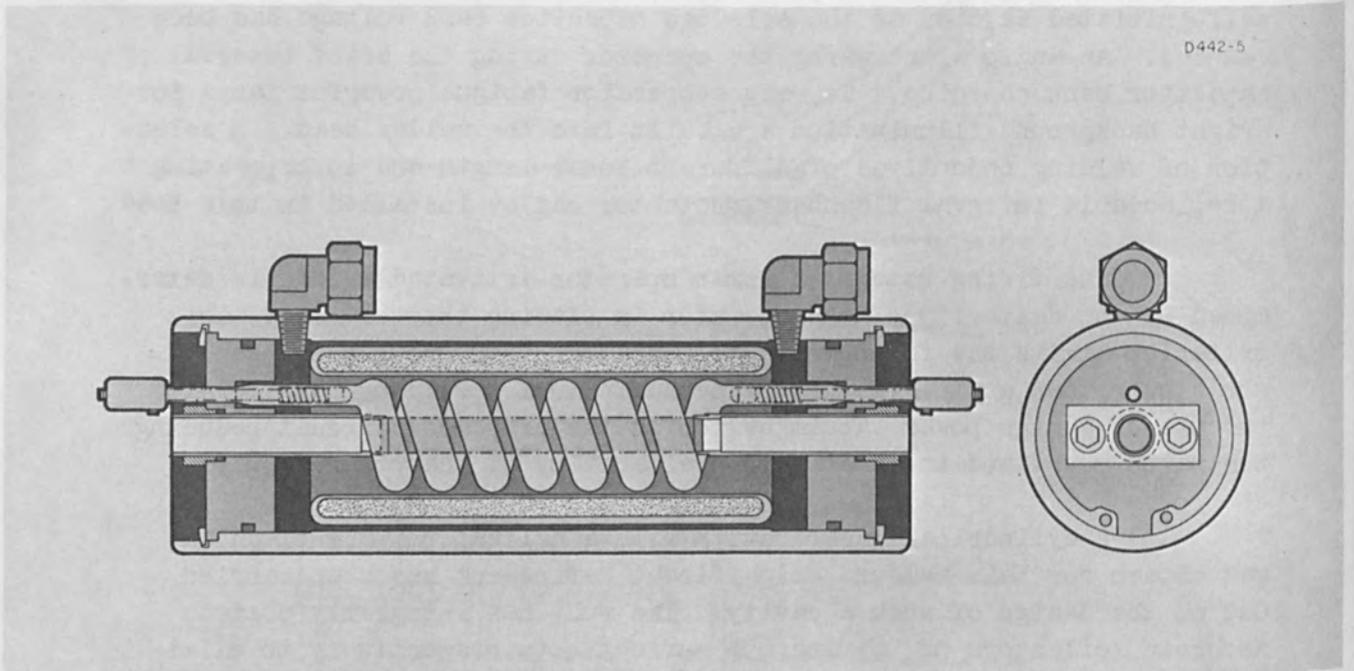


Figure 8.

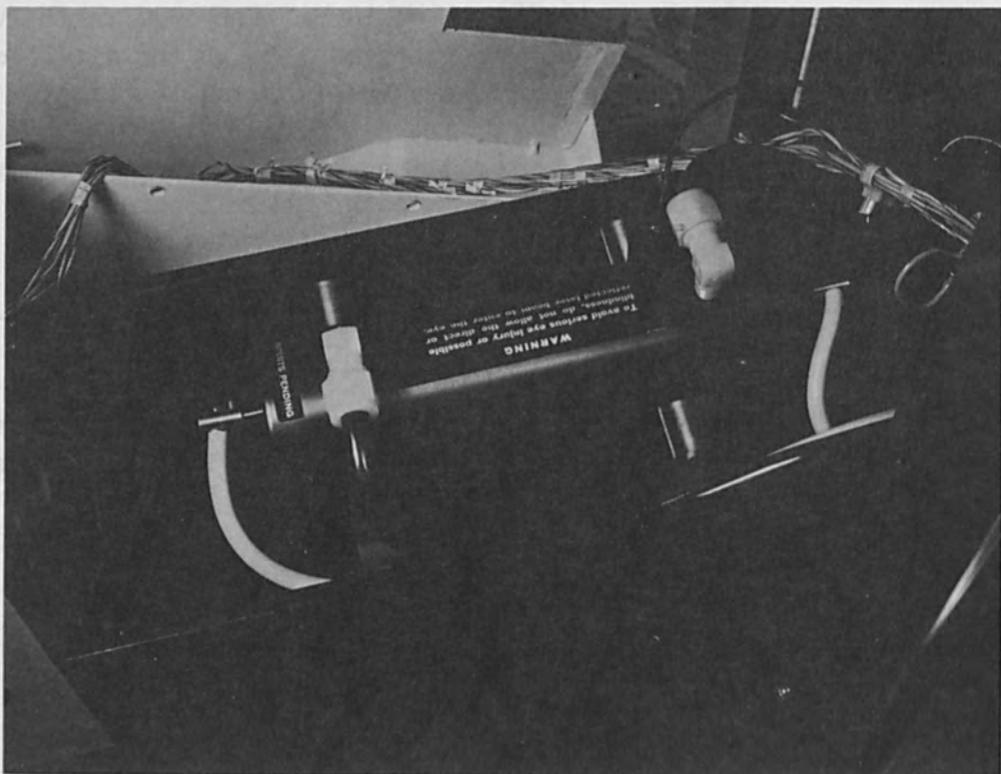
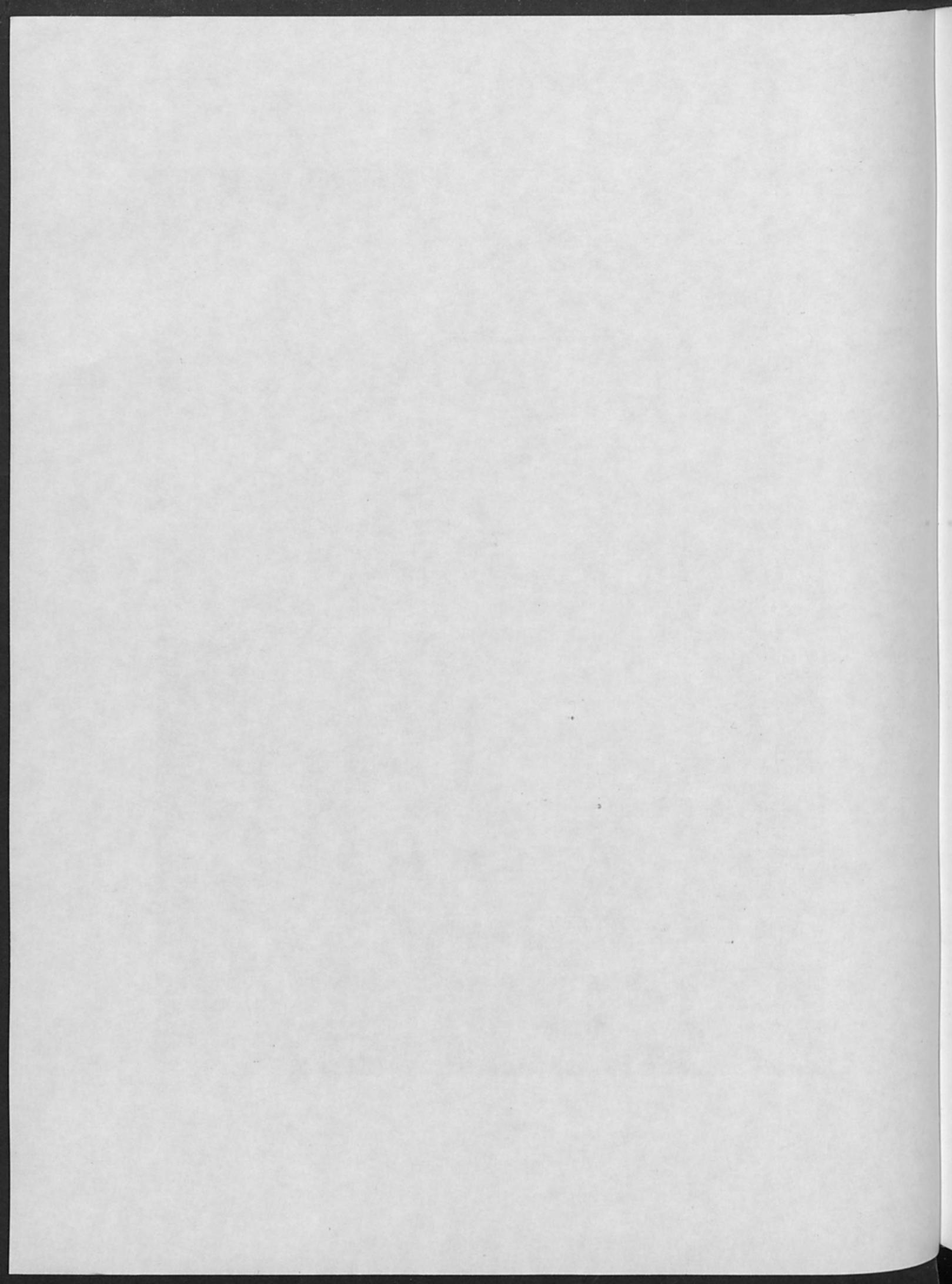


Figure 9.



Figure 10.



THE ANNULAR HOLLOW CATHODE; ITS OPERATION AND APPLICATIONS

by

J. W. Davis, A. P. Walch, and E. A. Pinsley
United Aircraft Research Laboratories
East Hartford, Connecticut

ABSTRACT

Annular cathodes have been developed which emit a disc-like beam of electrons radially inward to heat a cylindrical workpiece located along the axis. Designs and conditions for stable and efficient operation have been established. Operating characteristics and applications are described.

I. INTRODUCTION

Electron beam heating over extended focal lines and curves can be accomplished with conventional thermionic electron beam guns. However, these devices are susceptible to filament deterioration, non-uniform emission, electrode misalignment and arcing in the presence of ambient pressure excursions due to outgassing. Such devices are generally restricted to operation at pressures below 10^{-4} Torr. Accordingly a program was undertaken to investigate various types of cold cathode discharges as sources of high energy electron beams since the use of nonthermionic cathodes precludes many of the foregoing problems. An outgrowth of this program has been the development of cold cathodes that generate, control and focus electron beams in a glow discharge environment. One of these, the annular hollow cathode generates a high-energy disc-like beam which heats a narrow zone around the periphery of cylindrical workpiece with a high degree of azimuthal uniformity.

II. ELECTRON BEAM FORMATION IN AN ANNULAR HOLLOW CATHODE DISCHARGE

The annular hollow cathode discharge considered in this investigation is a high voltage (>1000 volts) dc glow discharge that operates in the pressure range from 10^{-2} to 1 Torr. Under these conditions the voltage-current characteristics of the discharge are similar to those of an abnormal glow discharge, i.e., voltage increases with increasing current. The structure and potential distribution of the region adjacent to the cathode of a cold cathode glow discharge are shown schematically in Fig. 1. In the operating range of interest practically the entire potential drop across the discharge occurs in the cathode fall distance, d_c^1 . Electrons are emitted from the cathode by a variety

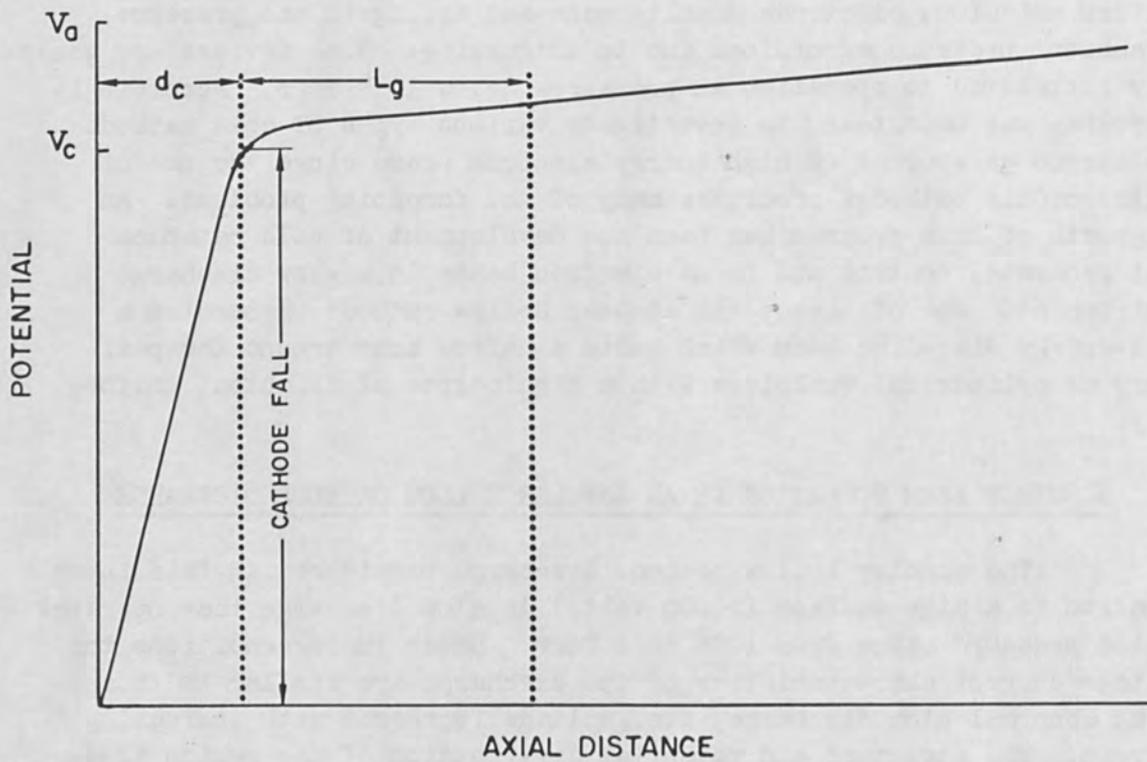
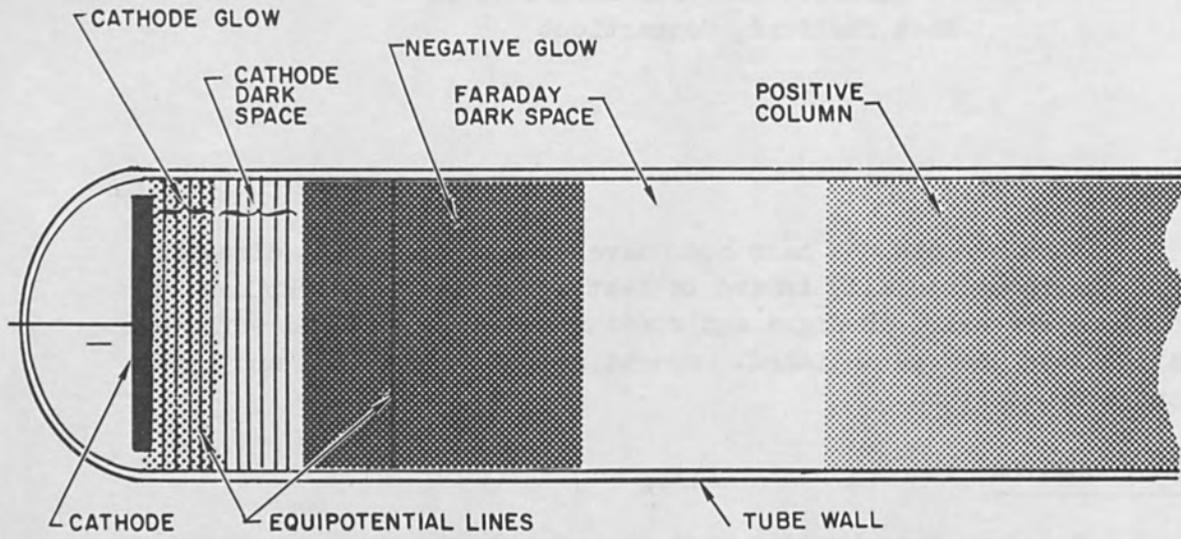


Figure 1. Structure and Potential Distribution of A Glow Discharge.

of nonthermionic processes, including photoemission and secondary emission due to bombardment by ions, metastables and fast neutrals. Electrons emitted from the cathode are accelerated to energy eV_c in the cathode fall, and their trajectories are relatively unperturbed well into the negative glow. In fact, the extent, L_g , of the negative glow² correlates very well with the range of electrons of energy eV_c in the same gas. The range of electrons is large compared to physical dimensions of the discharge appropriate for high energy beam formation.

The hollow cathode effect, observed when the usually planar cathode of a discharge is replaced by a hollow tube or by two closely spaced planar cathodes³ is characterized by a large increase in discharge current. A form of hollow cathode discharge which generates a collimated pencilform beam of energetic electrons was first reported by Van Passen and Allen;⁴ subsequent investigations have shown that sufficiently high power can be obtained to perform welding operations.^{5,6} The operation of this type of hollow cathode can best be understood by visualizing the equipotential lines in the vicinity of the cathode as shown in Fig. 2. The cathode configuration is usually a closed cylinder with an aperture, A , in one end. When operating in the desired electron beam mode at thousands of volts, most of the cathode fall region occurs outside of the cathode. However, the potential in the interior of the cathode may be of the order of 100 volts. Electrons emitted from the interior surface of the cathode by secondary processes are trapped in a potential well and produce many ion-electron pairs by volume ionization processes. Electrons escaping through the aperture are accelerated through the cathode fall to form a well-collimated beam of high energy electrons. In general, to maintain this type of operation the size of the aperture with respect to the cathode fall distance, d_c , must be in the range $0.01d_c < A < 0.1d_c$. For $A < 0.01d_c$ the aperture is not sufficiently large to perturb the potential distribution and the discharge does not penetrate to the inside of the cathode cavity. For $A > 0.1d_c$, the intensity of the discharge within the cavity builds up to such a high level that the cathode fall distance further reduces, and the entire potential drop occurs inside of the cathode. This results in transition to an arc-like mode of operation, characteristic of the more common hollow cathode discharges. (The transition is, however, also a function of the size of the cavity, H , as described later.) In the arc mode there is no well-collimated electron beam, and most of the discharge power is dissipated at the cathode.

Experiments have been performed with annular cathode geometries in the form of a figure of revolution such as that shown in Fig. 2. In this type of configuration the aperture is a slit around the inner periphery of a hollow annular chamber and a disc-like beam of electrons is emitted radially inward. It has been found, however, that a hollow cathode with an extended aperture of this type has a much more restricted

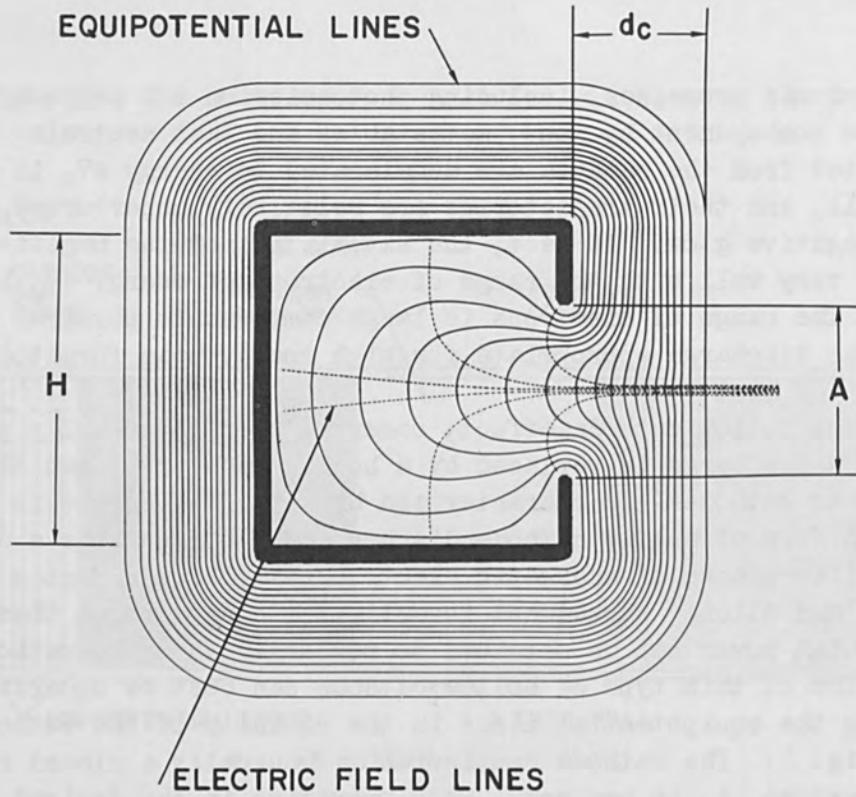


Figure 2. Potential Distribution in the Vicinity of A Hollow Cathode Operating in the Electron Beam Mode.

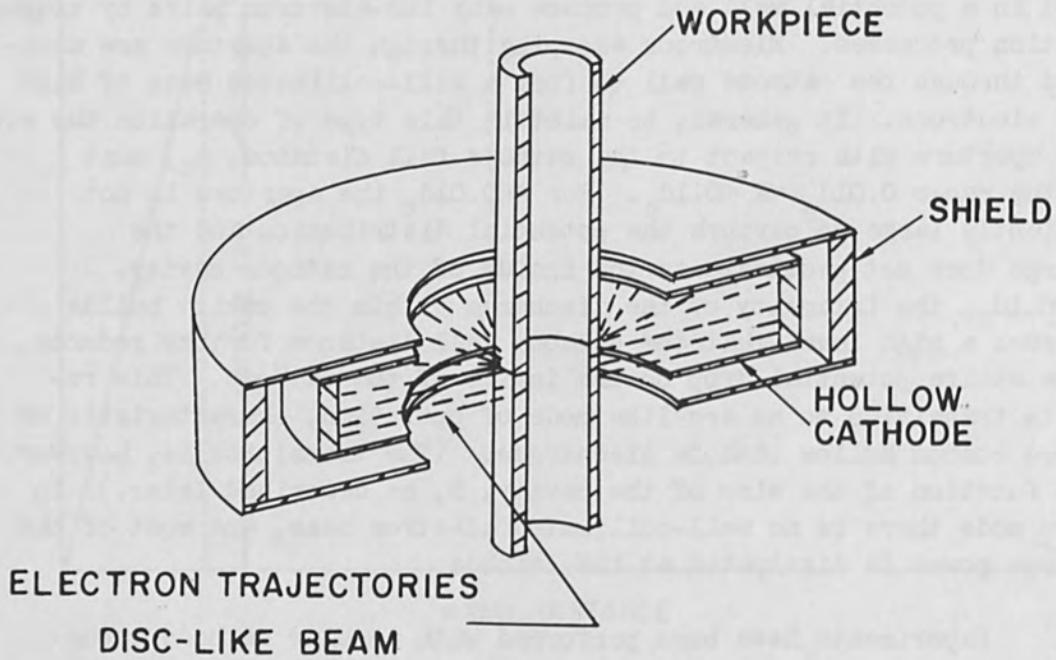


Figure 3. Sectional View of A Wide Aperture Annular Hollow Cathode.

range of operation in the electron beam mode than the hollow cathode with a simple circular aperture; i.e., it has a much higher tendency to transition into the arc mode of operation. Fortunately, however, this limitation exists only for narrow aperture cathodes, i.e., for cathodes with $A/H < 0.7$. For cathodes in which $A/H > 0.7$ (wide aperture cathodes) the range of operation is even greater than that of the axial beam hollow cathodes. In fact, these wide aperture cathodes operate stably in the electron beam mode even when $A = d_c$. Although this observation at first seems to contradict the previously described conditions for operation in the electron beam mode, those conditions pertain to narrow aperture cathodes in which the trapping of electrons within the cavity is very efficient. As A/H increases the trapping of electrons within the cavity becomes less efficient, and the discharge within the cavity does not build up to the high intensity which results in a transition to the arc mode. Thus the wide aperture cathodes are capable of stable operation at high power levels, but they are somewhat less efficient than the narrow aperture cathodes.

Because wide aperture cathodes are very stable in the electron beam mode, most annular hollow cathodes are now designed with $A \approx H$ as shown in Fig. 3. Another essential feature of the annular cathode is the shield which is spaced close to all exterior surfaces of the cathode to exclude the discharge from this region. The shields effectively suppress extraneous electron emission and greatly increase the power efficiency of the cathode. Although dielectric shields are preferable, metallic shields are often used because of the ease of fabrication. In the assembly of a cathode, special precautions must be taken to shield all high voltage leads and supports to eliminate breakdown and arcing.

III. OPERATING CHARACTERISTICS

Shielded annular hollow cathodes of the configuration shown in Fig. 3 are generally constructed of stainless steel. However, cathodes can be fabricated from almost any material, and the choice of material can usually be made on the basis of compatibility with the environment and operating conditions prescribed by particular applications. A wide variety of cathode configurations have been tested, the sizes ranging from 1 in. to 6 in. inner diameter and the apertures, A , ranging from $3/16$ in. to 1 in. A series of scaled cathodes were tested at conditions prescribed by gas discharge scaling rules⁷ and found to deviate from the usual similarity relationships. These deviations are attributed to the presence of effects which are not allowed by similarity considerations, such as photoemission, emission due to the bombardment of the cathode by metastable and fast neutral atoms, and variations in the secondary emission coefficient and density distribution due to

changes in surface temperature of different cathodes operating at different power levels. The voltage-current characteristics of all cathodes tested exhibit the usual positive slope and dependence on pressure as illustrated in Fig. 4. The data in Fig. 4 were obtained with a 2-1/2 in. ID cathode having a 1/4 in. aperture. With larger cathodes (6 in. ID) current levels up to 1 ampere have been obtained.

Annular hollow cathodes have been operated at input power levels up to 10 kilowatts. In general, these cathodes need not be cooled except for operation at the highest power levels over long periods of time. The power in the electron beam has been measured calorimetrically and, for stainless steel cathodes, is generally between 65 and 70 percent of the input power. The power density distribution has been measured using two different techniques. In one method, a "terminated" cylindrical calorimeter is traversed along the cathode axis and through the beam such that the output from the calorimeter represents an integration of the beam power over the distance traveled. The power density distribution is then determined by differentiation of this curve. In the second method, a one mil dia tungsten wire is used as a resistance thermometer. The wire is traversed through the beam to measure the power density directly as a function of axial position. Results of the measurements with both techniques are in good agreement. A typical power density distribution determined using the wire probe is shown in Fig. 5. These data were taken with a narrow aperture (0.22 in.) cathode and the beam width, defined as the width of the distribution at half of peak power density, was found to be approximately equal to one half the size of the cathode aperture. The power density distribution of wide aperture annular hollow cathodes is similar to that shown in Fig. 5, except that the beam width is approximately equal to the size of the aperture. This difference further distinguishes the operation of narrow aperture and wide aperture cathodes. Peak power densities achieved with the wide aperture cathodes are generally on the order of several thousand watts/sq. in. The local power density for an annular hollow cathode operating at a given power level is a function of radius. Measurements at different radii indicate that the electron beam does not focus appreciably in the axial direction. This implies that the power density decreases approximately linearly with increasing radius measured from the centerline of the cathode.

With wide aperture hollow cathodes, stable operation at high power levels has been attained for extended periods of time in many different gases, including hydrogen, helium, argon, nitrogen, oxygen, and air. After an initial transient time for a cathode to come to an equilibrium temperature, the discharge voltage and current remain quite constant. The transient times vary from several seconds to several minutes depending on the cathode design and operating power level. However, drifts in operating conditions that may occur over long periods of time can be readily controlled by slight adjustments in the ambient

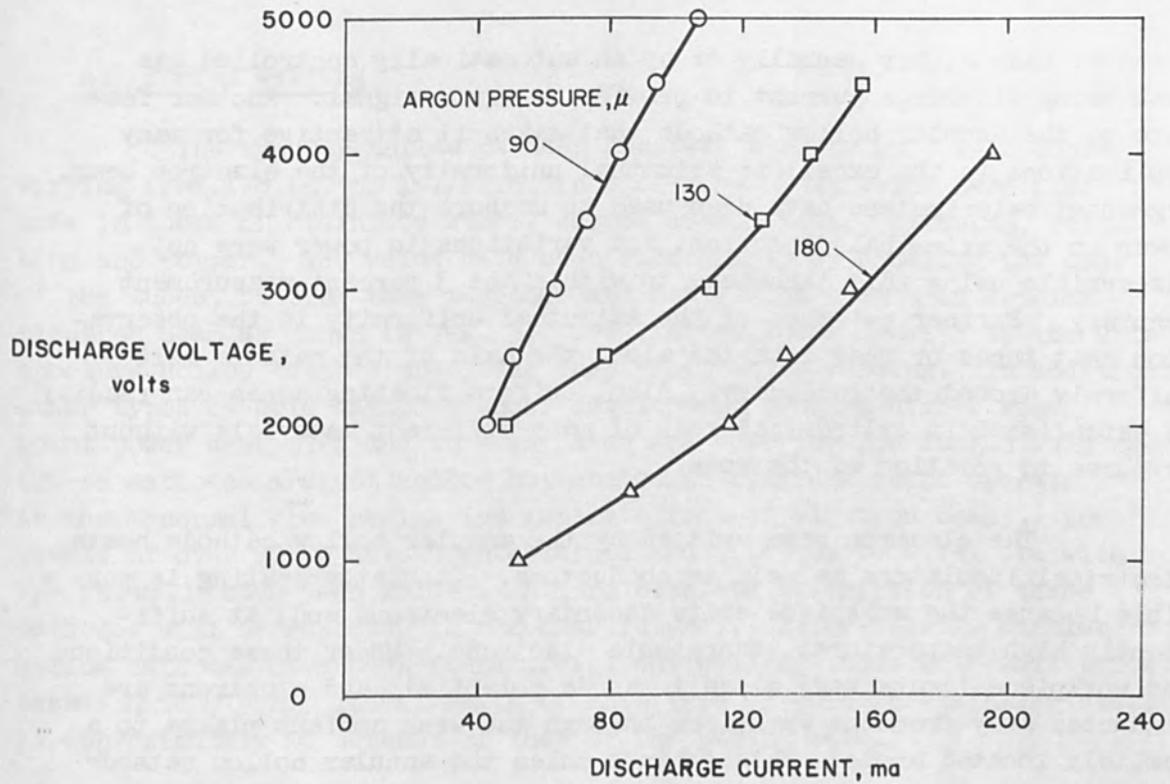


Figure 4. Voltage-current Characteristics of A Wide Aperture Annular Hollow Cathode; ID = 2-1/2 in., A = 1/4 in.

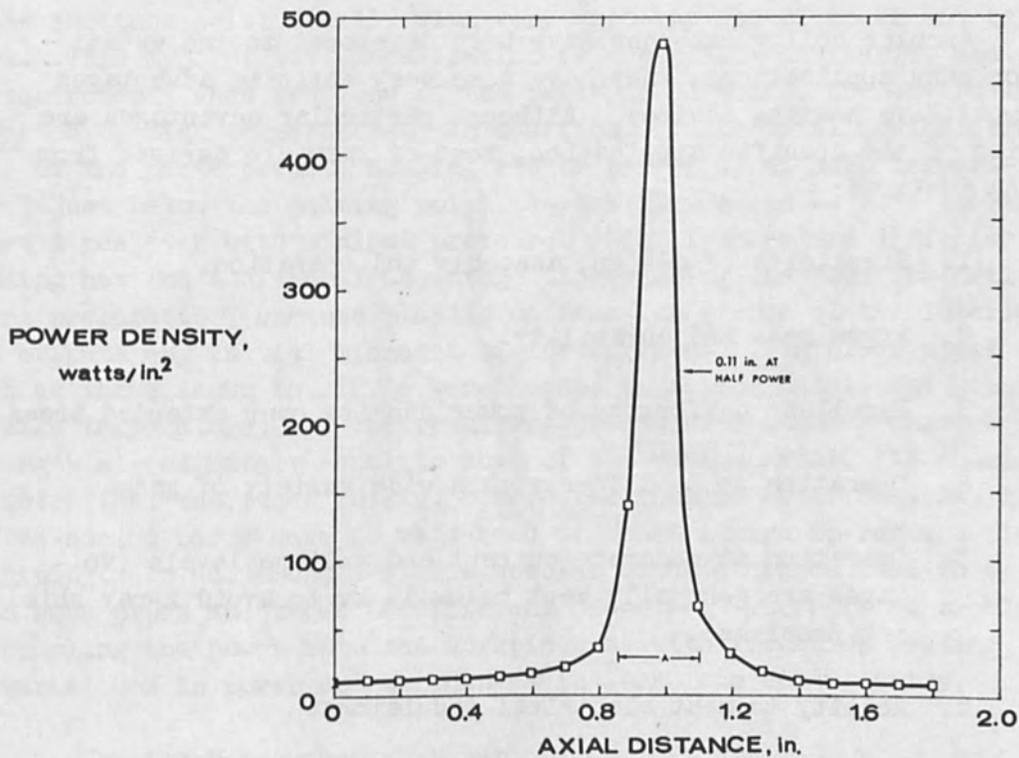


Figure 5. Typical Power Density Distribution Across the Beam of A 2.5 in. ID, Narrow Aperture Annular Hollow Cathode Operating at 3 kv and 31 ma in Argon at 11 Microns.

pressure made either manually or by an automatically controlled gas leak using discharge current to provide an error signal. Another feature of the annular hollow cathode that makes it attractive for many applications is the excellent azimuthal uniformity of the electron beam. Segmented calorimeters have been used to measure the distribution of power in the azimuthal direction, and variations in power were not discernible using this technique to within the 3 percent measurement accuracy. Further evidence of the azimuthal uniformity is the observation that tubes or rods inserted along the axis of the cathode melt uniformly around the periphery. Also, uniform floating zones can readily be established in cylindrical rods of many different materials without recourse to rotation of the rods.

The electron beam emitted by the annular hollow cathode heats electrical insulators as well as conductors. Insulator heating is possible because the workpiece emits secondary electrons and, at sufficiently high temperatures, thermionic electrons. Under these conditions the workpiece floats very close to anode potential, and electrons are conducted away from the workpiece through the weak ambient plasma to a remotely located anode. This feature makes the annular hollow cathode attractive for heating and melting dielectric materials with very high melting temperatures.

IV. APPLICATIONS

Annular hollow cathodes have been developed to the extent that, for many applications, they have some very definite advantages over competitive heating devices. Although particular advantages are a function of the specific application, most of them are derived from the following features:

1. Simplicity of design, assembly and operation.
2. Ruggedness and durability.
3. Excellent uniformity of power density over extended areas.
4. Operation up to 1 Torr with a wide variety of gases.
5. Operation at moderate current and voltage levels (voltages are generally kept below 15 kv to avoid x-ray shielding problems).
6. Ability to heat electrical insulators.

Some of the applications which have been performed with annular hollow cathodes are described in the following paragraphs.

A. Fusion Welding

The annular hollow cathode has been used to butt weld tubes varying from 1/8 in. to 1-1/2 in. in diameter. Butt welds have been made in tubes of stainless steel, carbon steel, Kovar, titanium, columbium and copper. The welds have been made simply by squaring the ends of the tubes, butting them together and heating them with an annular electron beam as shown in Fig. 3. The limited power density of the wide aperture hollow cathode precludes deep penetration welding. In addition, other types of cold cathodes under development produce higher focal point power densities and in some cases are superior for fusion welding. (These cathodes are not hollow but contoured cathodes which operate in the abnormal glow regime and generate focused electron beams. Power levels of over 20 kw have been obtained and pipes up to 4 in. dia with 1/4 in. wells have been welded. A more complete description of these cathodes will be reported in the near future). Nevertheless, annular hollow cathodes have been found useful for welding tubes with wall thicknesses up to 1/8 in. The strength of this type of weld in stainless steel is approximately 90 percent of that of the parent metal.

B. Diffusion Bonding

The annular hollow cathode discharges have also been used as a heat source for diffusion bonding pipes and tubes. The electron beam heats the tube joint locally with very good control of power and temperature. This type of heat localization is especially desirable, and often a requirement, when sections of the parts to be bonded are adversely affected by high temperature. In addition, if the metallurgical properties of the parts permit, bonding can be performed at high temperature, just below the melting point, so that bonds can be made in very short times even with minimal pressure. High temperature diffusion bonding has the additional advantage of obviating the need for meticulous joint preparation, because plastic deformation occurs at the interface and offsets any initial mismatch of the surfaces. Stainless steel tubes such as those shown in Fig. 6 were bonded in approximately one minute at high temperature, and the resulting joints were found to have a strength approximately equal to that of the parent metal. It should be noted that the tubes in Fig. 6 were not touched after bonding, i.e., the as-bonded parts have no weld bead or constriction to require finish machining. In addition, a single annular cathode can be used to diffusion bond pipes and tubes of different diameter. There is no problem of coupling the power into the workpiece as with induction heating, or of variations in power due to changes in workpiece conductivity.

C. Joining Nonconducting Materials

Annular hollow cathodes have been used to join alumina, zirconia

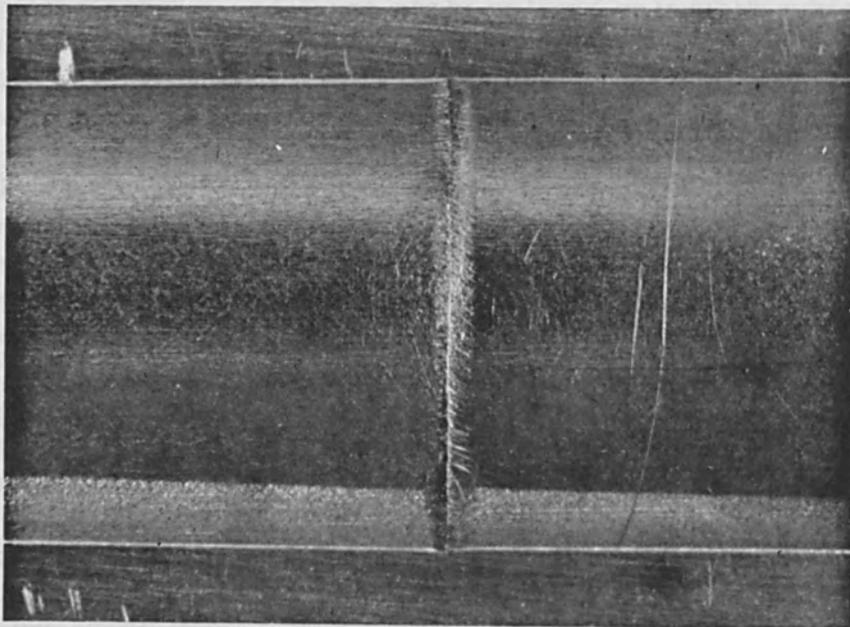
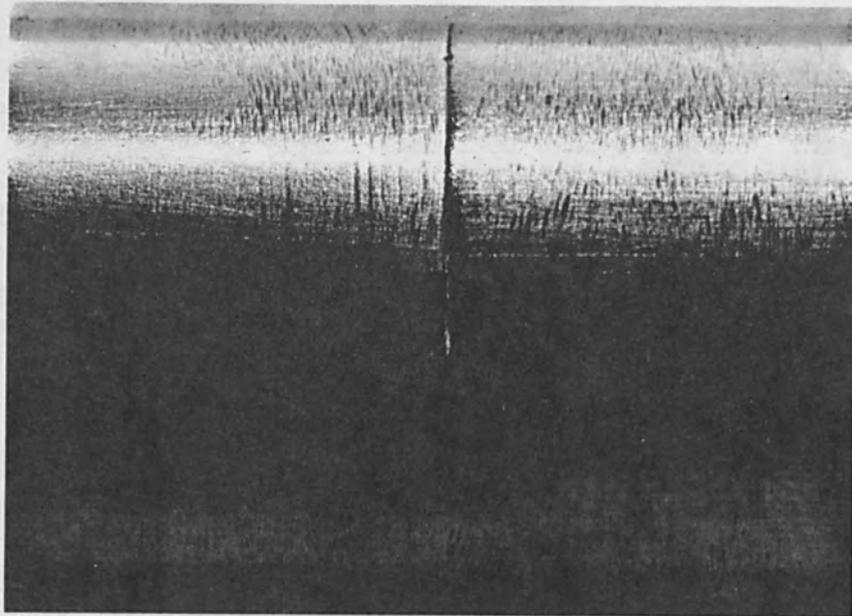


Figure 6. Diffusion Bond of 5/8 in. dia X0.060 in. Wall SS Tube Made with 2.5 in. ID Annular Hollow Cathode.

and other nonconducting refractory ceramics. A promising application is the joining of fused silica (quartz) parts, as illustrated in Fig. 7 which shows two 1 in. dia quartz tubes being fused together in an annular hollow cathode discharge.

D. Brazing

Annular hollow cathodes have been used for brazing at low pressures either in an inert gas or hydrogen atmosphere. For this application the annular hollow cathode again has the advantage of being able to establish a highly localized and well-controlled hot zone. This feature is especially desirable for brazing parts which have sub-assemblies with temperature limits which ordinarily might preclude furnace brazing. In addition to metal brazing, the annular cathode has also been used to braze ceramics to metals. This is a particularly attractive application because the electron beam heats the ceramic as well as the metal. In fact, it has been found that, in most cases, it is preferable to heat the ceramic and let the heat conduct to the metal, thus assuring that the ceramic is at the proper temperature when the braze material starts to flow.

E. Zone Refining

Electron beam zone refining of both metals and insulators can be performed with the annular hollow cathode. The hollow cathode cannot operate under high vacuum conditions; however, because it does operate stably at high ambient pressure levels, say several hundred microns, the partial pressure due to outgassing of the ingot being refined is usually a small fraction of the ambient pressure. Under this condition the outgassing does not have a significant effect on the stability of operation, as it does with conventional thermionic electron beam guns. Furthermore, it is often essential to operate with a partial pressure of a particular gas to prevent chemical change of the ingot. This is especially true with certain metallic oxides which reduce when heated in a vacuum, but which maintain their original stoichiometry when heated in an oxygen atmosphere.

F. Crystal Growing

Single crystal rods of sapphire and rare earth oxides have been grown by the floating zone technique using the apparatus shown in Fig. 8. In this process a polycrystalline rod, pressed from a powder, is held in the top clamp and fused to a dummy rod held in the lower clamp. The dummy may be a seed rod, if desired. However, most of the crystals have been grown without starting from a seed. A floating zone is established in the rod which is traversed downward through the annular hollow cathode; i.e., the zone moves up the rod. Typically, the rod is traversed

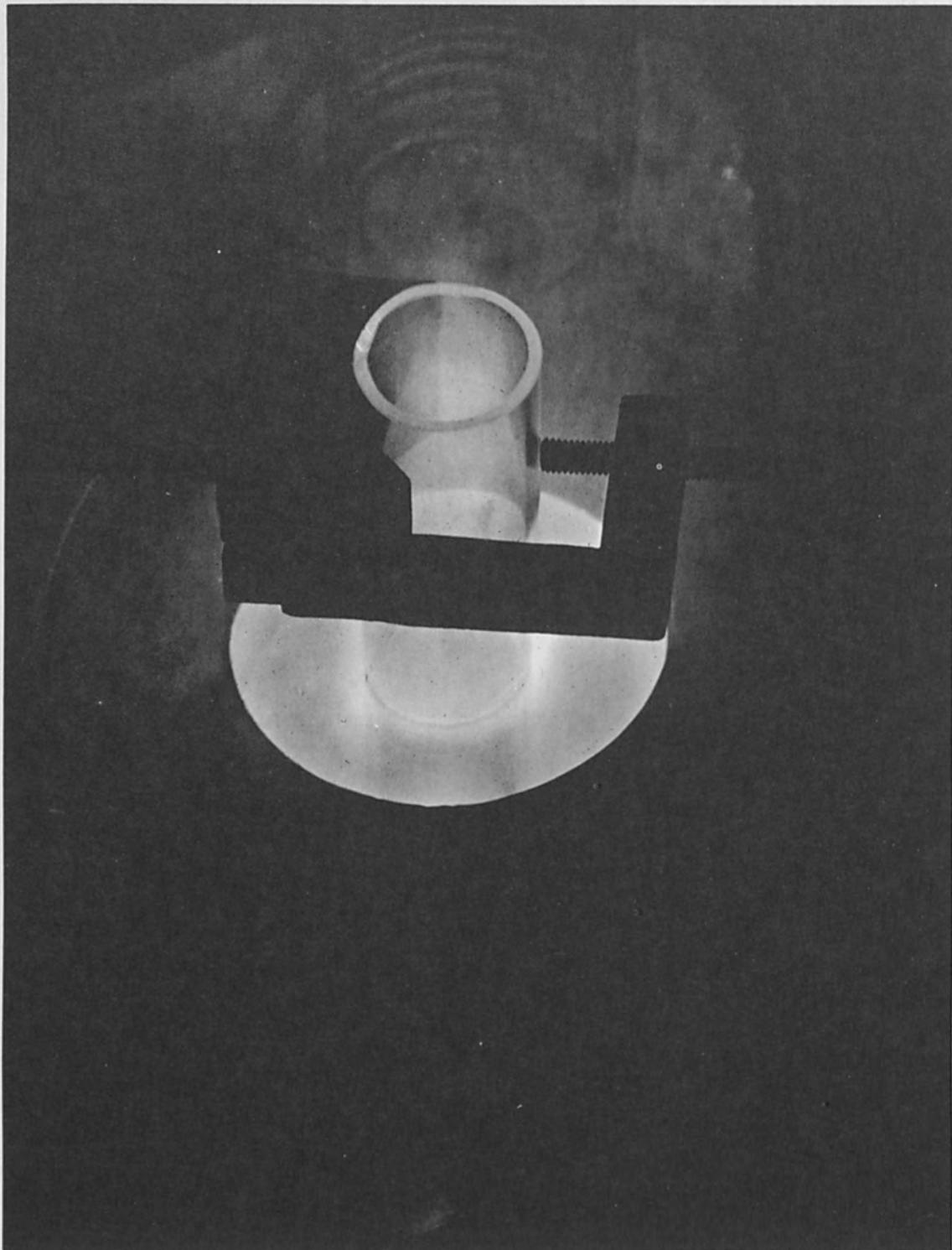


Figure 7. Joining 1 in. Dia Fused Silica Tubes in an Annular Hollow Cathode Discharge.

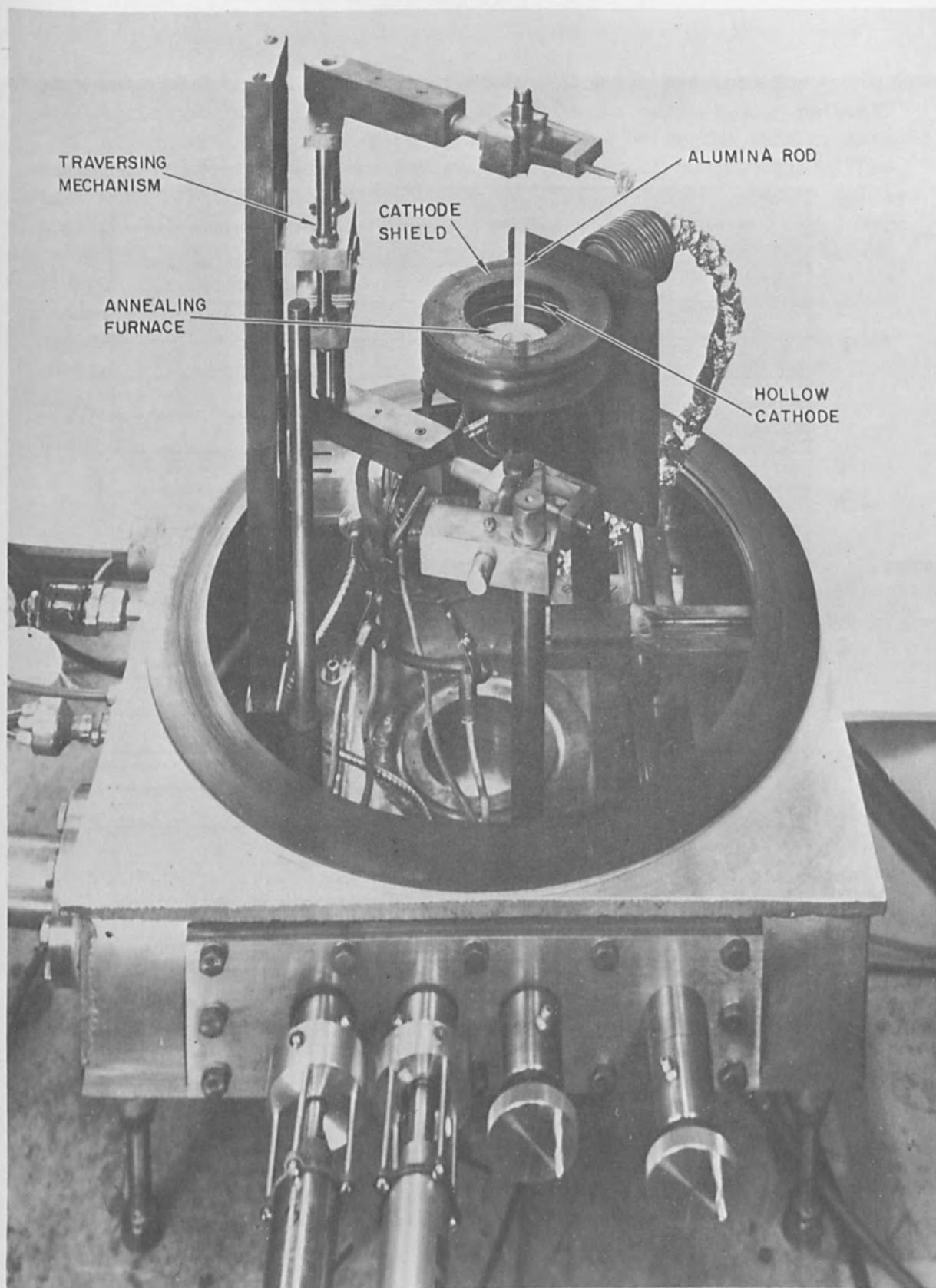


Figure 8. Annular Hollow Cathode Crystal Growing Apparatus.

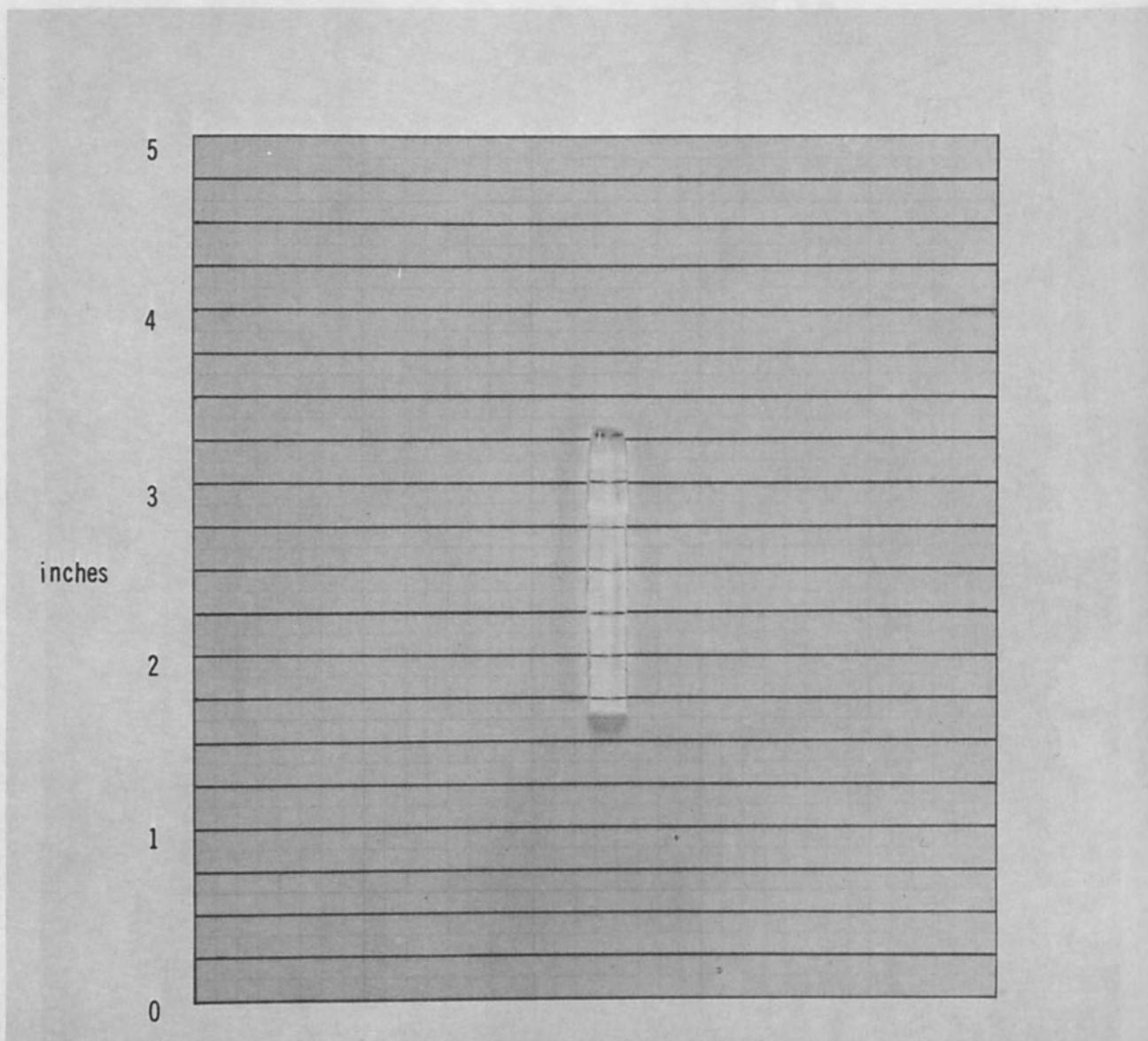


Figure 9. Sapphire Crystal Grown by Annular Hollow Cathode Floating Zone Technique.

at a rate of 2 cm/hr, with approximately 850 watts required for a 1/4 in. dia alumina rod at this speed. The newly formed crystal then passes into an annealing furnace which is located just below the annular cathode. Both the temperature and the design and location of the annealing furnace must be carefully selected to obtain proper annealing rates in the crystal. Crystals can also be grown by floating the zone in a downward direction. Regardless of the direction of traverse the annular hollow cathode has the advantage over conventional electron beam devices in that it operates stably at considerably higher pressure levels. In crystal growing especially, it is often necessary to operate with high partial pressures of certain gases to maintain stoichiometry, not only of the crystal itself, but often that of special dopants.

V. CONCLUSION

This work was sponsored in part by the United Aircraft Research Laboratories and in part by the Hamilton Standard Division.

The authors wish to acknowledge the contribution of Mr. F. J. Ferreira for his original concept and work on the annular hollow cathode.

REFERENCES

1. Druyvesteyn, M. J. and F. M. Penning: Rev. Mod. Phys. 12, 87 (1940).
2. Brewer, A. K. and J. W. Westhaver: J. Appl. Phys., 8, 779 (1937).
3. Von Engel, A.: "Ionized Gases," Second Edition, Oxford University Press (1965).
4. Van Passen, H. L. L. and R. J. Allen: Bull. Am. Phys. Soc. 7, 69 (1962).
5. Boring, K. L. and R. V. Pohl: "Proceedings of the Electron and Laser Beam Symposium," (A. B. El-Kareh, Ed.), p. 319 (1965).
6. Irving, R. R.: "The Iron Age," p. 61 (July 15, 1965).
7. Francis, G.: "Ionization Phenomena in Gases," Butterworths Scientific Publications (1960).

PRELIMINARY EVALUATION OF A NONVACUUM WELDER FOR
AEROSPACE APPLICATIONS

by

J. Lempert, J. F. Lowry, C. S. Williams, and
F. D. Seaman
Westinghouse Electric Corporation
Pittsburgh, Pennsylvania

ABSTRACT

A newly designed 12 KW nonvacuum electron beam welder in which the welding head can be moved horizontally at a speed in excess of 100 IPM is described. Results of a welding study on 2219 Al alloy are presented and a discussion of factors affecting porosity, strength, and depth of penetration is included.

I. INTRODUCTION

In many welding applications it is desirable to move the welding head rather than the work. In still others, it is desirable to weld with the beam shooting horizontally with the work piece vertical and moving in the vertical or horizontal directions. The nonvacuum electron beam welder described in this paper is a modification which takes advantage of the lightweight compact design of a prototype atmospheric welder reported upon in a previous paper⁽¹⁾ to provide the desired flexibility in welding techniques.

In the lightweight prototype unit described earlier, the electron beam was stationary and the work was moved by means of a power driven cart. In addition, the previous unit, like all the nonvacuum EB welders reported upon to date, could be operated only in the vertical position, that is, with the beam directed down and the work moving horizontally.

The equipment reported upon at this time is designed to provide the capability of moving the welding gun with the beam shooting in the horizontal or vertical directions. Conversely, the beam can be kept stationary in the horizontal or vertical position and the work moved either horizontally or vertically.

Electrical and mechanical features of the new unit are described. Performance data on an electron bombardment heated source of electrons obtained utilizing the earlier prototype nonvacuum welder are reviewed. A report on some of the welding results obtained using

the prototype unit for the welding of 2219 Al is also given. Finally, a brief report on a lightweight compact welder which is being designed for a space welding application on a development contract for the Marshall Space Flight Center, Huntsville, Alabama, is included. Such a unit has the potentiality of being adapted for use in conventional vacuum bell jar stations which do not have high voltage lead-ins.

II. MECHANICAL INSTALLATION

Views of the installation which has been built under contract for the Marshall Space Flight Center, Huntsville, Ala., are shown in Figs. 1 and 2. In order to obtain ease of maneuverability, the gun has been designed to fit integrally into the 15 Kva insulated power supply without the use of high voltage cables. X-ray protection has been obtained by mounting the entire installation in a 6' x 10' lead room having a ceiling height of 8'.

The general features of the design were described in the previous publication. A simplified schematic is shown in Fig. 3. Focusing is accomplished by the combination of an electrostatic and electromagnetic optical system. The gun is differentially pumped and designed to permit the section of the gun in which electrons are released to be maintained at a pressure of about 10^{-5} torr, with the exit orifice of the gun at atmospheric pressure. By efficient design of the orifice system, pumping requirements have been held down, permitting the use of the compact diffusion pumps which can be observed in Fig. 1.

The high degree of maneuverability achievable with this gun has been made possible by the combination of an extremely lightweight power supply with the efficient pumping system. The complete welding head including the power supply, the 12 KW electron beam gun, and the diffusion pumps weighs about 250 pounds.

The gun can easily be shifted from the vertical position shown in Fig. 1 to the horizontal position by lowering the entire counter-weighted assembly down the slanting support frame, at the same time rotating the gun and power supply in the appropriate direction. The diffusion pumps are connected to the gun using "O" ring seals which permit holding these pumps in the vertical position as the gun is shifted from vertical to horizontal. Welding conditions can be observed through lead glass windows at the front and side of the unit. Two control consoles, consisting of the overall gun control and the pumping controls, are connected to the installation by means of flexible cables and can be shifted so that control of the welding operations can be exercised from the side or the front of the unit. The gun in either the horizontal or vertical position is capable of a lateral excursion of 4 feet at speeds to 110 inches per minute using the positioner shown in Fig. 1. The flexible hoses which can be seen in Fig. 1 permit maintaining the required connection to the backing pumps as the gun moves from one end

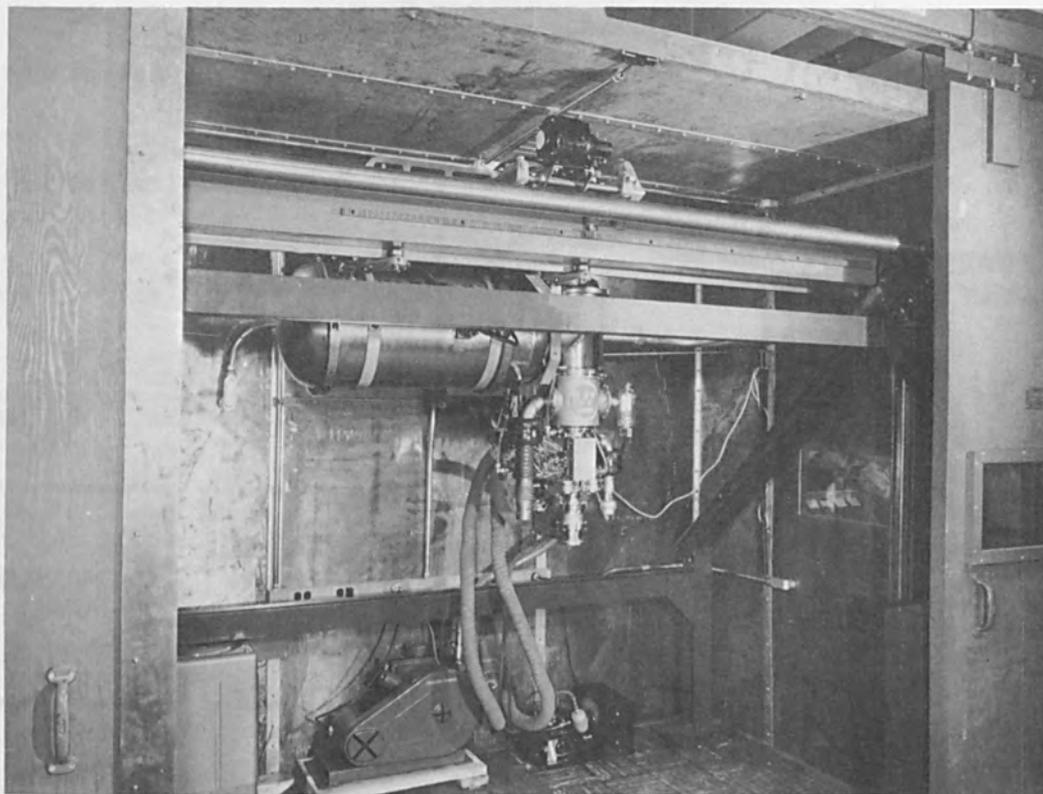


Figure 1. Photograph of New Welder in Vertical Position (Westinghouse Model Designation XWEB 15121).



Figure 2. Side View of Installation for Nonvacuum Welder Showing Control Consoles.

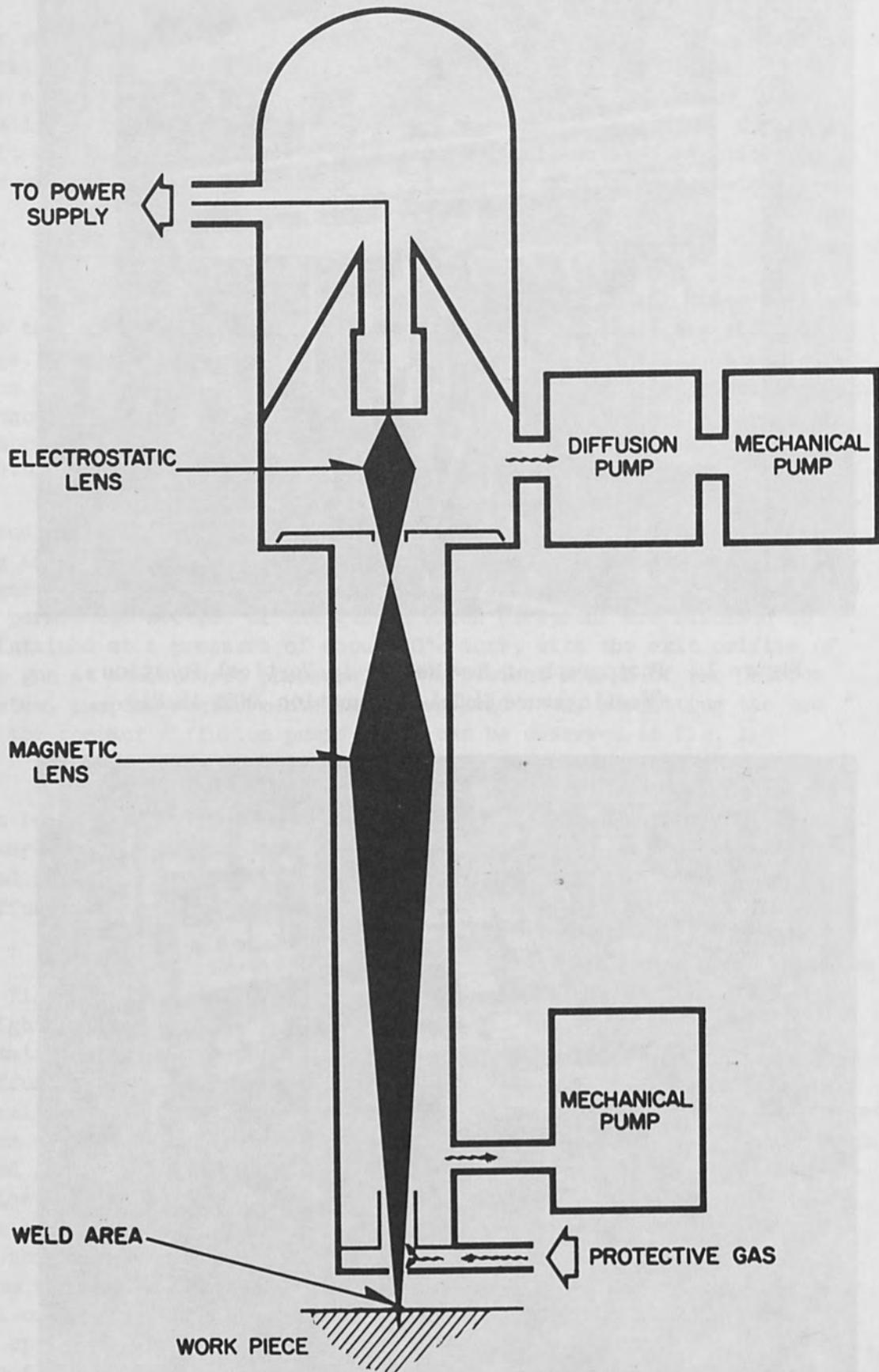


Figure 3. Simplified Schematic Nonvacuum Electron Beam Welder.

of its travel to the other.

A photograph of the prototype unit which was used in the welding investigation of 2219 Al, reported on in a later section, is shown in Fig. 4. A considerable portion of the characteristic data on the electron bombardment heated rod² or "bolt" used as a source of electrons for the nonvacuum electron beam guns was obtained during the course of the welding investigation.

III. ELECTRICAL CHARACTERISTICS

As reported in the previous paper an electron bombardment heated bolt is used as the source of electrons for the gun to take advantage of the longer life, improved electron optics and ruggedness inherent in such a design.

Since it is not necessary to make direct electrical contact with the bolt in order to heat it, its shape and size and the dimensions of its associated electron optical configuration can be optimized from the standpoint of the desired focusing properties and life characteristics. One of the bolt cathode assemblies used in the prototype unit was operated for 210 hours. The mean time between failures (MTBF) for bolt life on the assemblies used on the prototype welder was 79 hours. It is believed that a redesigned bolt assembly which has been developed for the welder shown in Fig. 1 will have substantially better life characteristics than the prototype unit.

Surprising stability of operation can be obtained when the appropriate settings are employed. As will be explained, it has proved possible to vary the gun beam current by simple adjustment of bolt diode voltage in very much the way that a conventional gun is controlled by adjustment of the input voltage to the filament. As in the case of a filamentary structure, electrical characteristics vary with life. Once proper settings on the filament have been employed, it has proved possible to use bolt voltage as the sole variable for controlling beam current when a given value of beam accelerating voltage is impressed on the gun. Just as excessive filament voltage will damage a conventional filament, so an excessive value of bolt diode voltage can overload the bolt.

The electrical characteristics of the bolt diode used in the prototype welder are shown in Fig. 5. At 5.35 amperes on filament and low diode voltages a rise in bolt diode current with the $3/2$ power of bolt voltage is expected, after which the current from the filament to the bolt should saturate when the applied voltage is high enough to cause all of the electrons leaving the filament to be collected. Actually, the consequences of back heating from the bolt to the filament are very much apparent in the two lower curves. When 15 to 20 watts

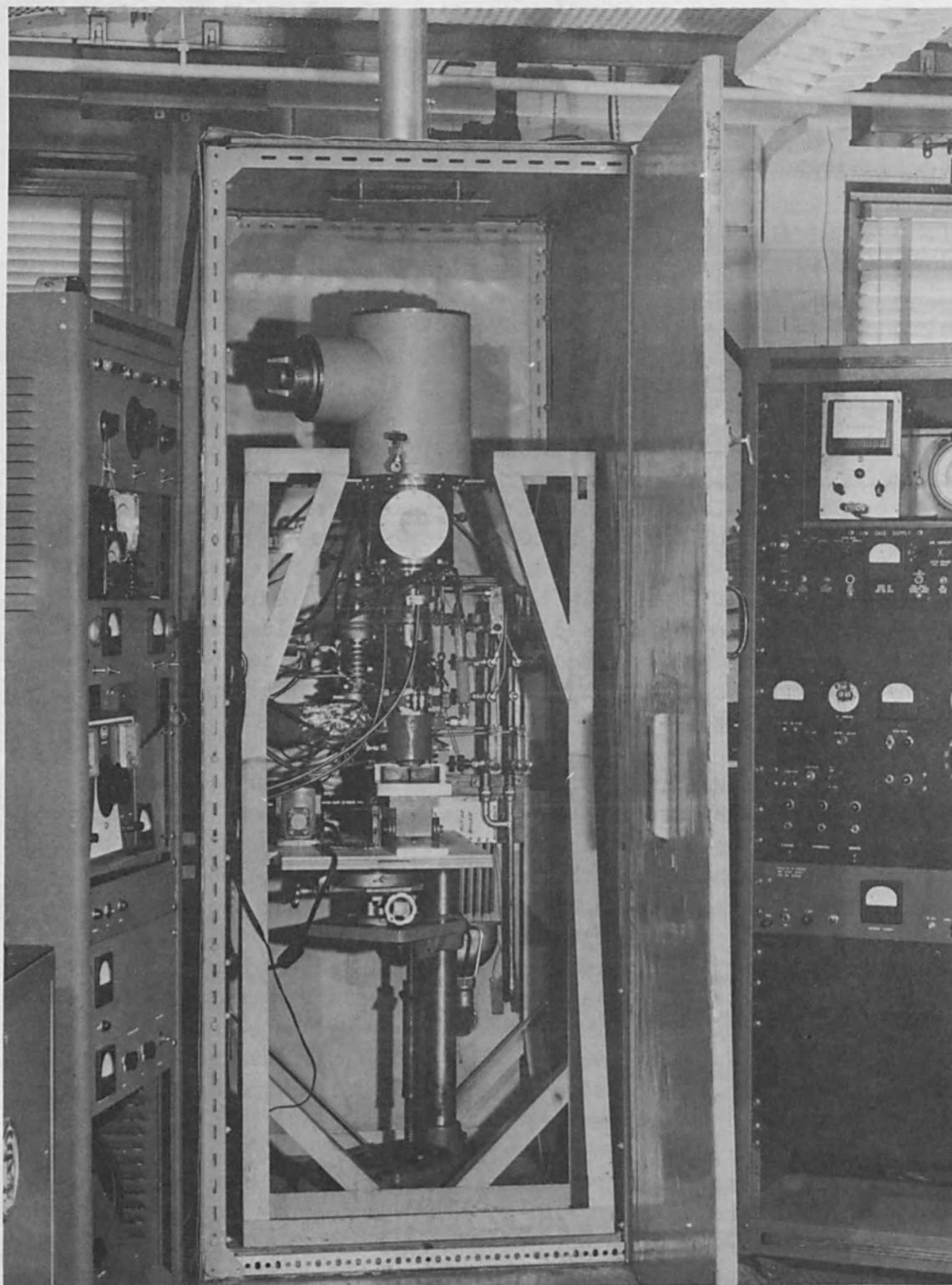


Figure 4. Photograph of Laboratory Prototype Nonvacuum Welder.

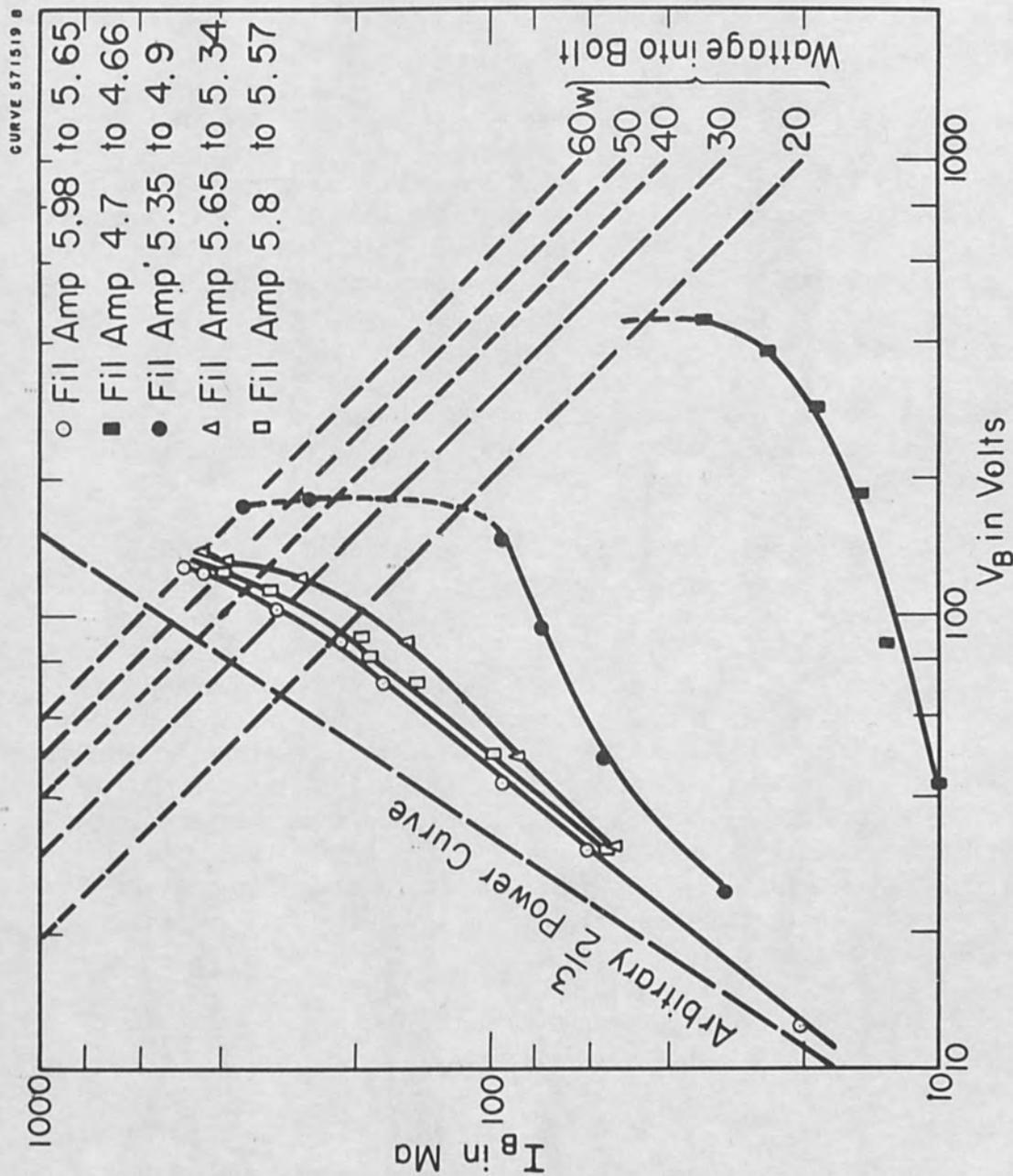


Figure 5. Bolt Diode Characteristics.

are impressed into the bolt, back heating causes the temperature of the filament to rise, resulting in an extremely rapid excursion of emission as the diode voltage of the system is increased. This rise in filament temperature increases its resistance and decreases the filament current to the values indicated in Fig. 5. (The legend gives the initial and final values of filament current under conditions of constant filament voltage.

Stable operation of the bolt assembly is obtained when filament currents between 5.4 and 6.0 amperes are used. When the bolt is operated at or near the space-charge limit, variations in filament current at a given diode voltage produce little change in bolt power. The preferred method for controlling the input power to the bolt is to set filament input so that operation near the space-charge limit is achieved. The desired power to the bolt is obtained by adjusting the bolt diode voltage to produce the required output beam current from the gun. In this way, beam current in the main gun is controlled by a single adjustment in much the way that the magnitude of the beam can be controlled in a conventional filamentary structure by changing the input to the filament.

The principal variations in the operating characteristics of the bolt assembly have been noted during life. Figure 6 shows the tendency of the characteristic curves to shift in the direction of lower perveance with time. Coincident with this shift has been a decrease in both the hot and cold resistance of the filament. These changes are caused by an increase in filament diameter and a decrease in bolt diameter which result from the estimated 150°C higher operating temperature of the bolt. As much as 10% increases in filament diameter have been noted with life. If filament current is held constant, temperature will decrease as filament diameter increases. The low temperature tends to shift the characteristic curve in the noted direction. In addition there is a real change in the perveance of the bolt diode due to the decrease in diameter of the bolt with life.

If filament temperature is to be held constant under condition of varying diameter the following parameters³ should be kept constant:

$$\frac{W_F}{LD}, \frac{I_F}{D^{3/2}}, \text{ and } \frac{V_F D^{1/2}}{L}$$

where W_F is filament wattage, I_F and V_F are filament voltage and current

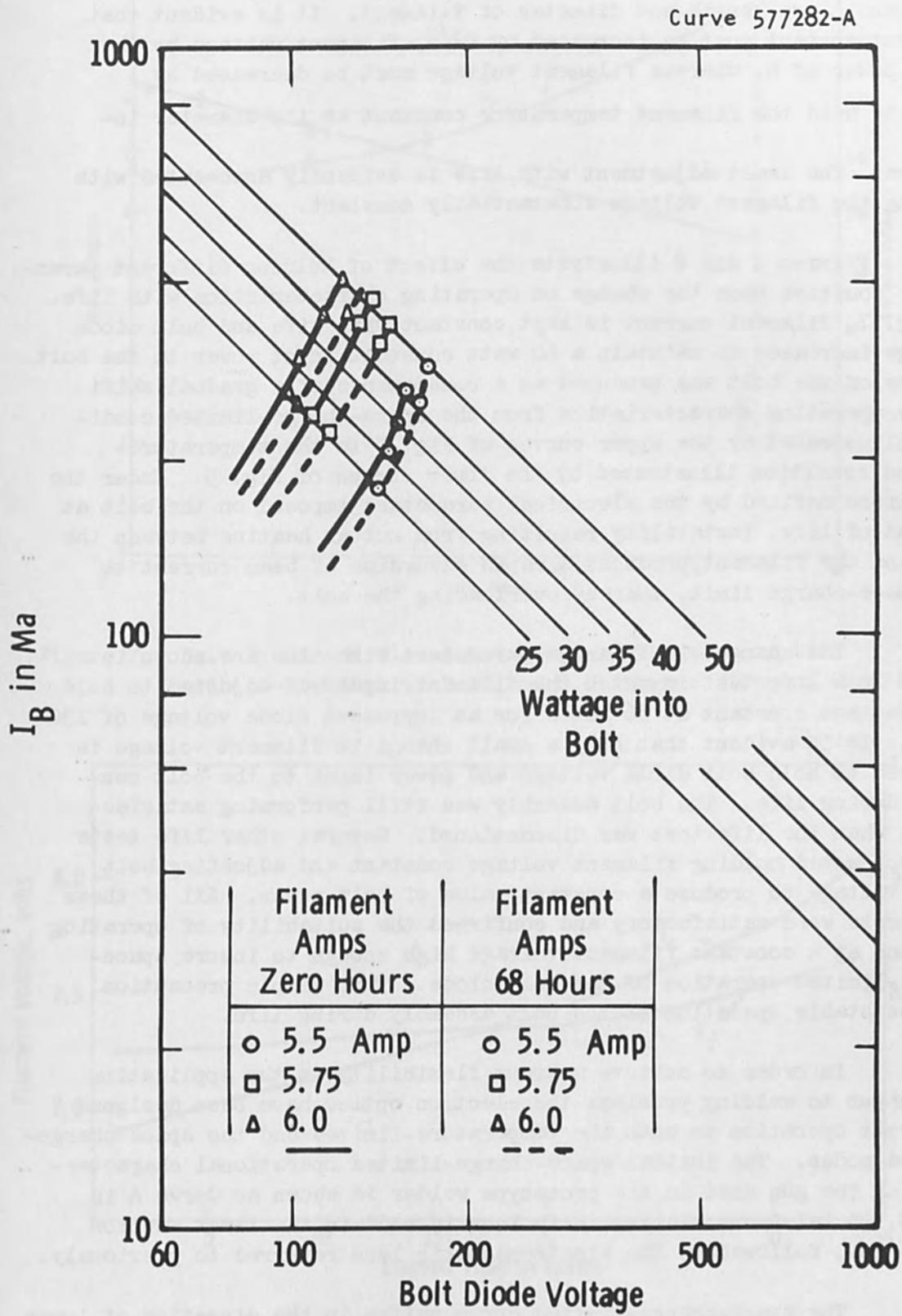


Figure 6. Shift in Bolt Diode Characteristics with 68 Hours Life.

and D and L are length and diameter of filament. It is evident that filament current must be increased by $D^{3/2}$, filament wattage by the first power of D , whereas filament voltage must be decreased by $\frac{1}{D^{1/2}}$ to hold the filament temperature constant as its diameter increases. The least adjustment with life is evidently associated with holding the filament voltage substantially constant.

Figures 7 and 8 illustrate the effect of holding different parameters constant upon the change on operating characteristics with life. In Fig. 7, filament current is kept constant with life and bolt diode voltage increased to maintain a 60 watt constant input power to the bolt. Failure of the bolt was produced as a consequence of a gradual shift in the operating characteristics from the space-charge-limited condition illustrated by the upper curves of Fig. 5 to the temperature-limited condition illustrated by the lower curves of Fig. 5. Under the conditions defined by the electrical parameters imposed on the bolt at the end of life, instability resulting from mutual heating between the bolt and the filament produces a rapid excursion of beam current to the space-charge limit, thereby overloading the bolt.

The changes in filament parameters with time are shown in Fig. 8 in a life test in which the filament input was adjusted to hold bolt wattage constant at 56 watts for an impressed diode voltage of 130 volts. It is evident that only a small change in filament voltage is required to hold bolt diode voltage and power input to the bolt constant during life. The bolt assembly was still performing satisfactorily when the life test was discontinued. Several other life tests were conducted holding filament voltage constant and adjusting bolt diode voltage to produce a constant value of bolt power. All of these life tests were satisfactory and confirmed the suitability of operating the guns at a constant filament voltage high enough to insure space-charge-limited operation of the bolt diode. This simple precaution assures stable operation of the bolt assembly during life.

In order to achieve maximum flexibility in the application of the gun to welding problems the electron optics have been designed to permit operation in both the temperature-limited and the space-charge-limited modes. The initial space-charge-limited operational characteristic of the gun used in the prototype welder is shown as Curve A in Fig. 9. A telefocus electrostatic lens is used in the first section of the gun, followed by the electromagnetic lens referred to previously.

The space-charge-limited curve shifts in the direction of lower perveance with life as indicated by Curves B and C of Fig. 9. The end of life in the practical application of the gun turns out to be the duration of time for which the gun will deliver the required beam power

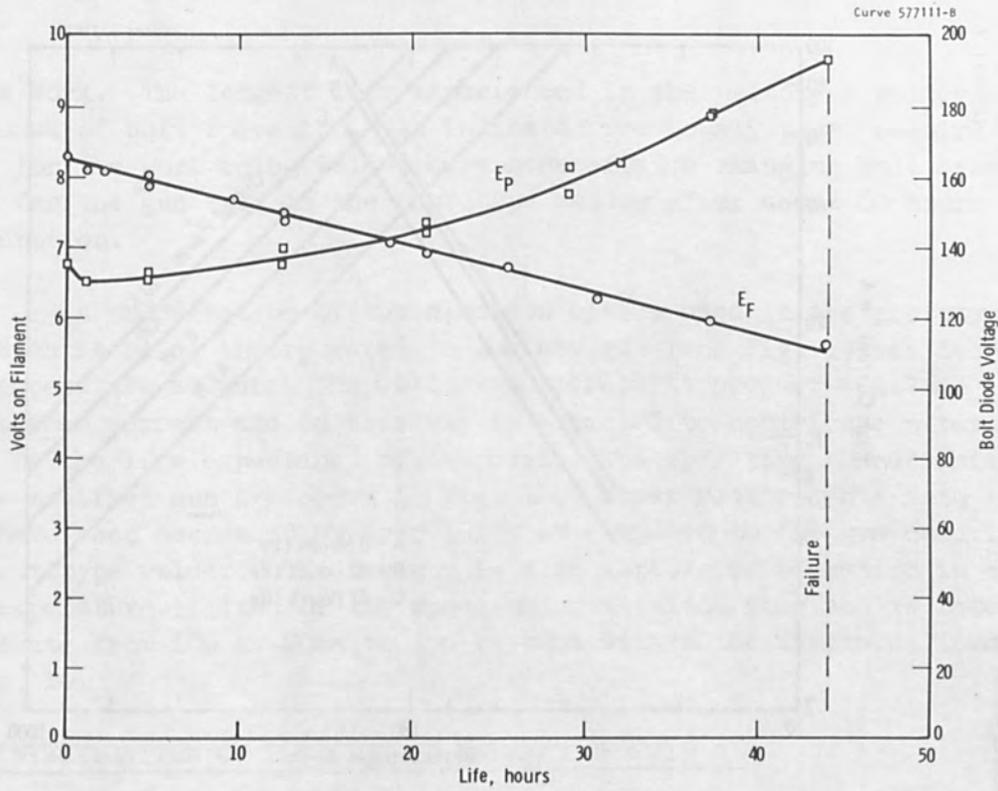


Figure 7. Bolt Diode and Filament Voltage as Function of Life with 56-57 Watts to Bolt and 6.0 Amperes Filament Current.

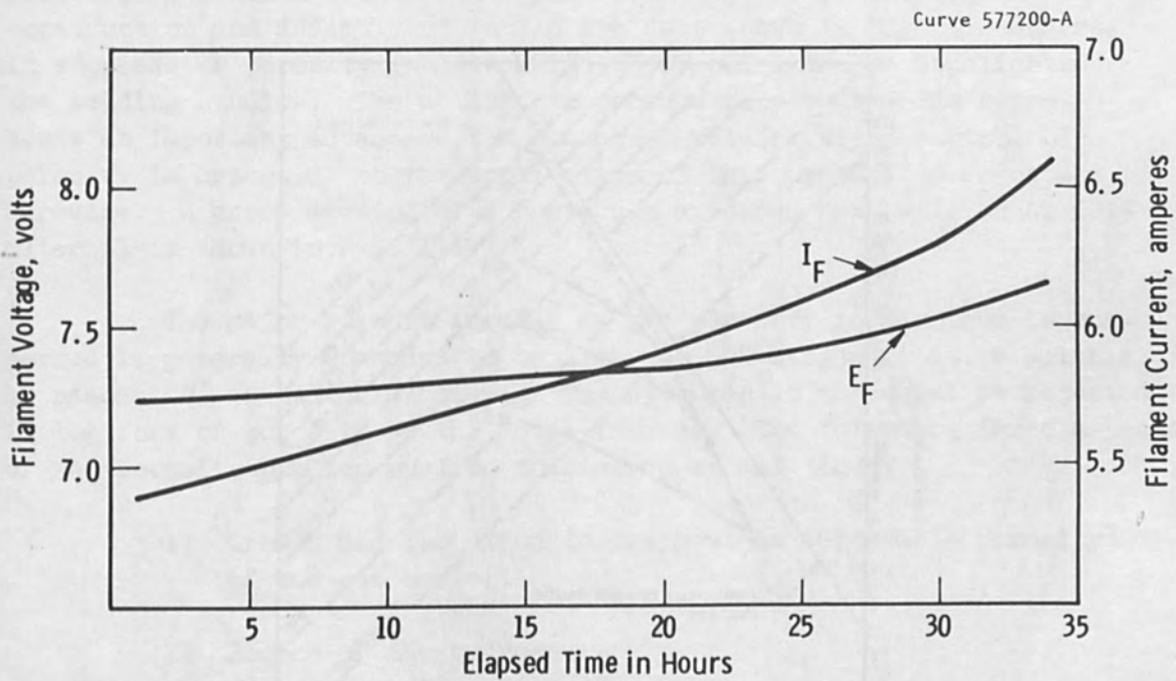


Fig. 8—Change of filament parameters with life, bolt power fixed at 56 watts, 130 volts on bolt

J. Lempert
I. m. l. 3-22-66

Curve 577107-

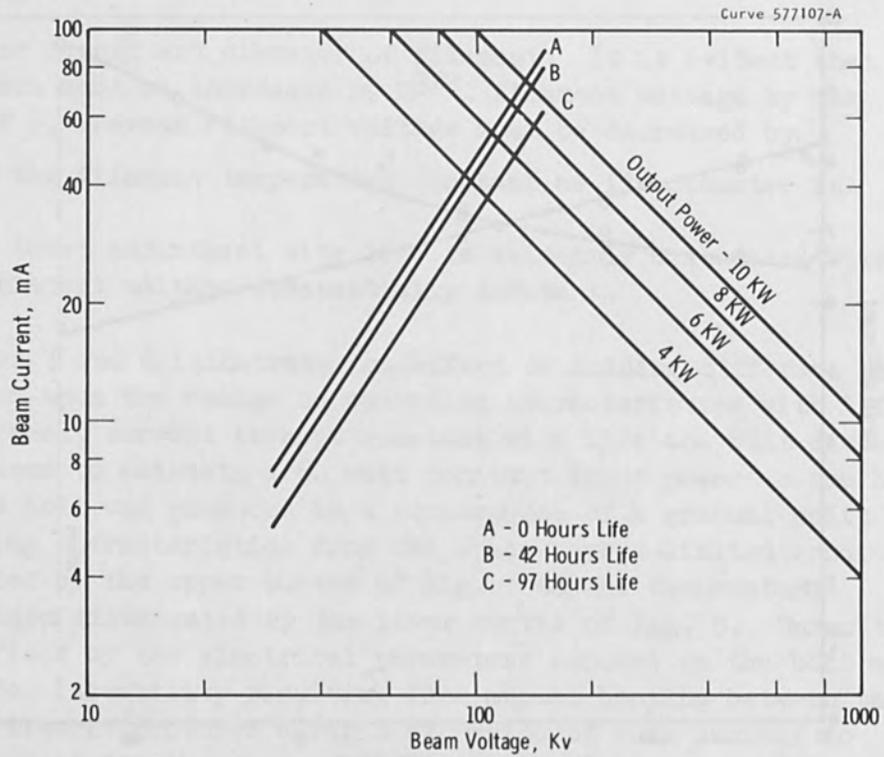


Figure 9. Shift in Beam Current Characteristics with Life, Prototype Laboratory Welder

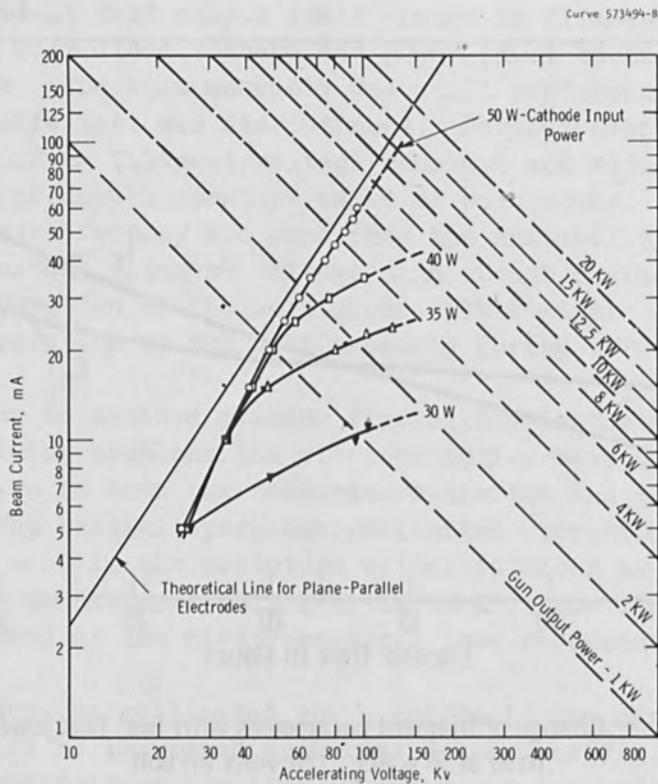


Figure 10. Characteristic Curves for Redesigned Electron Beam Gun.

to the work. The longest life experienced in the prototype welder was 210 hours of bolt operation. As indicated previously power requirements for the work being welded have necessitated changing bolt assemblies for the gun used in the prototype welder after about 80 hours of operation.

A modification of the electron optics used in the prototype gun which is being incorporated in the new gun (see Fig. 1) has decreased the temperature at which the bolt must operate to produce a given output beam current and in this way is expected to contribute materially to the life expectancy of the bolt. The operating characteristics of the modified gun are shown in Fig. 10. Power requirements into the bolt have been decreased by over 10% as compared to the gun used in the prototype welder. The new gun is also capable of operation in either the temperature-limited or the space-charge-limited mode and is rated to operate from 100 kv-20ma to 150 kv-80ma within the limits defined by Fig. 10.

IV. INVESTIGATION OF POROSITY IN WELDING OF 2219 Al

The earlier tests on welding of 2219 aluminum were reported upon in previous publications^{1,9}. The present section covers an investigation of a porosity problem in 2219 Al which has made use of the prototype nonvacuum welder while parallel work was proceeding on the construction and assembly of the 15 Kva unit shown in Fig. 1. Success in suppressing porosity in 1/4" 2219 - T 87 aluminum has highlighted the welding studies. The ability to produce nonporous welds represents an important advantage for nonvacuum welding since control of porosity is essential to the application of this process to aerospace hardware. A cross section of a nonvacuum electron beam weld on Al 2219 material is shown in Fig. 11.

The major culprit insofar as gas porosity in aluminum is concerned is generally conceded to be hydrogen. Hydrogen is quite soluble as nascent H^+ in molten aluminum. The hydrogen in the metal is rejected in the form of porosity as the metal freezes. The following three aspects of the porosity problem must be considered at all times:

- 1) Growth and flotation (entrapment as detectable porosity) of the gas nuclei.
- 2) Source of the hydrogen.
- 3) Mechanisms that govern the entrance of the hydrogen into the melt and its rejection from the metal in the form of gas nuclei.

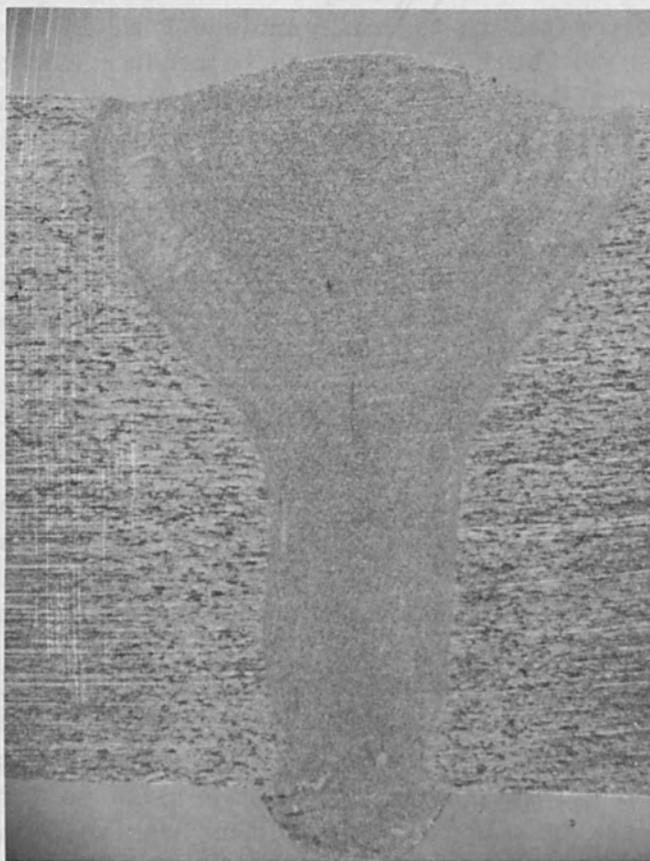


Figure 11. Nonvacuum Electron Beam Weld in 1/4" 2219 T87 Aluminum
Produced at 7 KW 120 Inches per Minute.

For any given process, #1 and #2 can be controlled. The following sections cover the highlights of the efforts to reduce porosity in categories 1 and 2. The work was performed on NAS8-11929 sponsored by the Welding Development Branch of NASA (Marshall Space Flight Center).

A. Utilization of Gas Nuclei Growth and Flotation to Reduce Detectable Porosity

As an initial probe, Westinghouse welding engineers examined a large number of bead-through test plates that had been made with only provisional auxiliary shielding and with questionable cleaning practice. The examination was concluded by measuring and counting pores in 2-2T radiographs*. These test plates did, however, represent a complete factorial experiment involving power level, speed, lens current (focal distance) and working distance. Power level was observed to have a marked effect on the amount of porosity observed. It was further observed if the process was slowed down even as little as 15% for any given power level less porosity could be detected. This latter effect was most noticeable at high power levels and in fact was shown to constitute a statistically significant two factor interaction involving power and speed.

Comparison of Figs. 12 and 13 illustrates the dramatic reduction in porosity achieved when the power level was dropped from 6 KW and the maximum possible speed of 60-70 IPM for which penetration through weld could be obtained (defined here as the bench mark speed) to 3.6 KW (bench mark speed 25 IPM). Recognizing that marginal shielding and cleaning conditions were in effect in both welding tests, it is apparent that the lower power procedure has a very great tolerance for cleaning and shielding. Undoubtedly the longer period available for pore nucleation and flotation explains this tolerance.

The same films also show that the use of high lens current (i.e., short focal distances which tend to produce "V" shaped welds) results in slightly improved weld quality (Note freedom from porosity in weld at top of Fig. 13). Freezing rate and direction, as they are affected by the shape of the penetrating beam, appear to be the cause of this observed reduction in porosity.

* 2-2T refers to a level of penetrometer sensitivity at which 2T hole is visible in a penetrometer representing 2% of the specimen thickness.

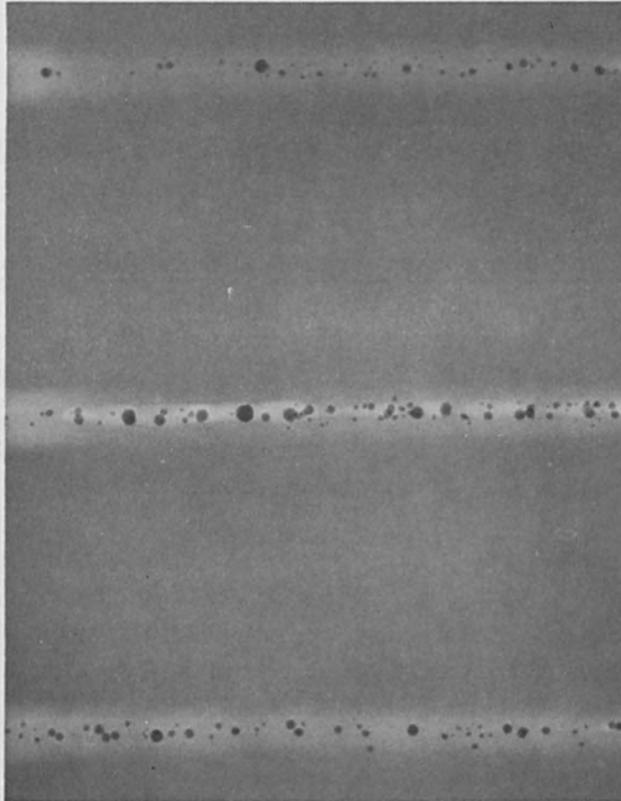


Figure 12. Radiograph Showing Severe Porosity of Early Specimens Welded with Nonoptimum Shielding and Cleaning, Using High Speed (60-70 IPM), High Power (6 KW) Procedure
Top-to-Bottom; Short, Long, Medium Focus.

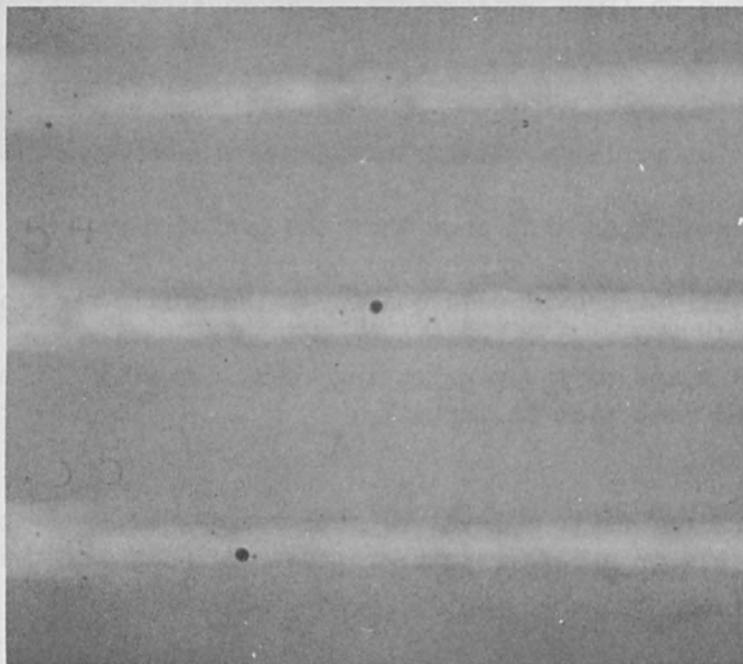


Figure 13. Moderate Porosity which Derives from 3.6 KW Procedure in Spite of Nonoptimized Shielding and Cleaning Practices
Top-to-Bottom; Short, Long, Medium Focus.

In spite of the success of the low power procedure, and the additional observation that a minimum cross section was maintained at the low power, a need to explore higher speed welds led to a study of means for eliminating the sources of hydrogen.

B. A Study of Means for Reducing Porosity through Elimination of Sources of Hydrogen in the Process

The initial objective was to achieve substantially pore-free welds with a 6 KW process (bench mark 60-70 IPM). The base metal was analyzed by vacuum fusion methods and found to contain less than .23 ppm of hydrogen. This small amount of hydrogen would not be expected to produce the porosity shown in Fig. 12. The surfaces of the metal and the atmosphere around the weld were considered to be the most likely sources of H^+ (probably in the form of H_2O).

The shields used for all tests described previously (experimentally identified as Series A) were constructed as an accessory to the prototype welder. Under these circumstances sealing around the nozzle of the welder was not very effective. The shields overheated at their midpoint. This destroyed the seal by causing a lengthwise bow. In addition to the bowing problem, severe oxidation of the diffuser material (steel wool and metal screen) necessitated frequent changes of the shield so that the experimental variance was high in Series A.

A later series, B, was primarily concerned with porosity as a function of shield configuration. The nozzle was therefore redesigned to accept mounting plates for the various shield test configurations. Effectively it served as a "test bed" which could be tightly sealed to the welder. This test bed was water-cooled to assure its dimensional stability during welding tests. Shield configurations that provided a simple, easily studied gas curtain around the beam were built up from modular components. These were fastened to the test bed with solder. Thus a modification in length or width could be accomplished in a few hours. A bottom view of an outer perimeter and filler - gas manifold soldered on the cooled mounting plate is shown in Fig. 14.

The following shield configuration parameters were evaluated:

- Length of Shield: 2 levels
- Width of Shield: 2 levels
- Volume of chambers (gallery) into which gas was introduced: 2 levels
- Level of gas flow (for one type of gas): 2 levels

The total number of tests was reduced to 32 through the use of a $1/4$ factorial design.

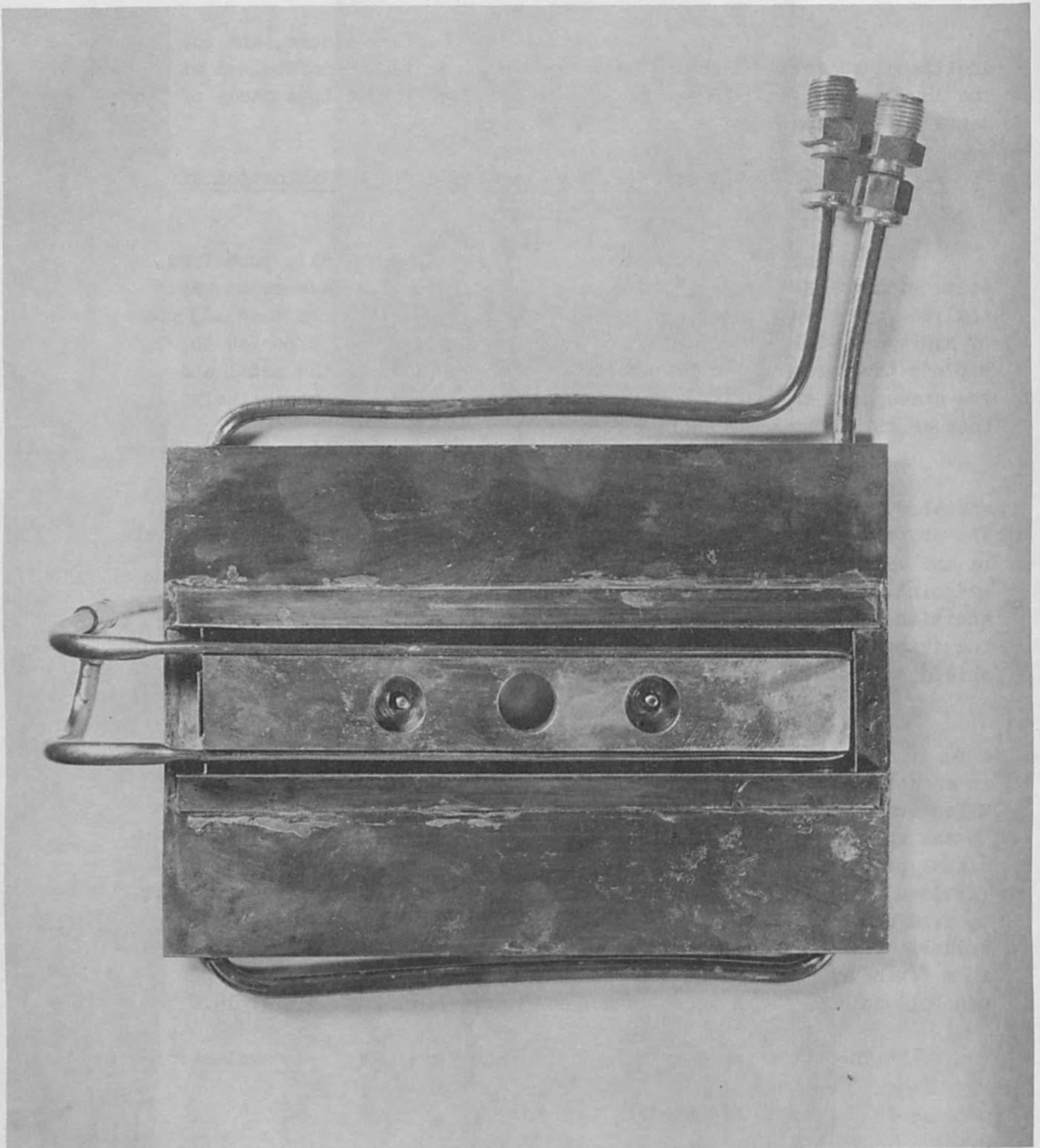


Figure 14. Assembled Gas Curtain Test Shield (Bottom View).

The improvement resulting from bringing the parallel inert gas manifolds together toward the center line is qualitatively illustrated in Fig. 15. This welding procedure is comparable to that used in producing the welds of Fig. 12. The improvement resulting when the manifolds were moved together suggests a thermal turbulence in the volume of gas surrounding the beam and weld. When the manifolds were close together the parallel streams of inert gas were able to converge ahead and behind the weld so that all gas drawn by convection into the weld zone was furnished by these streams.

Placing a diffuser between the manifolds (i.e., in the corridor between the manifolds) should help assure a supply of inert gas directly ahead of the process. The results obtained from the use of a diffuser in the corridor between the gas manifolds, particularly at the 85% bench mark speed, is illustrated in Fig. 16. The cleaning procedure (mechanical) used on this plate, consisting of the removal of less than .001 inch total from both sides, is not considered adequate in the light of later data. Alcohol was used after cleaning, a practice which has since been discontinued.

At this point in the optimization effort, blanketing of the weld area apparently was rather effective because the substitution of argon (with its high cross section for electron scattering) for helium severely altered the penetrating qualities of the process. Thus it was felt that the shield was working and the search for sources of porosity was switched to the surfaces of the work pieces.

Chemical cleaning did not decrease the volume of porosity but appeared to change the nature of the porosity by increasing its size (Fig. 17). The total result was an apparent decrease in the number of pores. It may be that the test plate was not adequately dried after the initial chemical cleaning so that succeeding welds on the same plate may have been somewhat less exposed to moisture. The alcohol rinse which had been used on the earlier mechanical cleaning was discontinued on the chemically cleaned welds. This constitutes a change from the original mechanical cleaning procedure and should reduce sources of H^+ and tends to produce a more favorable picture of chemical cleaning than might be the actual case.

Underbead shielding added to either cleaning method improved the appearance of the underbead and produced the welds shown in Fig. 18. These are among the more acceptable welds (except for the high speed weld) even though their numerical rating was not significantly different than those observed for welds without underbead shielding. Underbead protection (20 cfh of helium) also reduced or eliminated an "abrasive" surface condition that had been observed on some underbeads. This condition was sometimes linked with undue porosity in slow welds.

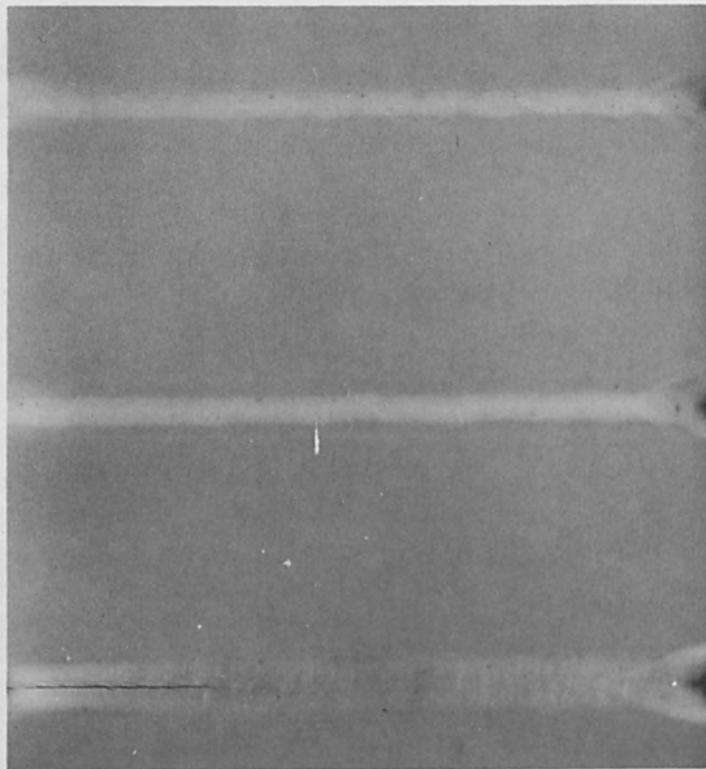


Figure 15. Optimum Shield Configuration (Closely Spaced Gas Manifolds)
Reduced Porosity from 6 KW Welding Procedure to Feasible Level
Top-to-Bottom; 60 IPM (Bench Mark); 50 IPM; 40 IPM

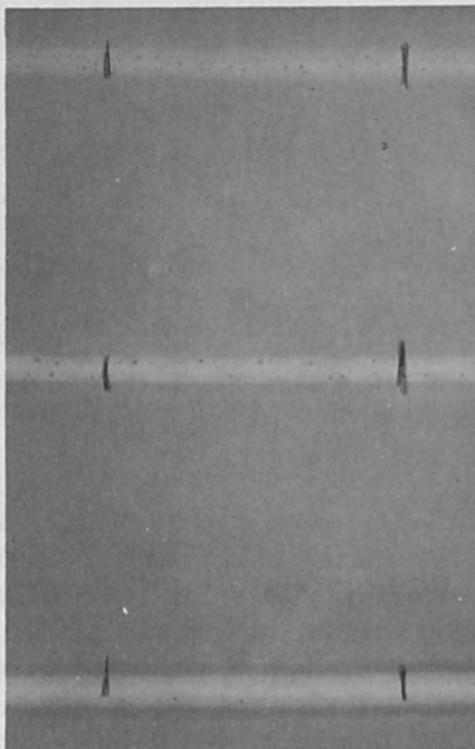


Figure 16. Porosity Level Obtained Using Shield which Incorporates Both
Diffuser and an Optimum Gas Manifold Configuration
Top-to-Bottom; 60 IPM (BM); 50 IPM; 40 IPM.

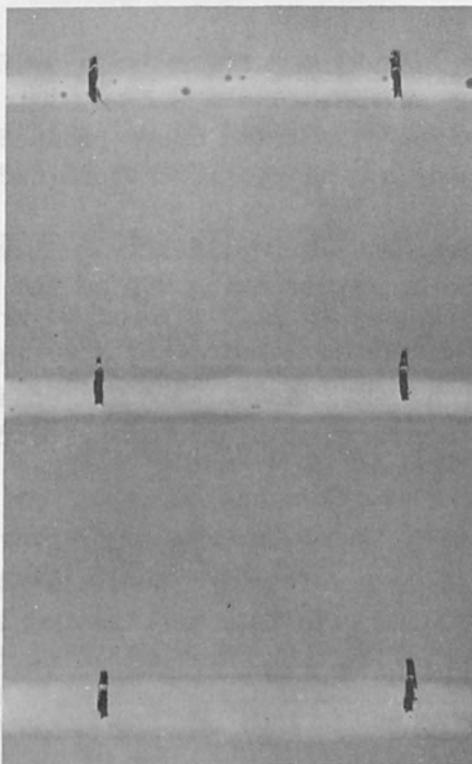


Figure 17. Porosity Level Obtained using Optimized Shield, Diffuser, and Chemical (Instead of Mechanical) Cleaning Plus and Underbead Shield
Top-to-Bottom; 60 IPM (Bench Mark); 50 IPM; 40 IPM.

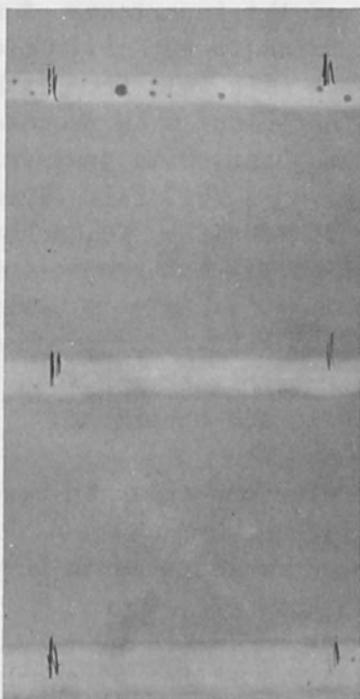


Figure 18. Porosity Level Obtained using Optimized Shield, Diffuser, Mechanical Cleaning and Underbead Shielding
Top-to-Bottom; 60 IPM (Bench Mark); 50 IPM; 40 IPM.

C. Conclusions

- (1) Several of the remedies tried above produced welds that could be accepted under actual specifications such as ABMA-PD-R-27A.
- (2) The greatest degree of acceptability achieved in 6 KW welds was made at speeds representing 85% of the bench marks speed (about 50-60 IPM).
- (3) Full penetration welds made at 3.6 KW tended to be less critical as to porosity than higher power processes.

D. Future Work

At the present time sufficient power is available to permit welding speeds above 140 IPM with full penetration in 1/4 inch aluminum alloy plate. Further prosecution of the shielding technique studies will be carried out to accommodate these unusual but highly desirable, welding conditions.

V. IMPROVED TENSILE STRENGTHS IN 2219 ALUMINUM

The first tensile tests, reported¹ last year, indicated a strength of 41 Ksi for single-side, single-pass welds in 1/4 inch 2219 aluminum, 56 Ksi ultimate. Improvements in bead contour, resulting from an increased understanding of the importance of such procedure detail as focus, have raised this value to 43 Ksi. This latter strength represents the strength of a bench mark weld from either the highly tolerant 3.6 KW process or the more efficient 6 KW process. Welds produced by fusing from each side of the joint, with just sufficient energy to overlap at the center, exhibited 45 - 45.7 Ksi. These weld tests were carried out at welding speeds of 120 IPM. Welds have also been produced from one side at 120 IPM (Fig. 11) but projections of the time-temperature (or speed-width) relationship as it affects strength indicated that it would be more profitable to test welds produced at or above 200 IPM. Such tests will be carried out this summer when the 12 KW power of the new welder can be used to weld the specimens.

The tentative conclusions that can be drawn from the work to date are:

- (1) Welding procedure refinements have raised strengths from 41 Ksi to 45 Ksi.
- (2) The potential for higher strengths lies in the extrapolation of speed-width, time-temperature relationships. This can be accomplished by:
 - (a) Utilizing higher power densities, for example, as

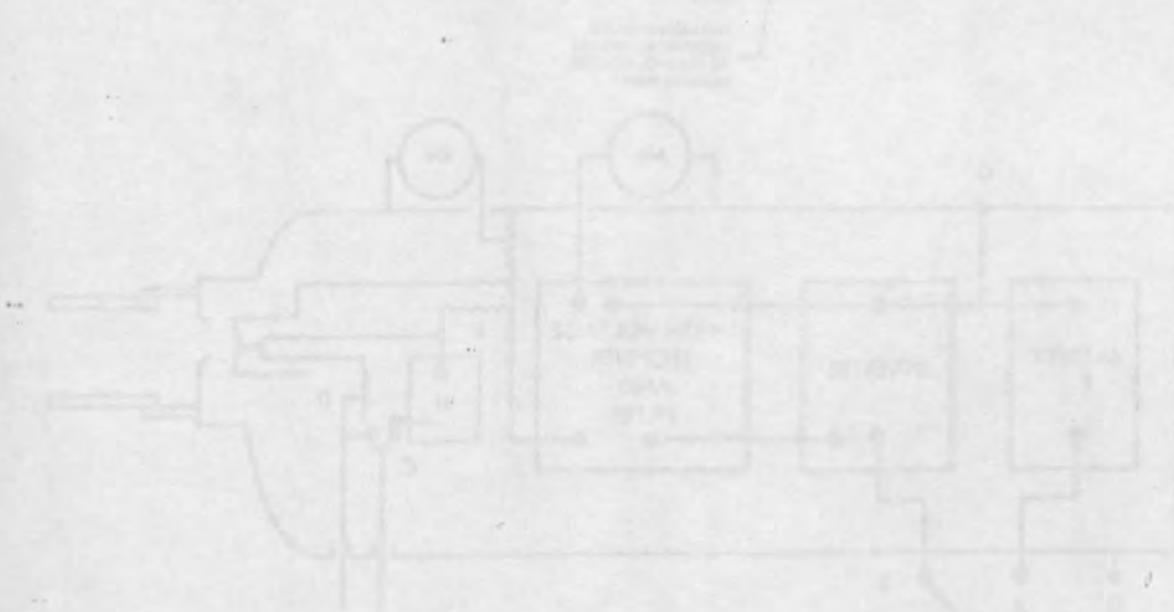
by decreasing the gun to work distance.

- (b) Utilization of higher power equipment at higher welding speeds.

VI. SELF-CONTAINED ELECTRON BEAM WELDER FOR SPACE APPLICATION

The capability of transmitting an electron beam directly into the external environment is a basic requirement for both the nonvacuum welder and the recently initiated space welder. A welder for use in space, however, has the advantage that the electron accelerating and focusing region of the gun does not have to be protected against the environment into which it transmits the beam as is the case with the nonvacuum welder. Portability, low weight, and compactness are important in both applications.

A program has been undertaken to develop a self-contained electron beam welding gun under contract with the Marshall Space Flight Center, Welding Development Section, Huntsville, Alabama. The general concept of the proposed equipment is illustrated in Fig. 19. Design objectives for the gun are a total weight including batteries of less than 75 pounds and a beam power output of 2 KW at 20 KV.



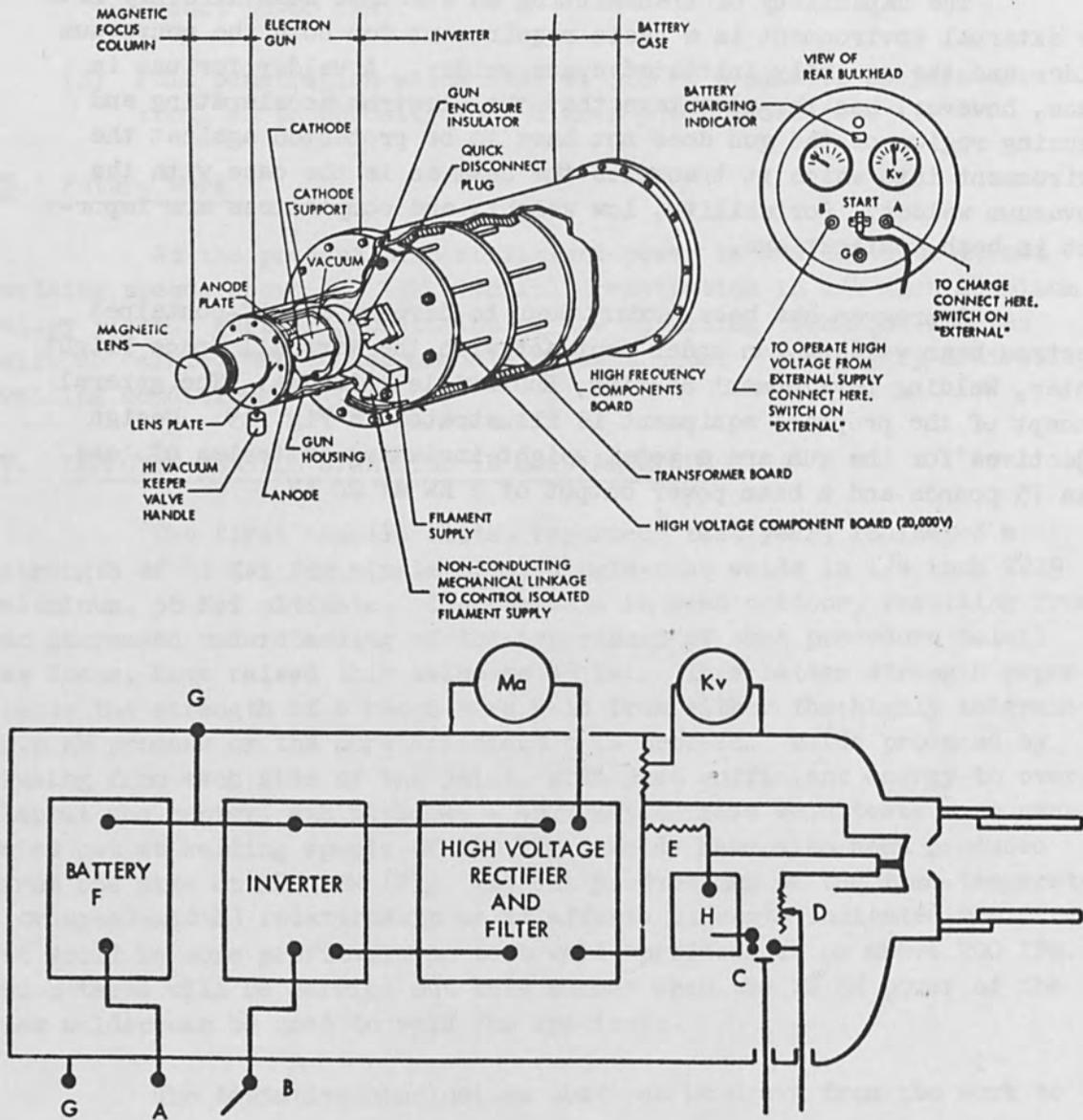
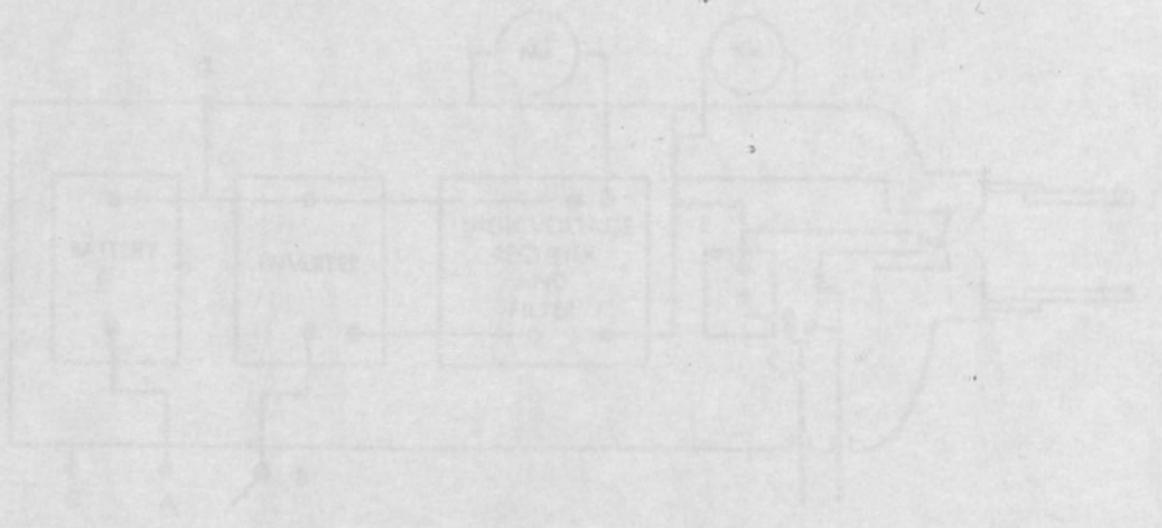
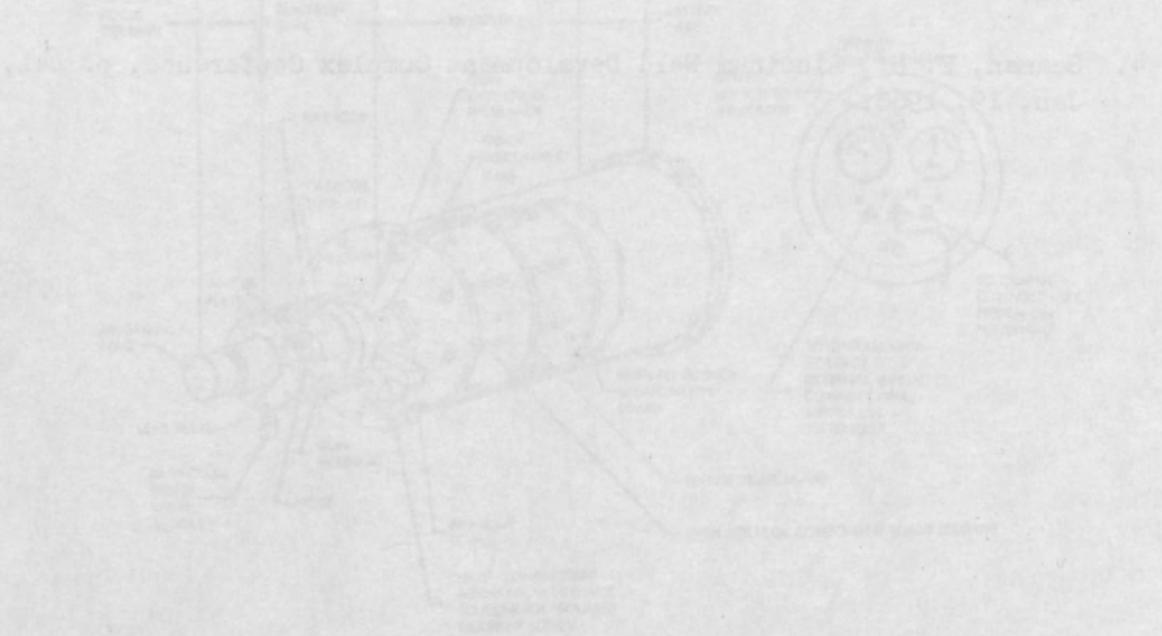


Figure 19. Concept of Portable Self-Contained Welder for Space Applications.

REFERENCES

1. Lempert, J., Lowry, J. F., Seaman, F. D., and Williams, C. S.
Proc. of the 7th Annual Electron and Laser Beam Symposium, March
31 - April 2, p. 393, 1965.
2. Bas, E. B., Zeit. für Angew. Phys., 7, p. 337, 1955.
3. Jones, H. A., and Langmuir, I., Gen. Electric Review, 30, p. 310,
1927.
4. Seaman, F. D., Aluminum Weld Development Complex Conference, p. 241,
Jan. 19, 1966.

1. Langner, J., Lee, J. L., Sargent, J. D., and Wilson, C. L. *Proc. of the Annual Meeting of the American Physical Society, 1951*, p. 100.
2. Lee, J. L., Sargent, J. D., and Wilson, C. L. *Phys. Rev.*, **73**, 100 (1951).
3. Jones, B. A., and Langner, J. D. *Phys. Rev.*, **73**, 100 (1951).



TRANSITION OF ELECTRON BEAM WELDING
FROM HARD VACUUM TO SOFT VACUUM

by

H. A. James and J. L. Solomon
Sciaky Bros., Inc.,
Chicago, Illinois

ABSTRACT

An extensive series of investigations into the effects of higher pressures (lower vacuums) than 10^{-4} Torr in the electron-beam welding chamber has just recently been concluded.

As a result of this work it can be shown that the same welding results which characteristically are obtainable at 10^{-4} Torr can also be obtained at pressures as high as one thousand times 1×10^{-4} Torr and even higher.

One of the more direct effects of this new medium-vacuum welding capability will be the increased use of electron-beam welding processes for high-production-rate applications, both in ferrous and nonferrous materials.

I. INTRODUCTION

A breakthrough, in any technology, is a seldom and long coming phenomena. And, when applied to a technology that was in itself newly created by Dr. J. A. Stohr in 1956, it is even more astounding. The breakthrough in the case of welding with the electron beam in soft vacuum is really only a backward step, which takes advantage of circumstances that were passed by while the industry hurried toward the use of the electron beam for welding at atmospheric pressure.

II. SOFT-VACUUM WELDING CONSIDERATIONS

That which we admired in the beginning, that is, the welding within a vacuum environment, became more and more of a handicap as the practical aspects of welding by this process became more apparent. The first and most obvious extension of thought was to move the process out of the hard vacuum and into the atmosphere. This has been accomplished with some success and with some limitations. With this solution, the need for a vacuum environment for the gun was not eliminated. It was still necessary to provide a hard vacuum system for the electron beam gun

wherein a beam of electrons could be caused to originate and to be initially accelerated. This obvious, but difficult, solution would then place the work at atmospheric pressure. This raised new problems in that it was necessary to use extremely high accelerating voltages in order to push the electron stream through the dense atmospheric gases. At the same time, the point of impact of the beam was moved out of its protective vacuum chamber and along with it the source of high energy x-ray radiation was moved out into the atmosphere. Further, the narrow weld geometry which was produced within the vacuum environment deteriorated due to the scattering effect on the electron beam due to atmospheric gases at ambient pressure. The pressure range between hard vacuum (.1 micron) and one atmosphere (760,000 microns) was ignored until, upon taking a careful second look, it was found that a very useful situation existed which had always been with us and available to us. The significance of this can be understood by first considering the following:

Inherently, the electron beam must originate and be accelerated within a reasonably "hard" vacuum for the following reasons:

Initial acceleration of the electrons must be accomplished within a vacuum environment so that the beam geometry remains as compact as possible without dispersion within the gun. The source of the electrons in an electron beam gun is the filament, which is operated at approximately 2,000°C. The prime mover of these electrons is the accelerating voltage impressed between the filament and the anode. This voltage may be from 10,000 to 60,000 volts in some guns. These elements must be physically separated by a small distance, in order to prevent discharges of the high potential between the elements. A less obvious requirement of the electron beam gun is the integrity of the space between these elements, as this space manifests itself by determining the internal conductance or perveance of the gun and which, in turn, determines the current flow through the gun. One can assume that there is always some quantity of gas in this interelectrode space at any vacuum level; nevertheless, it has been shown that this gas must be maintained at a reasonable minimum. If the gas population increases, a point will be reached for a specific electrode spacing that could cause electrical breakdown, increased beam current, and oxidation of the white-hot filament. For this reason, we are faced with the necessity of always maintaining the gun; i.e., its electrodes and filament, at a vacuum of .1 micron, or better.

On the other hand, this says nothing about the vacuum requirements surrounding the piece-part being welded. There the requirements are twofold: the weld size must be maintained small, and the metallurgical integrity must be maintained (not improved), or at least suit the service and design level desired. We may, therefore, look for the resolution of the purity problem to a field where the answer already exists. Certainly, years of arc welding experience have taught that the purity level available in commercial purity inert gas is adequate for most applications, and this purity level may be higher than is required for a great

many applications. In fact, it might be desirable to have a process in which we can choose the purity level of the environment. Electron beam welding in soft vacuum provides this in that we may make use of an impurity level extending from 10 ppm to 1,000 ppm, at will.

The problem of keeping the weld small is, of course, related to the problem of keeping the beam concentrated, as was pointed out by Dr. Stohr in his original work. As a criterion, we may look at welds produced under high vacuum conditions (.1 micron) and then affirm that any departure from welding at high vacuum is only desirable as long as we can maintain the characteristics of electron beam welds as produced at high vacuum; otherwise, the cost of the process is not justified.

At one (1) atmosphere, ambient gases exert a pressure of 14.7 psi., which corresponds to 760 mm Hg, or 760,000 microns Hg. The vacuum level traditionally used for electron beam welding is .1 micron, or less. .1 micron corresponds to an impurity level of .14 parts per million. Welding at somewhat higher pressures, in the medium or "soft" vacuum range, on the order of 7.6 to 760 microns, results in a somewhat greater impurity level. At 7.6 microns, the impurity level is 10 parts per million, at 76 microns the impurity level is 100 parts per million, and at 760 microns the impurity level is 1,000 parts per million.

Argon and helium (inert gases), now commercially available from most sources, have an impurity level of 500 parts per million. The impurities in the inert gases are primarily oxygen and nitrogen, with the remainder being oil and water vapor. The impurities in a soft vacuum are oxygen and nitrogen of ambient composition.

Therefore, it can be seen that a vacuum level of 380 microns corresponds to the purity level of commercial inert gases; i.e., 500 parts per million. The impurity level of the best quality of inert gases is about 50 parts per million, which corresponds to a vacuum level of 38 microns.

In any event, it is the specific intention of commercial soft vacuum equipment to produce electron beam welds having the same characteristics as criteria electron beam welds produced wholly within a standard vacuum chamber.

The desirable weld characteristics of standard in-vacuum welds are:

1. Purity of vacuum melt metals.
2. Welds created with a minimum of energy and, therefore, minimal heat-affected zones.

3. Extremely small diameter heat source.
4. Approximation of true parallelism of the weld sides (minimum of "V"-shape) to minimize distortion.
5. High welding speeds.

Primarily, it can be said that the electron beam process is most advantageously used where any other welding heat source does not suffice, such as the welding of refractory materials having extremely high melting points, which must be welded at reasonably fast welding speeds not easily attainable with the limited energy concentration combustible gases, or arc welding. It was demonstrated very early that these materials could not only be readily melted by the electron beam heat source, but they could be welded at an extremely high rate of speed; thus minimizing the growth of the grain in the heat-affected zone adjacent to the weld.

Another valuable contribution of the electron beam melting and welding process was the ability to produce a weld melt having parallel sides. This meant that angular distortion, so common to other welding processes, was eliminated -- or critically minimized. This made possible the welding of parts that were finished machined. In addition, the narrowness of the melt produced by the electron beam meant that welds could be placed in areas so small that all other heat sources were contradicted, and where only mechanical fastenings or brazing could previously be considered.

All of these attributes contributed to the initial acceptance of the electron beam as a welding heat source for critical industries, such as Aerospace, Aircraft and Atomic Energy. For a long time the process was deemed to be desirable, but too exotic for the commercial industry, even though commercial weldments could also benefit from the characteristics of electron beam welds. However, for commercial purposes, the emphasis of these features was shifted somewhat as common engineering materials do not necessarily need the high purity melt that is provided when welding at .1 micron. Further, the general welding of these materials could be achieved by conventional heat sources. Nevertheless, many of the attributes of electron beam welding are needed for commercial use. On the other hand, in the past, the cost of the equipment on a cost-per-piece part basis was the limiting factor, as far as introducing the "general" purpose electron beam welders into commercial use.

The configuration of an electron beam weld is determined by many influencing factors other than the geometry of the beam itself. However, it must be remembered that the geometry of the beam will be

superimposed upon the geometry of the weld melt, which is determined by other factors.

In any case, a beam made wider by defocusing or by dispersion makes the weld wider. Although, it is desirable to weld at higher pressures, it is imperative that the pressure level chosen does not add to the width of the weld.

Once the beam has left the gun, the gun exerts no further influence upon the geometry of the beam. Its geometry may then be affected and caused to increase in size by:

1. Mutual repulsion of the electrons due to their identical charges.
2. In addition to the above, dispersion due to collision of the extremely small electrons with the gas molecules is a second cause of beam enlargement. Therefore, the fewer the number of gas molecules in the path of the beam, the less the dispersion; and, of course, the lower the pressure, the greater the mean free path of the gas molecules; thus, occasioning fewer collisions. A plot of the mean free path of low voltage electrons vs. pressure is shown in Fig. 1.

It is interesting to note from existing data that a high voltage electron beam passing through only $3/8$ " of gas, at atmospheric pressure, requires approximately three (3) times the power to produce a weld of a given depth, as would be required by a beam issuing from the same gun type within vacuum where the gun-to-work distance could be literally several inches. The difference in the energy requirement obviously comes from the fact that "scatter" and loss of power density occurs when attempting to penetrate a gaseous medium at ambient (760,000 microns) pressure.

If the beam is allowed to expand in this manner, it will, of course, be less concentrated. Therefore, it becomes important to minimize dispersion of the electron beam by reducing the pressure by a reasonable amount. It is interesting to note that at standard air pressure, the average or mean free path of a gas molecule is on the order of five millionths of a centimeter. With pressure reduced to 1 micron, the mean free path increases to 5 centimeters or 1 million times greater. At 100 microns, the mean free path of a molecule of gas is 10 thousand times greater than at one (1) atmosphere. The effect of pressure on the mean free path and quantity of air molecules is shown in Fig. 2. As the beam dispersion is directly proportional to the density of the gas, and as weld penetration is inversely related to the beam size, it can be seen that the beam dispersion will decrease as the pressure decreases, with

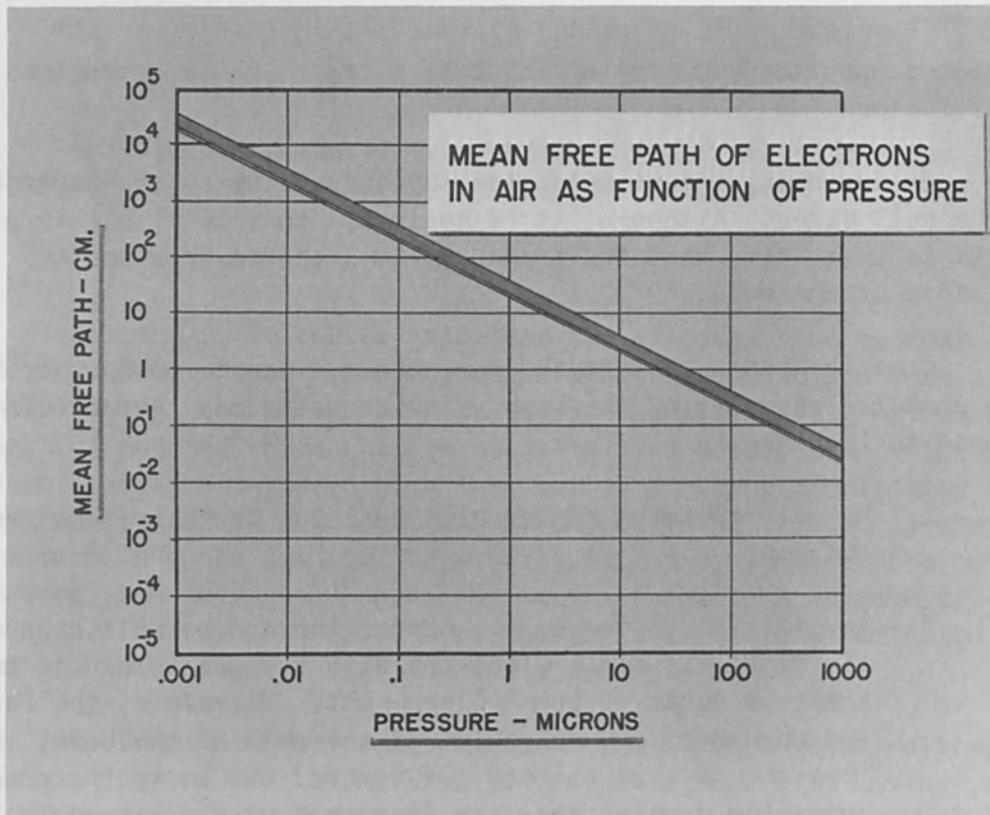


Figure 1.

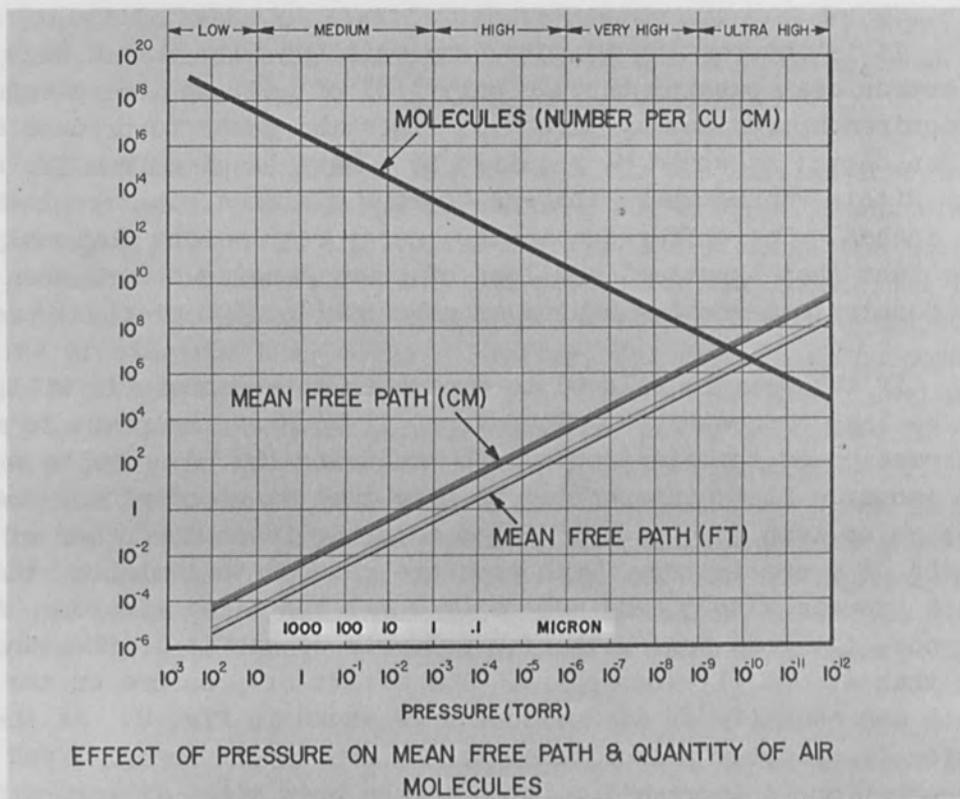


Figure 2.

the result that weld penetration will increase. It can be concluded, therefore, at some intermediate pressures below atmospheric pressures that the melting efficiency of the beam can be approximately that obtained under conditions of hard vacuum, or at least adequate to produce welds that are characteristic of the process.

We are, therefore, interested in producing a beam within a high vacuum chamber and bringing it into a volume held at some moderately low pressure. This is shown schematically in Fig. 3. At some point in the column, the beam must pass through a transition zone from the vacuum environment of the gun chamber to a low pressure gaseous environment held at 100 microns (or other experimentally determined pressure). As the electrons move to the point of egress, the beam encounters first, extremely rarified gas, and later, gas at 100 microns. The requirements for initially accelerating and concentrating the electrons within the vacuum are known factors, as established by Dr. Stohr nearly ten years ago. These known factors lend themselves to theoretical and empirical analyses of voltage requirements and the contingencies of electron optics. One new factor was introduced, which was the design and implementation of the ideal transition zone. This involved the design of orifices sized to isolate the high vacuum gun environment from the molecular wandering of gas from the environment of higher pressure.

The choice of operating pressure in the tooling chamber, relative to the pressure in the gun chamber, was made on the basis of:

1. The size of the orifice between the two chambers. That is to say that there is an optimum relationship between orifice diameter, orifice tube length, beam diameter and the mean free path of the gas molecules at the selected pressure.
2. The simplicity of the orifice system then relates to the size of the pumping equipment attached to the gun chamber. The pumps must handle the migration of gas from the tooling chamber, in order to maintain the pressure of .1 micron around the gun. Naturally, the lower the pressure in the tooling chamber, the smaller and less expensive the pumps that must be used to support the gun chamber.
3. On the other hand, the choice and cost of pumps in the gun chamber is offset by the availability and cost of mechanical pumps that will produce a given welding vacuum in the tooling chamber within a given time. For example, a 140 CFM mechanical roughing pump will pump a one cubic foot volume to a pressure level of 1000 microns in about

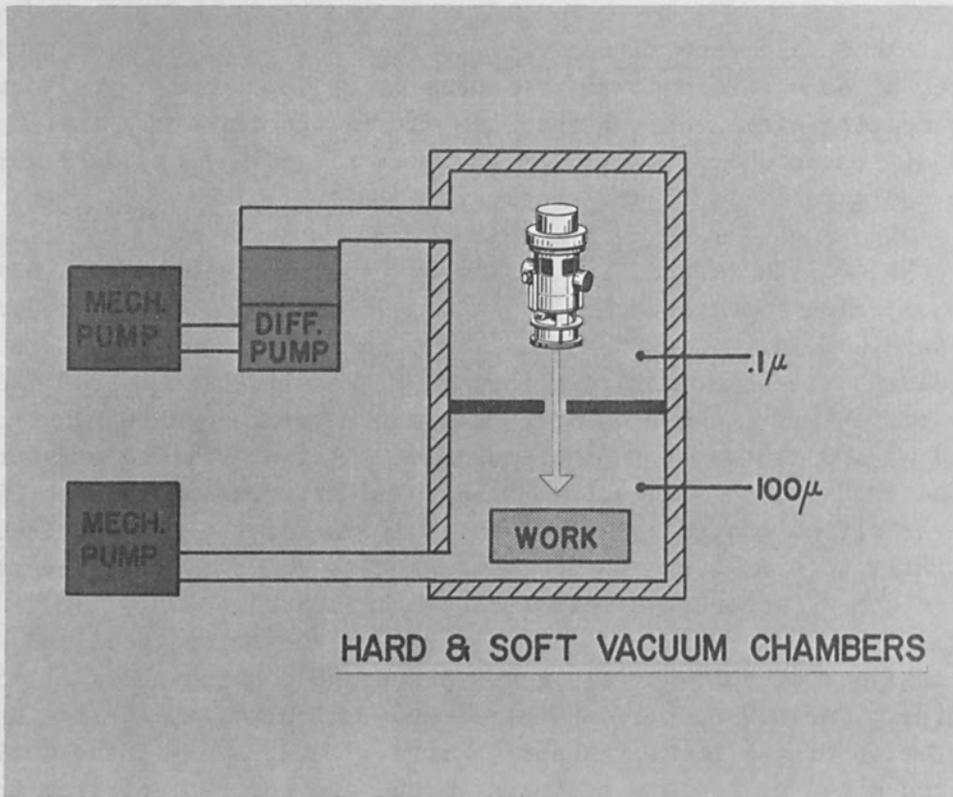


Figure 3.

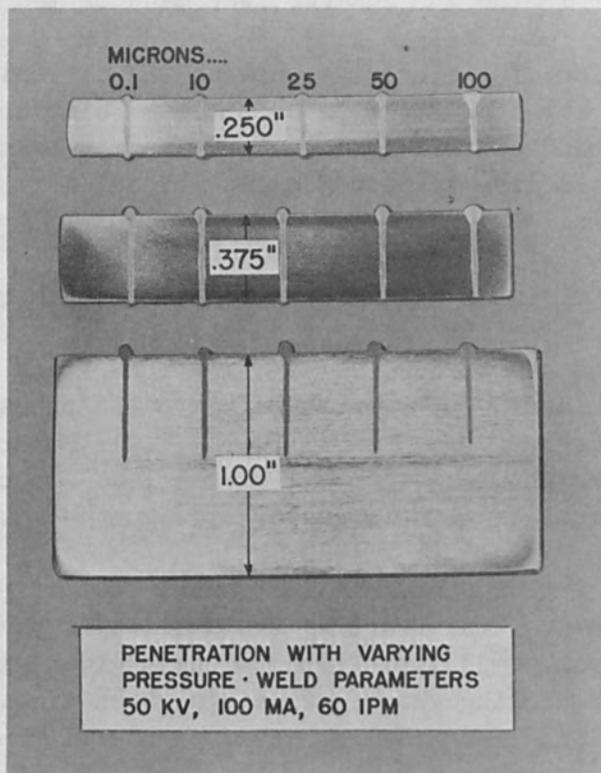


Figure 4.

3.22 seconds, or it will pump to 100 microns in about 5 seconds. However, it will take 7.2 seconds for the same pump to reach 27 microns. A roughing pump can theoretically reach about 10 microns after a long time period in the asymptotic region of the pumping curve. On the other hand, a Roots Blower, combined with a mechanical pump, can reach 1 micron Hg of pressure in 3.3 seconds.

4. Of primary importance in sizing the pumps applied to the work chamber is the production rate for a given piece of work. The entire work cycle consists of: (a) load, (b) index part to entrance position, (c) seal valve to part, (d) pump, (e) weld, (f) lower part from valve, (g) eject part. The other factors are fixed; therefore, pumping time must not be allowed to become a limiting factor.

Although welding experiments have been carried out at pressure levels up to 1,000 microns, the most judicious choice would appear to be 100 microns, based on the optimum balance of the above considerations. The configuration of an electron beam at 100 microns can be seen in Fig. 5.

Two additional considerations must be added to the above four in order to finalize the choice of operating pressures:

5. Weld purity has been shown earlier to be acceptable with pressure as high as 380 microns and of exceptional purity with pressure as high as 38 microns. 100 microns still appears to be an advantageous compromise, with one more point to be considered.
6. Experiments have shown that the beam configuration that produces the preferred weld geometry is not appreciably altered by pressures as high as 50 microns. Figure 4 shows welds that have been produced at discrete steps of pressure from hard vacuum of .1 micron to 100 microns. Specifically, the steps are: .1, 10, 25, 50, and 100 microns Hg. Calling attention first to those partial penetration welds made in the 1" thick block, it will be observed that penetration depth is reasonably constant, within 5%, from .1 to 50 microns. The penetration depth decreased by 11% at 100 microns, which demonstrates that some of the beam power is beginning to be dissipated in the increased population of gas. The actual values of penetration are:

	<u>% of Hard Vacuum Penetration:</u>
.1 micron, .485"	100%
10 microns, .475"	98%
25 microns, .470"	97%
50 microns, .460"	95%
100 microns, .430"	89%

Some variation in penetration can be expected in the process even when all welds are made in hard vacuum. However, it is significant that the change in penetration is in the expected direction, decreasing, with increasing pressure. Up to 50 microns, this decrease is within the expected tolerance of the system; however, at 100 microns, the 11% change in penetration indicates that the curve of the graph of penetration vs. pressure has taken a radical change in slope. It can also be noted that the "nail head" or the size of the top of the weld at the top surface has also increased at 100 microns, although, the width of the stem of the weld measured below the head is approximately the same (.026") throughout the pressure range. This is true of welds made at increasingly higher pressures up to 1,000 microns, which has been the limit of the tests.

All welds were made with the same power and the same focal setting. The same depth of penetration could easily be re-established by simply increasing the applied power. Although the depth could be re-established, the width of the "nail head" would remain large at the higher pressures, as this is a function of elastic angular beam scatter, as well as the loss of beam power due to inelastic scattering.

For these reasons, 50 microns of pressure is the design level which meets all of the tests of practicability and economy. Certainly, the mode of welding at soft vacuum will be used throughout the range of 1 to 1,000 microns -- depending upon the demands of the product being welded.

Welds having 100% penetration are less sensitive to the effects of various pressure levels. For example, the same welding parameters used to produce the partial penetration welds in the 1" material were used to produce 100% penetrating welds in both .250" and .375" under the same variations in pressure level.

While the foregoing may seem reminiscent of electron beam welding at atmospheric pressure, a comparison of the magnitude of pressure differential between the gun and the exit nozzle, at atmospheric pressure, must be considered.

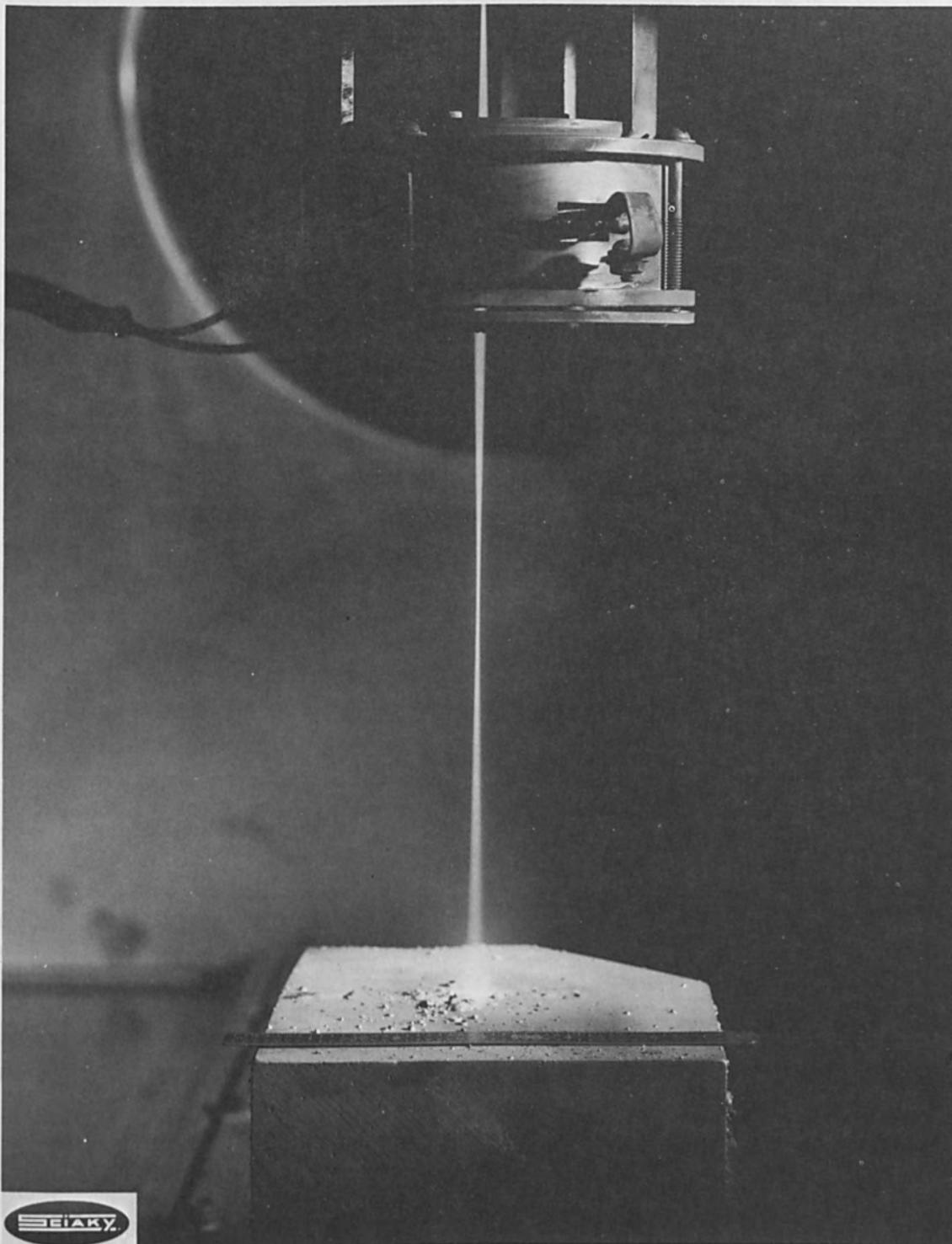


Figure 5. Electron Beam at 100 Microns Pressure.

When considering the use of an electron beam in atmosphere, the high voltage (180 KV) beam is going from an environment of at least .1 micron to an environment of approximately 760,000 microns; whereas, in the case of welding in "soft" vacuum, we are discussing a low voltage (60 KV) electron beam developed at a pressure of .1 micron, which is projected into a second chamber to weld at approximately 100 microns. The first case (electron beam in air) represents a difference in pressure of 7.6 million times, and the second case represents a pressure difference of only 7.6 thousand times.

It might be thought, due to the above discussion, that any gas at all would cause beam scatter and enlargement. This is not necessarily true -- neglecting the gun for the moment, it must be remembered that the electron beam itself never exists in total vacuum while welding, even when the system is totally enclosed within a vacuum chamber at .1 micron, for when the melting begins, vapor is discharged from the molten metal at reasonable high velocity, which increases the pressure locally at the point of the beam impingement to some high unknown value.

It is known that in both a theoretical sense and in a practical sense, characteristically narrow and deep electron beam welds are produced (by proper equipment) at all pressures lower than .1 micron, and that narrow welds are not usually produced at 760,000 microns (ambient atmospheric pressure). Thus, the pressure limits established by existing data are: .1 micron as adequate, and 760,000 microns as inadequate. Therefore, it was required first to determine whether or not some intermediate pressure within this range (between .1 and 760,000 microns) would permit an electron beam to produce characteristic welds.

Because it is required (as explained in the discussion) to maintain the gun at .1 micron, or less, it is necessary that a double vacuum system be used. This experimental equipment is in existence at Sciaky Bros., Inc., laboratory and is shown in photograph Fig. 6. Essentially, this equipment consists of a small vacuum chamber contained within a larger vacuum chamber. The chambers are fixed in their relation to each other. The pressure in each may be varied independently. The electron beam gun is contained within the inner chamber and is maintained at all times at a pressure not exceeding a .1 micron. The material being welded is placed in a second chamber which is held at approximately 100 microns. The work is fixtured to a moving platen, which moves beneath the electron beam issuing from the gun chamber; thus, providing the relative motion for controlled welding speed.

Vacuum pumping for a hard vacuum (.1 micron) involves the pumping of both the volume and the surface area of the contents of the vacuum chamber. The volume is pumped by the mechanical roughing pump

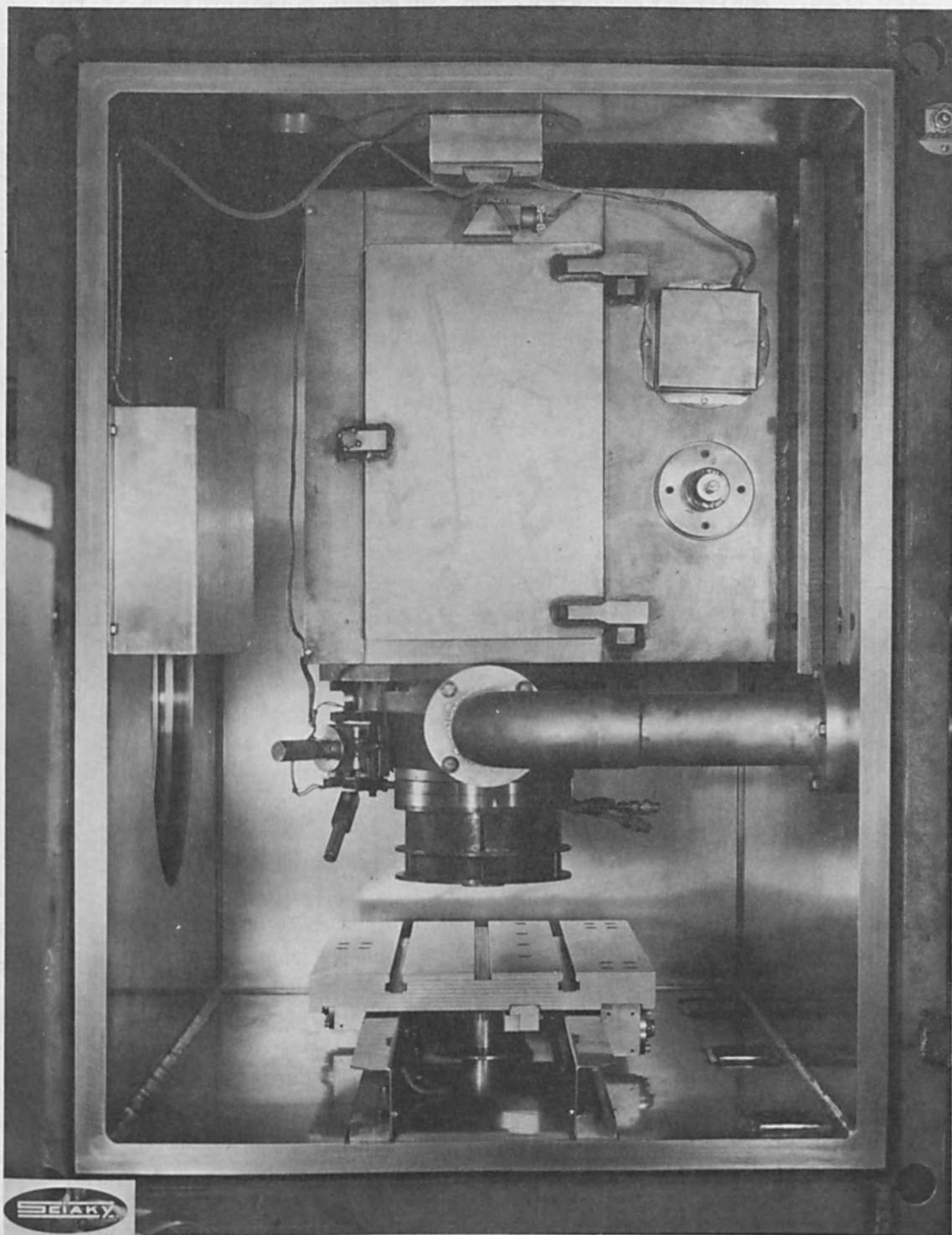


Figure 6.

and the surface area of the tooling and of the piece-part can be thought of as being pumped by the diffusion pump, (see Fig. 7). The use of these two pumps implies the use of large valves, piping, etc. It also implies cost. Basically, however, the important consideration is the time required to pump to the hard vacuum level.

When one makes use of the arc welding process using an inert gas as the shielding mechanism for the hot weld metal, it is easy to compute the cost of the weld based on an established price per cubic foot of gas and based on a known flow of gas per minute. In the electron beam process the inert atmosphere is obtained for the incidental and trivial costs of a change of pumping oil once or twice a year and the electrical power used by the diffusion pump heaters. Nevertheless, this inert vacuum environment must be paid for in terms of "time". It is necessary to reduce the pressure in the chamber of a given size to the welding level at a certain finite speed based on the size of the vacuum pumps employed. The size of these pumps varies directly with the speed of pumping and with the level to which one must pump in order to conduct the process.

Fortunately, it is almost axiomatic that piece-parts which are produced in large quantities are also reasonably small in size. It is uneconomical to open and close the chamber and to pump down the chamber for each individual piece-part. Therefore, the first attempts to adapt the electron beam welding process to commercial use were by the batch-loading of a large number of small parts in multiple fixtures into the vacuum chamber where they were then sequentially welded by manipulating each part by some form of remote positioning device to produce the desired weld path. This might mean either a simple or complex form of programming to move the beam from piece to piece and to trace the weld path. When one considers that the individual piece-parts must be independently fixtured, it is obvious that a batch of 100 parts might require 100 fixtures of great dimensional accuracy. Remote manipulation of a piece part, either simple or complicated, locked up within a vacuum chamber is a difficult procedure, and at times, a frustrating problem. As you know, the electron beam is extremely small; therefore, alignment of the beam to the weld joint is extremely critical. Optical alignment of each of the 100 parts in the batch load is time consuming and cannot be called a high production technique. What is saved in vacuum pump-down time is lost in the multiplicity or problems involved in remote manipulation and remote alignment.

Currently, general purpose or batch-loaded electron beam welding vacuum chambers must be pumped to a vacuum level of .1 micron, starting from atmospheric pressure of 760,000 microns. The finite time for such a pumping operation will be from five minutes for a small chamber, to 15 to 20 minutes for extremely large chambers. During this pump-down

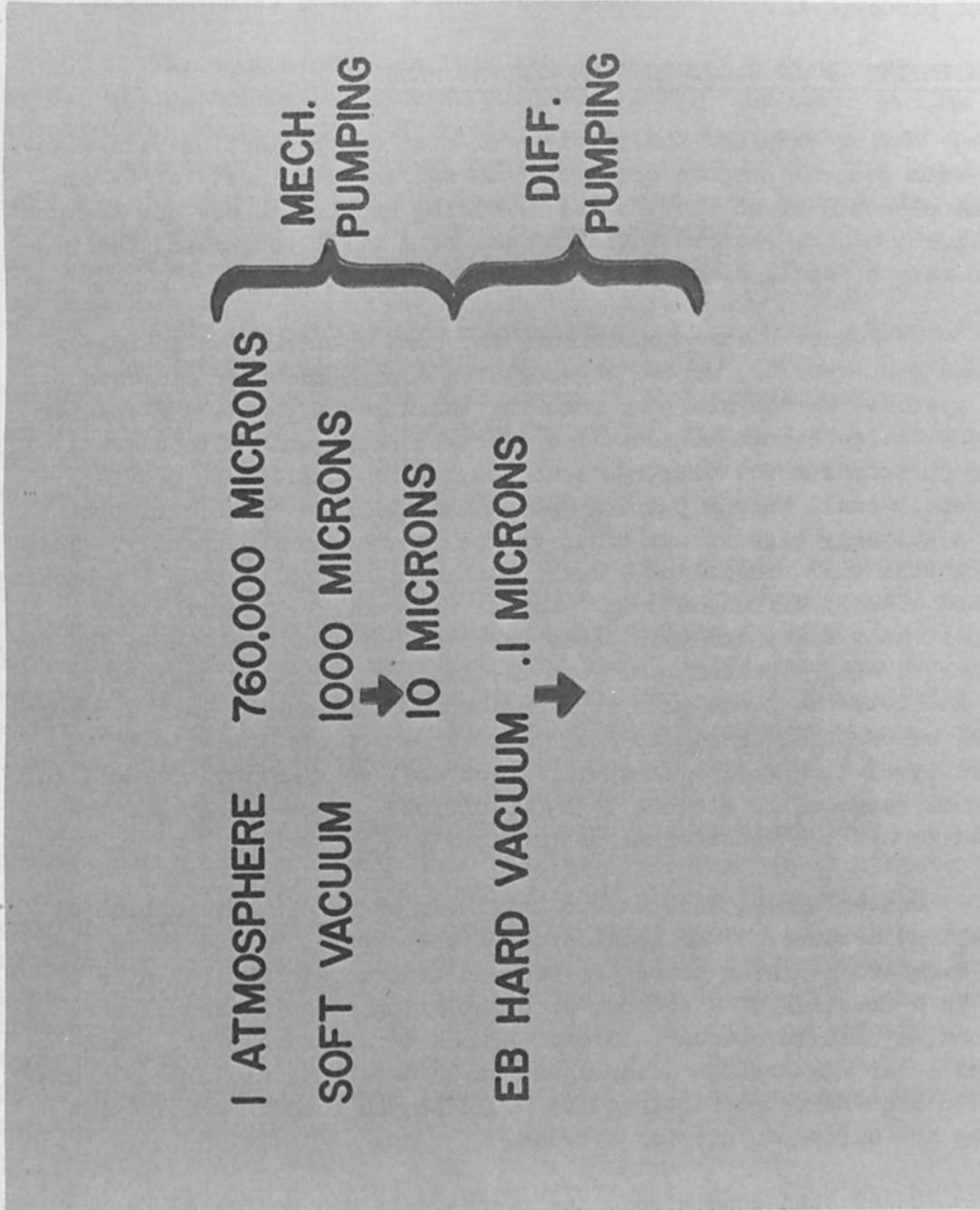


Figure 7.

time the equipment is, for all practical welding purposes, inactive. It must be remembered that this pumping time is expended each and every time the chamber is opened and closed to take the entry of a single or batch of piece-parts.

III. ELEMENTS OF AN ELECTRON BEAM WELDING MACHINE

Having examined the problems of cost and production rate associated with general purpose electron beam welding machines, let us examine the elements of an electron beam welding machine to see how the cost of equipment may be reduced while, at the same time, increasing the production rate of small size production parts.

We cannot change appreciably the cost or design of the electron beam gun; however, we can do something about the other elements of the system. We can start by reducing the size of the vacuum chamber to be just large enough to house the electron beam gun. With a small chamber surrounding the electron beam gun, it is possible to provide an extremely small vacuum pumping system that will be capable of producing a suitably high vacuum of .1 micron in the gun region only. This vacuum system will consist of a small mechanical roughing pump (or backing pump) and a small diffusion pump. We can now consider handling one piece-part at a time, and this piece-part can have its own vacuum chamber, or rather, a simple vacuum cavity in the tooling, which is just large enough to clear the piece-part with an absolute minimum of excess volume. Further, we can then decide to pump this extremely small cavity to a pressure level much greater than .1 micron and, as a matter of fact, in the vacuum range of 10 microns to 1,000 microns -- depending upon the application and the considerations previously discussed.

Conveniently, this vacuum level can be reached by mechanical pumps which, because of the small size of the cavity, can be of small size. Standard roughing pumps can be used to evacuate a given size of cavity in a fraction of a second, or combination roughing and blower pumps can be used to evacuate larger volumes in the same time. This provides other non-obvious advantages, as no expensive and time consuming valve sequencing is required, as would be the case in combination roughing and diffusion pumping systems.

Under such conditions, the piece-parts may be loaded in only one "nest" or fixture, in atmosphere, one at a time, and automatically sequenced through the machine. As compared to the multiple tooling of batch-loading, another savings is realized. It can be said, truthfully, that this concept limits the size and type of parts that can be welded, and the type of joints that can be used. However, machines using this concept have been designed to handle a variety of piece-parts: (1) 36"

of linear weld at a production rate of about 120 pph; (2) annular welds varying from 1" to 3" in diameter, on the same piece-part at 120 pph; (3) spot welds at 1,000 pph; (4) small annular welds at 750 pph; (5) large casting weighing 75 pounds welded at 10 pph, etc.

The equipment falls into the category of the special purpose welder as opposed to the general purpose welder. The basic machine is standard, as shown in Fig. 8, having always the same elements of gun, gun vacuum system, high voltage power supply sized to fit the immediate job, and the necessary instrumentation and electrical interlocks. The tooling on the other hand is designed and built to fit a specific piece-part, or group of very similar piece-parts.

Although quickly interchangeable tooling can be added to this type of equipment, it is primarily intended to satisfy that segment of industry that can justify the dedication of the machine to mass-production of a specific piece-part during the design life of that product. The standard portion of the machine may be adapted and made use of with a new set of tooling for a model change.

The techniques employed in implementing this unique concept are supported by the known behavior of gases under certain conditions. It can be seen from the above that the gun chamber is maintained at a sufficiently high vacuum to protect the gun filament, and the gun elements, while the work is subject to a different vacuum level, which is about 1,000 to 10,000 times greater than the gun chamber vacuum. How does one maintain this individuality of the two chambers?

If the air at the weld area were at a pressure of one atmosphere, which is approximately 10 million times greater than the vacuum level of .1 micron in the gun chamber, then one could expect that the pressure differential would certainly be great enough to push a viscous flow of air into the gun chamber (Fig. 9) through the orifice system which permits physical particles of the electron beam to move from the gun chamber to the weld area. In order to avoid this in-rush of air at atmospheric pressure into the sacred confines of the hard vacuum gun chamber, it is necessary to provide a series of differentially pumped, orifices between the two domains.

A simpler solution is permissible when operating at the low pressure levels due to the fact that a gas at about 100 microns, does not undergo viscous, i.e., directional flow, but rather the penetration of the gas from the outer "soft" vacuum chamber into the "hard" vacuum chamber is a random phenomena, (Fig. 10). The motion of the molecules of gas at medium (soft) vacuum is in the form of "molecular wandering". The individual gas molecules have a difficult time in finding the orifice, as there is little directional mass flow. Of course, some molecules of

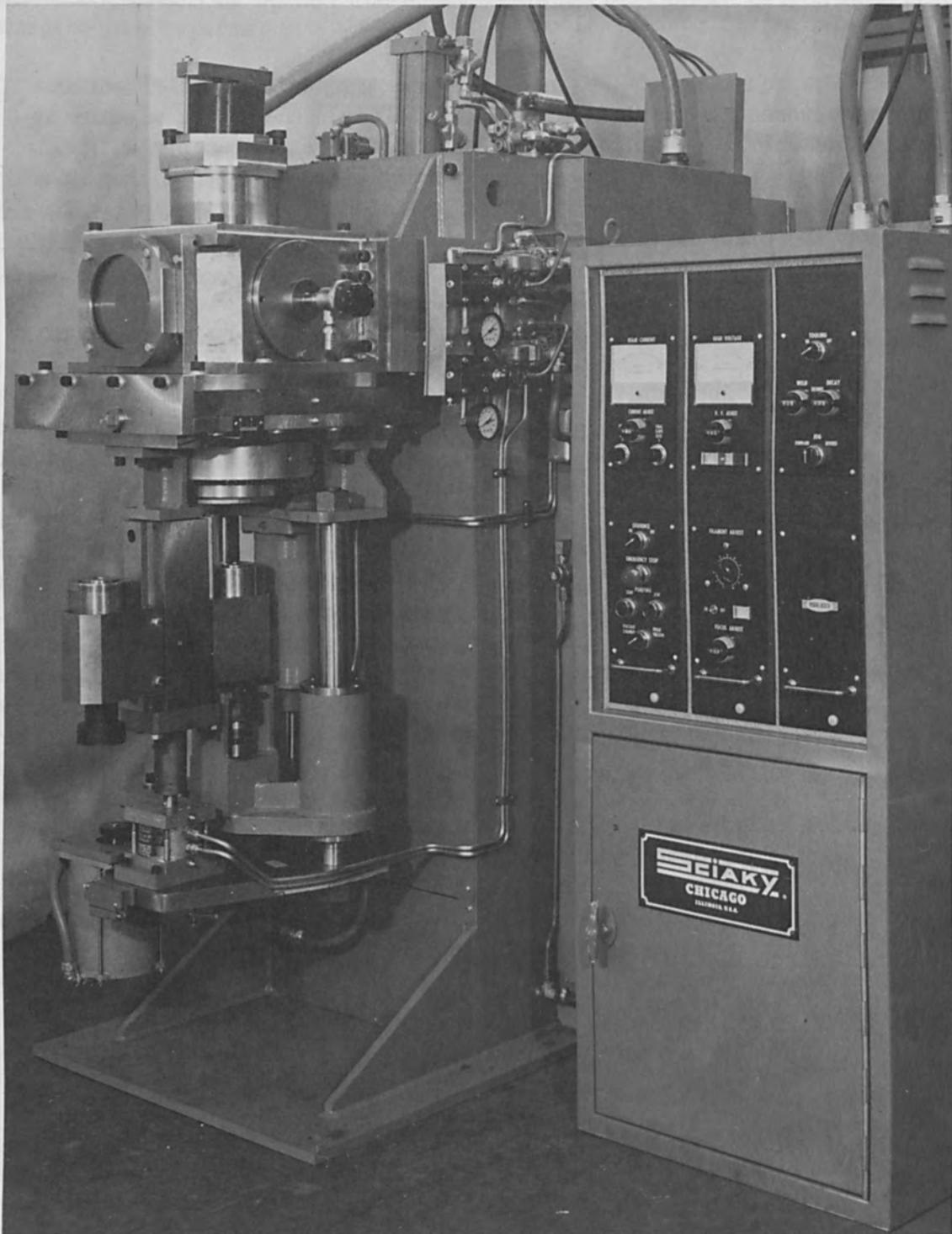


Figure 8.

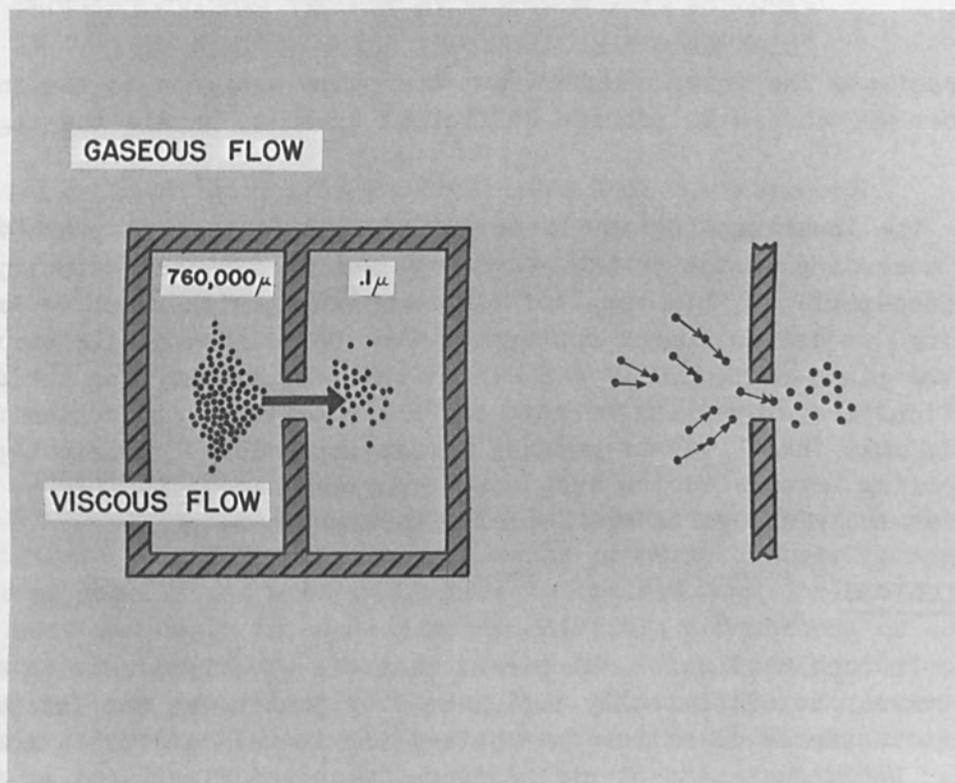


Figure 9.

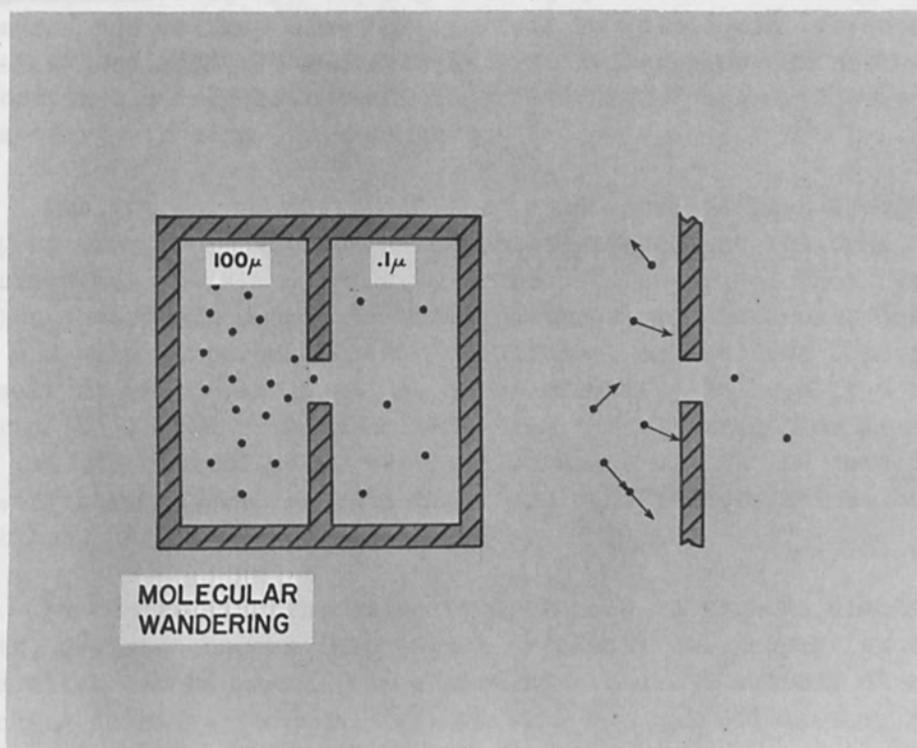


Figure 10.

gas do penetrate the inner chamber, but the pumps assigned to the inner gun chamber are chosen to provide sufficient speed to handle the migration.

The lower tooling cavity is pumped by a mechanical pump, which is sized according to the cavity volume necessitated by the geometry of the piece-part. Within the "soft" vacuum range, with which we are now working, we are no longer concerned with the pumping of the surface area of the piece-part, and we are almost entirely immune from the attendant difficulties of pumping certain surface conditions and contaminants that would make "hard" vacuum pumping almost impossible. Lubrication of the tooling becomes easier and, most important, accessibility is greatly improved as the piece-parts are loaded in ambient atmosphere.

V. CONCLUSIONS

In conclusion, it is apparent that the move from welding at hard vacuum can be economically implemented by pausing at the intermediate soft vacuum level rather than attempting to weld at full atmospheric pressure. Welding at atmospheric pressure, which at first look appeared desirable, has its own field of application but, in general, suffers from increased weld size, x-ray generation, lower weld quality, and generally high cost. Welding at soft vacuum extends the process to the commercial application field by providing lower equipment cost, high production rates, simplicity of tooling, and weld quality and characteristics which are competitive with hard vacuum electron beam welds.

PROPERTY CHANGES INDUCED BY ELECTRON BEAM EXPOSURE
IN VARIOUS SEMICONDUCTOR AND DIELECTRIC MATERIALS*

by

David R. Sivertsen, Olin B. Cecil, and Rolf R. Haberecht
Materials & Controls Research & Development Laboratory
Texas Instruments Incorporated
Dallas, Texas

ABSTRACT

Electron beam energy from 50 to 150 KeV was focused on selected substrate materials to create discrete property changes in high resolution geometrical areas. Changes in resistivity, dielectric constant, junction depth in semiconductor materials and methods of controlling such changes are examples of the investigations conducted. A spectrum of beam power levels ranging from practically zero power density to greater than 1 megawatt per square centimeter was examined as a means of relating the electron beam energy to the structural, mechanical and electrical changes observed. Various analytical techniques are used to support explanations of observed phenomena.

I. INTRODUCTION

Microminiaturization of electronics has generated considerable effort in electron beam technology, stimulated mainly by the need to focus energy in very small areas.

Fabrication of microminiature circuitry has been aided in the industry by use of electron beam technology to scribe, cut bond, and weld various materials and material overlays. The resolution, rapid deflection according to predetermined patterns, modulation of beam energy, and other important characteristics of an electron beam in vacuum make it particularly suitable for producing changes in a variety of materials. In the work described here, the electron beam is to be used to produce controlled preferential changes within the material by exposing bare materials (semiconductors and dielectrics) without masks and overlays.

By controlling the structural changes in various electronic materials, desired changes in physical or electrical properties can be achieved which can be useful in developing circuit functions of extremely high packing density. Several factors will be explored as they relate

*This work supported by U.S. Air Force Avionics Laboratory under Contract AF 33(615)-2882.

to characterization of the electron beam and its effect on several dielectric and semiconductor materials. In addition, the resultant change in physical or electrical properties on selected materials will be examined.

II. EQUIPMENT USED AND EXPERIMENTAL PROCEDURE

The technique employed here, that is, utilizing an electron beam to produce preferential property changes, is an outgrowth of corporate-sponsored work at Texas Instruments wherein a spectrum of materials was examined and the change in resistivity noted as a permanent alteration of the material. Examples of part of this work, shown in Table I, indicate that substantial resistivity changes can be obtained in a broad variety of materials.

Haberecht's work¹ on high-resolution preferential changes in these materials indicated that specialized test techniques are needed to detect and measure resultant changes in electrical or structural properties.

A. Zeiss Electron Beam Machine

The electron beam equipment used at TI is a Carl Zeiss machine consisting of the electron beam sources, vacuum pumping system, high-voltage power supplies, and mechanical equipment for moving the sample work in the x and y directions. The work table holding the samples could be moved manually or by belt-driven synchronous motors. If a constant, calibrated traverse rate past the electron beam was desired to control exposure, a magnetic clutch arrangement provided a positive stop-start motion for precise control of exposure time. The equipment could provide repetitive pulsing by setting the pulse generator to the desired frequency. In addition, a "one-shot" pulser was fabricated according to the circuit of Figure 1. It provided a single energy pulse from the electron beam.

B. Resistivity Test Equipment

The resistivity test equipment employed a two-point contact measuring system instead of the standard four-point probe.

While the four-point resistivity method eliminates contact resistance, it has the disadvantage of larger probe contact area and difficulty in locating the probes in very small areas. The two-point resistivity equipment shown in Figure 2 uses a stereomicroscope to locate the contacts in the exposed area of the sample surface. A voltage is applied across the contacts with a regulated dc power supply.

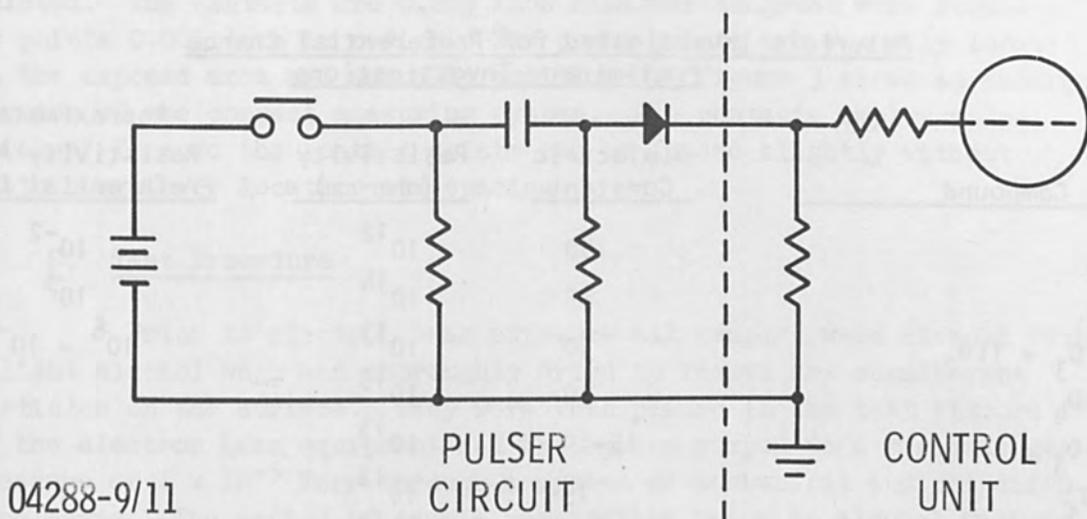


Figure 1. Circuit for "One-Shot" Pulser Used with Zeiss Electron Beam Machine.

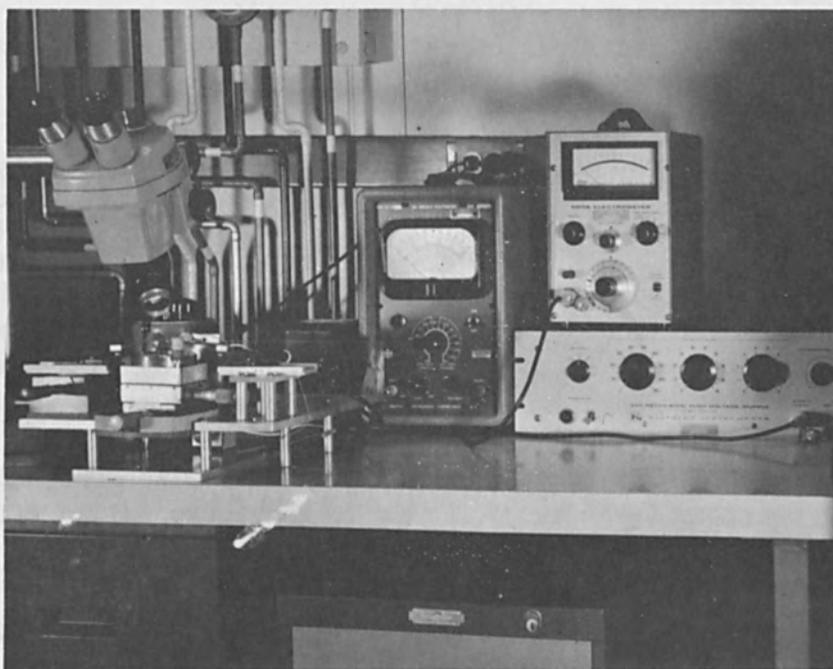


Figure 2. TI-Built Two-Point Resistivity Measurement Equipment.

TABLE I

Materials Investigated for Preferential Change
Preliminary Investigations

<u>Compound</u>	<u>Dielectric Constant</u>	<u>Resistivity (ohm-cm)</u>	<u>Approximate Resistivity After Preferential Change</u>
CeO ₂	100	10 ¹²	10 ⁻²
TiO ₂	85	10 ¹⁴	10 ³
MgTiO ₃ + TiO ₂	30	10 ¹³	10 ⁶ - 10 ⁻¹
TiCeO ₄	30	10 ¹³	
NiTiO ₃	15-19	10 ¹³	
Cr ₂ O ₃	15	10 ¹²	1
2MgO·TiO ₂ - MgTiO ₃ Eutectic	12-14	10 ¹³	
MgTiO ₃ - 100T83	12-14	10 ¹³	10 ⁶ - 10 ⁻¹
MgTiO ₃ - 120T83	12-14	10 ¹³	10 ⁶ - 10 ⁻¹
TICON A	12-14	10 ¹³	
TICON B	12-14	10 ¹³	
TICON C	12-14	10 ¹³	
Y ₃ Fe ₅ O ₁₂	12	10 ¹³	10 ⁻⁴
Fe ₂ O ₃	12	10 ¹²	10 ⁻⁴
SPINEL FERRITES	12	10 ⁸	10 ⁻²
Al ₂ O ₃	9	10 ¹⁴	
MgO FUSED	6	10 ¹⁵	Physical Change and/or adapta- bility to prefer- ential plating
MgO SINGLE	6	10 ¹⁵	
STEATITE	6	10 ¹⁴	
MgO·SiO ₂			
FORSTERITE	6	10 ¹⁴	"
2MgO·SiO ₂			
MgCr ₂ O ₄		10 ¹⁴	10 ⁻²
NiO		10 ¹³	1
GaAs (High Purity)		10 ⁷	50
(Medium Purity)		10 ⁵	30
AlSb		10 ²	2 - 10 ⁻³
InSb		30	20
PbTe		5	3

The voltage and current measurements are made and the resistance calculated. The contacts are 0.003 inch diameter tungsten wire etched to points 0.001 inch in radius. The probe points are visually located in the exposed area and manually positioned. Figure 3 shows an enlarged portion of the contact measuring system. The contacts employ a "c-whisker" form so the contact points can be loaded slightly without disturbing their location and spacing.

C. Test Procedure

Prior to electron beam exposure all samples were cleaned with a light alcohol wash and thoroughly dried to remove any nonadherent particles on the surface. They were then placed in the test fixture of the electron beam equipment and the system pumped to a residual gas pressure of 1×10^{-5} Torr through a system of mechanical and oil diffusion pumps. The method of sample preparation prior to alcohol wash is noted for each material in Table II.

TABLE II

Source and Preparation of Sample Materials

<u>Material</u>	<u>Vendor</u>	<u>Preparation</u>	<u>Crystal Orientation</u>
Single Crystal TiO ₂	Linde Corporation	Lapped and Polished	"C" Axis ⊥ to Face
Single Crystal Si	Texas Instruments Incorporated	Lapped, Polished & Chemical Etch "N" Type, Sb-Doped	110
Single Crystal Ge	Texas Instruments Incorporated	Lapped, Polished & Chemical Etch	110, 111

Before exposing samples to the electron beam, power density is monitored by focusing the beam in a Faraday cup. Pulse height, width, and frequency are measured, and peak and average power are calculated. Table III shows that average beam power is not necessarily related to the actual peak power developed. The pulsed power is instantaneous and consequently produces a transient thermal effect whose nature depends substantially on the waveforms of pulsed power developed.

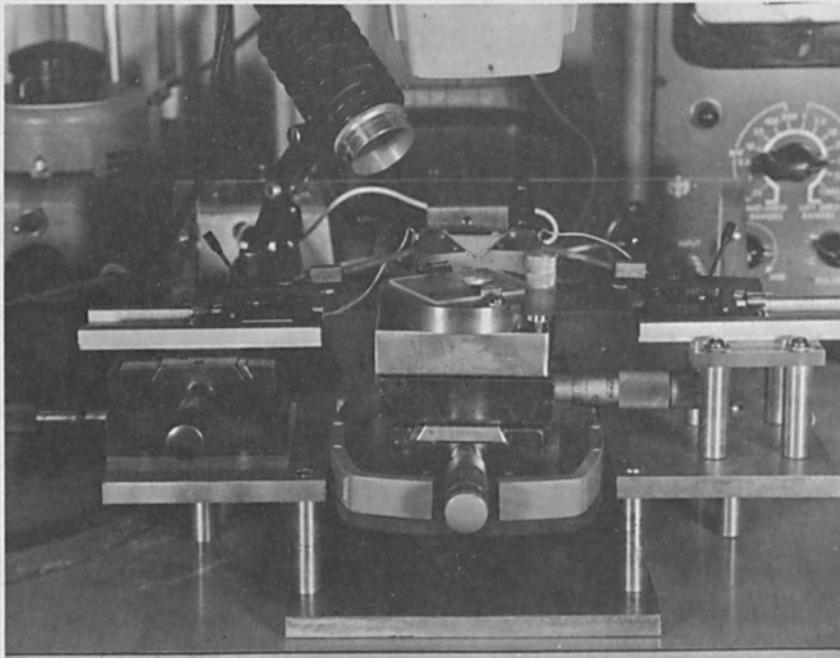


Figure 3. Close-Up View of Two-Point Probe System.

TABLE III

Relationship of Beam Power to Pulse Shape (Beam Voltage 100 keV)

<u>Average Power</u>	<u>Peak Power</u>	<u>Pulse Height</u>	<u>Pulse Width</u>	<u>Pulse Period</u>
0.61 W	430 W	4.3 V	6 Micro-seconds	4200 Micro-seconds
1.2 W	860 W	8.6 V	6 Micro-seconds	4200 Micro-seconds

D. Residual Gas Analyzer

Consolidated Electrodynamics Corporation model 21-614 residual gas analyzer was connected to the electron beam chamber to measure gases that might result from interaction of the beam and the sample substrate. The procedure for these measurements is as follows:

1. Focus the beam into the Faraday cup and adjust pulse height, width, and frequency to provide desired beam power.

2. Adjust the residual gas analyzer according to the manufacturer's recommended procedure. Set the ionizing current (which controls sensitivity) at a value commensurate with the quantity and type of gas to be analyzed.

3. Background gases are scanned, and a test scan is made while the sample is exposed to the electron beam.

4. The scan is "peaked" on the gas to be observed, and the mass selector locked. The gas analyzer is now set so that it is in effect a selective sensing gas flowmeter; only gas having this pre-selected mass will be sensed.

5. The procedure used with the residual gas analyzer also includes standardizing the system with a measured quantity of test gas, such as oxygen or argon. The gas is leaked into the vacuum system while the residual gas analyzer is "peaked" on this gas. The reading on the residual gas analyzer output recorder will correspond to the quantity of injected gas.

6. Changes in recorder readings during electron beam exposure of test samples can be integrated to give the total quantity of discharged gas.

E. Dielectric Test Measurements

Dielectric measurements were originally attempted using a parallel-plate capacitor technique. However, this method did not lend itself to measurements on microgeometrical areas of 0.0015 inch x 0.500 inch or 0.0015 inch diameter with a penetration depth of 0.01 inch to 0.002 inch. We settled instead on a technique used to measure dielectric losses of micro-strip transmission lines. By making our samples of dielectric quite thin (10 mils) and placing a ground plane contact and micro-strip counterelectrode over the area that has been exposed to the electron beam, it is possible to calculate dielectric constant over a range of frequencies. Figure 4 compares parallel-plate and micro-strip techniques for dielectric measurements. Figure 5 shows the dielectric test equipment with a test sample in place.

Our calculations followed a procedure used by Wheeler² by which strip lines of varying geometry were examined. Figure 6 is a graphical representation of Wheeler's work showing the relationship among wave resistance, shape ratio, and dielectric constant. Steps in calculating results on a thin plate of Al_2O_3 are as follows:

1. Calculate shape ratio from dimensions of micro-stripline contact and substrate thickness, using the formula

$$\text{Shape ratio (a/b)} = \frac{\text{width of contact}}{\text{thickness of material}} .$$

2. Determine wave resistance for air (z_1) for shape ratio from Figure 6. (For example, at $a/b = 0.595$, $z_1 = 233$ ohms)
3. Measure capacitance per centimeter of stripline contact on sample substrate.
4. Using formula, determine z_2 , the wave resistance of the sample dielectric material

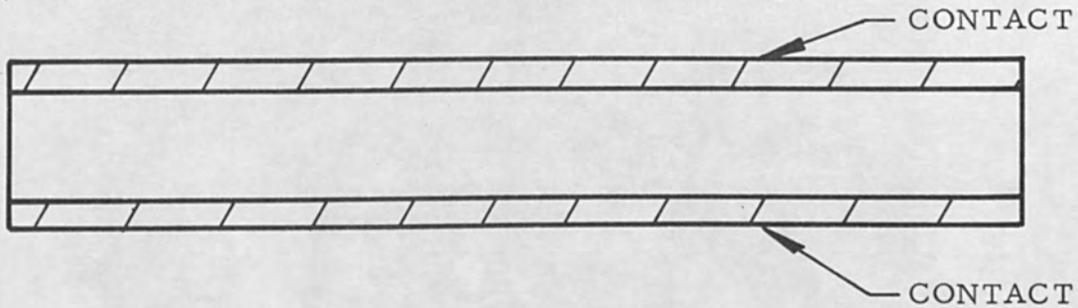
$$z_2^2 = \frac{z_1}{C v_0}$$

where z_1 is wave resistance of air in ohms
 C is capacitance measurement in farads
 v_0 is velocity of light in cm.

5. Determine dielectric constant, k , of material by locating the intersection of the shape ratio and wave resistance of material (z_2) lines with respect to the constant dielectric lines. For example,

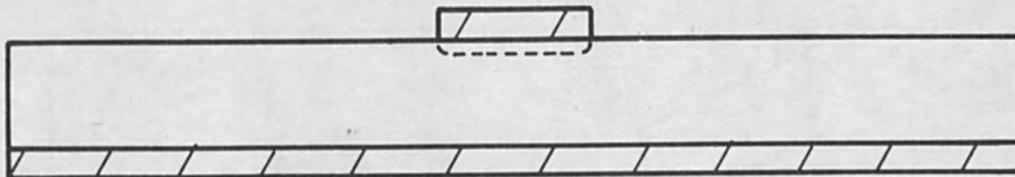
PARALLEL-PLATE CAPACITANCE METHOD

(Entire area to be covered by contact must be exposed to beam.)



MICRO-STRIPLINE METHOD

(Only a small area of sample need be exposed to beam.)



04107-2/7

Figure 4. Configurations of Parallel-Plate and Micro-Stripline Methods of Measuring Dielectric Constant.

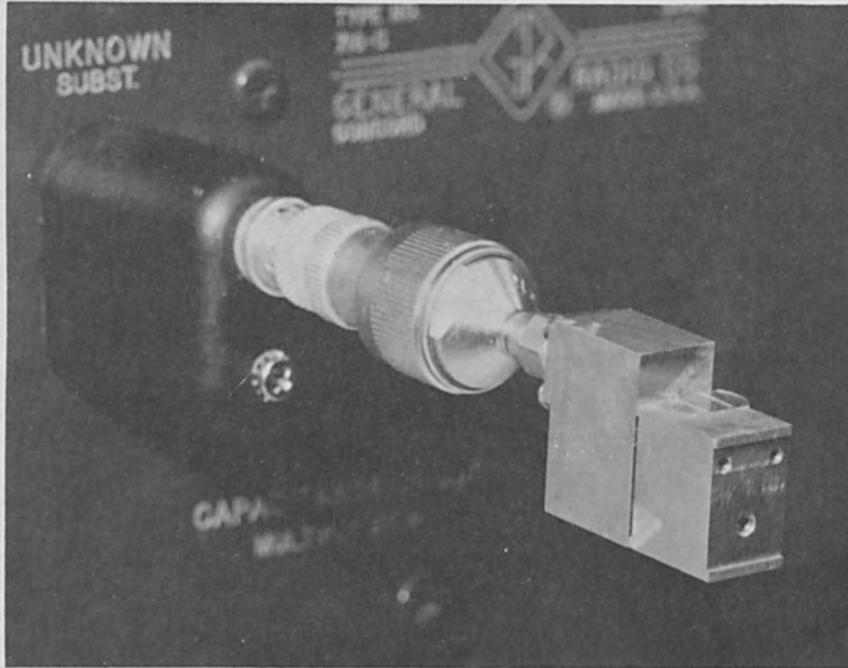
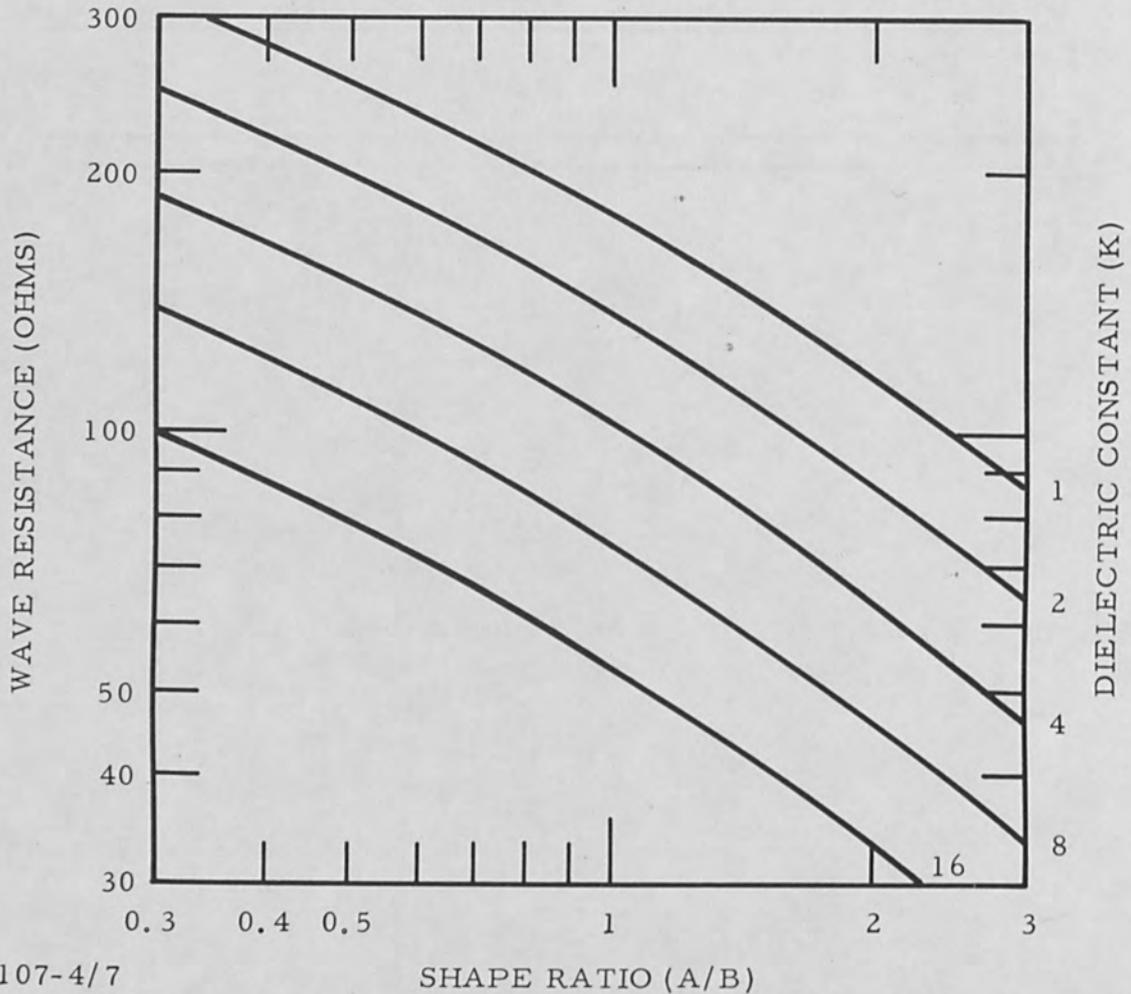


Figure 5. Capacitance Bridge and Test Fixture for Measuring Dielectric Constant.



04107-4/7

Figure 6. Chart for Computing Dielectric Constant (From Wheeler²).

if z_2 is 100 ohms, and shape ratio is 0.595, k is 8.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Exposure of various materials to high-energy electron pulses produces a temperature effect on the material that is predominantly transient. Temperatures on the surface and in the interior of the substrate material are not allowed to attain equilibrium during electron beam exposure. There are undoubtedly secondary effects associated with the electron beam due to the vacuum environment, such as degassing or phase changes permitted by the reduced pressure. To understand the mechanism that predominates during electron beam exposure, an attempt was made to characterize temperature conditions during these short exposures.

A. Thermal Analysis of Electron Beam Exposure Cycles and Transient Temperatures Developed

In applying electron beam technology to welding, melting, or joining metals, other investigators have attempted to estimate surface temperatures developed by an electron beam.^{3,4,5} Some work has also been done in the thermal analysis of temperatures developed by electron beams in stationary and moving x-ray targets. This prior work by others has been restricted to one-dimensional analysis with respect to time or simply to estimating maximum surface temperatures as they relate to metals.

Our work with dielectric and semiconductor materials indicated that a more sophisticated thermal analysis is required to characterize the effects of electron beams on a spectrum of materials having a range of thermal conductivities.

The goal of our work has been to express temperature phenomena in terms of a two-dimensional model (radius and penetration) with respect to time. An analytical mathematical approach was attempted which, besides applying to only a narrow range of cases, proved extremely complicated.

The following assumptions were made:

1. Thermal conductivity of the sample remains constant with temperature.
2. Radiation losses are linear with respect to temperature.
3. Sample material rests on an infinite heat sink during exposure to the beam.

4. Beam power is focused on a slab of infinite radius insofar as thermal effects are concerned.

This technique has several advantages:

1. The effect of temperature on materials with different thermal conductivity can be determined.
2. Temperatures developed are the maximum that would be expected.
3. Pulse geometry can be altered and related to different temperature profiles it produces.
4. The power distribution in the beam spot can be varied to simulate actual experimental conditions.

Figure 7 is a schematic drawing of the model used to represent conditions that would be observed with exposures of pulsed electron beam energy. To illustrate the method, an isothermal plot of the heat distribution developed in TiO_2 is shown just prior to cut-off of beam energy (Figure 8).

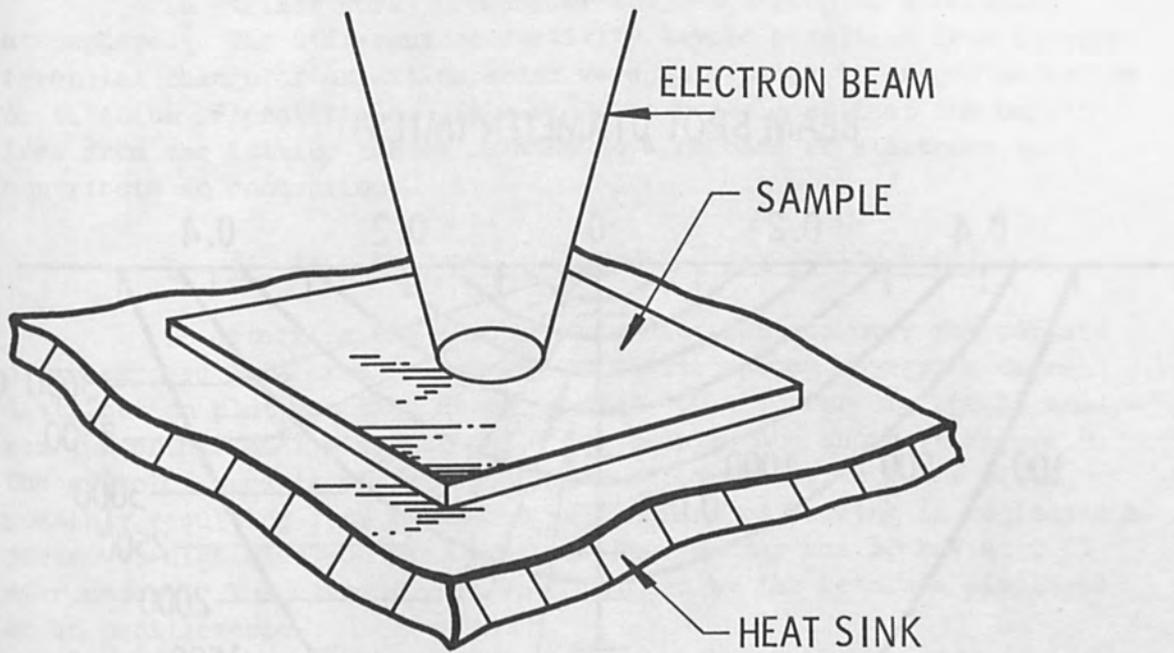
B. Effect of Electron Beam Energy on Dielectric Materials

Several approaches were used to study the effects induced by electron beam energy on the various dielectric materials. A series of tests utilizing the residual gas analyzer were used to determine the oxygen outgassing characteristics of Pyrex, TiO_2 , quartz, and sapphire. The literature was surveyed for clues to how TiO_2 would be affected by the thermal heating and vacuum involved in electron beam exposure. Following exposure the structure of the TiO_2 was examined by selected analytical methods such as electron microprobe analysis and electron spin resonance.

1. Background

Some researchers have reported that bulk TiO_2 is transformed into a highly conductive blue-black material in the presence of a reducing atmosphere.⁵

Huber was able to alter the oxidation state of evaporated titanium films and thereby obtain a corresponding change in the films' electrical conductivity.⁶ The conductivity changes were not preferential, but were a conversion of the titanium film surface to another oxidation state. Conductivity of these oxidized titanium films varied from 1 to 10^{13} ohm-cm.



04288-10/11

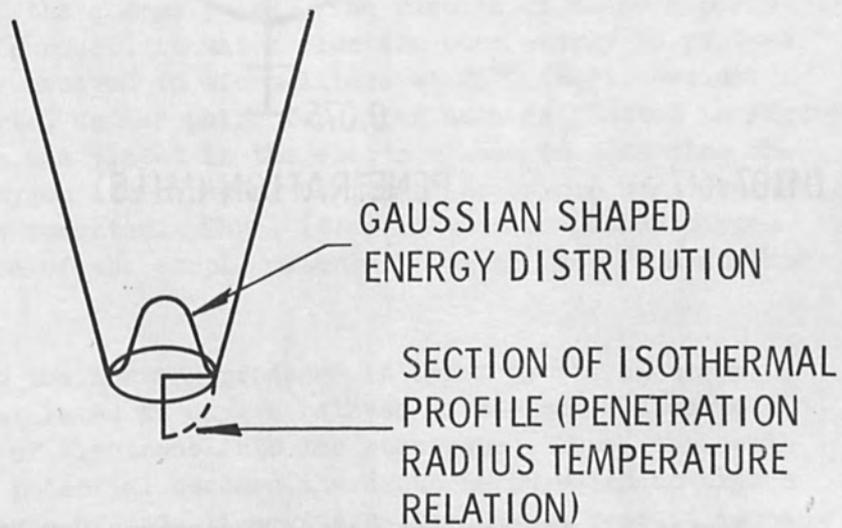
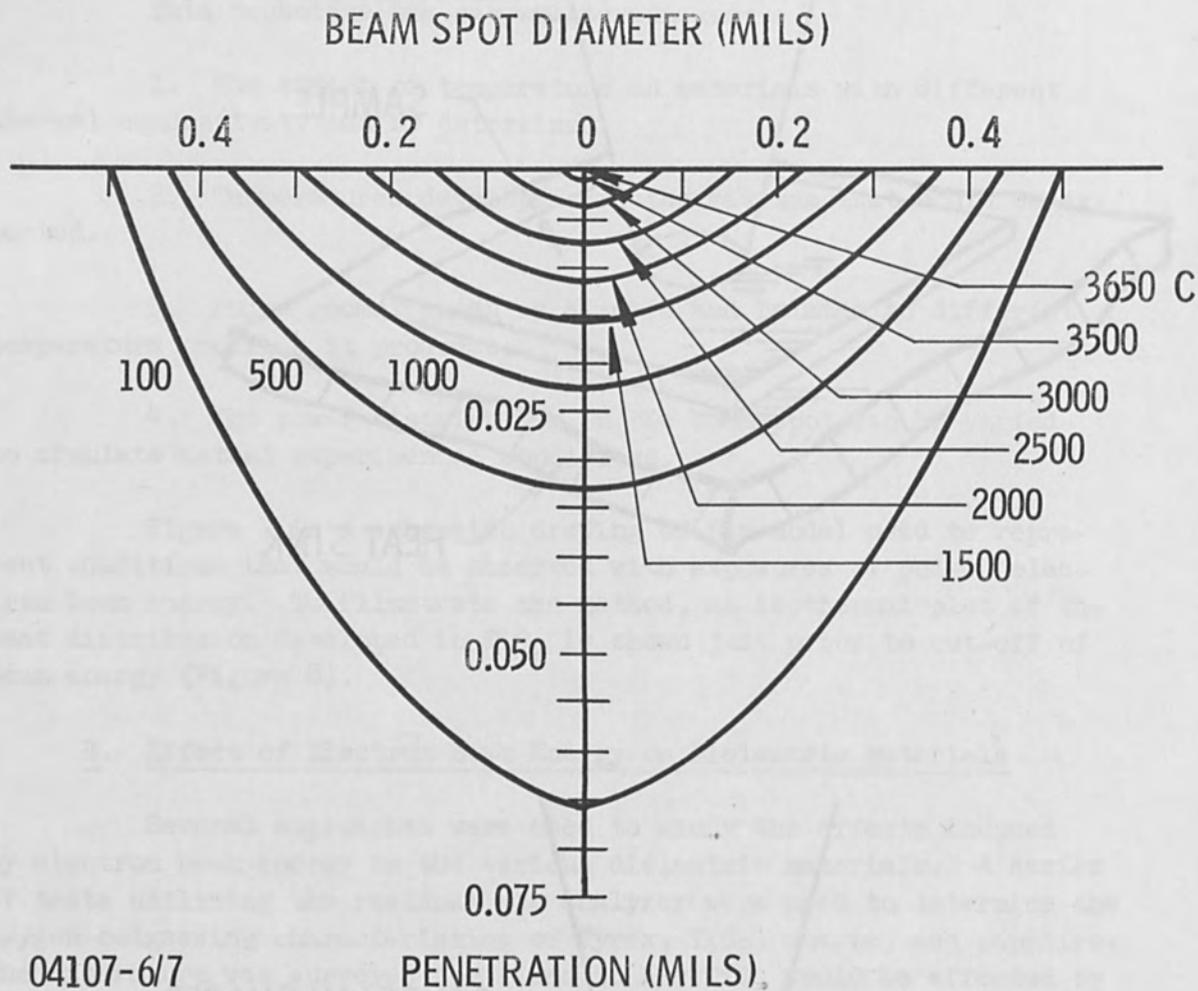


Figure 7. Diagram of Mathematical Thermal Analysis Model.



04107-6/7

PENETRATION (MILS).

Figure 8. Maximum Estimated Temperature Development by a 5-Microsecond Electron Beam Pulse (Measurement Made 0.5 Microsecond Before Beam Cutoff).

In earlier work, Cronmeyer subjected TiO_2 to a reducing atmosphere.⁷ The different conductivity levels resulting from non-preferential change of oxidation state were attributed to oxygen vacancies or titanium interstitials. In effect, he postulated that the oxygen loss from the lattice can be related to a release of electrons that contribute to conduction.

2. Electron Microprobe (Current Distribution Plot)

To determine the conductivity distribution over the surface of an exposed TiO_2 area (subjected to electron beam energy) a current distribution plot was made using a model 400 electron microprobe analyzer (Materials Analysis Company). The results are shown in Figure 9. The affected area is of fairly uniform appearance except for voids, possibly resulting from excessive melting and refreezing in regions subjected to high electron beam energy. Beam energy was 30 keV at 0.03 microampere. The current pattern developed by the beam was displayed on an oscilloscope.

3. Oxygen Outgassing Test

A series of oxygen outgassing tests were conducted using the residual gas analyzer. The electron beam was pulsed for a 1-second cycle in various dielectric materials. Beam power level was changed by changing the wave shape of the energy pulse. The results of these experiments are shown in Figure 10, in which electron beam energy is plotted against volume of gas evolved in micro-liters at 25°C (Hg). Besides the Pyrex, TiO_2 , quartz, and sapphire for which data is plotted in Figure 10, a tungsten target was placed in the electron beam to determine the amount of adsorbed oxygen liberated by heating. No oxygen was liberated at the surface of the tungsten. Thus, it seems that outgassed oxygen comes from the lattice of the sample materials rather than from surface adsorption.

In studying the heating produced in Pyrex by the electron beam, Lineweaver⁸ postulated an oxygen release mechanism related to the depth of penetration of electrons into the substrate. Thus, the depth of the most negative potential becomes the depth below which no oxygen is removed. In the case of soda-lime glasses the oxygen release mechanism described by Lineweaver would be related to the Na_2O concentration in the glass, which effectively ruptures the normal Si-O-Si bond to provide nonbridging oxygen atoms which can migrate to the surface and be discharged.

In a study of electron beam effects on the melting of tungsten metal, Schwarz⁹ postulated that vaporization of the tungsten would reduce the mass density and hence permit a greater electron penetration.

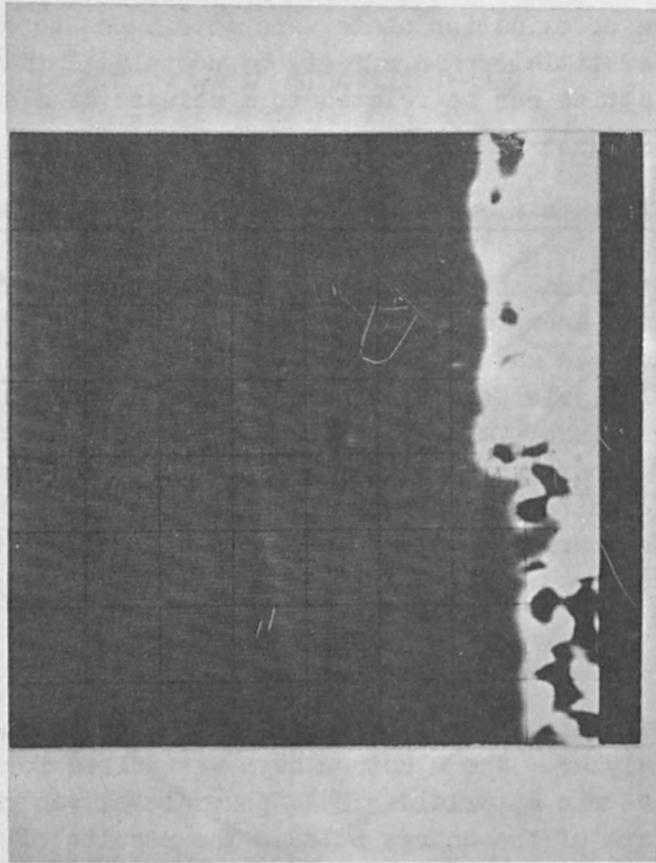
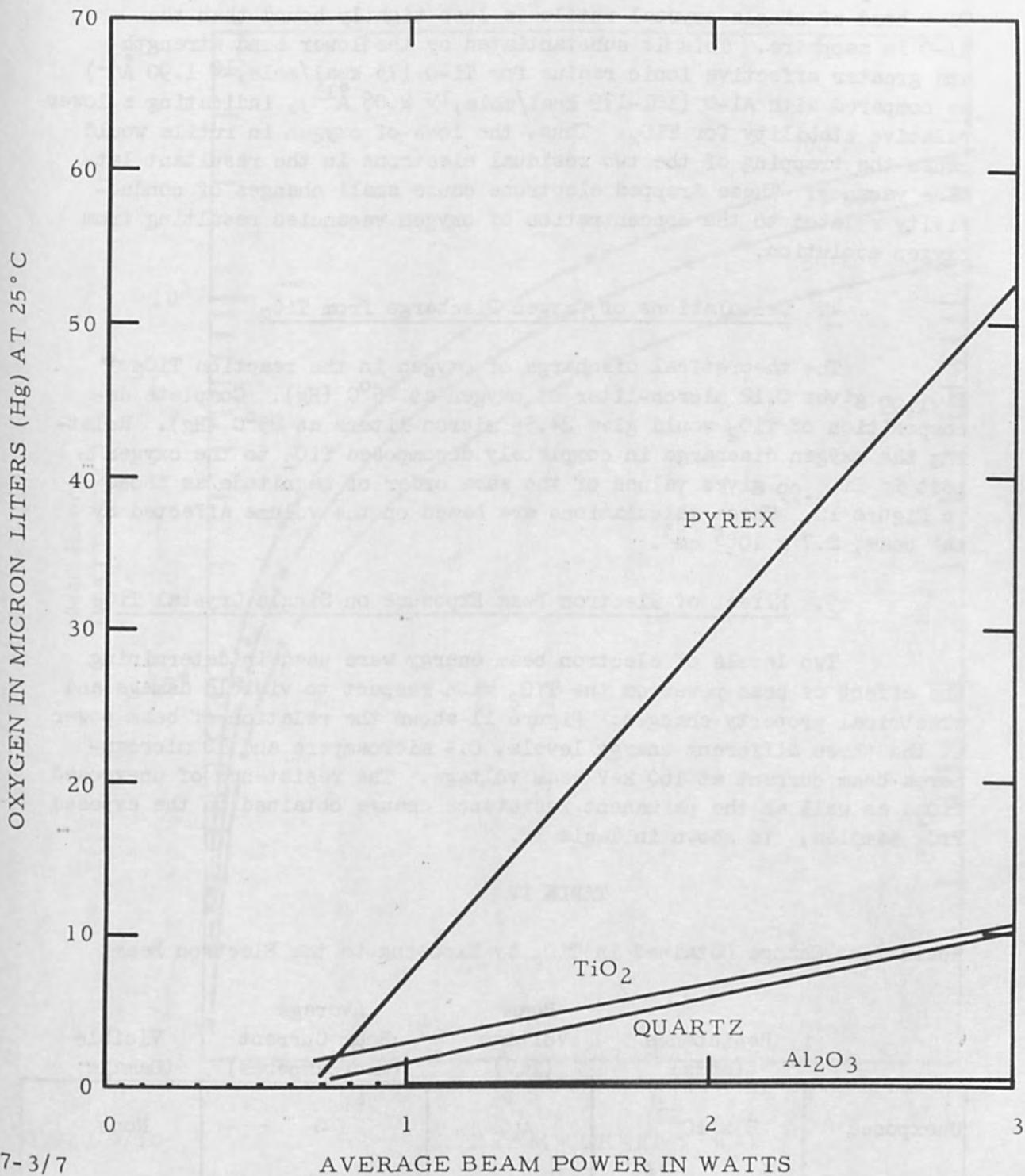


Figure 9. Electron Microprobe of Single Crystal TiO_2 Showing Conductivity of Area Exposed to the Electron Beam (Light Area is Conductive Material).



04107-3/7

Figure 10 Oxygen Outgassing Characteristics of Various Dielectric Materials Exposed to the Electron Beam

In our work, the preferential changes induced in TiO_2 by the electron beam appear to have an oxygen release mechanism similar to that postulated by Lineweaver for Pyrex glass. The oxygen in the Ti-O bond of single crystal rutile is less tightly bound than the Al-O in sapphire. This is substantiated by the lower bond strength and greater effective ionic radius for Ti-O (73 kcal/mole,¹⁰ 1.90 Å¹¹) as compared with Al-O (101-179 kcal/mole,¹⁰ 2.08 Å¹¹), indicating a lower relative stability for TiO_2 . Thus, the loss of oxygen in rutile would cause the trapping of the two residual electrons in the resultant lattice vacancy. These trapped electrons cause small changes of conductivity related to the concentration of oxygen vacancies resulting from oxygen evolution.

4. Calculations of Oxygen Discharge from TiO_2

The theoretical discharge of oxygen in the reaction $TiO_2 \rightarrow TiO_{1.99}$ gives 0.12 micron-liter of oxygen at 25°C (Hg). Complete decomposition of TiO_2 would give 24.55 micron-liters at 25°C (Hg). Relating the oxygen discharge in completely decomposed TiO_2 to the oxygen lost in $TiO_{1.99}$ gives values of the same order of magnitude as those in Figure 10. These calculations are based on the volume affected by the beam, $2.7 \times 10^{-5} \text{ cm}^3$.

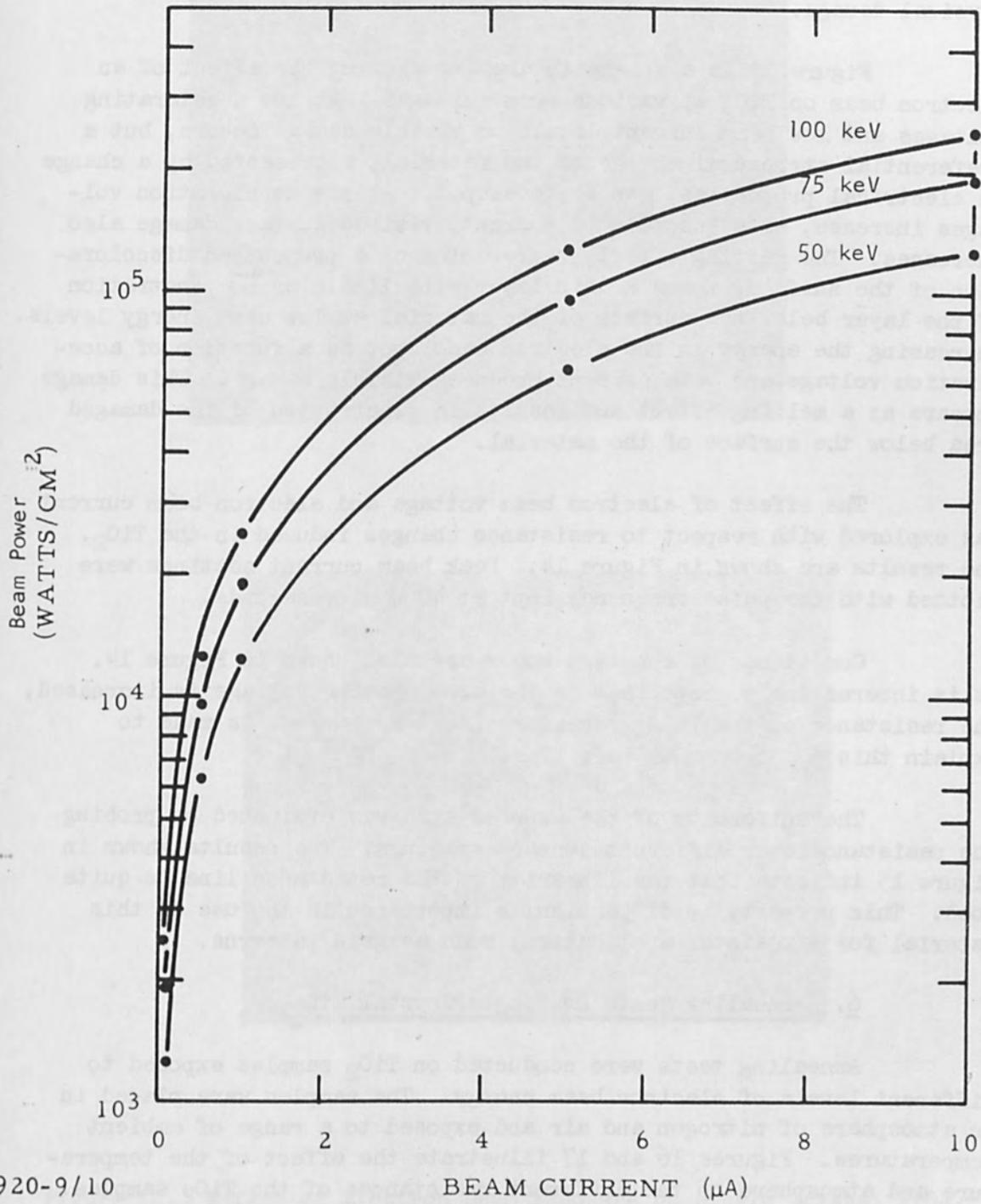
5. Effect of Electron Beam Exposure on Single Crystal TiO_2

Two levels of electron beam energy were used in determining the effect of beam power on the TiO_2 with respect to visible damage and electrical property changes. Figure 11 shows the relation of beam power at the three different energy levels, 0.4 microampere and 10 microamperes beam current at 160 keV beam voltage. The resistance of unexposed TiO_2 , as well as the permanent resistance change obtained on the exposed TiO_2 samples, is shown in Table IV.

TABLE IV

Resistance Change Obtained in TiO_2 by Exposing to the Electron Beam

	Resistance (ohms)	Beam Voltage (keV)	Average Beam Current (microamperes)	Visible Damage
Unexposed	2×10^{12}	0	0	None
Exposed	2.2×10^6	100	0.4	None
Exposed	2.0×10^4	100	10	Pronounced Damage



03920-9/10

Figure 11 Relation of Beam Power to Beam Current at Different Beam Voltages

Photographs of the morphology of the samples are shown in Figure 12. By proper selection of the beam power conditions it appears that controlled changes can be obtained in TiO_2 without inflicting appreciable physical damage.

Figure 13 is a schematic drawing showing the effect of an electron beam on TiO_2 at various beam voltages. At low accelerating voltages and low beam current levels no visible damage occurs, but a preferential structural change of the material, represented by a change in electrical properties, can be developed. As the acceleration voltages increase, as a function of current, visible surface damage also increases. The melting effect, represented by a pronounced discoloration of the surface, forms a thin layer with little or no penetration of the layer below the surface of the material at low beam energy levels. Increasing the energy in the electron beam spot as a function of acceleration voltage and beam current produces visible damage. This damage appears as a melting effect and results in penetration of the damaged area below the surface of the material.

The effect of electron beam voltage and electron beam current was explored with respect to resistance changes induced in the TiO_2 . The results are shown in Figure 14. Peak beam current settings were plotted with the pulse frequency kept at 4000 microseconds.

Conditions of constant power are also shown in Figure 14. It is interesting to note that as the accelerating voltage is increased, the resistance of the TiO_2 increases also. No attempt is made to explain this phenomenon at this time.

The uniformity of the exposed line was evaluated by probing the resistance over different contact spacings. The results shown in Figure 15 indicate that the linearity of the resistance line is quite good. This property is of particular importance in the use of this material for a resistor application, such as grid patterns.

6. Annealing Tests on Single Crystal TiO_2

Annealing tests were conducted on TiO_2 samples exposed to different levels of electron beam energy. The samples were placed in an atmosphere of nitrogen and air and exposed to a range of ambient temperatures. Figures 16 and 17 illustrate the effect of the temperature and atmosphere on the different resistances of the TiO_2 samples. Examination of these figures indicates the following:

1. The higher the resistance of the line exposed to the electron beam the more likely it is to be changed by heat and an oxidizing atmosphere.

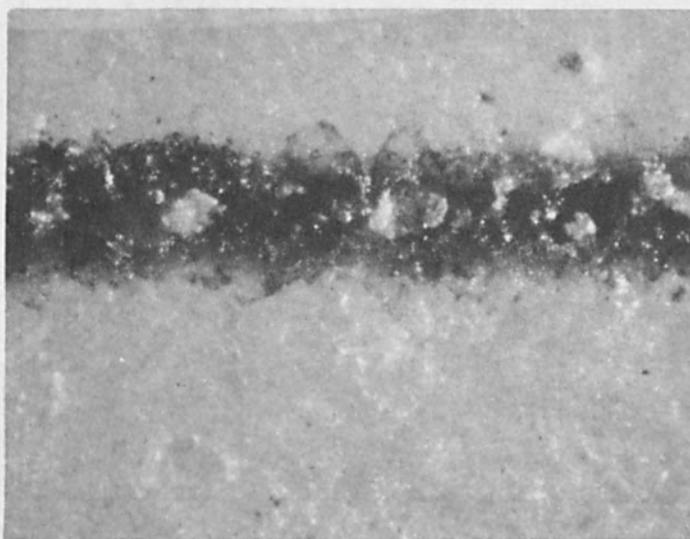
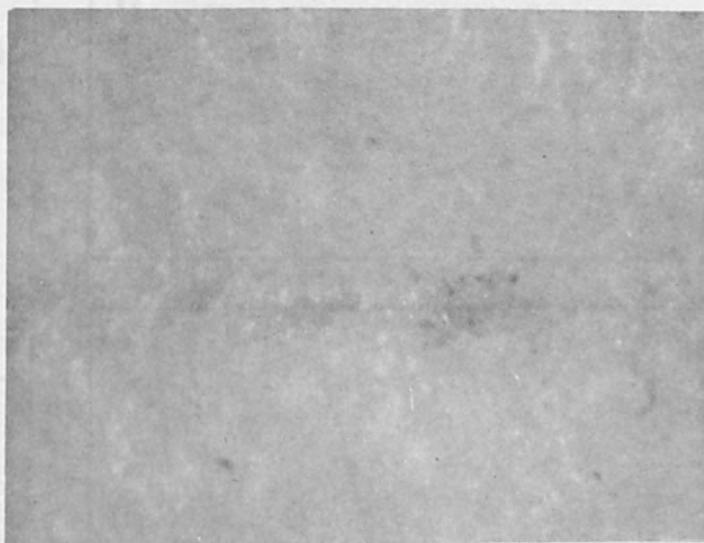
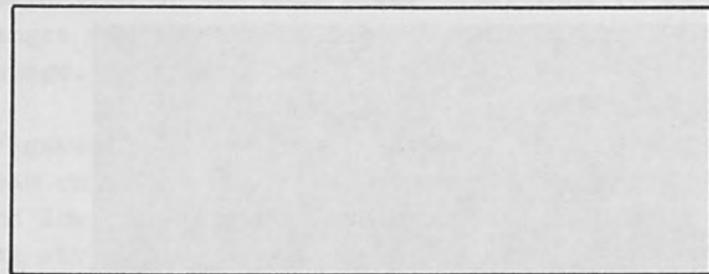
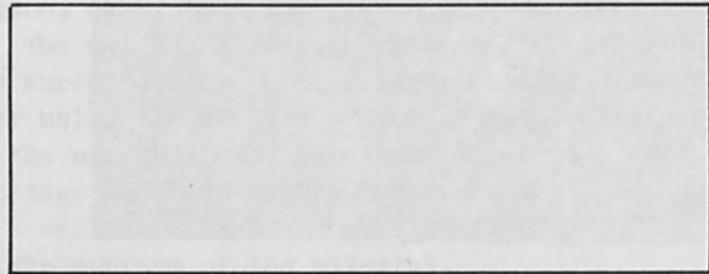


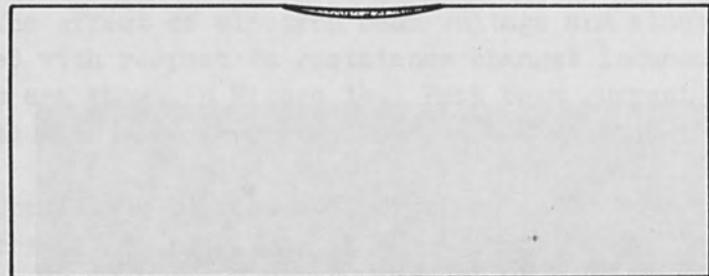
Figure 12. Effect of Electron Beam Exposure on Single Crystal TiO_2 .



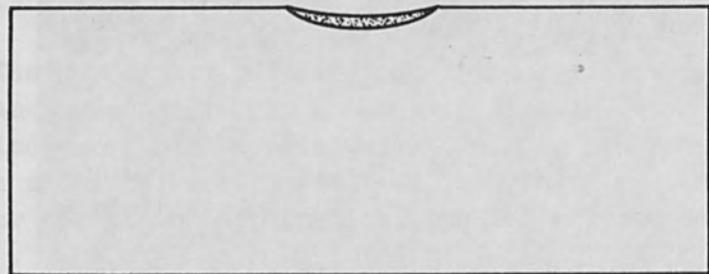
UNDAMAGED - NO
ELECTRON BEAM
EXPOSURE



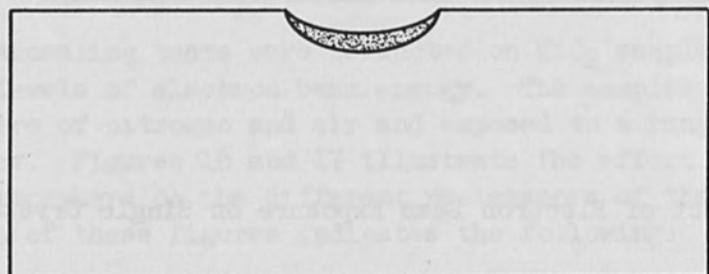
0.5 TO 20 KEV f(I) NO
VISIBLE DAMAGE



50 KEV f(I) VISIBLE
SURFACE DAMAGE



75 KEV f(I) VISIBLE
SURFACE DAMAGE -
MELTING EFFECT



100 KEV f(I) VISIBLE
SURFACE DAMAGE -
MELTING EFFECT

03920-6/10

Figure 13 Effect of Electron Beam on TiO_2 at Different Beam Voltages
(Schematic Representation)

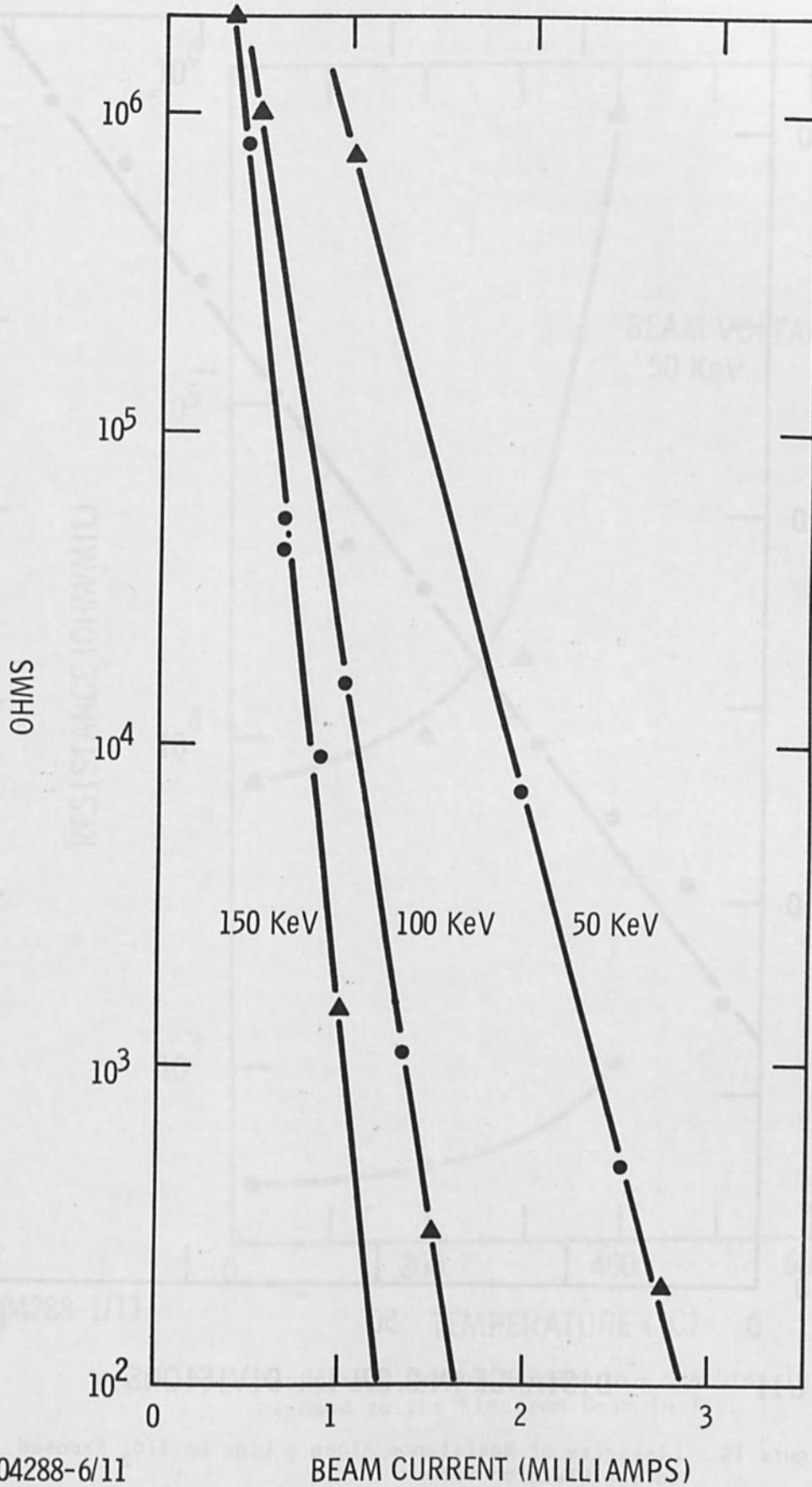


Figure 14 Electron Beam Current Related to Resistance Produced in TiO₂ at Various Voltage Settings

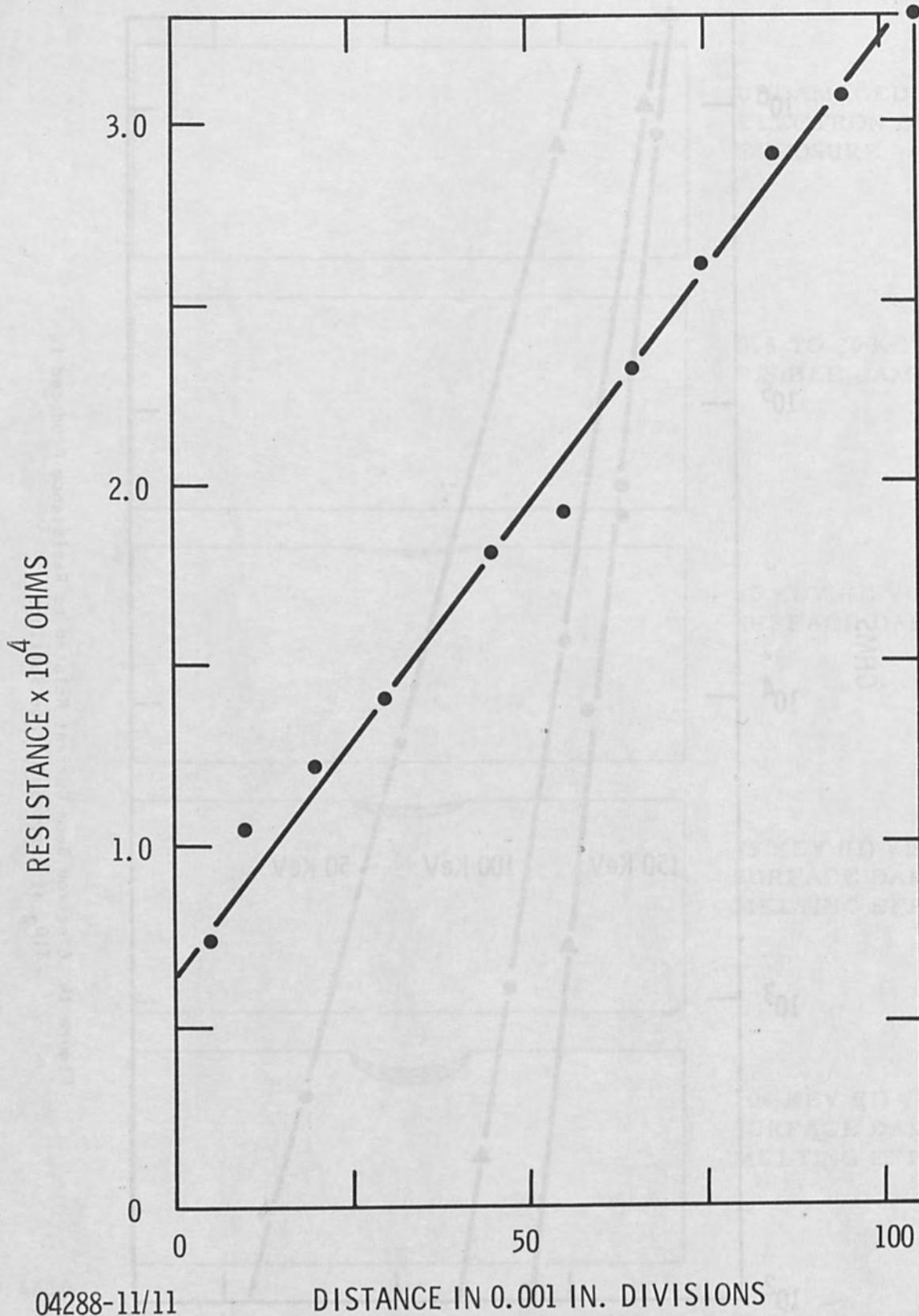
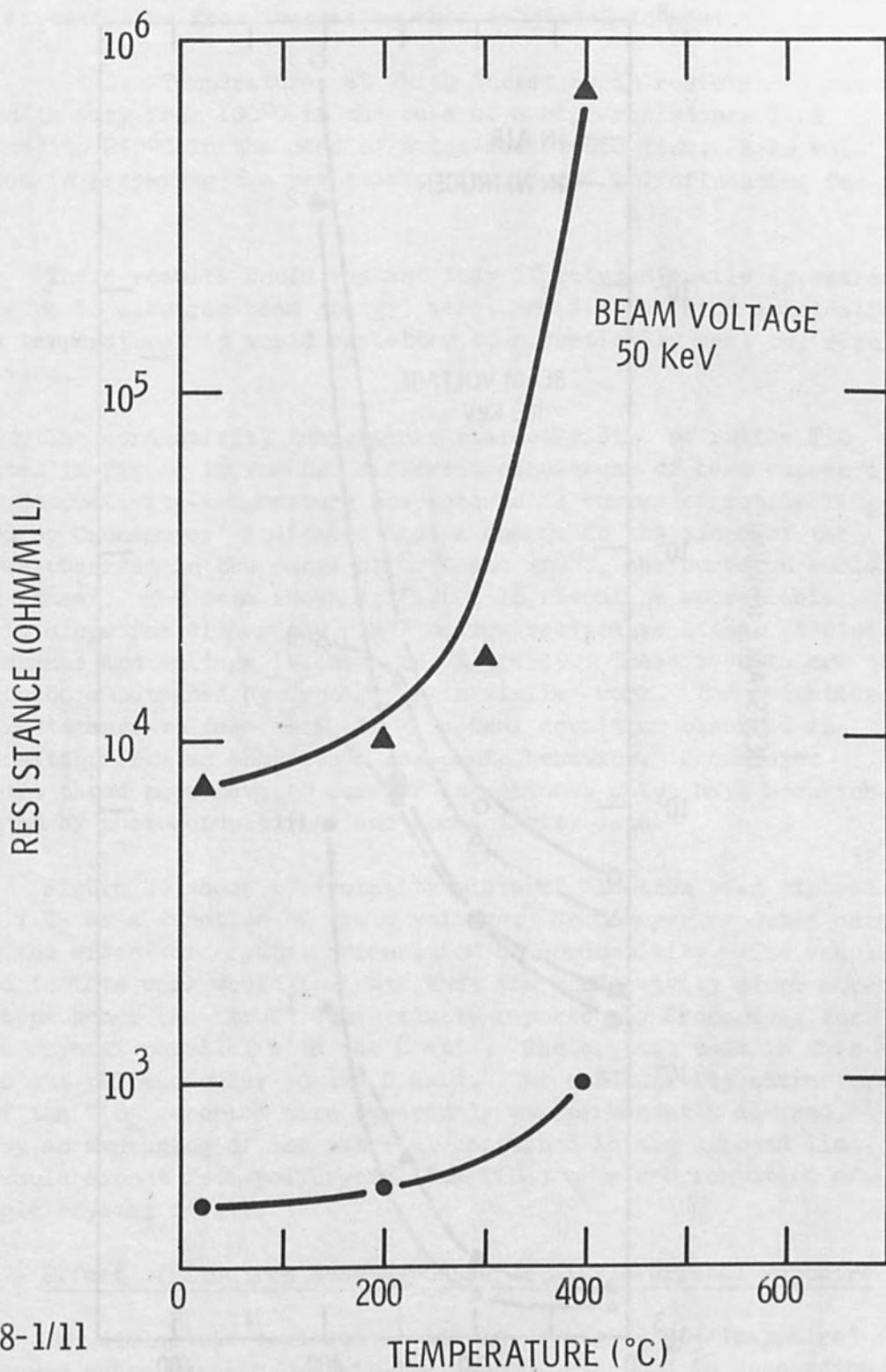
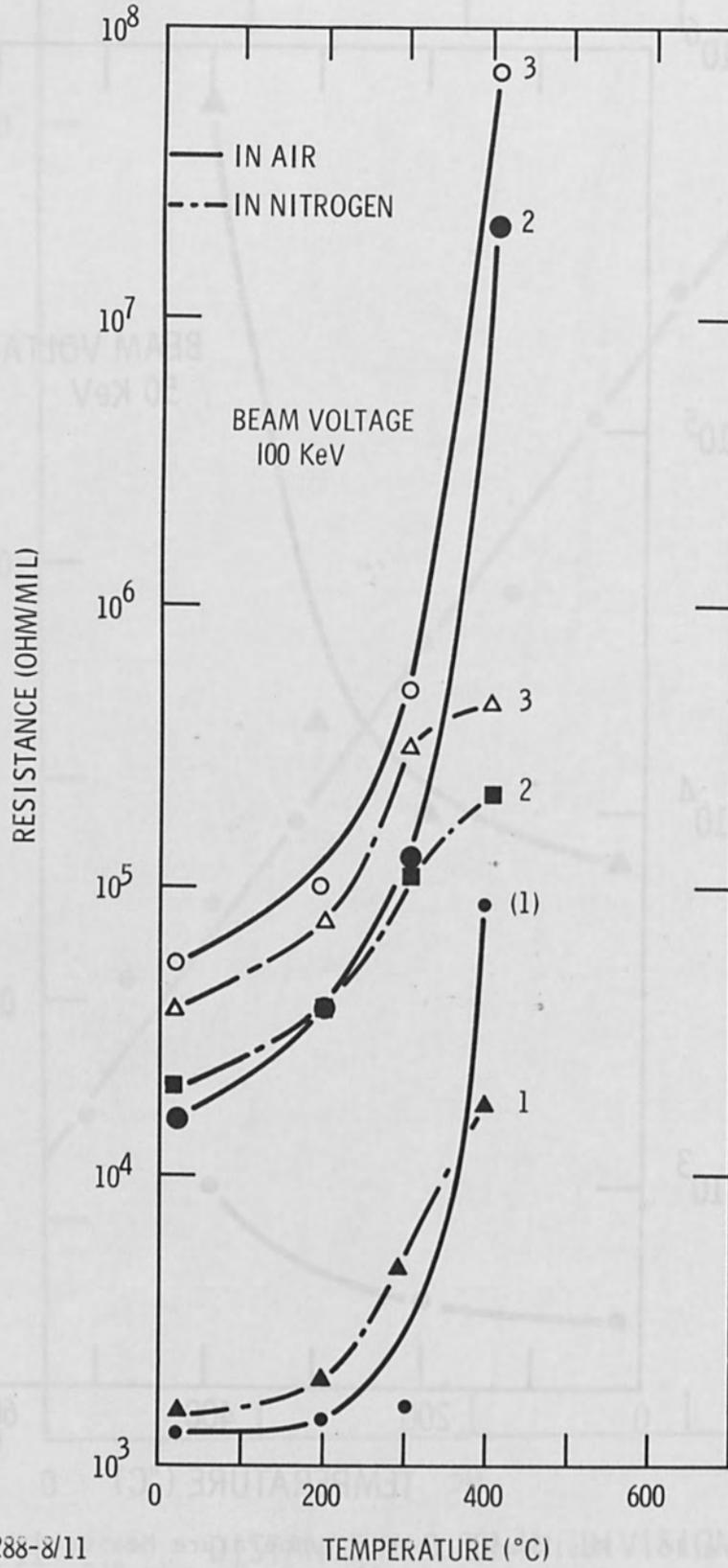


Figure 15 Linearity of Resistance Along a Line on TiO₂ Exposed to the Electron Beam



04288-1/11

Figure 16 Resistance Versus Temperature Measured in Lines Exposed to the Electron Beam in TiO_2



04288-8/11

Figure 17 Resistance Versus Temperature in Air Measured in Lines Exposed to the Electron Beam in TiO_2

2. The nitrogen helps slightly in keeping the resistance of the exposed line from increasing when subjected to heat.

3. Temperatures at which increases in resistances can be observed to vary from 100°C in the case of a high-resistance line (10^4 ohms) to 250°C in the case of a low-resistance line. Beam voltage used in preparing the resistance line is not an influencing factor.

These results would suggest that if reduced rutile (prepared by exposure to electron beam energy) were used for resistor application at high temperature, it would be better to hermetically seal the resulting device.

The conductivity temperature characteristic of rutile TiO_2 is plotted in Figure 18 for two different conditions of beam exposure level. Conductivity-temperature characteristic curves of rutile TiO_2 reported by Cronemeyer⁷ indicated that a change in the slope of the curve was observed in the range of 60°C and 160°C, the curve on cooling was linear. The data shown in Figure 18 reveal no appreciable change in slope for either the high- or low-resistance lines. A plot of resistance and voltage is shown in Figure 19. These results are similar to those obtained by Cronemeyer in similar work. The deviation in the resistance-voltage slope is a typical condition observed in the transition from an ohmic to a non-ohmic behavior. Cronemeyer attributes these phenomena to carrier injections, which have been substantiated by photoconductivity and conductivity data.

Figure 20 shows conductivity plots of electron beam exposed line in TiO_2 as a function of field voltage. Cronemeyer reported data showing the effect of crystal orientation on conductivity. The results achieved in this work would indicate that the conductivity slope more nearly approaches the "true" conductivity reported by Cronemeyer for a single crystal parallel with the C axis. The crystal used in this work was cut perpendicular to the C axis. The conductivity characteristic of the TiO_2 reported here apparently was permanently altered, either by an averaging of the material contained in the exposed line, as one would expect from polycrystalline TiO_2 , or a reorientation of the single crystal itself.

C. Effect of Electron Beam Exposure on Single Crystal Sapphire

The dielectric constant of single crystal Al_2O_3 (sapphire) was measured using the micro-stripline method described in Subsection II.E.

The sapphire samples were first exposed to electron beam

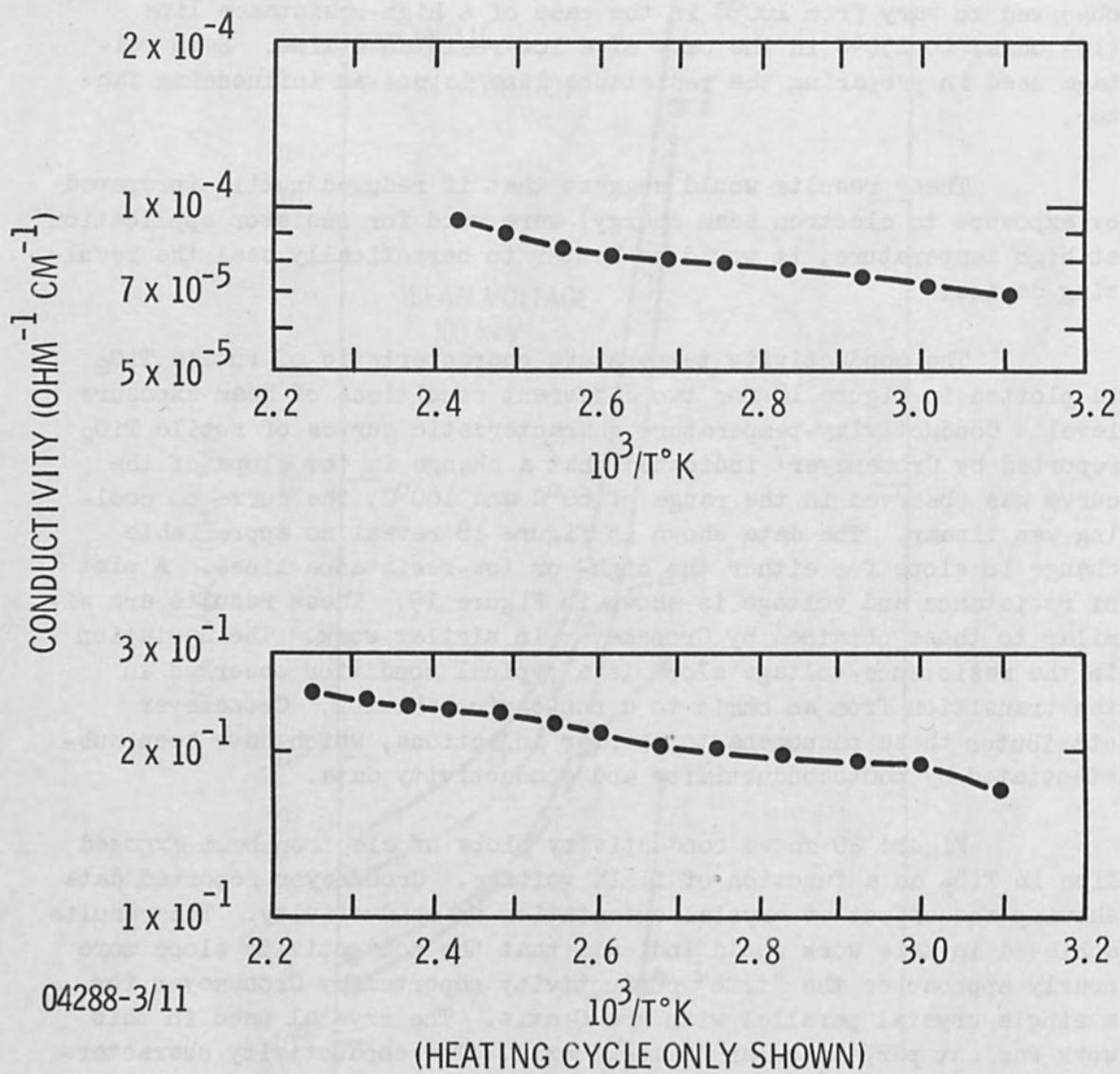


Figure 18 Conductivity-Temperature Characteristic of TiO_2 Exposed to the Electron Beam

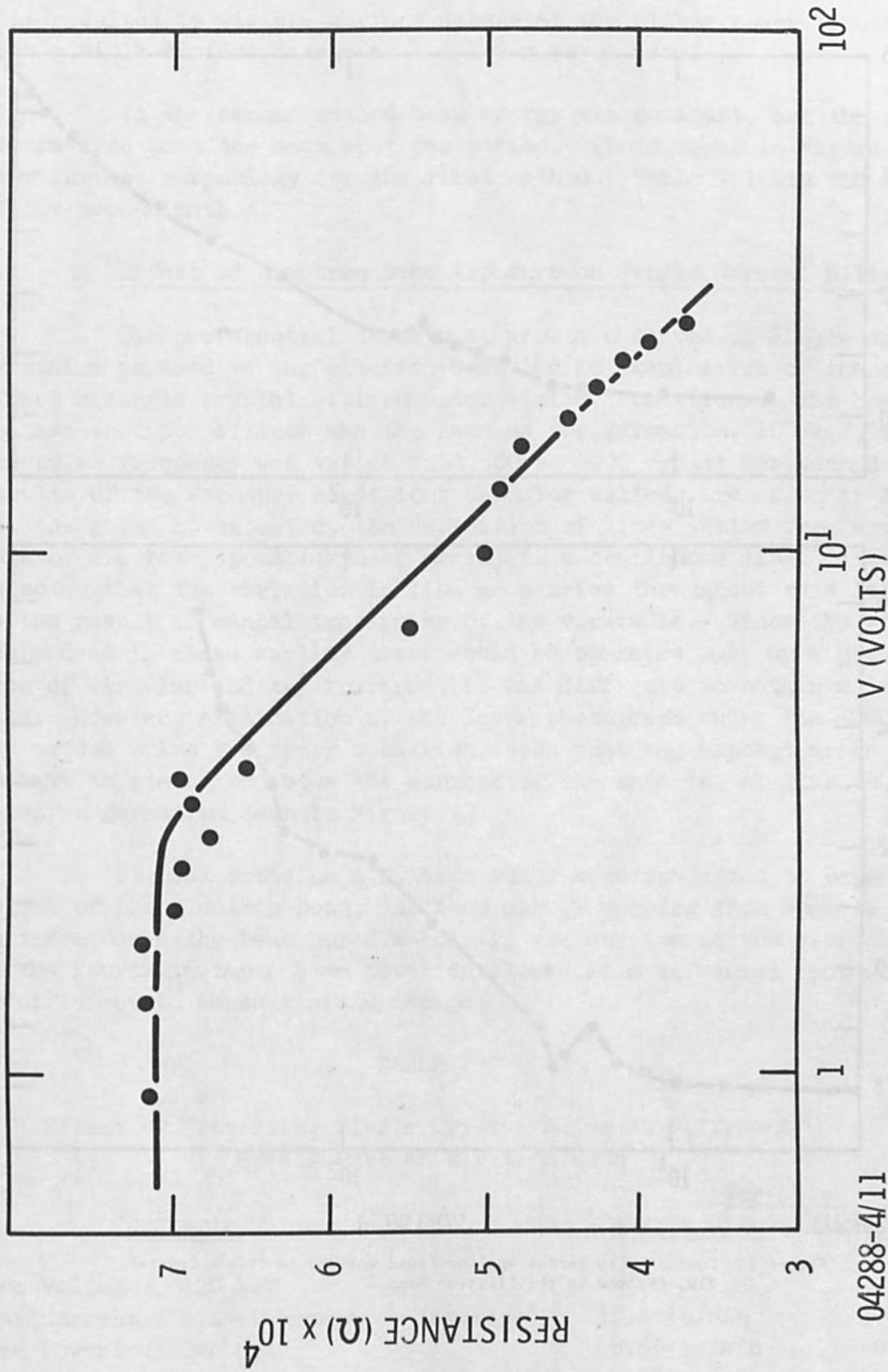
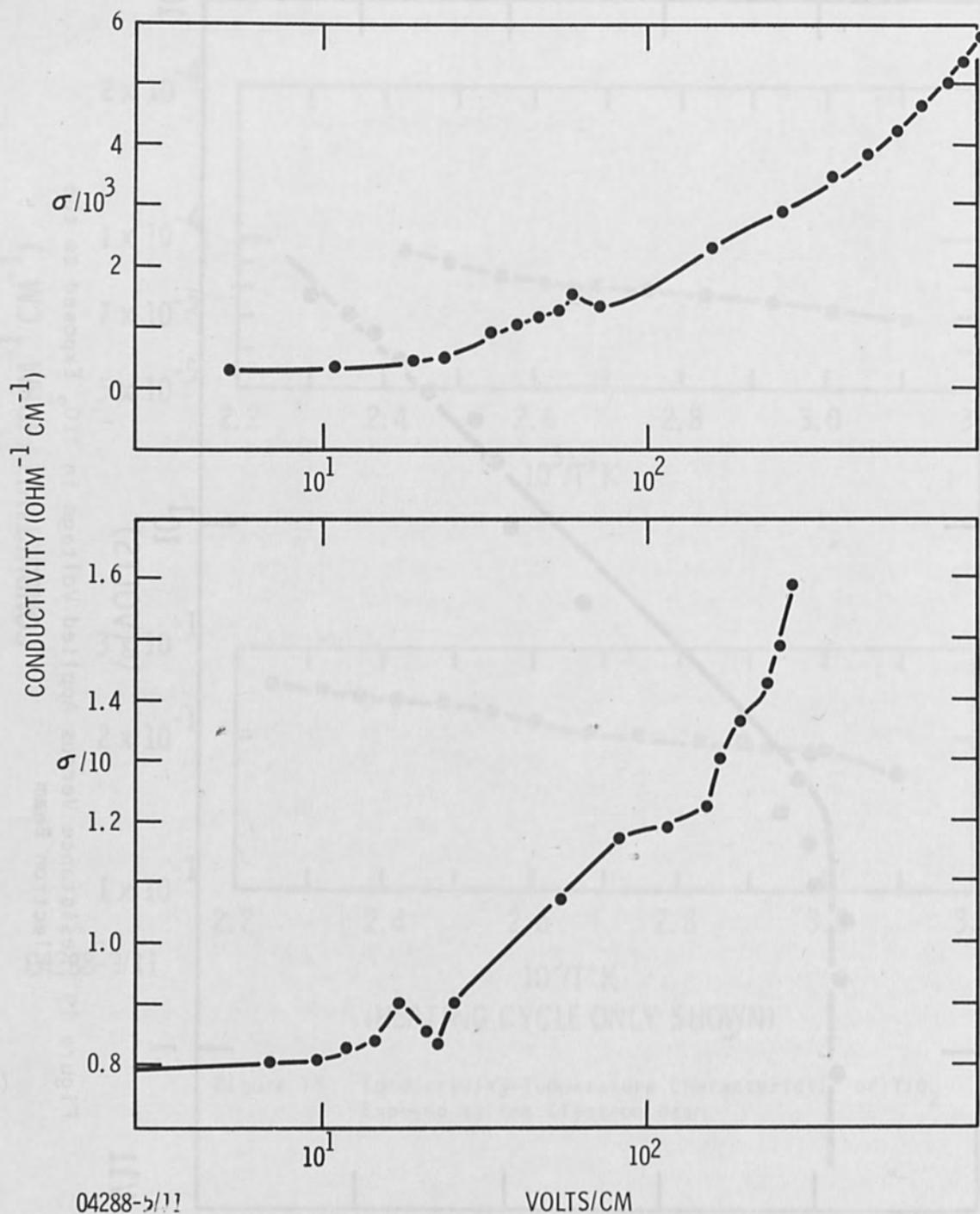


Figure 19 Resistance Versus Applied Voltage in TiO_2 Exposed to the Electron Beam



04288-5/11

Figure 20 Conductivity Versus Applied Field Voltage in Single Crystal TiO_2 Exposed to the Electron Beam

energy by one of two methods. In the first the sapphire was traversed past the beam spot at a constant rate as beam energy was increased. This resulted in visible surface damage at the higher power levels and less visible surface damage at the lower power level.

In the second method beam energy was constant, but the rate of traverse past the beam spot was varied. Photographs in Figure 21 show surface morphology for the first method. Table V lists the results of the second method.

D. Effect of Electron Beam Exposure on Single Crystal Silicon

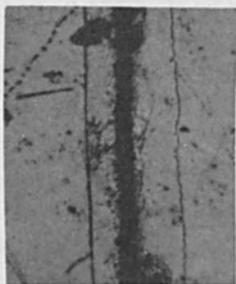
The preferential structural growth observed on single crystal germanium exposed to the electron beam led to examination of the beam's effect on single crystal silicon under similar conditions. The beam voltage used for silicon was the same as for germanium, 100 keV, and the pulse frequency was varied from 500 to 5000 cycles per second. Results of the exposure conditions used for silicon are shown in Figure 24. As might be expected, the definition of lines varies from a pattern of discrete spots (exposed areas) to a continuous line. It should be noted that the variation in line geometries throughout this length is the result of manual traversing of the worktable. Since the worktable used in these earlier tests could be operated only by a combination of circular and x-y traverse, it was difficult to obtain a straight line. However, examination of the lower photograph under the 500 cycle per second pulse frequency condition shows that the exposed areas appear to have an elevation above the surface of the material similar to the areas in germanium seen in Figure 22.

Several spots on a silicon slice were subjected to repeated pulses of the electron beam, the beam energy varying from spot to spot. In three cases the beam bored a hole in the surface of the silicon. In the fourth instance beam power developed in a defocused spot was insufficient to cause visible damage.

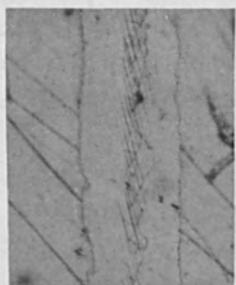
TABLE V

Effect of Traversing Single Crystal Al_2O_3 at Different Speeds Past an Electron Beam

<u>Constants</u>	<u>Variables</u>	
	<u>Traverse Speed</u>	<u>Dielectric Constant</u>
Beam Voltage: 120 keV		
Beam Current: 6.1 microamperes (average)	13.4 in/min	10.2
Beam Power: 243 MW/cm ²	8.65 in/min	9.5
Spot Diameter: 1.54 mils	4.13 in/min	8.7



(a)



(b)

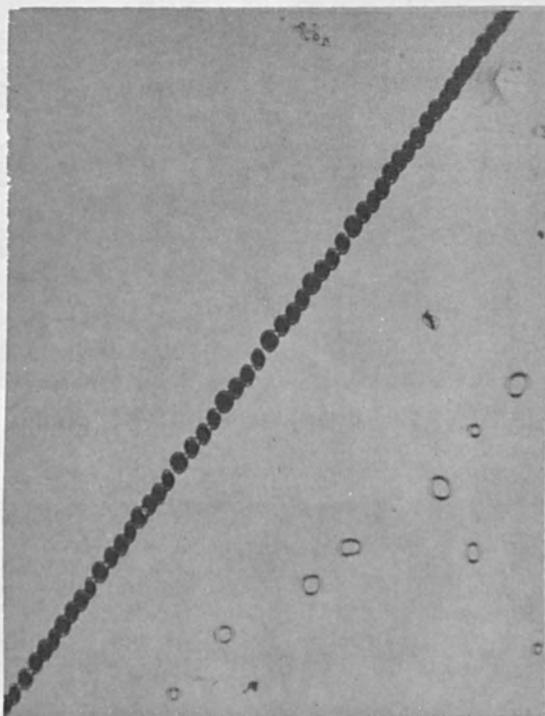
Constants

Beam Voltage: 120 keV
Beam Current: 6.1 microamperes (average)
Traverse Speed: 4.3 in/min

Variables

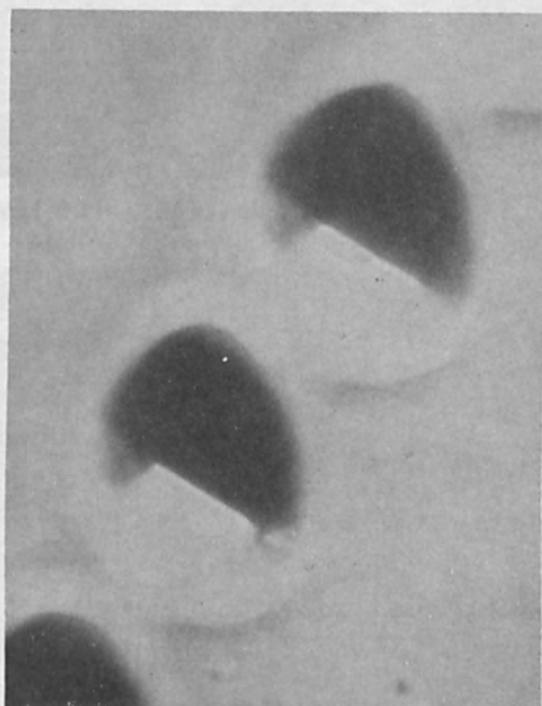
<u>Beam Power</u>	<u>Spot Diameter</u>	<u>Dielectric Constant</u>
(a) 0.375 MW/cm ²	15.6 mils	10.6
(b) 243 MW/cm ²	2.4 mils	10.0

Figure 21 Effect of Electron Beam on Sapphire at Different Power Levels



Ge <110> 0.005 ohm-cm
Beam Voltage: 100 keV
Beam Current: <1 μ A

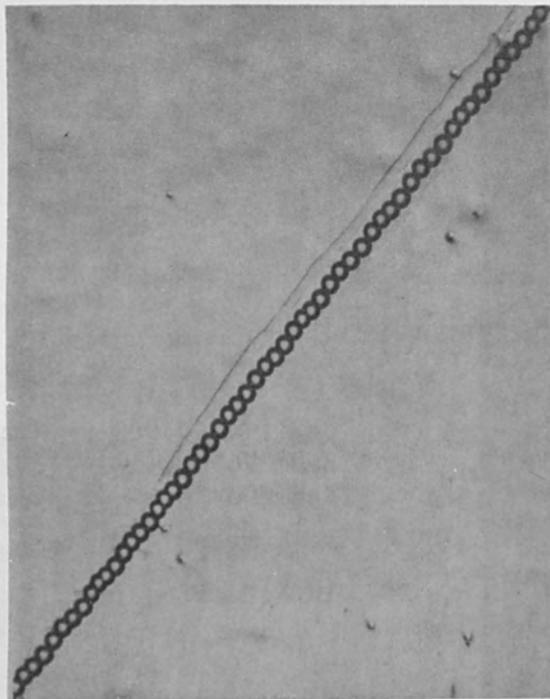
(67X)



Ge <110> 0.005 ohm-cm
Beam Voltage: 100 keV
Beam Current: <1 μ A

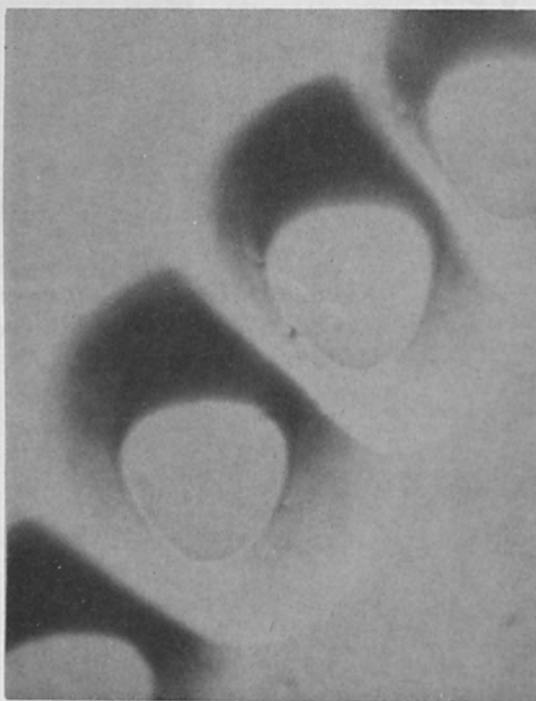
(1200X)

Figure 22 Effect of Electron Beam Exposure on Single Crystal Germanium



Ge $\langle 111 \rangle$ 0.005 ohm-cm
Beam Voltage: 100 keV
Beam Current: $< 1 \mu\text{A}$

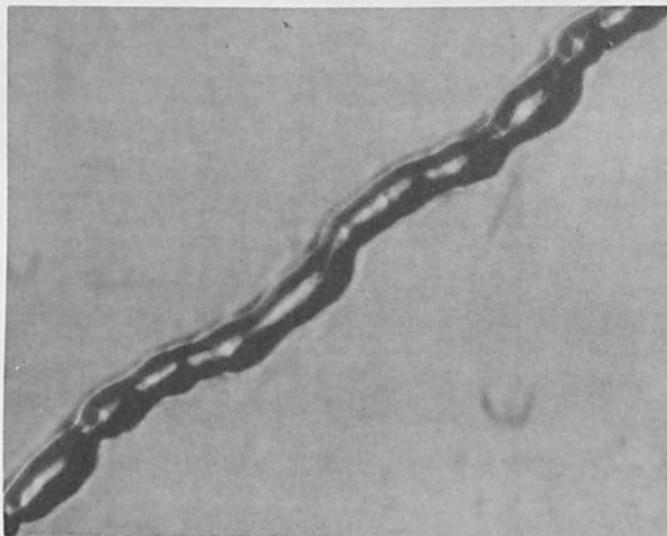
(67X)



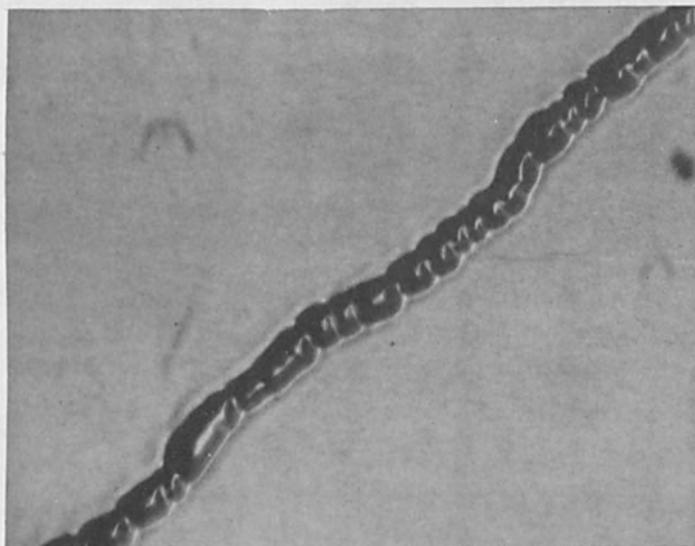
Ge $\langle 111 \rangle$ 0.005 ohm-cm
Beam Voltage: 100 keV
Beam Current: $< 1 \mu\text{A}$

(1200X)

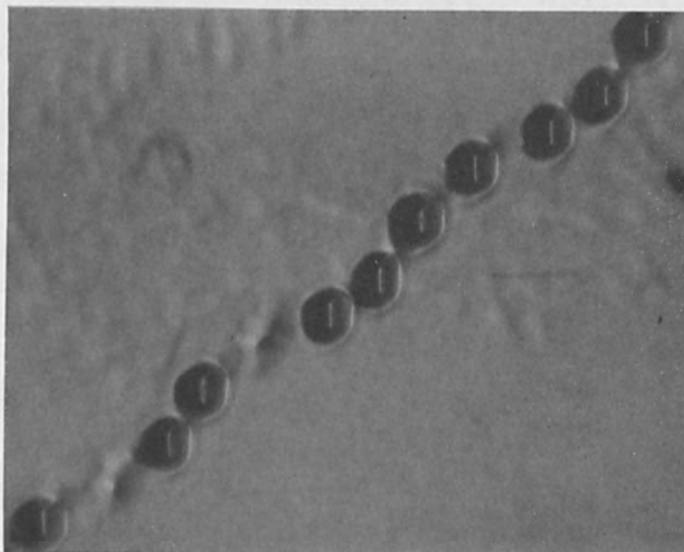
Figure 23 Effect of Electron Beam Exposure on
Single Crystal Germanium



Si $\langle 110 \rangle$
Beam Voltage: 100 keV
Beam Current: $< 1 \mu\text{A}$
Pulse Freq: 5000 cps
(620X)



Si $\langle 110 \rangle$
Beam Voltage: 100 keV
Beam Current: $< 1 \mu\text{A}$
Pulse Freq: 2000 cps
(620X)

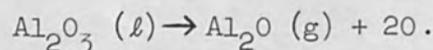


Si $\langle 110 \rangle$
Beam Voltage: 100 keV
Beam Current: $< 1 \mu\text{A}$
Pulse Freq: 500 cps
(620X)

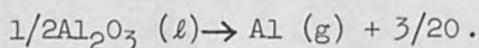
Figure 24 Effect of Electron Beam Exposure on Single Crystal Silicon
(All Conditions the same Except Pulse Frequency Varied)

Each condition of electron beam energy changed the dielectric constant of the material. Either of two theories could explain this.

In the first instance the temperatures developed during the exposure to electron beam energy are more than sufficient to vaporize the Al_2O_3 . The principal products in this reaction would be



In addition, we could also have, according to Ackermann¹²



By vaporization of the Al_2O_3 it would be possible in the short energy pulse to generate aluminum vapor with a possible condensation and diffusion into the lattice of the Al_2O_3 . This would further increase conductivity and reduce the dielectric constant.

In the second instance, Al_2O_3 exhibits anisotropic properties with respect to dielectric constant, as shown in Table VI.

TABLE VI

Anisotropic Properties of Single Crystal Al_2O_3
As Shown by Variation in Dielectric Constant¹³

<u>Crystal Orientation</u>	<u>Dielectric Constant</u>
Field Parallel to Optical Axis	10.55
Field Perpendicular to Optical Axis	8.6

Note - In such a circumstance, the change in dielectric constant could result from a melting and recrystallization of Al_2O_3 in a polycrystalline layer near the surface of the Al_2O_3 to produce an averaging effect in the dielectric constant with respect to the two orientations of the crystal.

E. Effect of Electron Beam Exposure on Single Crystal Germanium

Single crystal germanium slices were prepared, as noted in Table II, by lapping, polishing, and chemically etching to remove mechanical damage. The slice was then mounted on the worktable, and

no masks (such as KMER or SiO_2) or overlays were used. Beam voltage was 100 keV, and beam current was less than 1.0 microampere.

Two slices of germanium, oriented 110 and 111, were used. Figures 22 and 23 show that the beam energy was pulsed across the germanium surfaces in a fairly uniform line. Examination of the exposed areas at high magnification (1200X) revealed that the germanium had formed peaks. The geometry of the peaks is related to crystal plane orientation of the starting material. It is not clear from the appearance of the exposed areas whether the germanium melted and was forced into a peak or whether some other phenomenon had occurred.

Figure 25 shows these exposed areas. It appears that the constant high temperature allows the silicon vapor to escape before it can resolidify, thereby forming a hole. This phenomenon corresponds to effects commonly observed in melting or joining.

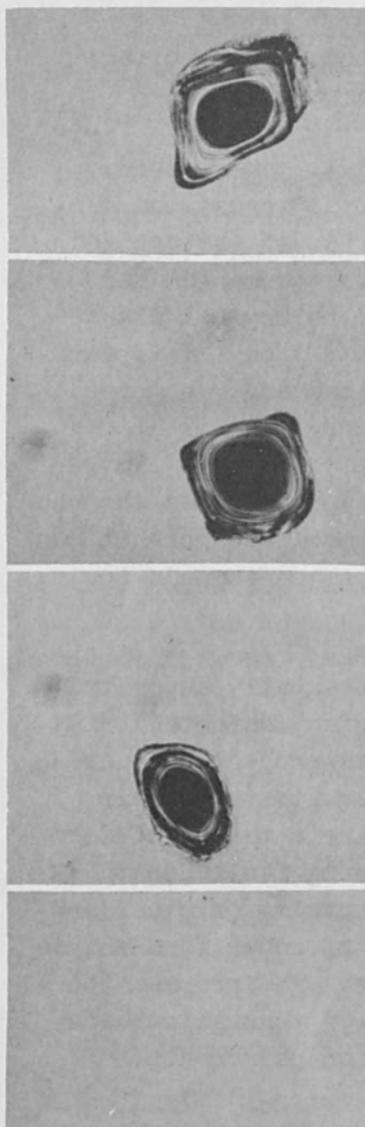
Under certain experimental conditions epitaxially deposited silicon contains defects which have been called surface defects, tripysramid, or raised triangle imperfections.^{14,15} These projections develop and propagate from the surface of the silicon substrate. Their characteristic geometry depends to some extent on their method of formation. In general, the projection result from stacking faults in the substrate crystal, contaminants on the substrate surface, or other surface imperfections. The defects exhibit a variety of nonuniform shapes depending on conditions associated with the surface. The propagation of the defects to form spikes or projections occurred through condensing of the silicon vapor by a preferential deposition phenomenon.

We believe that the peaks observed in our work may be formed through a related mechanism. Surface temperatures attained during short exposure to electron beam energy are substantially above the melting and vaporization temperature of either silicon or germanium. The initial energy pulse would be sufficient to create a surface defect. From there, vaporization and condensation followed by rapid cooling could freeze the silicon in a peak.

It is interesting also that spikes developing during epitaxial deposition of silicon are fed from an infinite source of silicon vapor. The peaks developing during our work are built up from the very limited amount of silicon vapor produced during the transient thermal effect induced by a pulsed electron beam.

IV. CONCLUSIONS

The work reported here demonstrates the usefulness of focused electron beam energy to produce controlled property changes in various electronic materials.



Beam Voltage: 100 keV
Beam Current: 6.41 microamperes (average)
Exposure: 2 sec
Spot Diameter: 12.8 mils
Hot Center: 4.25 mils

Beam Voltage: 100 keV
Beam Current: 6.41 microamperes (average)
Exposure: 1 sec
Spot Diameter: 2.84 mils

Beam Voltage: 100 keV
Beam Current: 1.86 microamperes (average)
Exposure: 0.1 sec
Spot Diameter: 2.84 mils

Beam Voltage: 100 keV
Beam Current: 3.87 microamperes (average)
Exposure: 2 sec
Spot Diameter: 21.3 mils

Effect of Repeated Pulses of Electrons
on Single Crystal Silicon

Figure 25. Effect of Repeated Pulses of Electrons on Single
Crystal Silicon.

Single crystal TiO_2 was exposed to electron beam energy and exhibited preferential changes in resistance ranging from 10^3 to 10^7 ohms. The electrical properties of the exposed area of the TiO_2 were evaluated and compared with results of other investigators. Cronmeyer⁷ used hydrogen-reduced single crystal rutile. The conduction mechanism observed exhibited a low activation energy (0.04 eV) and indicated that Ti^{+3} interstitials would be the principal source of conduction electrons.

A mass spectrometer technique reported by Lineweaver⁸ was adapted for our work and indicated that oxygen atoms were ejected from TiO_2 by the thermal energy developed by the focused electron beam. Oxygen outgassing characteristics were examined for a number of other dielectric materials.

Single crystal Al_2O_3 was exposed to an electron beam, and a change in dielectric constant was observed at different levels of beam energy. Dielectric constants were obtained using a micro-stripline technique. Calculations were made using Wheeler's data.²

A mathematical model was presented to express the peak temperatures developed during pulsing of electron beam energy. Temperature fronts were defined to show the distribution of thermal fronts with respect to depth and radius of the sample.

REFERENCES

1. R. R. Haberecht, Texas Instruments corporate-sponsored work, prior to 1965.
2. H. A. Wheeler, IEEE Trans. Microwave Theory and Techniques, MIT-13 (1965).
3. W. J. Oosterkamp, Philips Res. Rep. 3, 49 (1948).
4. G. E. Vibrans, MIT Report No. 268, Lincoln Laboratories (1962).
5. A. Müller, Nature 124, 30 (1929).
6. F. Huber, Microelectronics and Reliability 4, 283 (1965).
7. D. C. Cronemeyer, Phys. Rev. 8, 876 (1952).
8. J. C. Lineweaver, J. Appl. Phys. 34, 1786 (1963).
9. H. Schwarz, Proc. First International Conf. on Electron and Ion Beam Science and Technology (John Wiley & Sons, New York, 1965), p. 158.
10. Kuan-Han-Sun, J. Amer. Ceram. Soc. 30, 279 (1947).
11. L. Pauling, Nature of the Chemical Bond (Cornell Univ. Press, 1948), p. 346.
12. R. J. Ackermann and R. J. Thorn, "Reactions Yielding Volatile Oxides at High Temperatures," XVI International Congress of Pure and Applied Chemistry, Paris, 1957, and Argonne National Laboratory Report ANL-5824.
13. Laboratory for Insulation Research MIT, "Tables of Dielectric Materials," in Dielectric Materials and Applications, edited by A. von Hippel (Technology Press of MIT and John Wiley & Sons, New York, 1954), p. 301.
14. S. Mendelson, J. Appl. Phys. 36, 2525 (1965).
15. J. E. Lawrence and N. Tucker, J. Appl. Phys. 36, 3095 (1965).

ELECTRON-BEAM EVAPORATION OF SILICON DIOXIDE AS A STORAGE
LAYER FOR THE RCA AUTOMATIC-PICTURE-TRANSMISSION VIDICON*

by

D. W. Roe
Electronic Components and Devices
Radio Corporation of America
Lancaster, Pennsylvania

ABSTRACT

The electron-beam evaporation of silicon dioxide was developed to produce spot-free layers on top of photoconductive layers. These layers of highly resistive silicon dioxide serve to store a photo-excited charge image for long periods (minutes) in the RCA automatic-picture-transmission (APT) storage vidicon. This discussion treats of the probable causes of signal loss after a short operating time in qualified flight tubes and the solution to this problem.

I. INTRODUCTION

The development of the electron-beam evaporation of silica enabled the fabrication of the long-life RCA APT storage vidicon.* This storage vidicon is capable of storing a photo-excited charge image for long periods. The electron-beam evaporated silica layer is the storage layer in a two-layer target. The outer layer is a photoconductive layer which modulates the potential across the silica storage layer.

Initially, a major problem in the deposition of silica by electron-beam evaporation was the introduction of particles ejected from the silica source onto the evaporated silica layers. This problem was eventually resolved by the use of proper jigging and by adjustment of the evaporation parameters.

The problem of the short operating life of silica layers in qualified flight tubes compared with that in some of the developmental tubes was solved by passivation of the materials of the flight-tube gun mount. There are several theories or speculations about the failure mechanism in tubes and its elimination, but none can be proven.

This paper discusses the storage-vidicon operation, the silica-storage-layer evaporation equipment, the evaporation for flight tubes, the solution of the operating life problem in flight tubes, and finally the investigation of the life-failure mechanism in flight tubes.

*This work was performed for the National Aeronautics and Space Administration (NASA) by the Astro-Electronic Division (AED) of the Radio Corporation of America under Contract No. PNAS5-667, Mod. 80.

II. STORAGE-VIDICON OPERATION

The RCA storage vidicon, illustrated in Figure 1, utilizes a very sensitive photoconductor over which is deposited a thin-film, high-resistivity dielectric (silica).

The photoconductor transduces the image information into electrical signal information as in conventional vidicons. In this vidicon, the electrical signal is then electrically transferred from the photoconductor to the silica insulator surface where it is stored as an electrostatic potential. Readout is accomplished by using a low-velocity* beam, and the video signal is taken from the target ring of the vidicon. The photoconductor layer is generally 4 microns thick; the evaporated-silica layer is 0.4 micron thick.

The four modes of storage-vidicon operation are: (1) prepare, (2) expose, (3) develop, and (4) readout. The prepare mode requires 5 seconds, the expose-mode 1 second, the develop-mode 2 seconds, and the readout-mode 200 seconds to complete. In the prepare-mode, the residual charge is removed from the photoconductor layer and the storage layer. A potential above the first crossover of the secondary-emission curve is established on the electron-gun side of the insulator (usually plus 500 volts). The potential of the gun side of the insulator is lowered to about plus 450 volts just prior to exposure. During the exposure-mode, the interface between the insulator and the photoconductor is brought to the potential of the signal electrode in illuminated areas (usually plus 475 volts). During the develop-mode, the beam is turned off and the photoconductor is illuminated. This condition causes the interface between the photoconductor and the insulator to reach the signal-electrode potential and transfers the charge pattern to the gun side of the insulator. In the readout-mode, a low-velocity beam returns the surface potential to cathode potential (approximately 10 volts) and provides the video signal.

Reference (1) is a recommended source of electrostatic storage and recording data.

Reference (2) discusses the sensing characteristics of an electrostatic camera.

III. SILICA EVAPORATION EQUIPMENT

The equipment used in the deposition process includes an ion-pump system, two electron guns and associated power supplies, low-

* Low-velocity-beam electrons impact on target with energy such that less than one secondary per primary is dislodged. Conversely, high velocity defines the operation with a secondary-emission ratio greater than unity.

voltage power supplies for the heaters, a variable 5,000-volt power supply, a stainless-steel bell jar with lead-glass ports, and evaporation jigs. Figure 2 is a block diagram of the entire evaporation system; Figure 3 is a photograph of the system.

The evaporation jig comprises the usual shutters, a quartz oscillator thickness-monitor, a grounded thermocouple on top of the faceplate, copper heat sinks for the faceplates, special cages for the faceplates (with a wire mesh), and a tungsten heater for the quartz source.

The faceplate cages are constructed of a three-inch-diameter cylinder, one inch long. At one end of the cylinder, a wire mesh is attached; at the other end, a plate is attached to hold three faceplates. These cages are insulated from ground and held at a negative potential during the evaporation to repel negative particles. See Figure 4.

A tungsten sheet was used as the heater. The fused quartz plate was held in place by two strips of tantalum overlapping on opposite ends of the plate. A spot of "Aquadag" (carbon suspension) was painted in the center of the bottom of the quartz plate to help keep the beam from wandering to the tungsten heater or the tantalum strips. The tungsten strip is forced tightly against the center of the quartz plate by a ceramic tube between the strip and the baseplate of the evaporator. Before evaporation, the tungsten heater is brought to white heat in a vacuum; this temperature is maintained during evaporation.

IV. EVAPORATED SILICA LAYERS FOR DEVELOPMENTAL TUBES

The initial evaporated silica layers were tested in developmental tubes which differed from the flight-qualified tubes in that they were not ruggedized to withstand mechanical shock. Moreover, these initial layers were not of flight quality because they had large numbers of spots (pieces of foreign material, usually silica, which were visible in the TV monitor).

Experiments were performed to ascertain the effects of process parameters on the resultant quality and electrical performance of the silica insulator layer. The results of these tests appear in Table I. This table reveals that a wide variation in most of the parameters will provide silica layers which perform satisfactorily on initial testing in developmental tubes.

Accelerated life tests were performed on sample vidicons selected from those listed in Table I. These tests consisted of using a beam which has approximately 70 times the current density used in

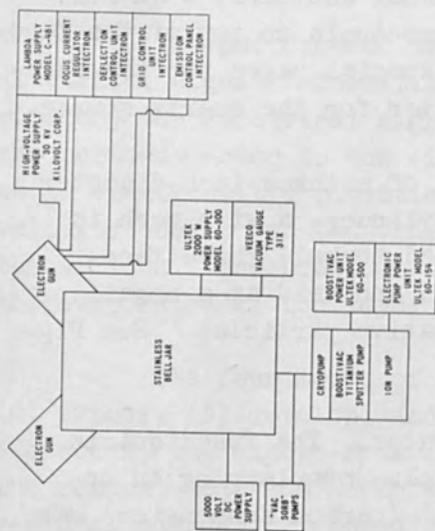


Figure 2

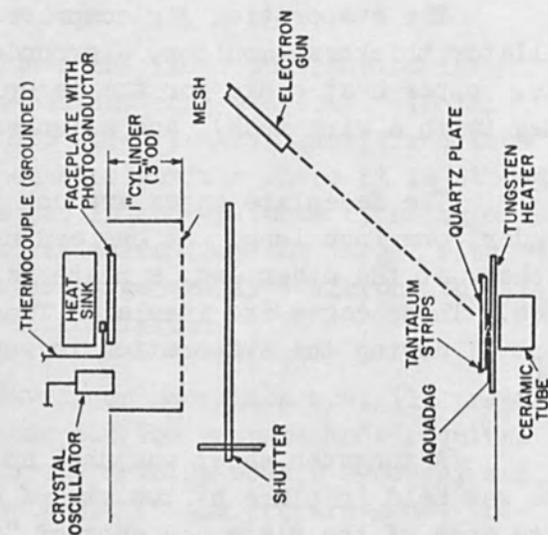


Figure 4

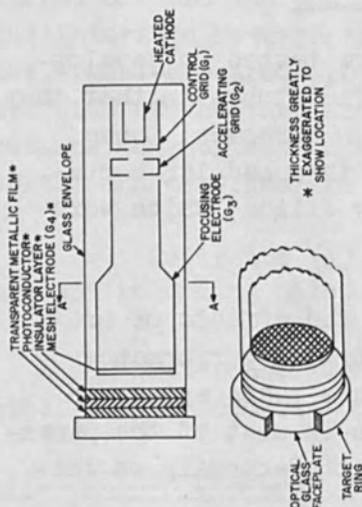


Figure 1

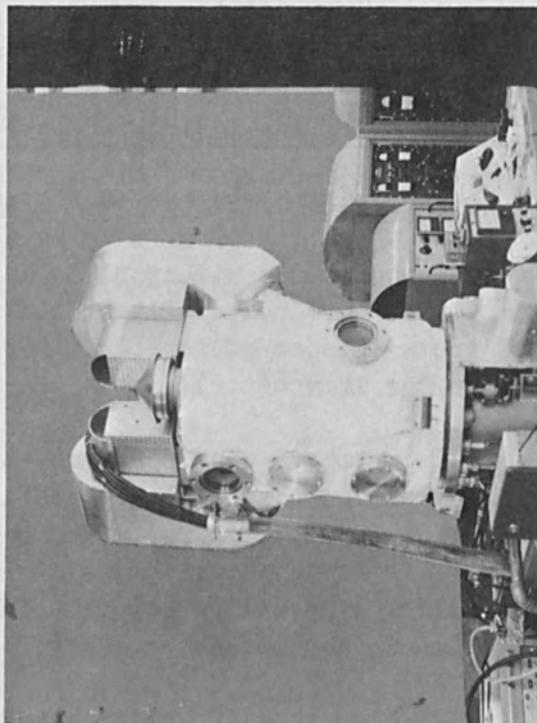


Figure 3

TABLE I. Silicon Dioxide Experiments

Tube No.	Deposition Rate (A/second)	SiO ₂ Film Thickness (angstroms)	SiO ₂ Source Thickness (inches)	Pressure During Evaporation (mm Hg)	Temp. of Faceplate Beginning-End (°C)	Beam Size (inches)	Comments		
							Signal	Resolution	Storage
15304	16	1,970	1/4	10 ⁻⁶	34 - 40	1/8	Good	Poor	Good
15306	25	4,090	1/4	10 ⁻⁶	39 - 42	1/8	Good	Fair	Good
19763	28	4,090	1/4	10 ⁻⁵ - 10 ⁻⁶	---	1/8	Fair	Fair	Fair
19912	10	4,550	1/4	10 ⁻⁵ - 10 ⁻⁶	30 - 32	1/8	Good	Poor	Good
19913	8	4,100	1/8	6 x 10 ⁻⁷	46 - 60	1/16	Very Good	Good	Good
20161	10	3,500	1/8	7 x 10 ⁻⁷	44 - 57	1/16	Very Good	Fair	Very Good
663	21	4,380	1/8	6 x 10 ⁻⁷	56 - 66	1/16	Good	Fair	Good
664	21	4,380	1/8	6 x 10 ⁻⁷	56 - 66	1/16	Very Good	Good	Fair
666	28	4,242	5/16	2 x 10 ⁻⁶	38 - 42	1/16	Good	Fair	Poor
23489	13	4,240	5/16	1 x 10 ⁻⁶	37 - 46	1/16	Cathode Slump		
673	71	4,240	1/8	2 x 10 ⁻⁶	28 - 34	3/8	Good	Good	Good
23488	25	4,170	1/8	2 x 10 ⁻⁶	30 - 31	1/8	Good	Good	Good
23494	6	6,750	1/8	4 x 10 ⁻⁶	47 - 65	3/8	Good	Good	Good
23493	26	4,100	5/16	6 x 10 ⁻⁷	51 - 51	1/8	Good	Poor	Good
C-10093	12	4,000	1/8	2 x 10 ⁻⁶	93 - 105	1/4	Good	Poor	Good
676	50	4,100	1/8	4 x 10 ⁻⁶	60 - 60	5/8	Very Good	Good	Good
677	95	7,570	1/8	6 x 10 ⁻⁶	92 - 110	1/2	Very Good	Good	Very Good

normal tube operation. Previous tests with another type of insulator, which failed rather quickly on life-test, showed that the loss in signal output could be correlated to the total quantity of charge handled by the target. Tube failure is defined as either a 50 percent loss in signal or a 50 percent loss in resolution at 200 TV lines. Results appear in Table II, where the accelerated life-test hours are shown as equivalent real hours.

TABLE II

Life-Test Results for Developmental Tubes

<u>Tube No.</u>	<u>Life-Test Equiv. Real Hours</u>	<u>Comments</u>
15304	15,000	OK - no change from beginning
15306	3,500	OK - no change from beginning
19763	10,000	OK - no change from beginning
19912	1,100	slight loss of resolution (10%)
20161	4,200	slight loss of resolution (10%)
23489	1,000	30% loss of signal Failed at 2700 hours
23493	600	25% loss of signal Failed at 2200 hours
23494	9,000	20% loss of signal

Several of these tubes gave faulty performances. Neither the loss of resolution* nor the loss of signal could be adequately explained. Moreover, all silica layers had a large number of spots.

V. SILICA EVAPORATION FOR FLIGHT TUBES

With results derived from some successful life tests of the developmental vidicons, the next objective was the fabrication of flight-quality silica vidicons with the APT-type ruggedized and flight-qualified mount. However, the problem of depositing spot-free silica surfaces still had to be solved.

After numerous trial depositions, the evaporation parameters which yielded spot-free silica layers were obtained. These parameters, listed below, must be closely controlled to obtain spot-free faceplates by electron-beam evaporation of silica.

1. A Large Electron Beam: to avoid thermal shock.

*Resolution is the number of lines per inch visible in the TV monitor.

2. Evaporation Rate: very slow, to avoid thermal shock.
3. Pressure: low, to avoid arcing and power shutdown.
4. Temperature of the Faceplates: low, to avoid photoconductor peeling.
5. Supply of External Heat to the Fused Quartz Plate: adequate to outgas the plate and to help avoid thermal shock.
6. Negative Potential on the Faceplate Cages: to repel the negatively charged silica particles which cause spots.

Natural fused quartz seems to effect a smoother evaporation than artificial quartz, possibly because of a lower gas content. However, more experiments with newer types of artificial quartz may invalidate this hypothesis. Gas bubbles in natural quartz do not cause extreme ion focusing of the beam provided the quartz is heated externally and the beam size is large.

Because longer evaporation times cause higher temperatures, the silica layer is usually 4,000 Angstroms thick to avoid excessive heating of the photoconductor layer.

VI. SOLUTION TO THE OPERATING LIFE PROBLEM IN FLIGHT TUBES

After the development of the silica evaporation to yield spot-free faceplates, life tests on flight tubes equipped with these faceplates revealed that the operating life of the tubes was too short. Tube failure was indicated by the gradual loss of signal with operating time until, finally, no signal could be detected.

Because there was long operating life in a few of the tubes having developmental mounts, attention was focused on the differences between developmental mounts and flight mounts. The flight mount differs in that it utilizes an added indium damper ring (although indium is present in both mounts) and a ruggedized mesh assembly. The same metals and alloys are present in both mounts and they go through the same processing. A quantity of vidicons was fabricated and tested in a matrix of experiments which deleted or modified some of the preceding items added to the flight mounts. Also included in the test matrix were ether washing of the gun structure and thermal passivation of the mount. Significantly, the test results demonstrated that this treatment of the flight-gun mount virtually eliminated the short-life problem. Table III contains life-test results on successfully modified tubes.

VII. INVESTIGATION OF THE LIFE-FAILURE MECHANISM

The failure mechanism which caused a loss of signal during the life tests is believed to be the result of a significant change in the secondary-emission ratio of the scanned silicon-dioxide surface. If this is true, physical and chemical changes occurred in the scanned surface and possible in the volume below the scanned surface.

An investigation of the failure mechanism was initiated, starting with the faceplate. Visual inspection of a faceplate removed from a life-tested tube that had failed revealed a different interference color in the scanned area than in the unscanned area. This observation indicated that a layer several hundred Angstroms thick had either been deposited or altered. The layer remained after exposure to air and it was insoluble in aqua regia. This insolubility indicated that it is improbable that a metal had been deposited, or that, if one had been deposited, it had been oxidized in the tube to form a low-secondary-emission oxide.

TABLE III

Life-Test Results for Flight-Qualified Tubes

<u>Tube No.</u>	<u>Life Test (Hours)</u>	<u>Equivalent Real Hours</u>	<u>Results</u>
J698	14	1,000	Still OK
C6437	15	1,000	Still OK
C6447	17	1,200	Still OK
C6453	38	2,600	*Loss of Resolution
J699	50	3,500	Still OK
C6468	104	7,300	Still OK
C6450	109	7,600	*Loss of Resolution
C6455	149	10,000	Still OK
C6452	176	12,000	Still OK
J709	190	13,000	Still OK
C6454	226	15,000	Still OK
C6439	243	17,000	Still OK
J712	380	26,000	Still OK
J702	411	28,000	Still OK

Two tubes were constructed with faceplates salvaged from tubes which had failed life testing. These faceplates were exposed to air prior to and during recapping of the tubes. Both tubes performed well in the

*The loss of resolution noticed in these tubes was again unexplained.

beginning, but one failed within 14 actual hours of life test and the other within 16 actual hours. These tests indicated that the scanned area was affected by air exposure and that the recovery was only temporary.

It is obvious that specific conclusions cannot be made on the basis of the foregoing observations. Therefore, only speculation about this recovery can be made, viz., that the evidence suggests a reoxidation yielding silicon dioxide from silicon or from silicon monoxide (or from both).

After a review of the observations of the scanned area, it is apparent that fundamental structural changes have occurred. Possible changes are:

1. Deposition of a new layer on the silicon dioxide.
2. Decomposition of the silicon dioxide into silicon monoxide or into silicon or both.
3. Reaction of the silicon dioxide, silicon monoxide, or silicon to form a new silicon-containing layer of different composition.
4. A physical change, such as crystallization of the amorphous silicon dioxide in the scanned area.

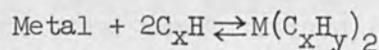
A. Layer Deposition

Speculation that a scanned area of the storage layer fails because of a deposited layer must be consistent with these views: a) that the electron-gun mount possibly contains a substance or substances promoting deposition of a layer; b) that the deposit in the scanned area is insoluble in aqua regia; c) that the scanned area is altered in some way by air exposure; d) and that thermal passivation of the electron-gun mount removes the substance or substances causing the layer formation.

Two different processes of deposition by the electron beam and one process by evaporation of compounds are possible:

1. Electron beam promoted:
 - a. Hydrocarbon decomposition to carbon.
 - b. Chemical-vapor transport of metals by either hydrocarbons or carbon monoxide.

Example:



(The metal is then oxidized to a low secondary emitter.)

2. Evaporation of compounds:

Example:

Sodium chloride which may be in the electron-gun mount and which may form sodium monoxide under electron bombardment.

The nature of the tube gases, which are involved in all the structural changes was determined by mass spectrometry. The analysis is given in Table IV. Different tube types with passivated mounts and with mounts which had not been passivated are included in the table.

TABLE IV

Mass Spectrometer Analysis of SiO₂ Tubes

	Tube A	Tube B	Tube C
<u>Gas Pressure</u>	<u>Normal Processing</u> <u>7.7 x 10⁻⁸ torr</u>	<u>Passivated</u> <u>4.7 x 10⁻⁷ torr</u>	<u>Passivated</u> <u>10⁻⁶ torr</u>
<u>Molecular Analysis</u>	CH ₄ abundant Possibly some CO and CO ₂ No argon	CH ₄ C ₂ H _x C ₃ H _x C ₄ H _x all fairly abundant	CH ₄ C ₂ H _x C ₃ H _x all fairly abundant

The dominant molecular mass in tube A which was short-lived is the hydrocarbon (CH₄) with carbon monoxide (CO) and carbon dioxide (CO₂) more prevalent than in tubes B and C.

Long-lived tubes B and C have no CO₂ or CO, but do have methane and higher hydrocarbons in fair abundance.

Deposition of a carbon layer by the decomposition of hydrocarbons with sufficient atoms to affect the secondary-emission ratio is not too probable. However, if such a layer does deposit, graphite has a secondary-emission ratio of one or of less than one up to hundreds of volts.⁽³⁾ On the assumption that there is a static 10⁻⁷ millimeter pressure of hydrocarbons in a tube, calculations show that there would be approximately one atom of carbon per 3,000 square Angstroms of surface area if all the hydrocarbons were decomposed in an area of one square centimeter. This amount of carbon is insufficient to explain the loss of secondary emission. Moreover, the carbon-layer theory does not

explain a layer several hundred Angstroms thick and with sufficient transparency to give a different interference color from that of the unscanned silicon-dioxide surface. However, a carbon layer less than 10 to 20 Angstroms thick might be deposited; such an amount might affect the secondary-emission ratio.

The presence, in the scanned area, of a layer several hundred Angstroms thick with a different interference color than that in the unscanned area might be explained by assuming a phase change caused by electron bombardment in the amorphous silica.* This change results in a form of silica with a different refractive index.

Chemical vapor transport of metals by hydrocarbons or CO might also be possible.^(4,5) If methane, or other hydrocarbons, or CO react to form volatile organometallics with such metals as chromium, nickel, iron, indium, copper, or tungsten, and to form carbonyls with chromium, nickel, iron and tungsten, they could conceivably transport metals to the silica surface. There they decompose and return to repeat the cycle. The metals would then be oxidized under electron bombardment. Such possible hydrocarbon-transporting agents were found in the tubes which failed early and the tubes which had a long operating life with passivated electron-gun mounts. A methane generator was fired in a tube with an air-baked mount calculated to give a methane pressure of 10^{-6} millimeters of mercury. This tube continued to give good performance after several hundred actual hours of operating life. The thermal treatment of the metal surfaces may have passivated them (made them unreactive). Carbon monoxide, however, was found only in non-passivated tubes, but no simple carbon-monoxide generator could be constructed to test the theory of carbonyl formation in a passivated tube.

B. Surface Decomposition

If the failure mechanism is decomposition of the silicon-dioxide surface into silicon monoxide or silicon, an oxygen-atmosphere bake at 125°C might reverse such disintegration. Therefore, an oxygen generator, calculated to give an oxygen pressure of 10^{-3} millimeters of mercury, was fired in a tube with a non-passivated flight mount after a life-test failure. The tube was then heated to 125°C for 24 hours and the getters refired. No change in the life-test area was noted, either because the temperature was not high enough or because the failure mechanism was not decomposition.

C. Surface Reaction

The possibility exists of a reaction to form a new silicon

*X-ray diffraction observations showed the evaporated silica layers to be non-crystalline.

compound. However, from a consideration of the gases and solids in the tube, no possible simple reactions could be deduced which would lead to useful experiments.

D. Phase Change

A phase change in the amorphous silica surface layer catalyzed by a foreign gas or solid is not too probable. A change to crystalline silica is not probable because the crystalline material would, no doubt, have a higher secondary-emission ratio. A change to another amorphous arrangement cannot be ruled out, however.

VIII. CONCLUSIONS RELATING TO LIFE-FAILURE MECHANISM

None of the previously discussed changes (A, B, C or D) can be completely ruled out on the basis of the evidence. The chemical-vapor deposition of metals by carbonyl formation from carbon monoxide is one of the more probable explanations for loss of signal. The formation of a graphite layer sufficiently thick to reduce the secondary-emission ratio might be possible if continuous outgassing of metal parts occurs.

Thus, passivation of the electron-gun mount may provide long operating life in tubes because of one or more of the following reasons:

- A. Formation of an impervious surface layer which prohibits chemical-vapor transport.
- B. Cleaning of a rather deep surface layer of hydrocarbons which avoids chemical-vapor transport or continuous outgassing.
- C. Removal of other unknown gases.

REFERENCES

1. Hutter, E. C., Inslee, J. A., and Moore, T. H., "Electrostatic Imaging and Recording", Society Motion Picture and Television Engineers Journal 64, 32-35 (1960).
2. Moore, T. H. and Kritzman, I. M., "Sensing Characteristics of an Electrostatic Recording Camera", I. E. E. E. Transactions on Electron Devices ED-11, 183-190 (1964).
3. Bruining, H., Philips Techn. Rev. 3, 80 (1938).
4. Schafer, H., Chemical Transport Reactions, New York, Academic Press, 1964.
5. Powell, C. F., et. al., Vapor Plating, New York, John Wiley and Sons, 1955.

A PULSED MACHINE-SCAN ELECTRON BEAM DEVICE

by

A. B. El-Kareh, R. B. Fair and C. R. Marsh
Electrical Engineering Department
The Pennsylvania State University
University Park, Pennsylvania

ABSTRACT

The results of an investigation concerning the feasibility of using a single, pulsed electron beam in both a machining mode and an imaging mode are presented. Experimental observations coupled with a theoretical analysis of the system are used in order to create a functional design of the machine-scan device.

I. INTRODUCTION

In order to meet the increased demand for reliability standards in microelement processing, it is necessary to observe carefully the specimen being manufactured before, during, and after electron bombardment. Many electron beam machines are equipped with an optical microscope. A prism mounted above the objective lens at a 45° angle to the beam axis bends the optical observation path through a 90° angle to the microscope. However, the resolving power of any optical system with such a large focal length is very poor. Also, a prism mounted inside the electron beam column will become coated with vaporized material from the workpiece.

Much higher resolving powers can be obtained with electron beams than with light microscopes. It may, therefore, be advantageous to incorporate the methods of electron image formation into an electron beam machine. Thus, by using a single beam suitably biased for either machining or for imaging, a continuous picture of the workpiece being formed may be created. Such a process is described by Yoder², and is reiterated below for convenience.

In order to prevent destruction of the workpiece through excessive heating, the electron beam is generally pulsed to high power density, and then a sufficiently negative bias is applied to the control cup of the electron gun in order to reduce the beam current or to cut it off completely. This process may occur many times per second. It has been suggested that during the period of time in which the beam is reduced in intensity, the unit could function as a scanning electron

microscope in order to produce an image of the target. Thus, the beam would scan the target in synchronism with the trace of an oscilloscope, and the signal created by collecting secondary electrons from the target would be used to modulate the intensity of the oscilloscope trace. Upon the completion of the formation of an image, the positive electrode at the workpiece for collecting secondary electrons would be biased negatively so that no electrons would be collected during the machining operation. The beam is once again increased in intensity for machining and the process repeats itself. Thus, used as a tool, the electron beam could perform as a source of heat for the purpose of welding, etching, drilling or cleaning, and then, by varying a single parameter, it could perform in the manner of the scanning electron microscope.

II. OPTICAL AND PHYSICAL REQUIREMENTS

A. Scanning Mode

The basic requirement of the machine-scan device operating in the scanning mode is that a spot must be produced at the focus roughly 1/10 to 1/20 the size of the machining spot. The current density of the primary electron beam must be carefully selected so as to allow image formation without specimen destruction. The only dimensional restriction on the design is that the axial length between the principal plane of the objective lens and the workpiece must be as small as is possible. This restriction exists in order to create a high demagnification of the crossover. A summary of the initial design requirements for operation in the scanning mode are given below.

<u>Item</u>	<u>Magnitude of Dimension</u>
Final beam voltage	$2-3 \times 10^4$ volts
Final spot size	1.5 - 2.5 microns
Average spot current density	2 amperes/cm ²
Working distance of final lens	2.5 cm
Total length of electron-optical system	56 cm
Number of lenses	2
Beam current	0.01 - 10 microamperes

B. Machining Mode

The basic requirement of the machine-scan device operating in the machining mode is that a sufficient power density must be produced at the focus in order to carry out welding and vaporization of most substances used in microelement fabrication. In order to obtain the value of a useful power density, Pittaway's³ analysis of the temperature distributions in thin foil targets due to a stationary beam will be used. The energy of the beam is assumed to be absorbed in the entire target,

and no account is taken of electron penetration or scattering.

The standardized temperature at the center of the beam focus is given by

$$\frac{TkD}{E_0} = \frac{1}{4\pi} \log_e (1 + 4\lambda), \quad (1)$$

where T is the temperature ($^{\circ}K$)

k is the thermal conductivity ($\frac{\text{watts}}{^{\circ}K \text{ cm}}$)

D is the thickness of the target (cm)

E_0 is the total energy of the source per second (watts)

and λ is a dimensionless parameter defined by

$$\lambda = \frac{k't}{d^2} . \quad (2)$$

k' ($\frac{\text{cm}^2}{\text{sec}}$) is the thermal diffusivity of the target, d (cm) is the standard deviation of the energy at the focus and t is the beam pulse length. The beam is usually pulsed to a high power density for short intervals in order to improve upon the thermal efficiency of the machining process. In his work, El-Kareh¹ has found it necessary for the high energy pulses to last from a few microseconds to several milliseconds.

For an accelerating voltage of 2×10^4 volts, $E_0 = 2 \times 10^4 I_b$. Rearranging terms in equation (1) and solving for I_b gives

$$I_b = \frac{4\pi (TkD)}{2 \times 10^4 \log_e (1 + 4\lambda)}$$

or

$$I_b = \frac{4\pi (TkD)}{2 \times 10^4 \log_e (1 + \frac{4k't}{d^2})} \quad (3)$$

Selecting a spot diameter of 2.5×10^{-3} cm, and assuming that the energy distribution of the beam falls off rapidly at a distance of two standard deviations from the spot center, d will equal 0.6×10^{-3} cm. Assuming a 100 micron thick piece of n-type germanium (3 ohm-cm at 25 $^{\circ}C$), and

setting $t = 1$ millisecond gives an I_b equal to 0.4 milliamperes. This amount of current corresponds to a power density of 4×10^5 watts/cm² which is the power density required to melt germanium in one millisecond. A summary of the initial design requirements for operation in the machining mode is given below.

<u>Item</u>	<u>Magnitude or Dimension</u>
Final beam voltage	$2-3 \times 10^4$ volts
Final spot size	25 microns
Spot power density	$4 \times 10^5 - 10^6$ watts/cm ²
Working distance of final lens	2.5 cm
Total length of electron optical system	56 cm
Number of lenses	2
Beam current	0.1 - 1.0 milliamperes

III. ELECTRON GUN DESIGN

The electron gun which was selected for use in the machine-scan device is the conventional tungsten hair-pin filament type with accompanying Wehnelt cylinder. This gun has been investigated extensively by Von Borries⁴ and by Haine and Einstein⁵. The main reasons for selecting this type of gun were for its simplicity, and for the fact that any gun geometry within wide limits gives a brightness of the electron source very close to the theoretical value⁵.

Von Borries has shown that under optimum conditions a beam crossover diameter of 8×10^{-3} cm can be obtained for an accelerating potential of 20×10^3 volts, and a filament height of 1 mm above the outside of the cathode shield. The corresponding shield aperture was given as 1.5 mm. Assuming that spherical aberration can be neglected for the small currents used in scanning², the spot diameter at the focus is given by

$$d_s = \frac{d_c \theta_c}{\theta_s}, \quad (4)$$

where

d_c is the crossover diameter,

θ_c is the crossover aperture half-angle, and

θ_s is the image aperture half-angle.

In order to obtain a spot size of 2.5×10^{-4} cm.

$$M^{-1} = \frac{d_c}{d_s} = \frac{8 \times 10^{-3}}{2.5 \times 10^{-4}} = 32. \quad (5)$$

Therefore, a demagnification of 32 is required of the electron-optical system. This amount of demagnification will cause the minimum disc of confusion to be very large when the beam current is increased for machining². Thus, the required power density at the focus could not be obtained.

In order to reduce the amount of demagnification required to produce the scanning spot, the diameter of the beam crossover must be reduced. Haine and Einstein⁵ have shown curves relating the cathode-image diameter to beam current and to filament height. Therefore, by scaling down these results, the necessary electron gun geometry may be obtained which would produce a 2.5×10^{-3} cm crossover diameter.

Since the filament is being heated to a temperature where space-charge limitation occurs, it is possible to scale voltages, distances, and currents according to the following equations¹:

$$\begin{aligned} V' &= AV \\ d' &= A^{3/4}d \\ I' &= A^{3/2}I \end{aligned} \tag{6}$$

where the prime mark denotes the new value, and A is the scaling factor. The crossover diameter scales according to the relation $d'_c = A^{1/4}d_c$. The results of scaling are shown below.

<u>Haine and Einstein</u>		<u>Scaled</u>
V_a	5×10^4 volts	2×10^4 volts
T	2850°k	2850°k
I_b	0.4×10^{-6} ampere	0.1×10^{-6} ampere
d_c	3.2×10^{-3} cm.	2.5×10^{-3} cm.
h	0.7 mm	0.35 mm.
d_g	1.75 mm.	0.88 mm.
ϕ_B	-445 volts	-178 volts

where

h is the filament height above the outside of the cathode shield

d_g is the shield aperture diameter

ϕ_B is the cut-off bias

I_b is the beam current

V_a is the accelerating voltage

It can be seen from the curves of Haine and Einstein that for $h = 0.7$ mm, the crossover diameter remains constant for a wide range of beam currents. Thus, for $I_b = 500$ microamperes, d_c remains equal to 2.5×10^{-3} cm.

It is now necessary to determine the anode-to-grid spacing for the required current cut-off of the triode system. According to Spangenberg⁶, the amplification factor $\mu = -\frac{V_a}{\phi_B}$ for an ideal triode has the specific value

$$\mu = \frac{\pi d_a}{h} \left(\frac{1}{\arctan \frac{r_g}{h} - \frac{1}{\frac{r_g}{h} + \frac{h}{r_g}}} - \frac{1}{\pi} \left(\frac{h}{d_a} + 1 \right) \right) \quad (7)$$

where d_a is the anode-to-grid spacing and r_g is the radius of the aperture in the grid (Wehnelt cup). Substitution of the scaled values into equation (7) gives $d_a = 0.6$ cm. This value of anode-to-grid spacing will allow the beam current to be cut off at $\phi_B = -178$ volts. Figure 1 shows the physical construction of the electron gun.

The cathode emission current can be obtained by knowing the emitting area and the current density for a given filament temperature. At a temperature of 2850°K the emission current density of a tungsten filament is 5 amperes/cm². Assuming that the emitting area of the tungsten cone is spherical, a 0.01 cm diameter filament will give a saturated emission current of approximately one milliampere. This value of current corresponds to the maximum beam current required in the machining mode. The life expectancy of this particular cathode is about 10 hours under continuous vacuum⁶.

IV. THE ELECTRON-OPTICAL SYSTEM

In any electron beam device utilizing the standard electron microscope gun, the beam crossover is used as the object whose image forms the working spot of the beam. Therefore, in order to focus the beam to 1.5 microns in diameter, a demagnification of $\frac{25}{1.5} = 17$ is required. If two lenses are used in the electron-optical system, the demagnification can be varied over a range of values. Therefore, a condensing lens will be placed as close to the workpiece as is possible.

The working distance of the objective lens will be limited primarily by the size of the secondary electron-collection apparatus the vicinity of the workpiece. A working distance of 2.16 cm has been

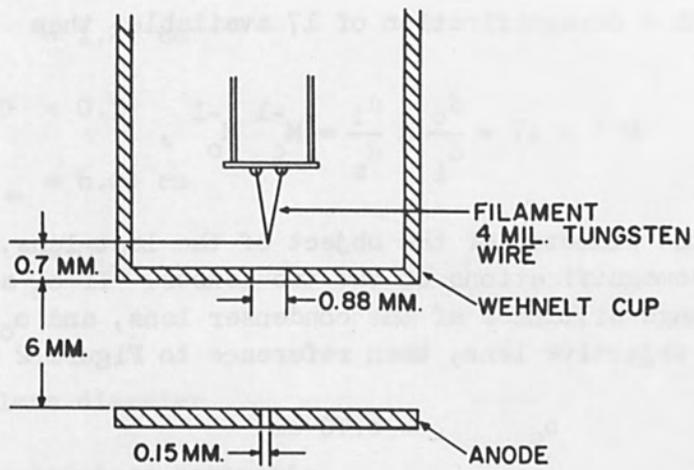


FIGURE 1 - ELECTRON GUN CONFIGURATION

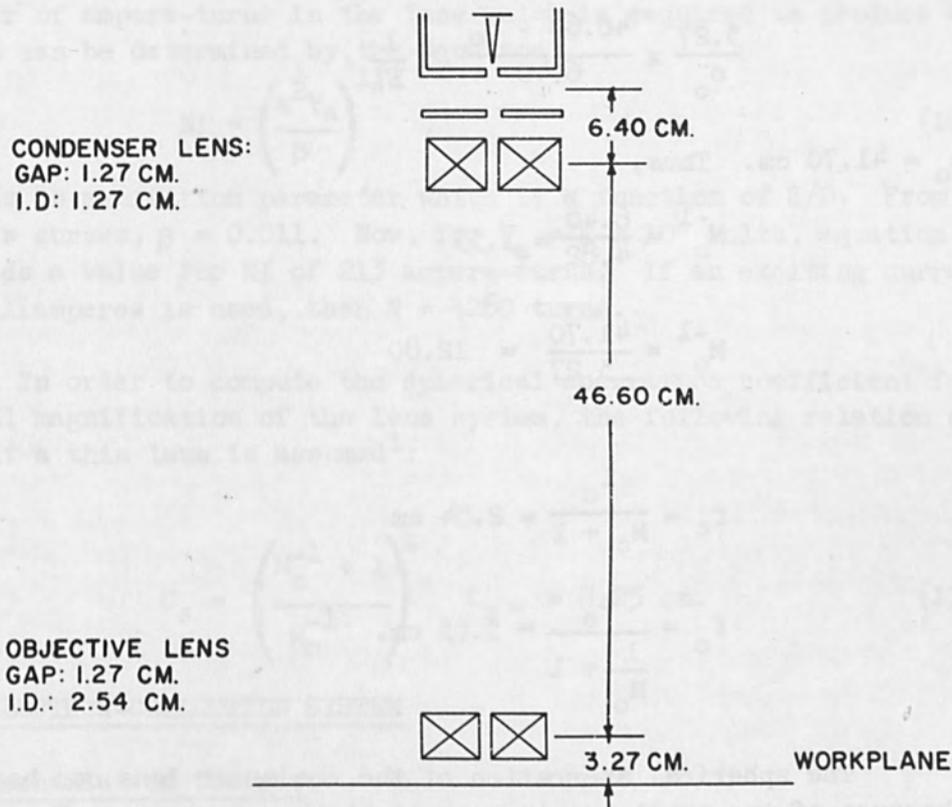


FIGURE 2 - ELECTRON OPTICAL CONFIGURATION

experimentally determined to be sufficient.

With a demagnification of 17 available, then

$$M^{-1} = 17 = \frac{d_c}{d_i} \times \frac{d_i}{d_s} = M_c^{-1} M_o^{-1}, \quad (8)$$

where d_i is the diameter of the object of the last lens, and M_c^{-1} and M_o^{-1} are the demagnifications of the two lenses. If o_c and i_c are the object and image distances of the condenser lens, and o_o and i_o are those of the objective lens, then reference to Figure 2 will give

$$o_c = 6.40 \text{ cm}$$

$$i_c + o_o = 46.60 \text{ cm}$$

$$i_o = 3.27 \text{ cm}$$

Therefore

$$\frac{3.27}{o_o} \times \frac{46.60 - o_o}{6.40} = \frac{1}{17}, \quad (9)$$

or $o_o = 41.70 \text{ cm}$. Thus,

$$M_c^{-1} = \frac{6.40}{4.82} = 1.33$$

$$M_o^{-1} = \frac{41.70}{3.27} = 12.80$$

and

$$f_c = \frac{i_c}{M_c + 1} = 2.54 \text{ cm}$$

$$f_o = \frac{o_o}{\frac{1}{M_o} + 1} = 3.03 \text{ cm.}$$

The spherical aberration of the condenser lens can be neglected because of the small aperture angle at the crossover. However, the spherical aberration of the objective lens is considerable, and the lens must be designed so as to minimize this effect. The following lens data has been determined in order to meet this requirement⁷:

$$D = 2.54 \text{ cm}$$

$$S = 1.27 \text{ cm.}$$

$$S/D = 0.5$$

$$C_{s\infty} = 6.04 \text{ cm}$$

$$k^2 = 0.5$$

where

D is the lens diameter

S is the pole piece separation.

$C_{s\infty}$ is the spherical aberration coefficient for an infinite object.

k^2 is the dimensionless excitation parameter.

The number of ampere-turns in the lens which is required to produce a k^2 of 0.5 can be determined by the equation⁷

$$NI = \left(\frac{k^2 V_a}{\beta} \right)^{1/2}, \quad (10)$$

where β is an excitation parameter which is a function of S/D . From Liebmann's curves, $\beta = 0.011$. Now, for $V_a = 2 \times 10^4$ volts, equation (10) yields a value for NI of 213 ampere-turns. If an exciting current of 50 milliamperes is used, then $N = 4260$ turns.

In order to compute the spherical aberration coefficient for the actual magnification of the lens system, the following relation can be used if a thin lens is assumed¹:

$$C_s = \left(\frac{M_o^{-1} + 1}{M_o^{-1}} \right)^4 C_{s\infty} = 8.25 \text{ cm.} \quad (11)$$

V. THE ELECTRON-COLLECTION SYSTEM

A. Mechanism of Contrast

Contrast in the electronic image of the scanning microscope is created by changes in secondary emission from point to point across the specimen. This variation in secondary emission yield is primarily due to the dependence of the number of secondary electrons liberated on the angle of incidence of the primary beam⁸. The energies of these

liberated electrons range up to 150 electron volts.

Another method of producing contrast is to use reflected electrons. These electrons tend to be emitted in the plane of the normal to the surface at an angle equal but opposite to the incidence angle of the primary beam⁸. Because these electrons have such high velocities, it is impossible to observe any part of the target which is not in a direct-line path with the collector.

It has been experimentally shown by several workers that higher contrast can be achieved when reflected electrons are used to image a target⁹. However, more detailed pictures can be obtained by using secondary electrons.

B. Electron Collection and Detection

In the machine-scan device the problem of choosing between the two methods of image formation is a simple one. If reflected electrons are used, the collector must be in a direct line with the target. However, during the machining mode, all surfaces in the vicinity of the target become coated with vaporized material. Thus, a scintillator or a first dynode of an electron multiplier close to the target will be rendered useless in a very short time. Therefore, the electron detector of the machine-scan device must be out of the line-of-sight with the target, and must be shielded from any light shining on the target due to the filament. Thus, slow secondary electrons, which are easily affected by electric and magnetic fields, must be used in the image formation process.

In order to carry the collected secondary electrons away from the workpiece to a shielded scintillator, a curved magnetic field is employed. The electrons enter one end of a copper tube with a coil of wire wrapped along its length. The action of the axial magnetic field within the tube causes the electrons to spiral around and through the core. At the exit end the electric field created by a high voltage scintillator (see Figure 3) accelerates the low energy secondary electrons, and gives them enough energy to excite the scintillator.

It is evident that this "toriodal core" device has an energy bandwidth associated with it. That is, for a given magnetic field strength, only a certain range of electrons are able to pass through the device without spiraling into the walls of the core. Therefore, only slow secondary electrons are detected by the scintillator for image formation. Figure 4 shows the completed design of the machine-scan device.

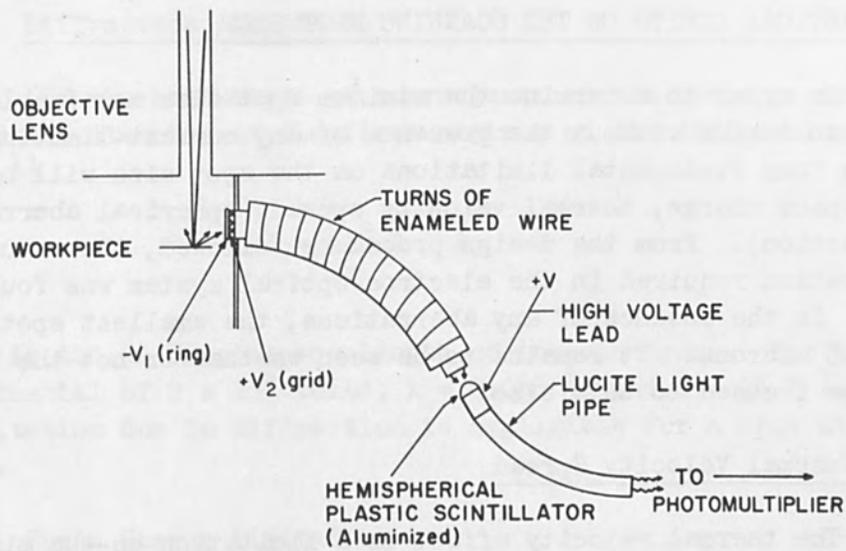


FIGURE 3 - SECONDARY ELECTRON COLLECTION AND DETECTION SYSTEM

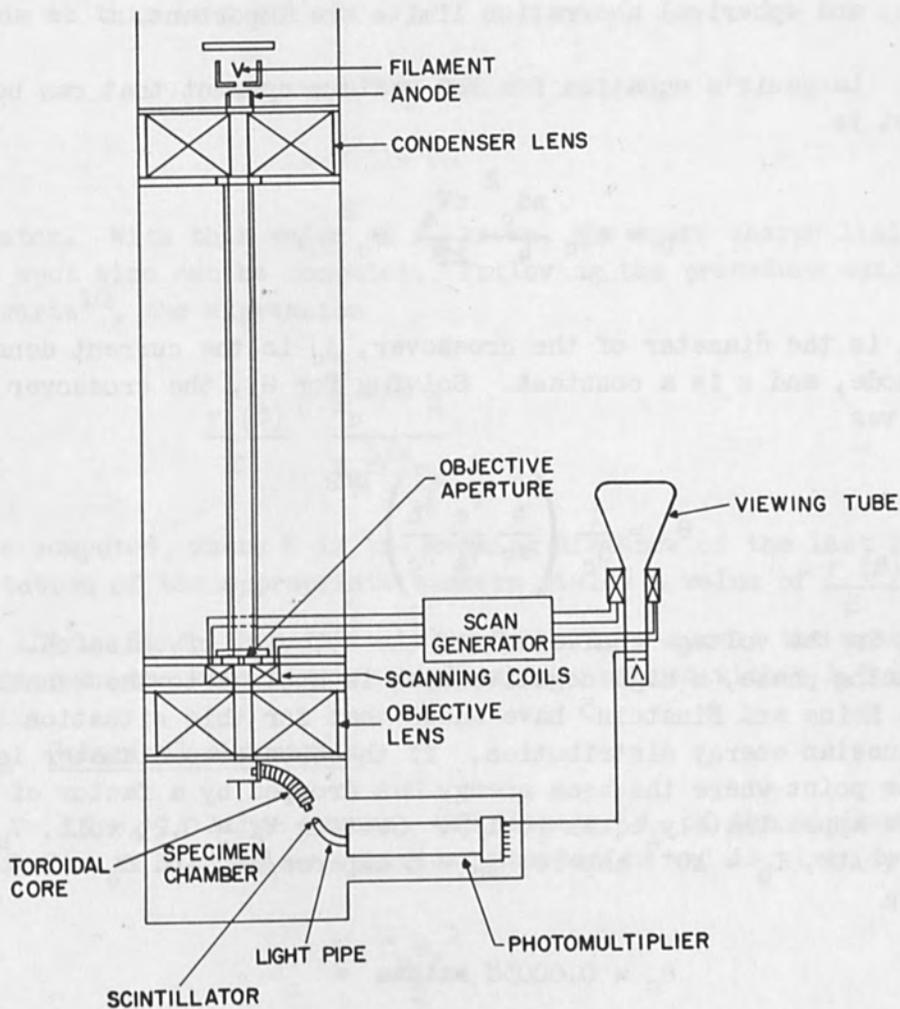


FIGURE 4 - MACHINE-SCAN ELECTRON BEAM DEVICE

VI. THEORETICAL LIMITS ON THE SCANNING SPOT SIZE

In order to determine the minimum spot size attainable in the machine-scan device without the presence of any current-limiting apertures, the four fundamental limitations on the spot size will be investigated (space charge, thermal velocity spread, spherical aberration, and diffraction). From the design procedure followed, the amount of demagnification required in the electron-optical system was found to be $M^{-1} = 17$. In the absence of any aberrations, the smallest spot attainable is 1.5 microns. It remains to be seen whether or not the beam can actually be focused to this size.

A. Thermal Velocity Spread

The thermal velocity effect is a limitation on the minimum half-angle of convergence, θ_s , of the beam which, in turn, is a limitation on the minimum radius of the beam at the objective lens. This permissible angle will be calculated and used to see if space charge, diffraction, and spherical aberration limits are important.

Langmuir's equation for the maximum current that can be focused to a spot is

$$I_b = c j_c \frac{\pi d_c^2}{4} \frac{eV_a}{kT} \theta_c^2, \quad (12)$$

where d_c is the diameter of the crossover, j_c is the current density of the cathode, and c is a constant. Solving for θ_c , the crossover half-angle gives

$$\theta_c = \frac{1}{d_c} \left(\frac{4}{\pi c} \frac{V_e}{V_a} \frac{I_b}{j_c} \right)^{1/2}, \quad (13)$$

where V_e is the voltage equivalent of the velocity of emission. During the scanning phase, a high negative bias is applied to the Wehnelt cylinder. Haine and Einstein⁵ have shown that for this situation the beam has a Gaussian energy distribution. If the crossover diameter is taken to be the point where the beam energy has dropped by a factor of $1/e$, then c is approximately equal to $1/3$. Setting $V_e = 0.25$ volt, $V_a = 2 \times 10^4$ volts, $I_b = 10^{-7}$ ampere, $j_c = 5$ amperes/cm² and $d_c = 2.5 \times 10^{-3}$ cm, then

$$\theta_c = 0.00038 \text{ radian}$$

or greater. Thus, $\theta_s = 17 \times 0.00038 = 0.0064$ radian.

B. Diffraction Limitation

If the half-angle of convergence at the focus is denoted by θ_s , then the diameter of the first minimum of the diffraction image is given by¹¹

$$d = \frac{1.2 \lambda}{\theta_s} \text{ \AA}, \quad (14)$$

where λ is the de Broglie wavelength of the electrons. For an accelerating potential of 2×10^4 volts, $\lambda = 0.088 \text{ \AA}$, and $d = 16.5 \text{ \AA}$. Therefore, the limitation due to diffraction is negligible for a spot size of 1.5 microns.

C. Space Charge Limitation

The objective lens in the machine-scan apparatus will be placed at a distance of 2.2 cm from the workpiece. Therefore, the radius of the beam at this lens will be

$$\begin{aligned} r_i &= 0.0064 \times 2.2 \\ &= 0.014 \text{ cm} \end{aligned}$$

or greater. With this value of r_i known, the space charge limitation on the spot size can be computed. Following the procedure outlined by Schwartz¹⁰, the expression

$$\frac{f(\beta)}{\beta} = \frac{I_b^{1/2} Z}{V_a^{3/4} r_i} \quad (15)$$

must be computed, where Z is the working distance of the last lens. Substitution of the appropriate numbers yields a value of $\frac{f(\beta)}{\beta}$ equal to 3×10^{-5} . Since this falls below Schwartz's curve, the space charge limitation on the spot size is negligible compared with a 1.5 micron spot.

D. Spherical Aberration

If θ_s is large enough, the diameter d_o of the minimum disc of confusion is limited by spherical aberration and is given by²

$$d_o = \frac{C_s \theta_s^3}{2} \quad (16)$$

C_s , the spherical aberration coefficient of the last lens, was computed to be equal to 8.25 cm. Thus, $d_o = 108$ A. This value is well below the acceptable spot diameter. Therefore, space charge, diffraction, and spherical aberration are not significant in limiting the scanning spot size.

VII. DETERMINATION OF THE BEAM POWER DENSITY

In the machining mode it is desirable to focus the image of the crossover to as small a spot as is possible, without changing the high demagnification of the electron-optical system. The only system parameter that can be changed effectively is the bias on the Wehnelt cup (which is made less negative for larger beam currents). In the scanning mode aberrations can be neglected and the spot size can be obtained using Lagrange's relation,

$$d_s = d_c \frac{\theta_c}{\theta_s} \quad (4)$$

But when the bias is decreased for machining, θ_c increases, and, therefore, spherical aberration cannot be neglected.

The spot size is increased by $\frac{C_s \theta_s^3}{2}$. Thus

$$d_s' = d_s + \frac{C_s \theta_s^3}{2} \quad (17)$$

With a demagnification of 17 available in the electron-optical system, equation (17) becomes

$$\frac{C_s \theta_s^3}{2} = d_s' - \frac{d_c}{17} \quad (18)$$

Assuming that d_c remains approximately equal to 2.5×10^{-3} cm, in order to obtain a 2.5×10^{-3} cm spot at the focus,

$$\frac{C_s \theta_s^3}{2} = 2.38 \times 10^{-3} \text{ cm.}$$

For $C_s = 8.25$ cm, then $\theta_s = 0.084$ radian. Referred back to the crossover, this gives $\theta_c = \frac{0.084}{17} = 0.0049$ radian.

In order to determine the amount of current that can be focused to a given spot size d_s' , Langmuir's equation can be used with equation (17) which defines d_s . Thus,

$$I_b = j_c \frac{\pi e V}{4kT} \theta_s^2 \left(d_s' = \frac{C_s \theta_s^3}{2} \right)^2. \quad (19)$$

For $d_s' = 2.5 \times 10^{-3}$ cm, $\theta_s = 0.084$ radian, $V = 2 \times 10^4$ volts, $j_c = 5$ amperes/cm², and $T = 2850^\circ\text{K}$, then $I_b = .051$ milliampere. This represents a power density of 2.1×10^5 watts/cm² which is usable but is below the minimum power density requirement.

If the demagnification of the system is reduced to a value of 10, then following a similar procedure, the maximum current attainable in a 25 micron spot is $I_b = 0.141$ milliampere. This amount of beam current represents a power density of 5.8×10^5 watts/cm² at 2×10^4 volts, which is within the desired range.

Finally, if the accelerating voltage is increased to 3×10^4 volts, the maximum attainable power density for $M^{-1} = 10$ is 8.7×10^5 watts/cm², with $d_s' = 2.5 \times 10^{-3}$ cm. The corresponding beam current is 0.21 milliampere.

From this analysis it can be seen that the machine-scan device can operate effectively as a welder and a vaporizer for most materials. In order to determine the extent of this device's usefulness, Pittaway's³ analysis of the temperature distributions in thin targets will be used again in order to calculate the beam pulse durations necessary to melt several materials.

VIII. DETERMINATION OF THE REQUIRED BEAM PULSE LENGTH

Starting with equation (1) which defines the standardized temperature at the center of the focus as

$$\frac{TkD}{E_0} = \frac{1}{4\pi} \log_e (1 + 4\lambda), \quad (1)$$

the beam pulse duration, t , can be determined by rearranging terms. Thus,

$$e^{\frac{4\pi TkD}{E_0}} = 1 + 4\lambda$$

or

$$t = \frac{\left(e^{\frac{4\pi T k D}{E_0}} - 1 \right) d^2}{4k'} \quad (20)$$

where all of the terms have been defined.

For an example, consider the same piece of n-type germanium as described previously. The following values will be used in equation (20):

$$k' = 0.23 \text{ cm}^2/\text{sec}$$

$$k = 0.4 \text{ watts}/^\circ\text{K cm}$$

$$T = 1232^\circ\text{K}$$

$$I_b = 0.21 \text{ milliampere}$$

$$V = 3 \times 10^4 \text{ volts}$$

$$D = 100 \text{ microns.}$$

If it is assumed that the energy distribution falls off rapidly at a distance of two standard deviations from the spot center, then, for a spot size of 2.5×10^{-3} cm, d will equal 0.6×10^{-3} cm. Equation (20) can now be solved for the pulse duration. Substitution of the appropriate value yields

$$t = 7.3 \text{ milliseconds.}$$

Following a similar procedure, the required pulse lengths for melting several other types of materials have been computed. All of the targets listed below are assumed to be 100 microns thick, unless specified otherwise.

<u>Material</u>	<u>Beam Pulse Duration</u>
Germanium	7.3 milliseconds
Tin	0.08 milliseconds
Aluminum	cannot melt through
Aluminum (10 microns thick)	5.5 milliseconds
Nickel	cannot melt through
Nickel (40 microns thick)	6.2 milliseconds
Silver	cannot melt through
Silver (10 microns thick)	0.97 milliseconds

The targets tabulated above which can be melted are, in fact, melted completely through their thicknesses. This is due to the fact that equation (20) was determined by assuming that the beam energy is absorbed in the entire target. Thus, a certain amount of machining can be performed (metal removing and drilling) on many thin metal targets. Thicker targets can also be melted, but the melting zone will not be very deep. A tabulation below provides a direct comparison between the parameters associated with the two modes of operation.

<u>Item</u>	<u>Scanning Beam</u>	<u>Machining Beam</u>
Crossover Diameter	2.5×10^{-3} cm	2.5×10^{-3} cm
Crossover Position (from cathode)	5×10^{-3} cm	4.8×10^{-2} cm
Crossover Half-Angle	0.00038 radian	0.0084 radian
Anode Aperture Diameter	1.5×10^{-3} cm	7.6×10^{-3} cm
Spot Size at Focus	1.5×10^{-4} cm	2.5×10^{-3} cm

IX. IMAGE FORMATION WITH THE TOROIDAL SEGMENT

A. Experimental Procedure

In order to test the effectiveness of the toroidal segment as a vital link in the imaging system of the machine-scan device, a prototype electron beam machine of simple geometry was constructed. This apparatus was then operated as a scanning electron microscope so that the proposed imaging system could be evaluated. The procedure followed in order to obtain an electronic image is described below.

A focused electron beam was made to scan a rectangular raster across a fine-meshed wire grid at the workpiece. Secondary electrons were collected at each instant by applying a positive voltage on the collecting grid of the toroidal segment. These electrons were then "spiraled" through the core, and were made to strike a high voltage scintillator. The resulting scintillations were detected by a photomultiplier, and the corresponding electronic signal was used to modulate the intensity of an oscilloscope trace scanning in synchronism with the primary electron beam.

B. Experimental Apparatus

Figure 5 is a diagram of the experimental scanning microscope and its supporting equipment. Figure 6 is a photograph of this system. The important aspects of this photograph are described below.

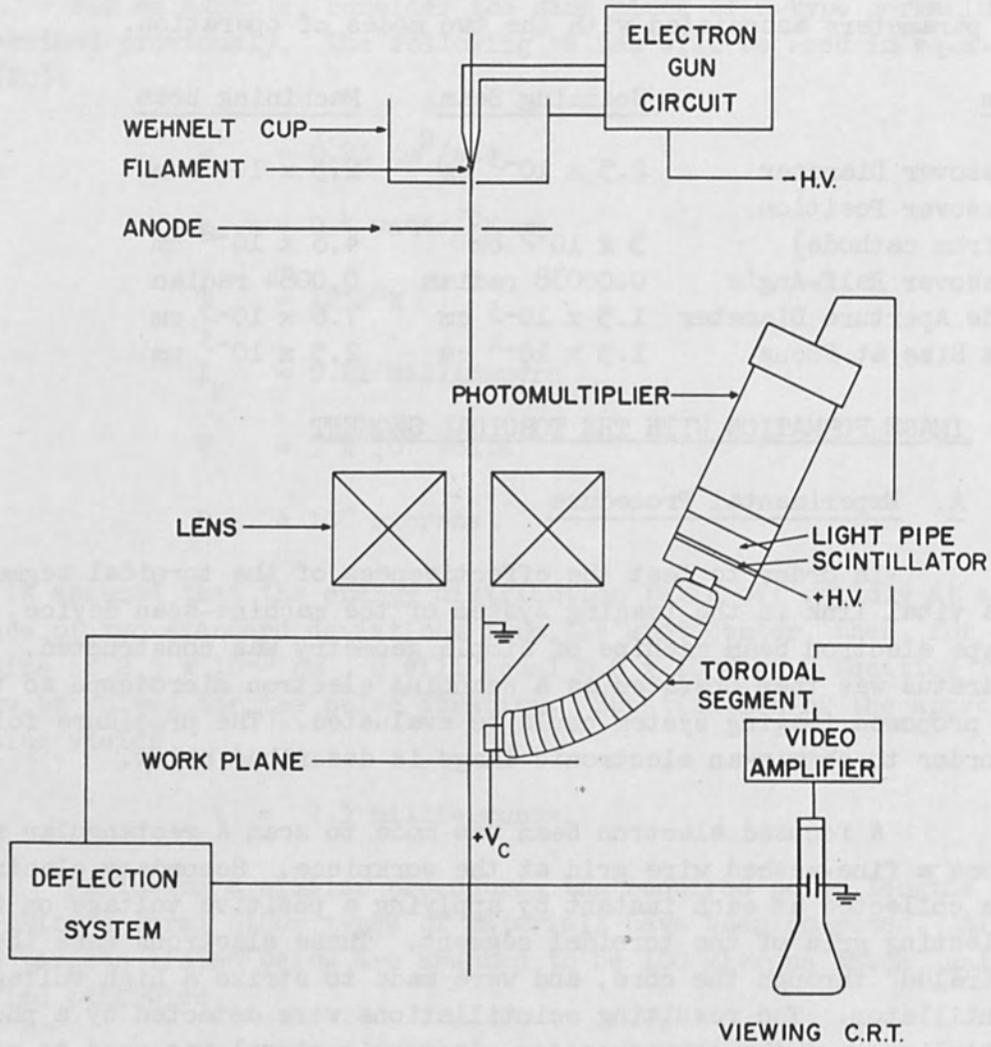


FIGURE 5 - EXPERIMENTAL APPARATUS FOR DETERMINING THE IMAGING PROPERTIES OF A TOROIDAL ELECTRON COLLECTOR

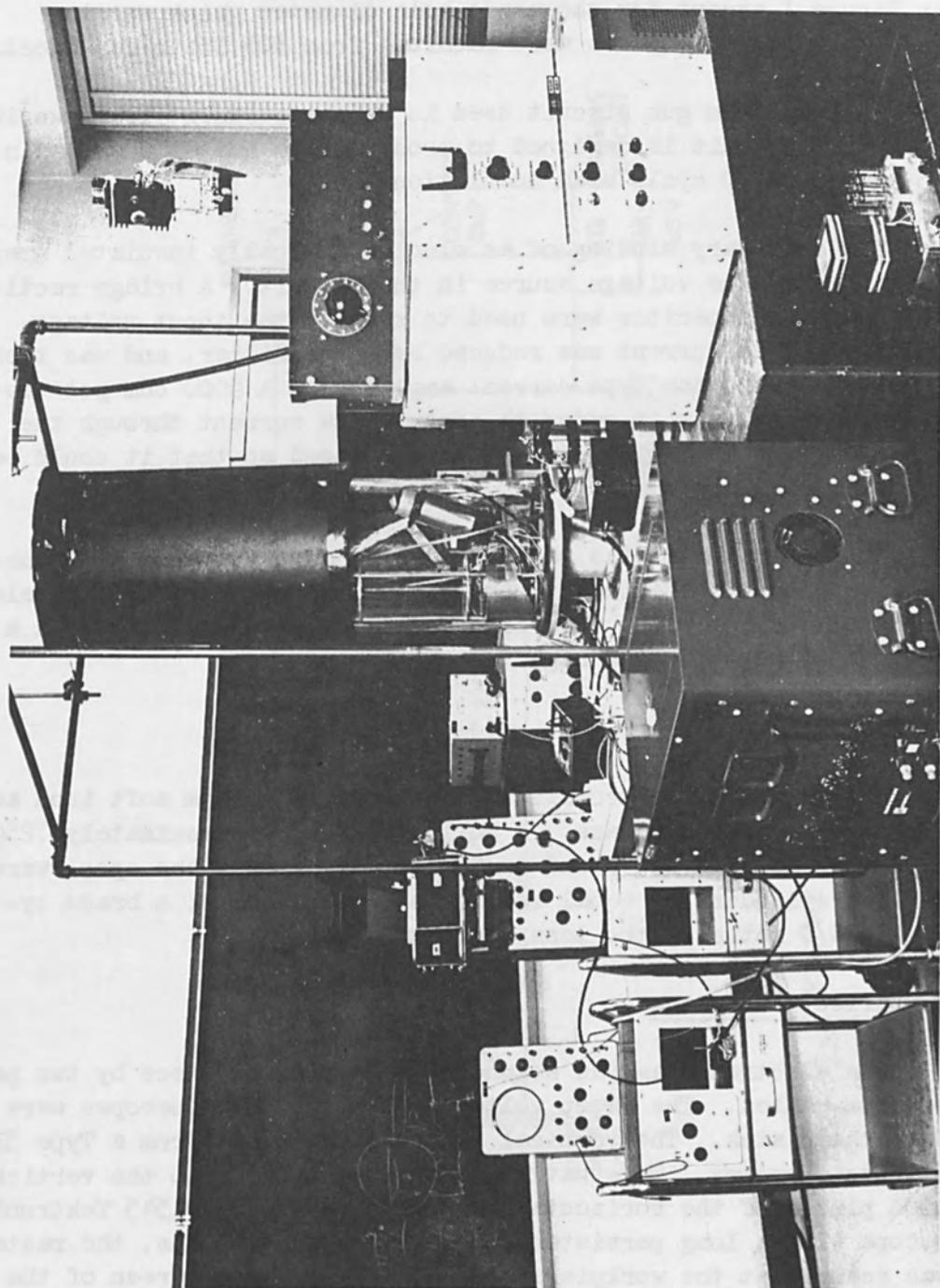


FIGURE 6 - PHOTOGRAPH OF THE EXPERIMENTAL APPARATUS

C. Electron Gun

The electron gun used in this experiment was the standard microscope gun as shown in Figure 1. All of the gun parameters are given in Figure 1 except for the anode-hole diameter which was increased to 1.25 mm. All parts were machined from 303 Stainless Steel.

The electron gun circuit used in this experiment is shown in Figure 7. This circuit is designed to produce a dc heater current in order to eliminate 60 cycle beam modulation.

The secondary winding of an electrostatically insulated transformer was used as the voltage source in the circuit. A bridge rectifier and smoothing capacitor were used to rectify the input voltage. The resulting ripple current was reduced by an RC filter, and was then amplified by a Darlington type current amplifier. A 5000 ohm pot was used as a voltage adjust in order to control the current through the TR-03 transistor. The entire circuit was packaged so that it could be floated at high voltage.

The control cup bias was provided by a potential divider network. By applying a sufficiently negative voltage on the cup, the electron beam could be cut off. This bias was manually adjustable with a long plexiglass rod while the high voltage was on.

D. Lens

The magnetic lens consisted of a case made from soft iron and a coil spool made of plexiglass. A coil containing approximately 7250 turns of Number 28 wire was wound on the spool. Inside the spool were two soft iron cylinders of equal length on either side of a brass cylinder. The S/D ratio of the lens was equal to 1.33.

E. Deflection System

The electron beam was scanned across the workpiece by two pairs of deflection plates. The sweep voltages from two oscilloscopes were applied to the plates. The vertical sweep was obtained from a Type 515A Tektronix Oscilloscope. This waveform was also applied to the vertical deflection plates of the horizontal sweep source - a Type 545 Tektronix Oscilloscope with a long persistence phosphor screen. Thus, the raster which was scanned at the workpiece also appeared on the screen of the Type 545 Oscilloscope. Potential divider networks were used to create symmetrical scans about the beam axis.

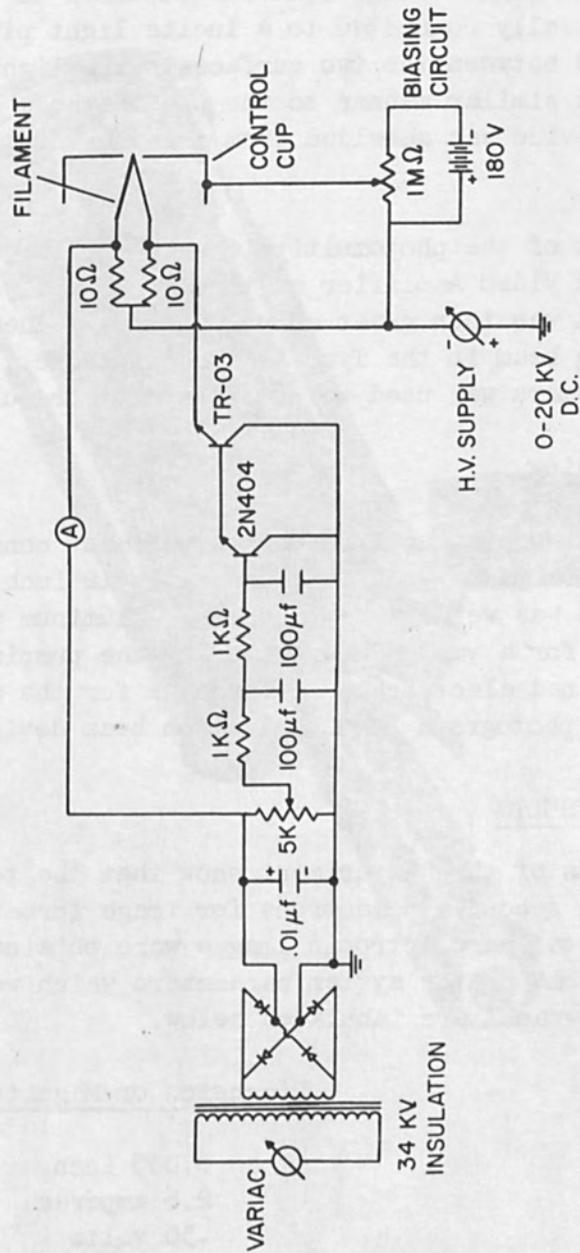


FIGURE 7 - THE ELECTRON GUN CIRCUIT

F. Imaging System

Figure 8 is a photograph of the toroidal segment and electron detection apparatus used in the imaging system. The electron detection device consisted of a scintillator, a light pipe, and a DuMont 6292 Photomultiplier. The scintillator was a NaI crystal coated with a 150 Å layer of aluminum. A high voltage lead was attached to this layer. The crystal was optically connected to a lucite light pipe by a drop of vacuum oil placed between the two surfaces. The light pipe was optically connected in a similar manner to the photocathode of the multiplier tube. The entire device was shielded from the ambient light in the vacuum chamber.

The output of the photomultiplier was amplified by a Model 450A Hewlett Packard Video Amplifier having a two megacycle bandwidth. The resulting signal was then great enough to create the required modulation of the scanning beam in the Type 545 Oscilloscope. A Tektronix C-19 Oscilloscope Camera was used to photograph the resulting images.

G. Vacuum Housing

The vacuum housing used in this experiment consisted of a 15 inch diameter aluminum plate upon which rested a 12 inch diameter glass bell jar. The plate was welded to a four inch aluminum pipe with a flange on the opposite end for a vacuum connection to the pumping system. The 15 inch plate contained electrical feedthroughs for the electron beam apparatus. Figure 9 is a photograph of the electron beam device and the vacuum housing.

X. DISCUSSION OF RESULTS

The results of this experiment show that the toroidal-core method of collecting secondary electrons for image formation is effective. Reasonably detailed electronic images were obtained using the described system. Some of the system parameters which were held constant for all of the micrographs are tabulated below.

<u>Item</u>	<u>Dimension or Magnitude</u>
Filament diameter	0.005 inch
Filament current	2.6 amperes
Grid bias	-30 volts
Scintillator voltage	7000 volts
Photomultiplier anode voltage	850 volts
Deflection sawtooth	100 volts p.p.

The beam spot size was estimated to be 2.5×10^{-3} cm in diameter. The scanned area at 5000 volts accelerating potential was

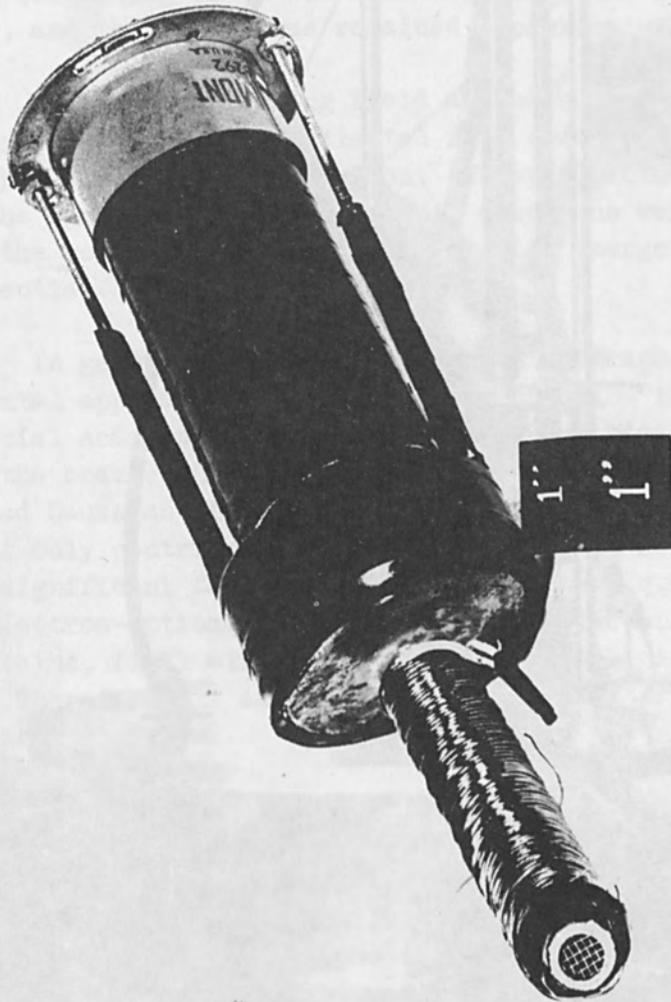


FIGURE 8 - PHOTOGRAPH OF THE TOROIDAL-CORE-PHOTOMULTIPLIER
COMBINATION

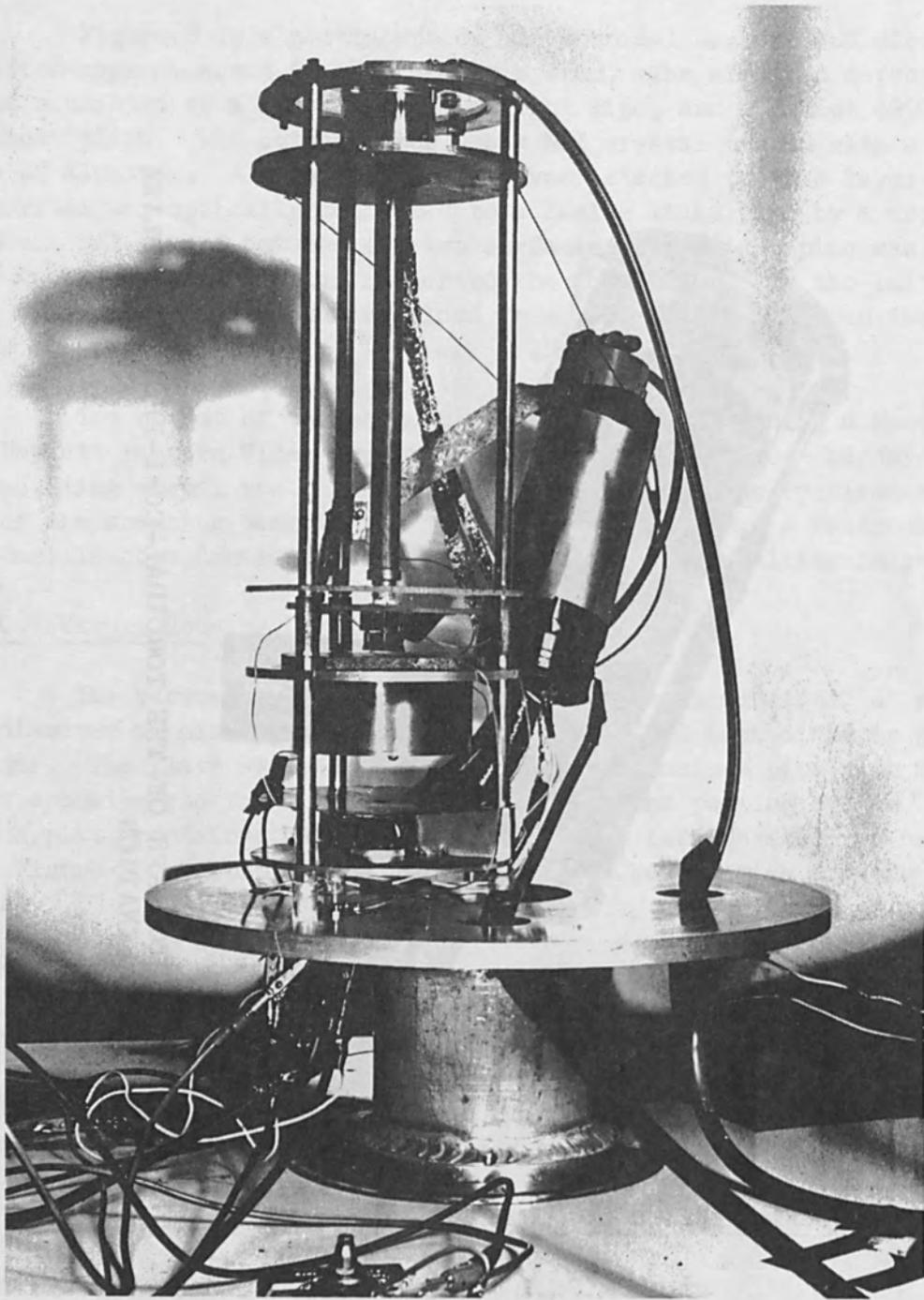


FIGURE 9 - PHOTOGRAPH OF THE ELECTRON BEAM APPARATUS
AND VACUUM HOUSING

two millimeters square. Thus, the maximum number of lines attainable in one frame without overlap was approximately 80 lines. Most of the micrographs averaged 100 lines each.

Picture contrast was affected when the coil current in the toroidal segment was changed. For higher exciting currents more electrons were able to pass through the core (that is, higher energy electrons and electrons more nonparaxial in nature). Since the secondary electron collecting field was held constant, the light areas appeared brighter, and the dark areas remained approximately the same shade.

When the collecting field at the workpiece was increased in intensity, electrons were collected from some areas which did not contribute to previous picture formation. However picture quality was lowered due to the fact that many "important" electrons were not able to pass through the core. These electrons were too energetic due to the increased grid potential.

In general the quality of the micrographs obtained with the experimental apparatus of Figure 5 was far below that obtainable with a commercial scanning microscope. No current limiting apertures were used in the beam column, so that the resulting spot contained a rather widespread Gaussian distribution of electrons. Thus, the peripheral electrons only contributed spurious signals and noise in the final image. Another significant factor in determining image quality is the alignment of the electron-optical system. Because of the simple construction of the apparatus, final alignment was not possible once the system was under vacuum. Therefore, it should be accepted that serious aberrations were present in the system.

XI. CONCLUSIONS

The machine-scan concept will be useful in applications requiring welding, vaporizing, annealing and even machining of very small devices. The incorporation of the scanning mode into the device has limited its operation as a machining tool of metals with high melting points. Nevertheless, when these metals are in the form of thin sheets, some degree of drilling and metal removing can be accomplished.

There is still much research to be done before a commercially acceptable machine-scan device is available. The following areas will be investigated:

1. Increase the overall frequency response of the electronic imaging system so that the number of frames per second can be increased significantly.

2. Investigate other types of electron guns, especially those where the beam crossover changes appreciably with grid bias.
3. Investigate the possibility of lowering the spherical aberration of the final lens in order to increase the maximum power density at the focus.
4. Design a suitable pulsing system that will accurately control each mode of operation.

REFERENCES

1. El-Kareh, A. B., "An Electron Beam Machine", RCA Review 24-1, p. 5, 1963.
2. Yoder, L. W., "A Scanning Beam Viewing Technique for the Electron Beam Machine", Thesis, Pennsylvania State University, March 1965.
3. Pittaway, L. G., "The Temperature Distributions in Thin Foil and Semi-Infinite Targets Bombarded by an Electron Beam", Proceedings of the Electron Beam Symposium Fifth Annual Meeting, p. 88, 1963.
4. Von Borries, B., "The Energy Data and Limits of Electron Microscopy", Optik, 3, p. 389, 1948.
5. Haine, M. E. and Einstein, P. A. "Characteristics of the Hot Cathode Electron Microscope Gun", British Journal of Applied Physics, 3, p. 40, 1952.
6. Spangenberg, K. R., Vacuum Tubes, 1st ed., p. 416, McGraw-Hill Book Company, Inc., New York City, 1948.
7. Liebmann, G. and Grad, E. M., "Imaging Properties of a Series of Magnetic Electron Lenses", Proceedings of the Physical Society of London, 64-B, p. 956, 1951.
8. Everhart, T. E., Wells, O. C., and Oatley, C. W., "Factors Affecting Contrast and Resolution in the Scanning Electron Microscope", Journal of Electronics and Control, 7, p. 97, 1959.
9. McMullan, D., "An Improved Scanning Electron Microscope for Opaque Specimens", Proceedings of the Institution of Electrical Engineers, 100, p. 245, 1953.
10. Schwartz, J. W., "Space-Charge Limitations on the Focus of Electron Beams", RCA Review, 18-1, p. 3, 1957.
11. Gabor, D., The Electron Microscope, p. 58, Hulton, 1946.

APPLICATIONS OF THE BARRIER ELECTRON VOLTAIC EFFECT

by

C. Munakata
Central Research Laboratory
Hitachi Ltd.
Kokubunji, Tokyo, Japan

ABSTRACT

Experiments of measuring the dynamic resistance of a p-n junction and the diffusion length for minority carriers are described.

A Si p-n junction whose geometry is $0.3 \times 0.3 \times 3$ (mm) was irradiated by a pulsed electron beam whose frequency, energy, beam current and spot diameter were 1000 c/s, 20 keV, $50 \mu\text{A}$ (mean value) and about $10 \mu\text{m}$ or less, respectively. The dynamic resistance was varied from $10 \text{ k}\Omega$ to 200Ω and the measured value coincided fairly well with the calculated. However, generally, the measured value was smaller than the calculated. The present method is convenient to measure the effective diffusion potential of the junction, and it was 0.73 V in the experiment.

Using a metal-semiconductor point contact and an electron beam, the diffusion length for holes in a Ge crystal was measured. The geometry and the mean specific resistance of the Ge pellet were $0.5 \times 1 \times 1.9$ (mm) and $2 \Omega\text{-cm}$, and the mean diffusion length was $32 \mu\text{m}$.

I. INTRODUCTION

When a semiconductor crystal pellet or a p-n junction is irradiated by an electron beam or accelerated electrons, there appears a voltage. This phenomenon is generally called an electron voltaic effect after the photo-voltaic effect. The electron voltaic effect can be, however, classified into two different types. One is the bulk electron voltaic effect and another is the barrier electron voltaic effect. The former is due to the resistivity inhomogeneity in a bulk semiconductor crystal¹ and this phenomenon is successfully applied to the measurement of the inhomogeneity in a semiconductor crystal². The latter is due to the barrier layer between a p-type and an n-type semiconductor crystal³ or between a metal and a semiconductor. The application of the barrier electron voltaic effect is the theme of the present paper.

As is well known, electrons and holes created by the impinging

electrons are separated by an intense electric field at the p-n junction or the metal-semiconductor junction. An electric current flows then through an external circuit. If there is no external circuit, the whole current produced by the barrier electron voltaic effect is counterbalanced by the forward current of the junction diode. When the external circuit has a finite electric resistance, the induced current due to the electron bombardment is partly counterbalanced by the forward current and partly flows into the external circuit⁴. Therefore, the current flowing into the external circuit is diminished when the forward current can flow very easily, which certainly comes true at a comparatively high forward bias voltage. This means that the current in the external circuit depends upon the forward conductance of the junction diode. By measuring the external circuit current due to the electron bombardment, the forward conductance then can be obtained. If the induced current is small compared with the forward current, the dynamic conductance or resistance of the diode can be obtained.

In the present experiment the pulsed electron beam scanned upon the diode, and the external circuit current was integrated electronically. The integrated value shows a step form which is similar to the form of the conduction band in a junction diode. When the forward bias voltage is high, the step becomes low since the external current becomes small. This step form is convenient to see the qualitative movement of the conduction band.

When the junction diode is scanned by the electron beam, the diffusion length for minority carriers is obtained from the induced current which varies according to the movement of the beam⁵. This method of obtaining the diffusion length is certainly simple, but not so convenient since a p-n junction must be prepared. In this experiment a metal-semiconductor junction was used instead of the p-n junction. Since this technique needs no p-n junction, it is easily applicable to any single pellet of a semiconductor crystal. Though the present technique is similar to the Morton-Haynes method⁶, it is convenient for measuring a short diffusion length since the spot size of the electron beam can be made small and the beam is minutely deflected not mechanically but electrostatically.

II. PRINCIPLE OF THE METHOD OF OBTAINING THE DYNAMIC RESISTANCE OF A JUNCTION DIODE

Since the barrier electron voltaic effect is quite analogous to the barrier photo-voltaic effect except that the primary electron beam carries a current into the specimen diode, the equivalent circuit of the electron voltaic cell is quite the same as that of a photo-voltaic cell⁷ as is shown in Fig. 1. The diode capacitance is, however, neglected

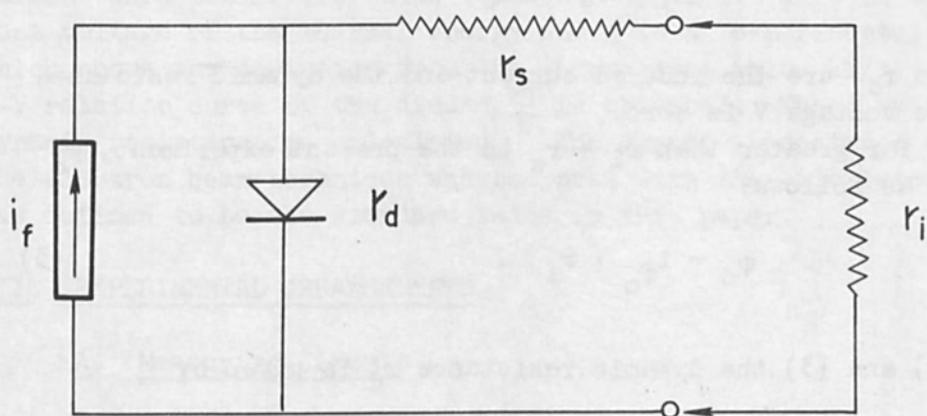


Figure 1. Equivalent Circuit of the Electron Voltaic Cell. The Diode Capacitance is Omitted.

because its effect is not important in the present experiment. In Fig. 1 i_f means a pulsive induced current due to the barrier electron voltaic effect, and r_d , r_s and r_i are the dynamic resistance, the series resistance of the diode and the input resistance of an amplifier, respectively. From the equivalent circuit the voltage drop ϕ through r_i due to i_f is easily calculated by

$$\phi = i_f \frac{r_i r_d}{r_i + r_s + r_d} \quad (1)$$

When the bias voltage is zero, the voltage ϕ_0 is given by

$$\phi_0 = i_{f_0} \frac{r_i r_{d_0}}{r_i + r_s + r_{d_0}} \quad (2)$$

where i_{f_0} and r_{d_0} are the induced current and the dynamic resistance when the bias voltage V is zero.

Since r_{d_0} is far greater than $r_i + r_s$ in the present experiment, ϕ_0 is expressed as follows:

$$\phi_0 = i_{f_0} \cdot r_i \quad (3)$$

From eqs. (1) and (3) the dynamic resistance r_d is given by

$$\begin{aligned} r_d &= \frac{(r_i + r_s)}{m \left(\frac{i_f}{i_{f_0}} \right) - 1} \quad , \\ &= \frac{r_i}{m \left(\frac{i_f}{i_{f_0}} \right) - 1} \quad (r_i \gg r_s), \quad (4) \end{aligned}$$

where m is ϕ_0/ϕ and r_i is assumed to be far larger than r_s . This relation holds also for the integrated value and m is calculated from the step height of the integrated value which is shown in Fig. 6. The current ratio i_f/i_{f_0} is called in this paper a collection efficiency of the induced current at the junction.

From eq. (4) the dynamic conductance of the junction is given by

$$1/r_d = [m \left(\frac{i_f}{i_{f_0}} \right) - 1] \left(\frac{1}{r_i} \right) \quad (5)$$

On the other hand, since the current voltage relation of a normal p-n junction diode is given by

$$I = I_0 \left(e^{(V)/(\eta kT/e)} - 1 \right) , \quad (6)$$

the dynamic conductance $1/r_d$ is expressed by the following equation:

$$(I + I_0) = \eta \frac{kT}{e} \left(\frac{1}{r_d} \right) . \quad (7)$$

In the above equations I is the diode current and I_0 is the saturation current when the reverse bias voltage is applied. kT/e is the equivalent voltage of the thermal energy and η is an experimental coefficient⁸ which shows the deviation from the theoretical value kT/e . From the I-V relation curve of the diode $\eta \frac{kT}{e}$ is obtained. Then, by eq. (7), the dynamic resistance is calculated. The dynamic resistance obtained by the electron beam technique was compared with the calculated value which was defined to be the standard value in this paper.

III. EXPERIMENTAL ARRANGEMENTS

A. Measuring Circuit

The block diagram of the measuring system including the simplified sketch of the electron beam system is shown in Fig. 2. The accelerating voltage is denoted by U . The half-fixed bias circuit⁹ was used. U was 20 kV. The electron beam was modulated by a 1000 c/s rectangular wave which was supplied from the pulse generator to the Wehnelt electrode through an isolating pulse transformer P.T. The beam was focused by two electron lenses and the spot diameter was considered to be less than $10 \mu\text{m}$. The mean beam current I_B , which was measured with a d.c. meter, was varied from $50 \mu\text{A}$ to 1 mA . Since the beam current was pulsive the induced current was also pulsive. The signal was amplified with a tuned amplifier and it was then rectified by a phase sensitive demodulator¹⁰. The rectified signal was integrated and the integrated value as well as the rectified signal was displayed on a screen of a synchroscope. The beam was deflected by the saw-tooth wave voltage from the synchroscope. The electronic integrator was reset synchronously with the saw-tooth wave voltage.

The d.c. bias voltage was supplied to the specimen diode through the input transformer of the tuned amplifier. The series resistance of the d.c. bias voltage source was less than 125Ω , which was neglected since the input resistance of the amplifier was $4.8 \text{ k}\Omega$.

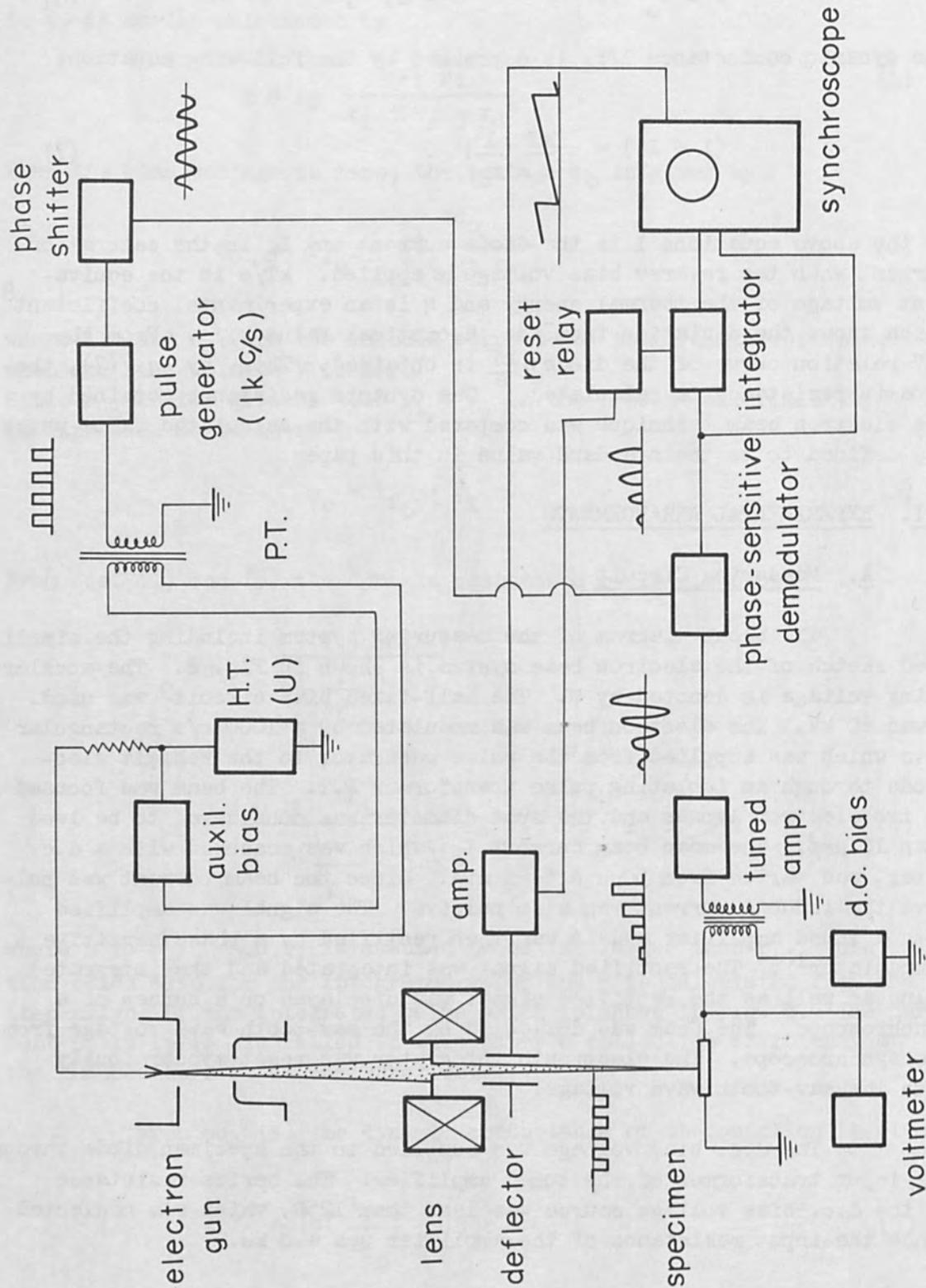


Figure 2. Block Diagram of the Measuring System.

B. Specimen Arrangement

A grown type Si p-n junction diode was used. The geometry of the diode is shown in Fig. 5. This diode was set on a transistor mount and the p-type side was grounded. The n-type side was connected with the amplifier and the bias voltage source.

C. Specimen and Circuit for Measuring the Diffusion Length

An n-type Ge pellet, whose thickness, width and length are 0.5 mm, 1 mm and 1.9 mm, respectively, was used.

The pellet was fastened upon a transistor mount with a non-conductive paste, and a tungsten wire was made touched upon the pellet. The contact was proved in advance to be non-ohmic.

The photo of the specimen is shown in Fig. 3. As is seen from the photo, the surface is not smooth. This surface was electrolytically etched and then chemically etched. The resistivity varies periodically from $1\Omega\text{-cm}$ to $3\Omega\text{-cm}$ in the pellet and the mean periodic length is about $240\mu\text{m}$. The mean resistivity is $2\Omega\text{-cm}$. Because of the reason mentioned above, it is imagined that the diffusion length will be short as well as position-dependent. The metal side of the point contact was grounded and the semiconductor side was connected with a d.c. voltmeter in this case. A d.c. electron beam was moved near the point contact and the induced voltage rather than the induced current was observed as a function of the beam position.

IV. RESULTS

A. Measurement of Dynamic Resistance

1. Current-Voltage Relation of the Diode

The I-V relation of the diode is shown in Fig. 4 (linear scale) and Fig. 5 (semilogarithmic scale). From the straight line in Fig. 5, $\eta kT/e$ was obtained to be 45 mV. Since kT/e is 25.7 mV, the coefficient η is 1.75. As is seen from the line in Fig. 5, the saturation current I_0 should be very small. The dynamic resistance r_d is then calculated by eq. (7) by neglecting I_0 .

2. Integrated Value

Typical examples of the integrated step curves are shown in Fig. 6. When I_B was zero, the integrated curve was exactly flat.

With the increase of the forward bias voltage, the height

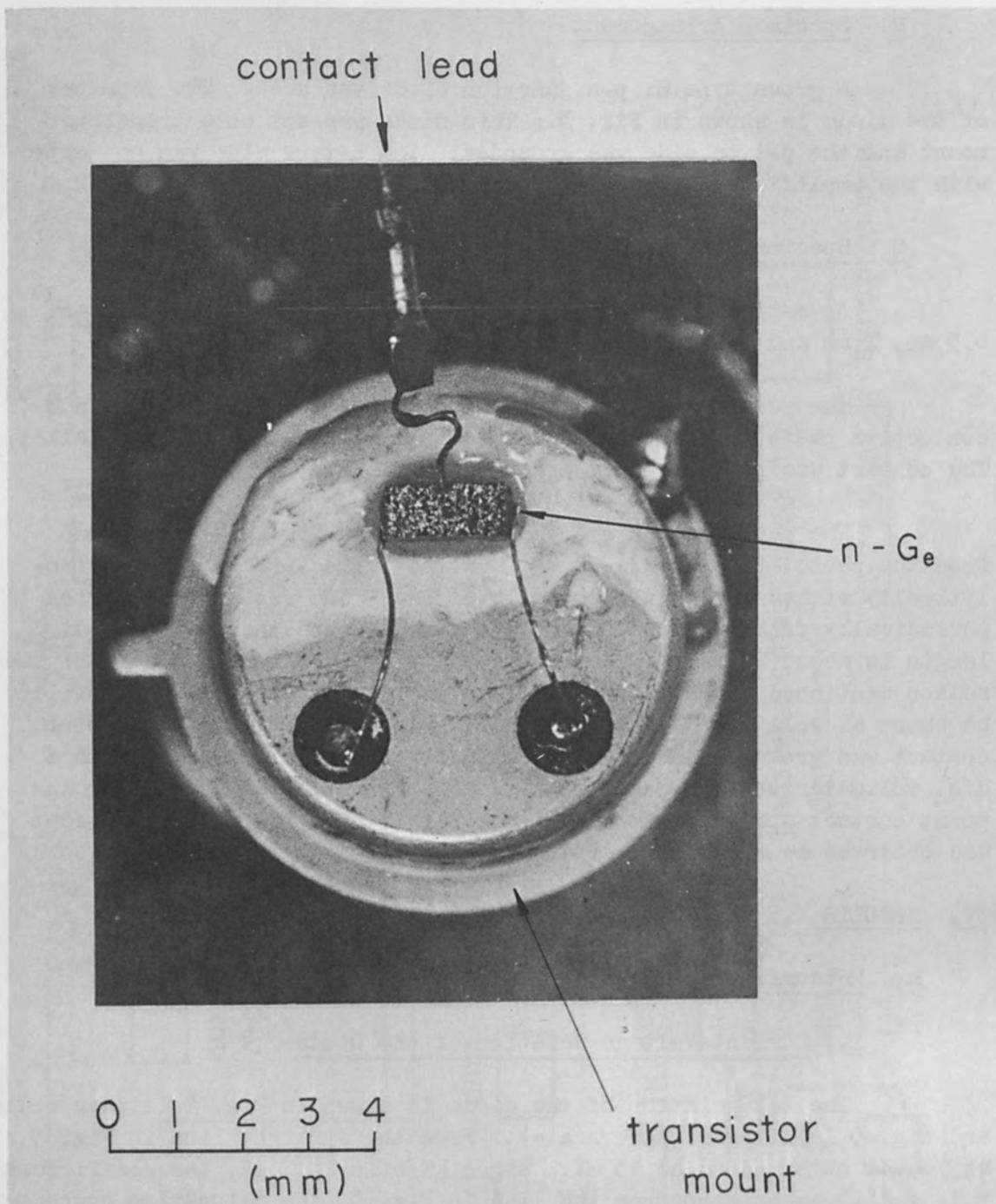


Figure 3. n-Type Ge Pellet and the Point Contact. This Was Used to Measure the Diffusion Length for Holes.

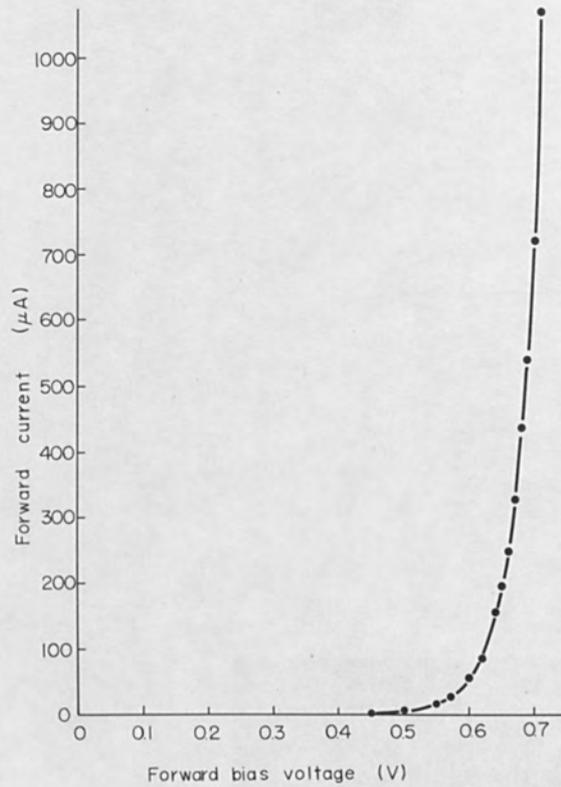


Figure 4. Characteristic of the Si Diode (I).

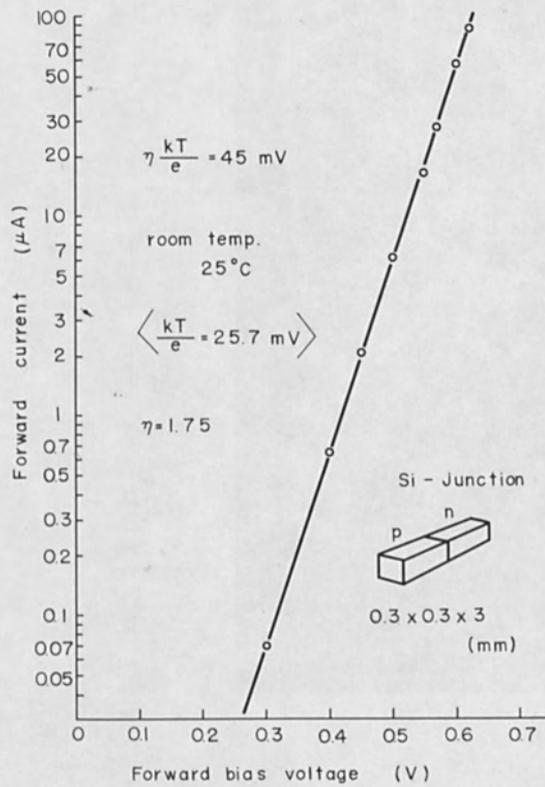
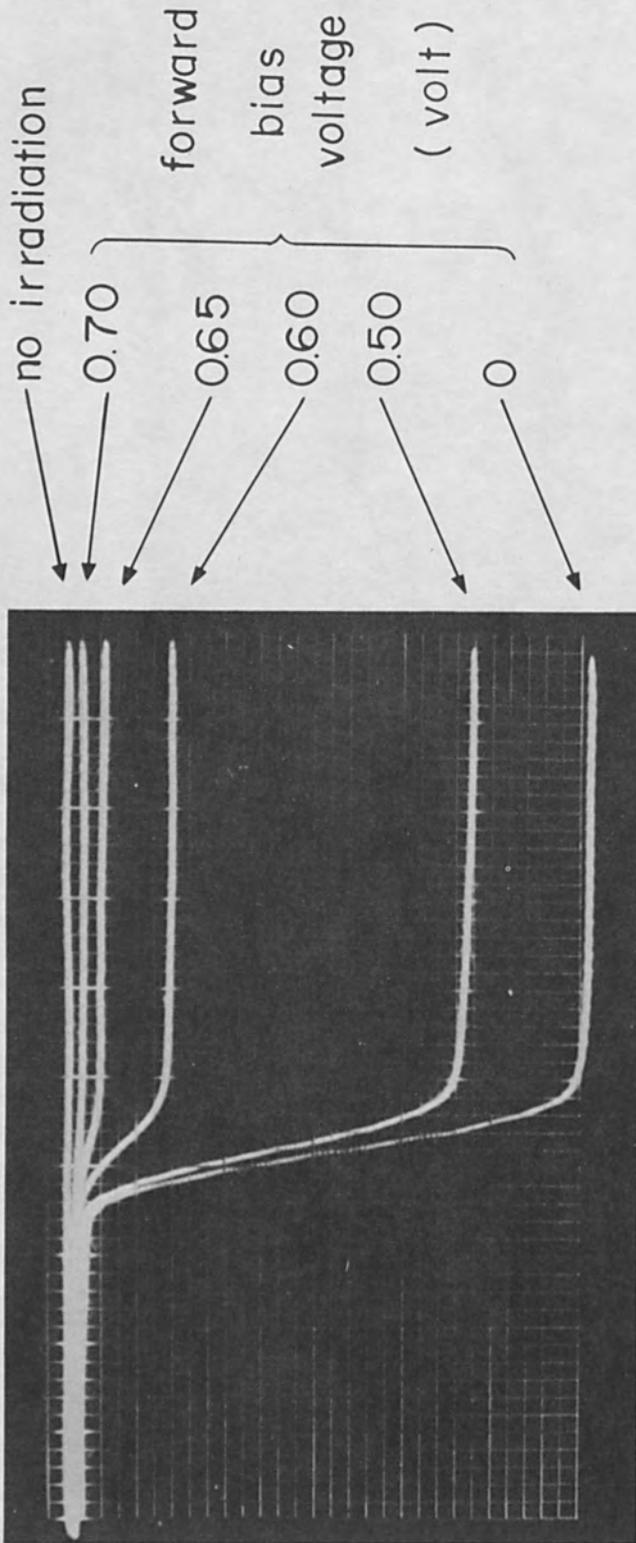


Figure 5. Characteristic of the Si Diode (II).



$I_B = 100 \mu\mu A$
 $U = 20 kV$
 $r_i = 4.8 k\Omega$
(sweep : 0.1 sec/cm)



Si p-n junction

Figure 6. Step Curves. The Curve was Obtained by Integrating the Induced Current.

of the step becomes smaller. This corresponds to the decrease of the barrier potential at the junction, because the barrier potential is cancelled by the forward bias voltage, which is schematically shown in Fig. 7. When the forward bias voltage is close to the diffusion potential, the barrier against electrons and holes becomes flattened, and the forward current is very large. Therefore, the dynamic resistance is small.

In the experiment the step height is proportional to the voltage ϕ in eq. (1). Accordingly, the step height becomes small when the dynamic resistance r_d is small. At the forward bias voltage of 0.7 volt, the step became nearly flattened as is shown in Fig. 6. This means that the dynamic resistance r_d was far smaller than the input resistance of the amplifier. From the forward characteristic curve in Fig. 4, it is seen that the forward current increases steeply near the voltage of 0.7 volt. Thus the voltage with which the step becomes flattened gives an effective diffusion potential of the diode and it was 0.73 volt in the experiment.

Since the voltage ratio m is easily obtained from the step heights, r_d is calculated by eq. (4).

3. Dynamic Resistance

The dynamic resistances are shown in Fig. 8. In this case, however, i_f/i_{f_0} was assumed to be unity. The full line shows the standard value of the dynamic resistance which is calculated from the I-V characteristic of the diode. The coincidence between the experimental and the standard value is fairly good. However, the experimental value is smaller than the standard when the forward current I is larger than $10 \mu\text{A}$. Conversely, it is usually larger than the standard when I is smaller than $10 \mu\text{A}$.

The same data in Fig. 8 are plotted in a linear scale in Fig. 9. As is seen from eq. (7), the gradient of the line gives $\eta kT/e$. The experimental value of $\eta kT/e$ thus obtained is 42 mV and it is smaller than the standard.

4. Collection Efficiency or Current Ratio i_f/i_{f_0}

It is imagined that with the forward bias voltage, the minority carriers created by the electron beam will not be efficiently collected since the electric field at the junction becomes small.

By using eq. (5) and assuming that the deviation of the experimental value of r_d from the standard is wholly due to the collection efficiency, the efficiency i_f/i_{f_0} was calculated. The calculated

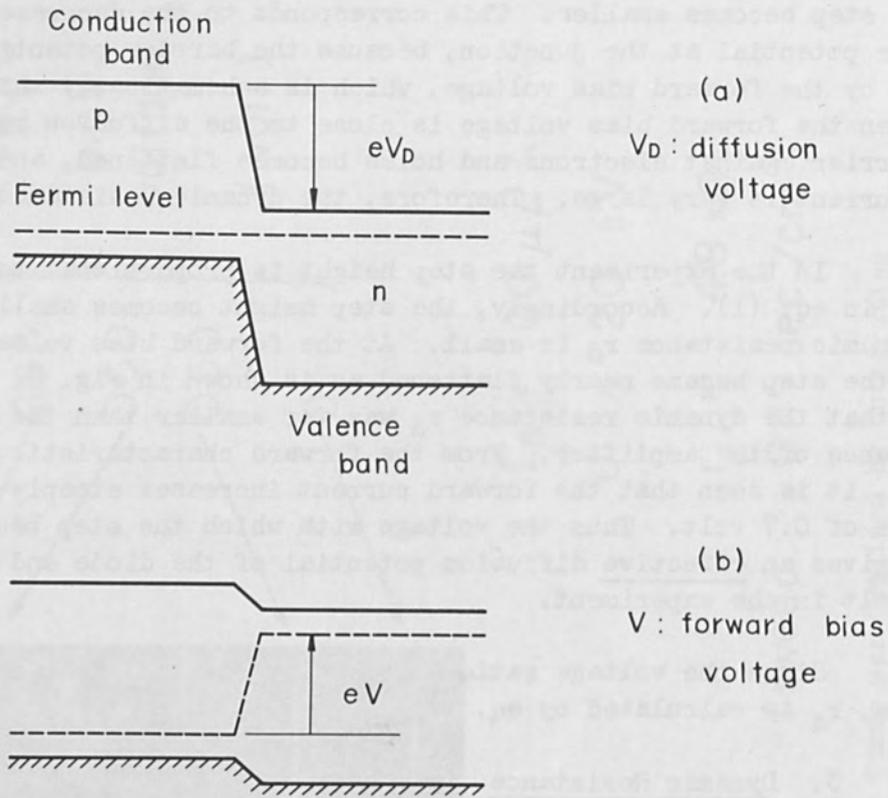


Figure 7. Conduction Band, Valence Band and Fermi-level in a p-n Junction Diode.

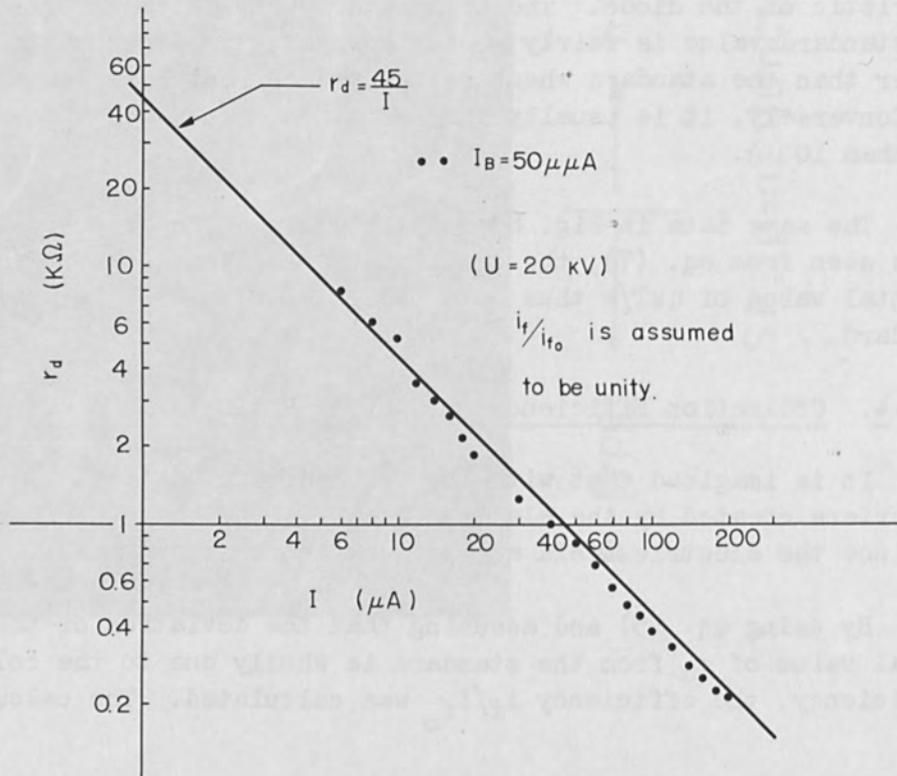


Figure 8. Dynamic Resistances.

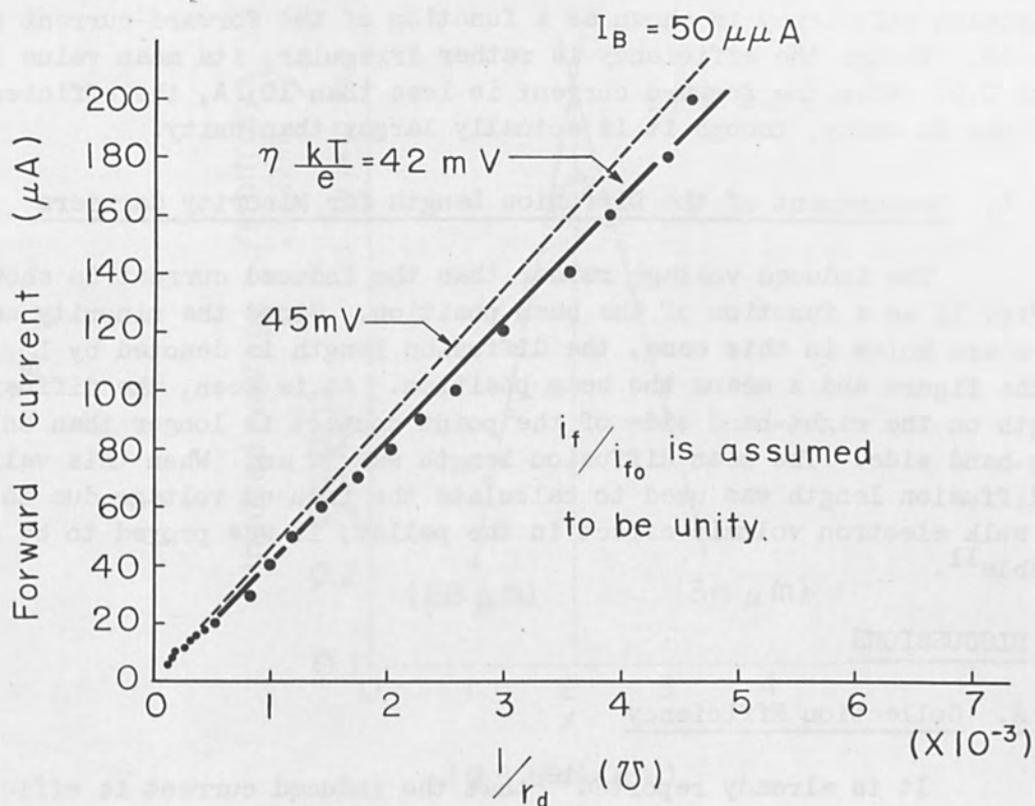


Figure 9. Dynamic Conductances.

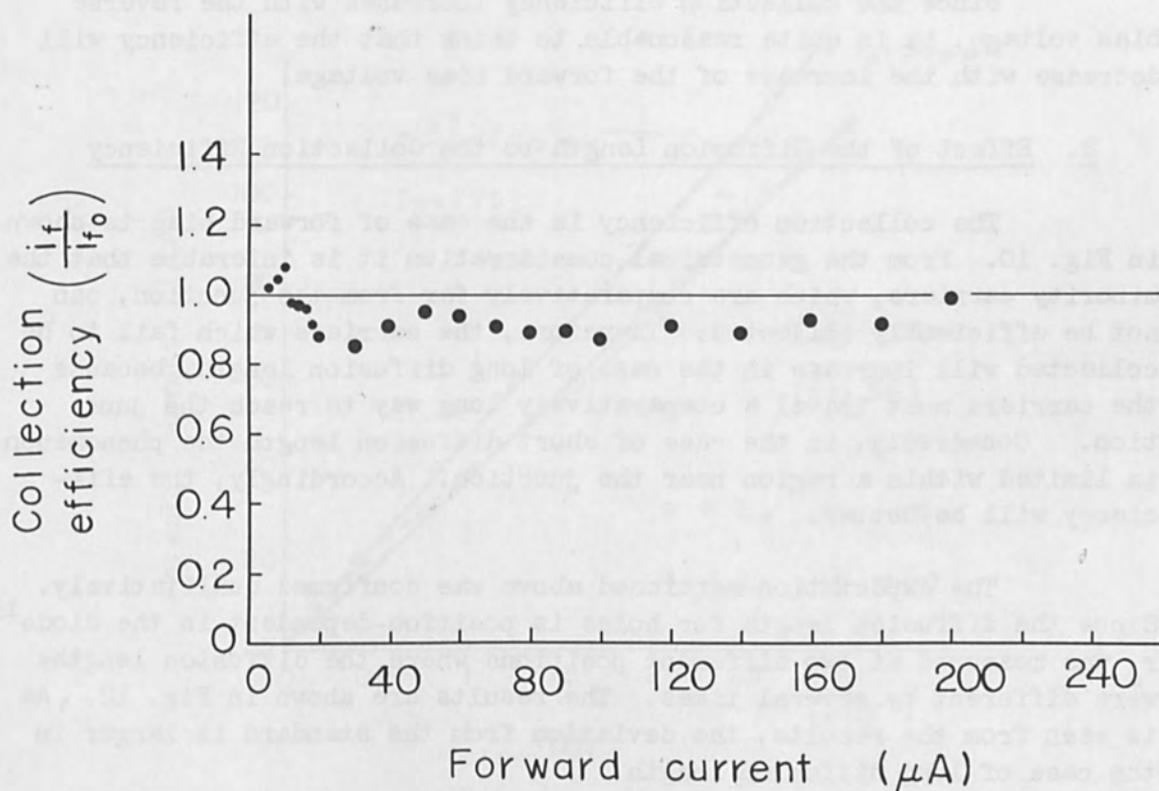


Figure 10. Collection Efficiency.

collection efficiency is shown as a function of the forward current in Fig. 10. Though the efficiency is rather irregular, its mean value is about 0.9. When the forward current is less than $10\mu\text{A}$, the efficiency is close to unity, though it is actually larger than unity.

B. Measurement of the Diffusion Length for Minority Carriers

The induced voltage rather than the induced current is shown in Fig. 11 as a function of the beam position. Since the minority carriers are holes in this case, the diffusion length is denoted by L_p in the figure and x means the beam position. As is seen, the diffusion length on the right-hand side of the point contact is longer than on the left-hand side. The mean diffusion length was $32\mu\text{m}$. When this value of diffusion length was used to calculate the induced voltage due to the bulk electron voltaic effect in the pellet, it was proved to be reasonable¹¹.

V. DISCUSSIONS

A. Collection Efficiency

It is already reported¹² that the induced current is efficiently collected with a high reverse bias voltage and the saturated induced current is larger than that collected with zero bias.

Since the collection efficiency increases with the reverse bias voltage, it is quite reasonable to think that the efficiency will decrease with the increase of the forward bias voltage.

B. Effect of the Diffusion Length to the Collection Efficiency

The collection efficiency in the case of forward bias is shown in Fig. 10. From the geometrical consideration it is inferable that the minority carriers, which are comparatively far from the junction, can not be efficiently collected. Therefore, the carriers which fail to be collected will increase in the case of long diffusion length, because the carriers must travel a comparatively long way to reach the junction. Conversely, in the case of short diffusion length the phenomenon is limited within a region near the junction. Accordingly, the efficiency will be better.

The expectation mentioned above was confirmed qualitatively. Since the diffusion length for holes is position-dependent in the diode¹³, r_d was measured at two different positions where the diffusion lengths were different by several times. The results are shown in Fig. 12. As is seen from the results, the deviation from the standard is larger in the case of long diffusion length.

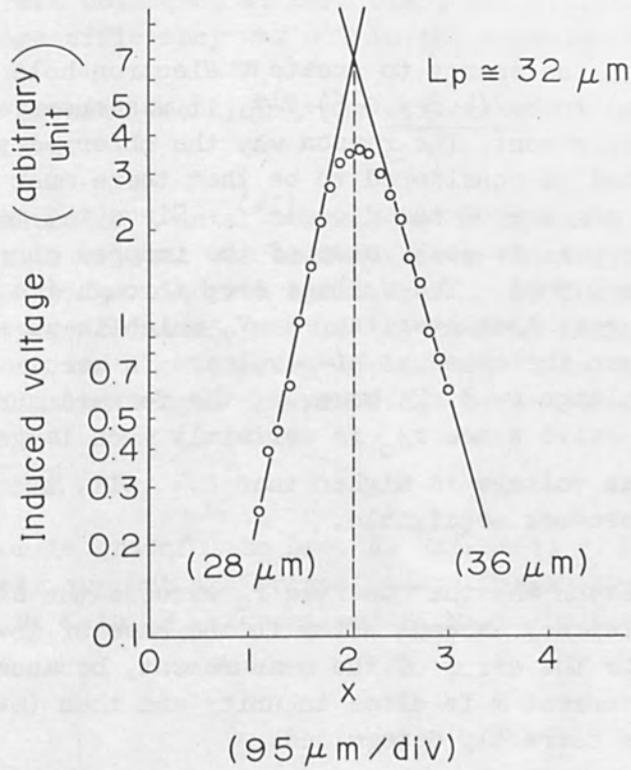


Figure 11. Induced Voltage. The Diffusion Length L_p was Obtained from the Curve.

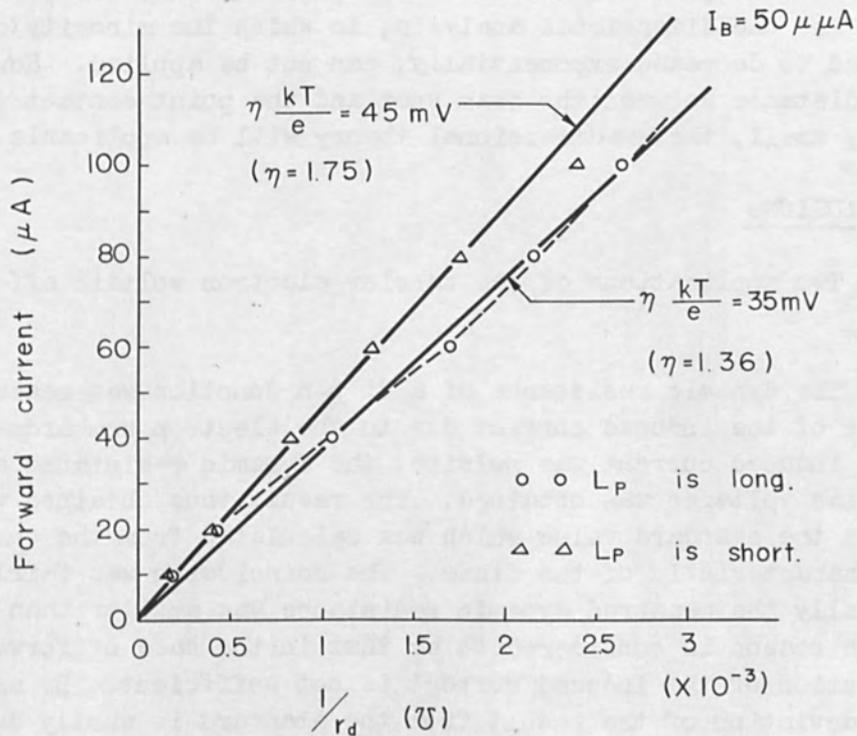


Figure 12. Dynamic Conductances in the Cases of Long and Short Diffusion Length.

C. Value of the Induced Current and Its Effect on the Measurement of r_d

Though the energy to create an electron-hole pair in a Si crystal is reported to be $(4.8 \pm 0.6)eV^{14}$, it was observed to be more than 5 eV in this experiment. The reason why the observed value is higher than the reported is considered to be that there must be a non-sensitive layer upon the surface of the diode⁽¹⁴⁾. Since the mean beam current was $50\mu A$, the peak-to-peak value of the induced current was considered to be less than $0.5\mu A$. The voltage drop through the input resistance of the amplifier is then less than 3 mV, which is an effective forward bias voltage when the external bias voltage is zero. With such a small forward bias voltage as 3 mV, however, the forward current is very small and eq. (3) is valid since r_{d0} is certainly much larger than r_i . When the forward bias voltage is higher than 0.1 volt, the effect of the induced current becomes negligible.

The reason why the observed r_d exceeds the standard, or the collection efficiency exceeds unity in the case of low forward bias, is considered to be the error of the measurement, because in the region of small forward current m is close to unity and then $(m-1)$ is comparatively difficult to be correctly determined.

D. Errors in the Measurement of the Diffusion Length

In the present method of measuring the diffusion length the beam spot and the point contact are both points. Therefore, strictly speaking, the one-dimensional analysis, in which the minority carriers are assumed to decrease exponentially, can not be applied. However, when the distance between the beam spot and the point contact is comparatively small, the one-dimensional theory will be applicable.

VI. CONCLUSIONS

Two applications of the barrier electron voltaic effect are reported.

The dynamic resistance of a Si p-n junction was measured by making use of the induced current due to the electron bombardment. Since the induced current was pulsive, the dynamic resistance at various forward bias voltages was obtained. The result thus obtained was compared with the standard value which was calculated from the current-voltage characteristic of the diode. The coincidence was fairly good, but generally the measured dynamic resistance was smaller than the standard. The reason is considered to be that in the case of forward bias the collection of the induced current is not sufficient. By assuming that the deviation of the result from the standard is wholly due to the

collection efficiency, which is defined to be the ratio of the collected current to the current collected at zero bias, the efficiency was calculated. The average efficiency was 0.9 in the experiment.

With the present technique, the effective diffusion potential can be easily obtained.

By making use of a metal-semiconductor contact, the diffusion length for holes in Ge crystal was measured from the varying form of the induced voltage by the electron bombardment. Diffusion lengths of 28 μm and 36 μm were obtained. Since the present method uses a fine focused electron beam and the scanning technique, it is convenient to measure a short diffusion length for minority carriers.

VII. ACKNOWLEDGEMENTS

The author is grateful to Drs. H. Watanabe, H. Kimura and T. Tokuyama for their support and suggestions. Thanks are also due to Messrs. Y. Ito, H. Sato, H. Higuchi and M. Migitaka for their useful discussions.

REFERENCES

1. C. Munakata: Japan, J. Appl. Phys. 4 697(1965).
2. C. Munakata: Japan, J. Appl. Phys. 4 815(1965)
C. Munakata: Japan, J. Appl. Phys. 5 (1966) (in press)
3. W. Ehrenberg, Chi-Shi Lang & R. West: Proc. Phys. Soc. 64 424(1951)
P. Rappaport: Phys. Rev. 93 246(1954).
4. S. M. Ryvkin: Photoelectric Effects in Semiconductors (Consultants Bureau, New York 1964) Chap. 15.
5. F. S. Goucher, et.al.: Phys. Rev. 81 637(1951).
6. L. B. Baldes: Proc. I.R.E. 40 1420(1952).
7. A. van der Ziel: Solid State Physical Electronics (Prentice-Hall-Maruzen, Tokyo 1959) Chap. 17 p. 383.
8. Y. Watanabe: Semiconductors and Transistors (in Japanese) (Ohmshya, Tokyo 1959) p. 109.
9. C. Munakata & H. Watanabe: J. Electronmicroscopy 11 47 (1962).
10. B. Chance et al.: Waveforms (McGraw-Hill, New York 1949) p. 512.
11. C. Munakata: (to be published).
12. K. Takeya & K. Nakamura: J. Phys. Soc. Japan 2 223(1958).
13. H. Higuchi & H. Tamura: Japan J. Appl. Phys. 4 316(1965).
14. V. M. Patskevich et al.: JETP (U.S.S.R.) 33 804(1957).

A METHOD OF EXAMINATION OF SEMICONDUCTOR OXIDES
WITH A SCANNING ELECTRON MICROSCOPE

by

D. Green
Westinghouse Research Laboratories
Beulah Road, Churchill Borough
Pittsburgh, Pennsylvania

ABSTRACT

The electron beam of a scanning electron microscope was used to induce conductivity changes in a thin layer of silicon oxide. The current flowing from a biased pad on the oxide to the substrate was used to form the microscope image. In this way micrographs showing variations in oxide characteristics were obtained.

I. INTRODUCTION

The scanning electron microscope has been used by various authors in the examination of semiconductor devices^{1,2}. In general there are two modes of examination which are used. In the first mode the device is scanned by the primary electron beam of the microscope and a cathode ray tube which is scanned in synchronism, displays variations in the amounts of secondary electrons collected from various parts of the sample. As the trajectories of the secondary electrons are influenced by electric fields on the surface of the device, this mode of examination shows potential distributions on the surface of the device as well as details of its surface topography.

In the second mode of examination the sample is again scanned with the primary beam which consists of electrons sufficiently energetic to penetrate into the semiconductor material and generate electron hole pairs. If these are generated within a diffusion length of a junction they can be collected as an electron beam induced current. This can then be used to modulate the brightness of the display tube. In this way a micrograph showing details of the diffusions in the device can be obtained.

Neither of these modes of examination yield any information about the electrical properties of the oxides, which are widely used as surface passivating layers and dielectrics in semiconductor devices. However it is well known that thin oxides of various materials which are normally insulators will become conductive under the influence of low energy electron beams. This effect has been studied by several authors

(3) and in the work reported here it was used to obtain information about the uniformity of the semiconductor oxides.

II. EXPERIMENTAL PROCEDURE

The scanning microscope which was used in this study was a two lens instrument (Fig. 1) in which the primary beam energy could be selected in the range 5 - 50 keV with beam currents up to approximately 10^{-9} amps. The diameter of the electron beam at the surface of the sample was of the order to 0.1 microns. A schematic diagram of the main parts of this instrument is given in Figure 2.

The samples were made by growing a layer of oxide on the chemically polished surface of a silicon slice. This was carried out in a wet oxygen atmosphere at 1100°C and yielded an oxide approximately 5,000 Å thick. These slices were covered with an evaporated aluminum film, also about 5000 Å thick, which was then selectively etched to form pads. The slices were then scribed into chips each having an aluminum pad on the oxide.

After mounting the sample in the microscope, a secondary electron image was used to focus the primary beam and to position the scanned raster on the sample. The secondary electron collection system was then disconnected from the video amplifier of the microscope.

The circuit for monitoring the oxide conduction current is shown in Figure 3. The evaporated aluminum pad was connected to a variable bias supply via a 100 kr resistor. A low noise differential amplifier was connected across this resistor, and in this way any current flow through the oxide could be detected. The output of this amplifier, which has a low frequency cut-off of 0.1 cps, was taken to an oscilloscope which allowed the conduction current to be monitored. The output from the oscilloscope was connected to the video amplifier of the microscope which caused the oxide conduction current to modulate the brightness of the display tube.

It was found that when the primary beam was not impinging on the sample no oxide conduction current was detected with this circuit for any bias values, indicating that the oxide behaved as a good insulator under normal conditions. However when the beam was scanned over the pad a current flowed and in this way micrographs were obtained which showed variations in this current.

III. RESULTS

The first series of micrographs (Figs. 4 and 5) show a sample which appeared to have a homogeneous oxide covered with a uniform aluminum

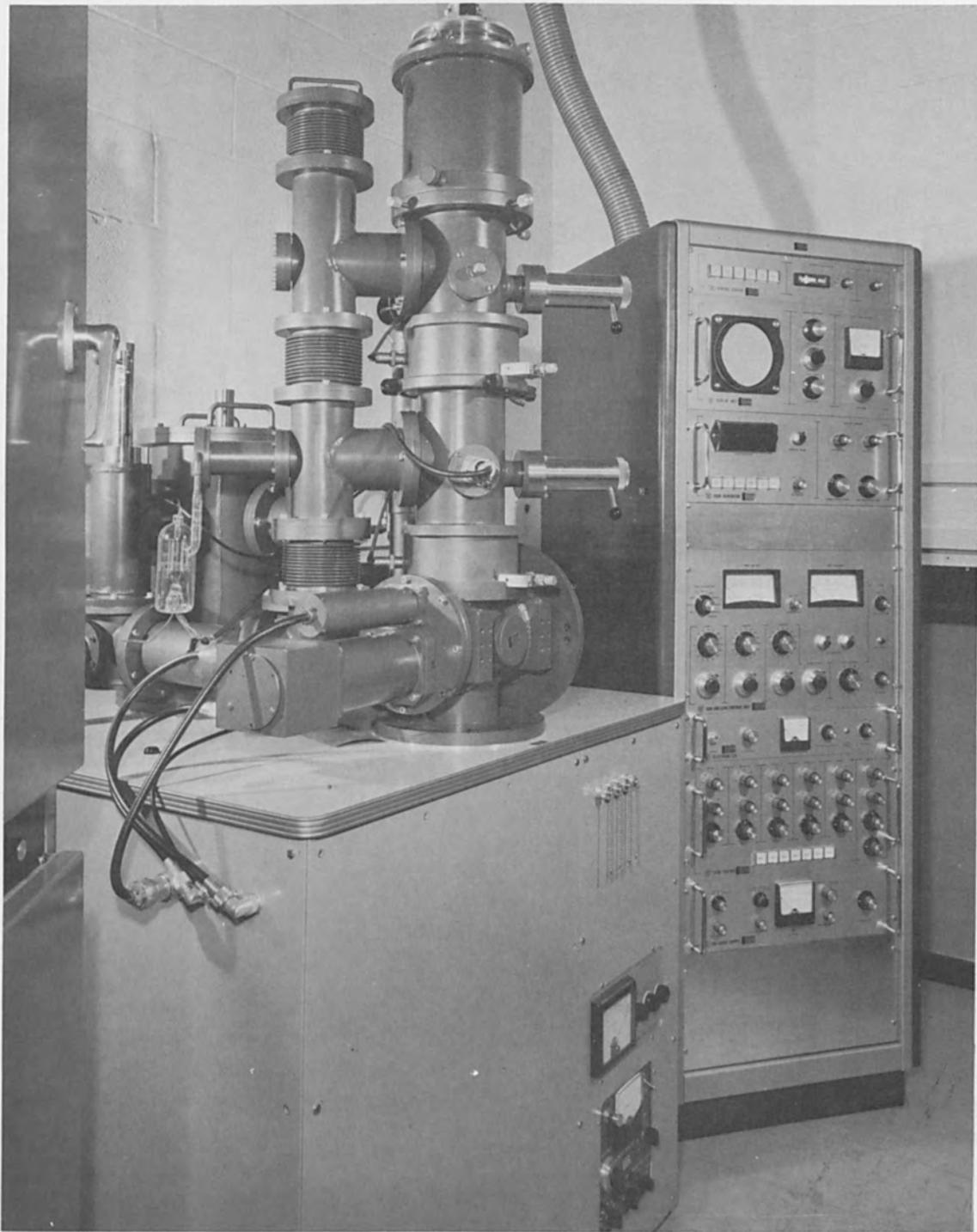


Figure 1. The Scanning Electron Microscope.

DWG. 194 A685

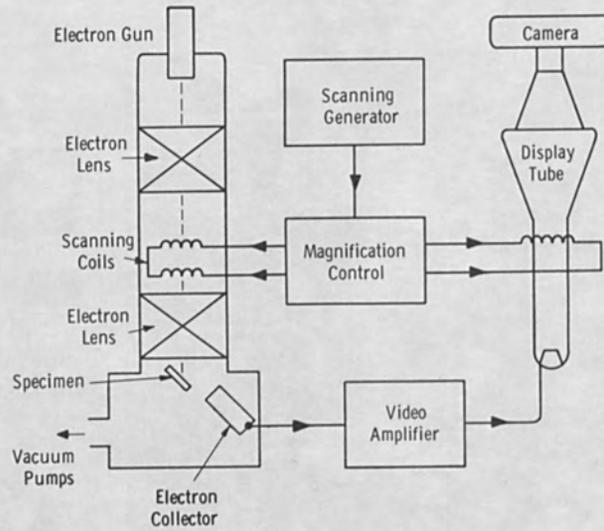


Figure 2. Schematic Diagram of the Scanning Electron Microscope.

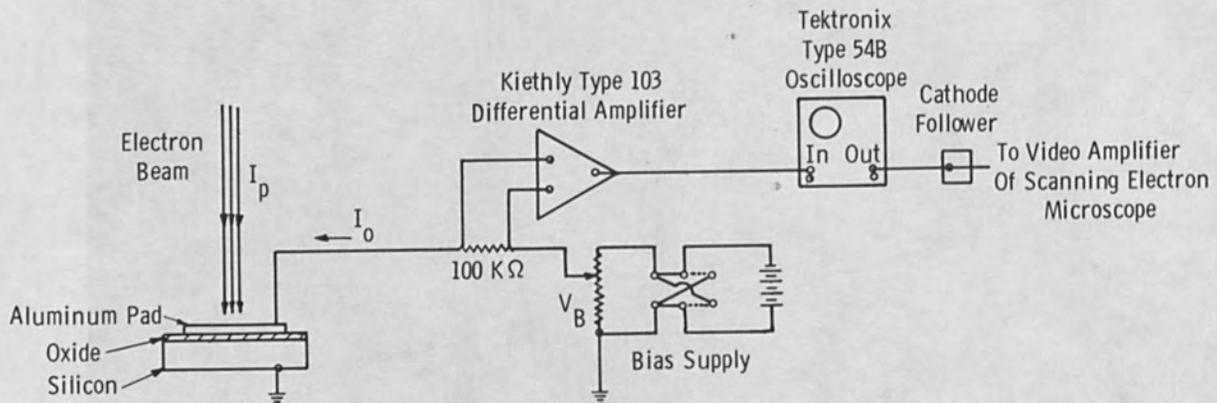


Figure 3. Circuit for Monitoring Oxide Conduction Current.

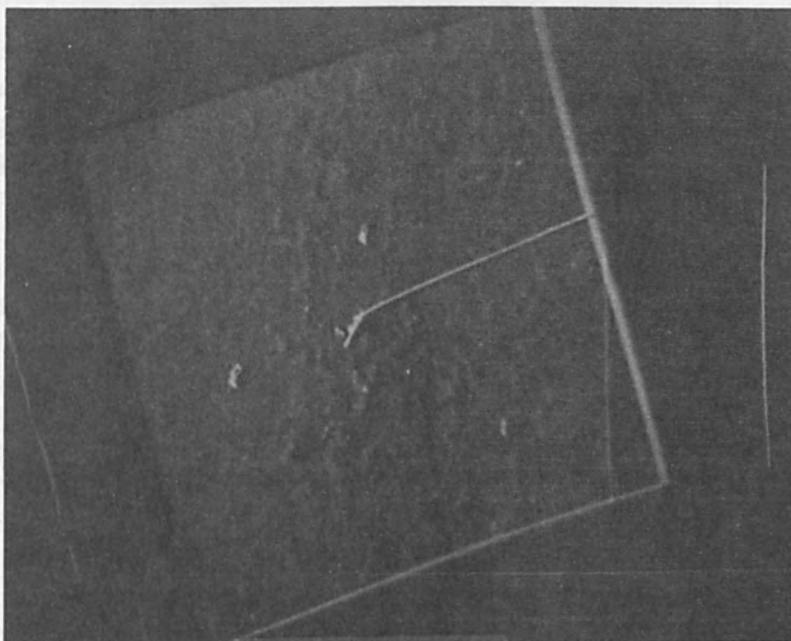


Figure 4. Micrograph of a "Uniform Oxide Using the Oxide Conduction Current Mode. (Mag. \approx 100x)

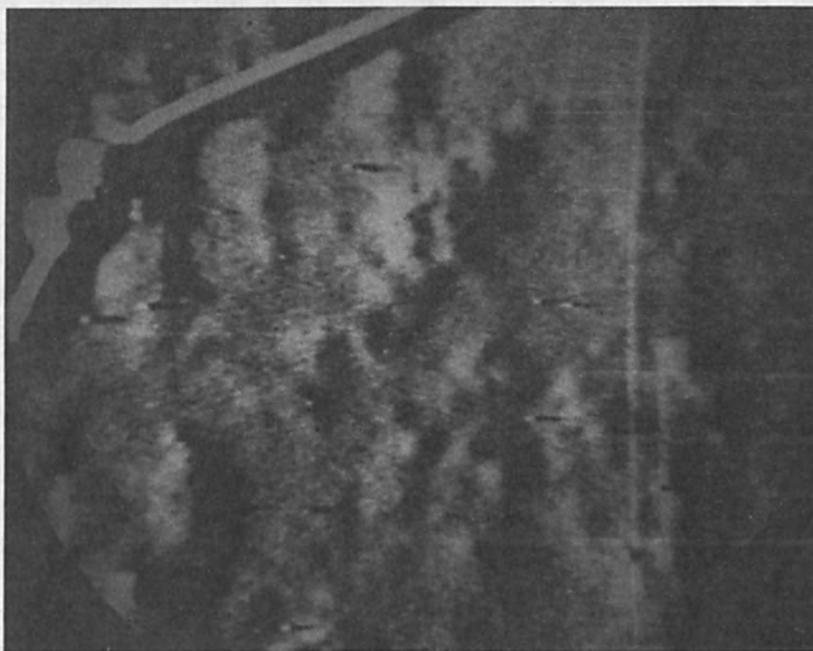


Figure 5. Micrograph of a Uniform Oxide at Higher Magnification. (Mag. \approx 500x.)

film when studied in an optical microscope and in the scanning microscope operating in the secondary electron mode. When the oxide conduction micrograph was taken details of inhomogeneities in the oxide became apparent. These could not be correlated to the slight "orange peel" of the chemically polished silicon surface and are interpreted as local changes in the oxide conductivity. The very sharp regions of contrast seen at about three points of Fig. 4 are dust particles on the aluminum and the lead shows up in silhouette.

When just one line of the raster is scanned across the sample a fixed conduction current waveform can be obtained on the oscilloscope (Fig. 6). In this, the region on the left represents no conduction until the pad is reached where the current sharply rises. When the lead is crossed there is a sharp dip in the current and at the right-hand end of the pad, the current falls again to zero. The overall "sag" of this waveform is due to the ac coupling of the amplifier.

Calibration of the gain of the system enables values of current through the oxide (I_o) to be obtained in terms of the primary beam current (I_p). This was done for this sample which had an oxide thickness of about 4000 Å. The primary beam energy was 15 keV and its current was 6.3×10^{-10} amps. Figure 7 shows this current ratio (R) expressed as a function of bias voltage (V_b). The values of R obtained in this way are in good agreement with previously published results of Pensak⁴.

When a limited region of the pad of a second sample was bombarded with the primary beam before the conduction micrograph was taken, this region showed a lower conductivity than the surrounding area. This is illustrated in Fig. 8 which represents the current obtained in a single scan across the biased pad. The center of the pad had been subjected to a total electron charge of 2.7×10^{-3} coulombs cm^2 and a sharp fall in current was observed in this region. The corresponding micrograph of the device is shown in Figure 9. Other areas of the pad which had been subjected to lower amounts of charge can also be seen. This micrograph was taken within a few minutes of bombarding the oxide, and it is not known to what extent this effect will decay with time.

In one sample which had been subjected to plastic bending at high temperature, an unusual structure was observed by this technique (Figs. 10 and 11). In this sample the pad had been placed so that it overlapped a diffused junction in the silicon substrate and the associated oxide step can be seen in the right of the micrograph. This oxide structure could not be seen optically and is interpreted as a local change in the oxide structure possibly caused by the earlier deformation.

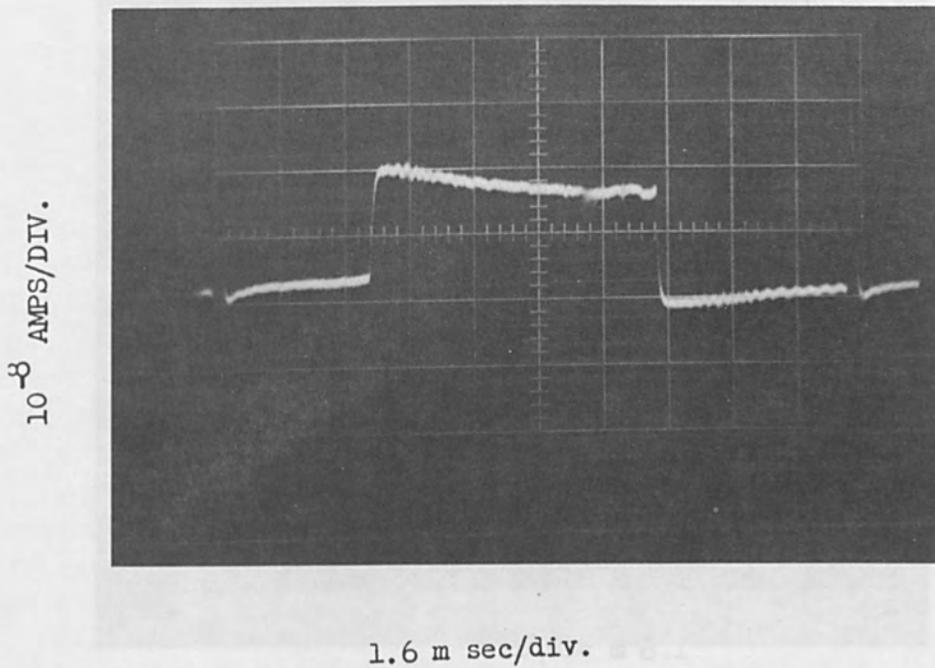


Figure 6. The Oxide Conduction Current Wave Form During a Single Line Scan.

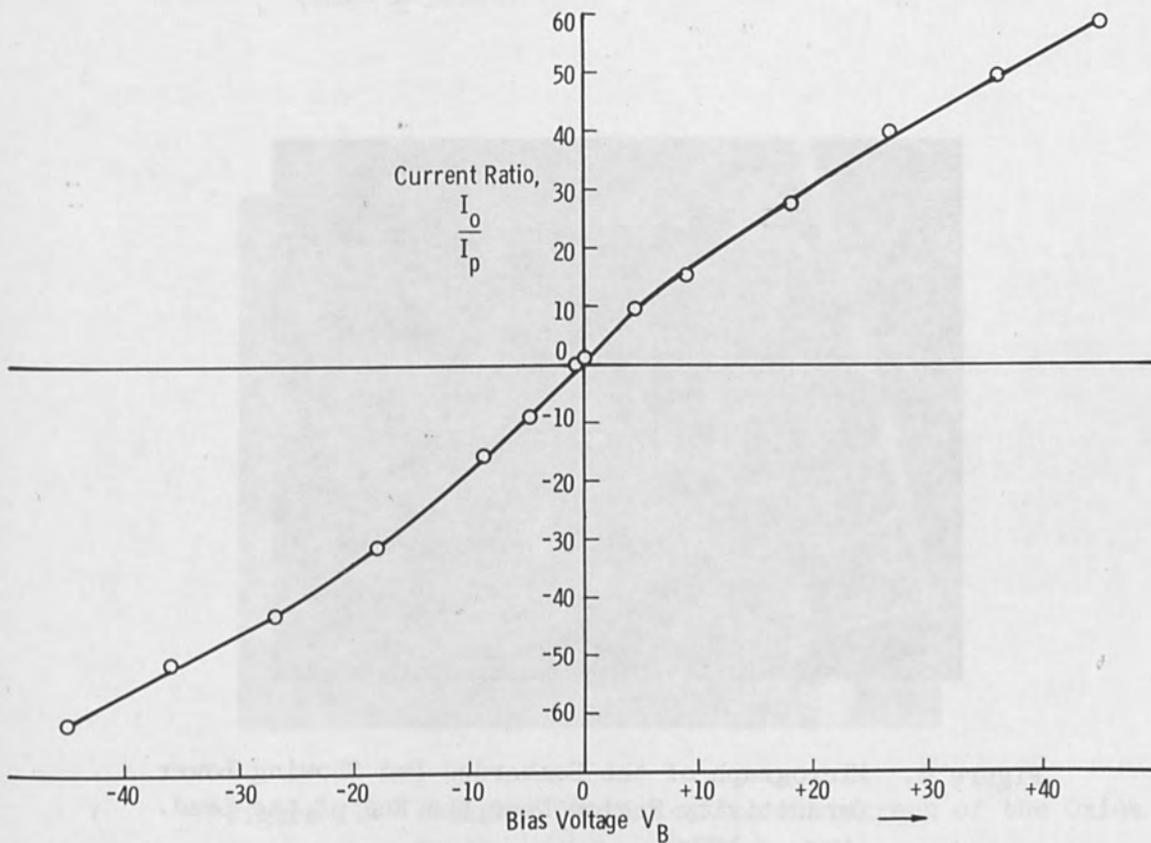
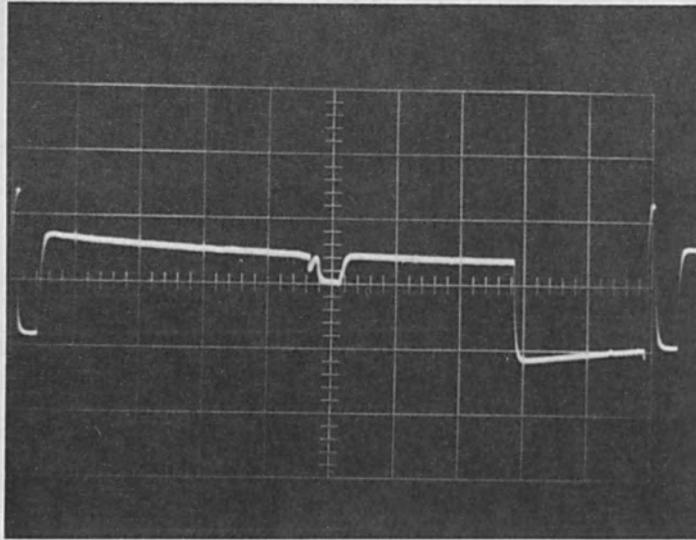


Figure 7. Ratio of Oxide Conduction Current of Primary Beam Current for Various Voltages. (Oxide Thickness 4,000 Å $I_p = 6.3 \times 10^{-10}$ Amps, Accelerating Voltage 15 KV).

10^{-8} AMPS/DIV.



1.6 m sec/div.

Figure 8. The Oxide Conduction Current for a Single Scan over a Bombarded Pad.

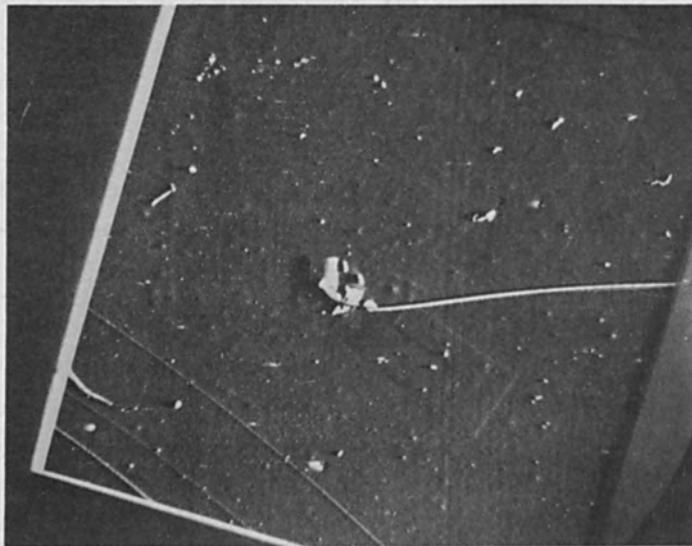


Figure 9. Micrograph of the Bombarded Pad Showing Lower Conductivity Region Near the End of the Lead. (Mag. $\approx 100x$).

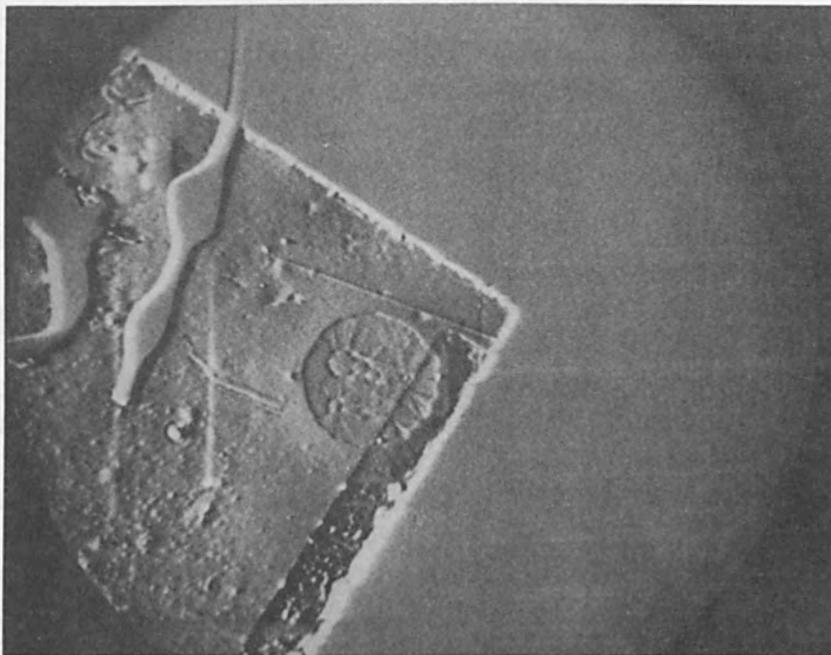


Figure 10. A Micrograph of an Oxide Fault in a Sample which had Previously been Subjected to Plastic Deformation. (Mag. \approx 300x.)



Figure 11. A Higher Magnification Micrograph of the Oxide Fault. (Mag. \approx 900x.)

The samples discussed above were all of thermal oxides grown on n-type silicon substrates. However similar effects have been observed with normal p-type substrates and also on various silicon web samples.

IV. DISCUSSION

This paper has outlined a method in which some information about the local electrical behavior of oxides can be obtained, with a resolution well below one micron. However great care must be exercised in interpreting micrographs obtained in this way. It is clear for example that variations in the thickness of the aluminum pad would cause conductivity variations due to the differing amounts of electron adsorption in the aluminum. On the other hand variations in the silicon substrate conductivity should have very little effect, as the ratio of the oxide resistance to that of the substrate is very high ($>10^4$).

It would appear that this technique could be used to assess the effects that various semiconductor device fabrication processes have on the oxides. It must be emphasized however that no attempt has yet been made to correlate inhomogeneities observed in the electron induced conductivity with electrical characteristics of the oxides under normal operating conditions.

REFERENCES

1. Everhart, T. E., O. C. Wells and R. K. Matta, Proc. IEE 52, 1642 (1964).
2. Czaja, W. and J. R. Patel, J. Appl. Physics 36 1476 (1965).
3. Fridrikhov, S. A. et al, Translated in: --Sov. Phys. - Solid State 6 1049 (1964).
4. Pensak, L., Phys. Rev. 75 472 (1949).

ELECTRON BEAM SYNTHESIS AND CRYSTALLIZATION OF InSb FILMS

by

N. M. Davis and H. H. Wieder
U. S. Naval Ordnance Laboratory
Corona, California

ABSTRACT

The Synthesis and growth of dendritic InSb films has been accomplished by electron beam micro-zone melting of vacuum deposited composite In and Sb films. These films are two phase in character, containing a nearly parallel array of InSb dendrites which are separated by narrow filaments of In. Their characteristic electrical properties are: an anisotropy of the galvanomagnetic coefficients relative to the long axis of the dendrites, an electron mobility of from 30,000 $\text{cm}^2/\text{v sec}$ to 40,000 $\text{cm}^2/\text{v-sec}$, and a large magnetoresistance coefficient. A detailed description will be given of the apparatus and techniques for micro-zone crystallization of InSb.

I. INTRODUCTION

Thin films of the intermetallic semiconductor indium antimonide have potentially important device applications which range from radiation transducers to galvanomagnetic devices such as Hall generators and magnetoresistors.

However, vacuum deposited films¹ of this compound have electrical properties which are in every respect inferior to those of high purity bulk crystalline InSb. Correlations made between the measured electron mobility and the size of the crystallites in such films suggest that scattering of electrons from grain boundaries is responsible for the observed low mobility. Films with few grain boundaries and large crystallites have a mobility of the order of 3×10^4 to 5×10^4 $\text{cm}^2/\text{v-sec}$ in comparison with 6.8×10^4 $\text{cm}^2/\text{v-sec}$ observed in the bulk material. Such films have been produced from vacuum deposited composite films of In and Sb reacted by melting by radiant heating and subsequent solidification of the molten film.^{2,3,4} Either stoichiometric InSb or two-phase (InSb + In) films may be grown by this method. The latter films exhibit a large change in their electrical resistance in a magnetic field and consequently may be employed in magnetoresistive devices. Both single phase and two-phase InSb films have a pronounced dendritic structure because the process of recrystallizing the films from the melt takes place over a period less than two minutes. Consequently, the thermal gradients and their time derivatives are large. This favors super-

cooling which results in rapid and relatively random nucleation and branching of the dendrites. An ordered dendritic structure should bring about an improvement in the magnetoresistance coefficient of two-phase films. Potentially, better control of the kinetics of the crystallization process should allow the growth of still larger crystallites and ultimately, the growth of single crystal thin films without the requirement of epitaxy.

The use of an electron beam for inducing the proper chemical reaction between the atomic constituents as well as for controlling the temperature and temperature gradients within the films, offers some unique advantages.^{5,6,7,8} It was to be expected that such a process would lend itself to micro-zone-refining of the film and, at the very least, produce ordered growth of the InSb dendrites.

The requirements imposed by the recrystallization process upon the electron beam system are relatively modest. An electron beam incident upon the surface of a vacuum deposited composite film of In and Sb is to produce a temperature of the order of 525°C, the melting temperature of InSb. The spot size is to be small enough so that a narrow molten zone may be generated and scanned across the film. Nevertheless, it must not be so intense that large amounts of the InSb disassociate allowing Sb to evaporate thus changing the composition of the film.

The temperature and rate of motion of the electron beam heated zone in the film must be adjusted so as to prevent either constitutional supercooling or dissociation of InSb already present in the film. Finally, the scan rate must be slow enough to promote growth of nucleated crystallites without inducing any additional nucleation.

Our preliminary studies, which comprise the subject of this paper, were concerned with the two-phase system InSb + In with the free In comprising up to 10% of the film weight.

II. THE ELECTRON OPTICAL SYSTEM

The electron optical system used for these investigations is illustrated in Fig. 1. It consists of a telefocus electron gun and an electrostatic deflection system mounted vertically within a demountable vacuum chamber with multiple access ports.

The electron gun is of telefocus design, requiring no additional electron lenses to produce a spot 200 μ m in diameter at a distance of approximately 15 cm from the cathode.⁹ The latter is a directly heated tungsten "bolt" cathode. It consists of a 0.025 cm diameter tungsten wire spotwelded to the end of a 0.05 cm diameter tungsten hairpin heater. The electron emitting surface is the end of the wire which is ground

and polished flat using standard procedures. The cathode is operated in the temperature limited condition; it is heated by 60 cps current flowing through the hairpin.

The Wehnelt cylinder, which effects the focusing action of the gun, is negatively biased with respect to the cathode. The anode, deflection system, and the specimen bombarded by the beam are at ground potential.

The electron beam is deflected by an electrostatic deflection system consisting of two orthogonal sets of plates such as used in a conventional cathode-ray tube electron gun.

The acceleration potential is provided by a 0 to 30 kV, 5 ma, dc power supply regulated against line voltage fluctuations by a saturable core reactor. A foot-switch operated, vacuum relay is used to apply the high voltage to the electron gun and also provides a safety interlock. The horizontal deflection potential is supplied by a 60 cps step-up transformer controlled by a variac; its output is adjustable from 0 to 900 V ac (peak). Vertical deflection of the beam is accomplished by a 1000 V regulated dc power supply whose output is programmed by a motor-driven, reversible, ten turn potentiometer to produce a triangular waveform with a period of 300 sec.

A 2 inch diameter silicone-oil diffusion pumped vacuum system is used to bring the chamber to its operating pressure of less than 10^{-6} torr.

III. FILM PREPARATION

Composite films of In and Sb in a (weight) ratio of 1.27 were vacuum deposited on microscope glass cover slips in an ion-pumped vacuum system at a pressure $< 2 \times 10^{-7}$ torr. The thickness of the composite films were chosen between 0.75 and $2.5 \mu\text{m}$. Their Debye-Sherrer x-ray spectra indicated the presence of InSb and free In in microgranular form. To a great extent they consist of inhomogenous layers of Sb, InSb, and In with nearly pure Sb in the vicinity of the substrate and pure In as the top layer. This is because as the In and Sb are heated in the same crucible, the Sb evaporates first because its vapor pressure is $\sim 10^3$ times that of In. In order to improve the adhesion of the evaporated films, it was found desirable to deposit a thin In layer initially upon the substrate.

After removing the specimen from the deposition chamber, the top In layer of the film was oxidized by heating it for 1 minute to 300°C in a chamber maintained at a pressure of 10^{-1} torr of air. This

produces an In_2O_3 surface layer 50 Å to 200 Å thick which appears as a light tan coloration of the film surface.

The oxidized film is then placed film side toward the electron gun, at the focal plane of the electron-beam system, Fig. 1. Since the film is insulated from ground by its glass substrate, a small, grounded metal spring contacts the film surface and prevents charging.

After the pressure in the chamber is reduced below 10^{-6} torr, the electron beam is deflected off the sample and its current allowed to stabilize. Typical beam currents used for the recrystallization of the film are 60 to 70 μA . The acceleration potential was maintained at 15 KV and the steady-state spot size at the specimen was determined to be about 150 to 200 μm in diameter.

The horizontal deflection plates were excited with a 60 cps sinusoidal waveform of sufficient amplitude to generate a line 0.5 cm in length. For our purposes, a sinusoidal waveform is superior to a triangular or a sawtooth waveform because of the additional energy input needed near the ends of the zone to compensate for the heat losses due to thermal conduction from the molten zone to the unheated portions of the film. The horizontal line generated by the beam was then slowly swept onto and across the sample by the slowly varying dc potential applied to the vertical deflection plates. Its rate is 50 $\mu\text{m}/\text{sec}$. With this speed and with the indicated beam currents, the molten zone produced by the beam in the film was found to be approximately 0.10 cm wide. The solid-liquid interface at the trailing edge of the molten zone is very nearly a straight line.

Direct, visual observation is available and the crystallization process can be observed without any optical magnification. InSb dendrites nucleate at the trailing edge of molten zone and grow in the direction of the thermal gradient along the direction of zone motion. In this manner, an area approximately 0.5 cm x 0.5 cm may be recrystallized with essentially parallel dendrites of up to 0.3 cm length and 0.02 cm width.

If rapid vertical scanning of a higher current beam is employed, the entire area scanned is molten at the same time. Upon cooling, the dendrites grow from the cooler edges toward the center of the heated region in a random pattern of short branched dendrites instead of the well-ordered structure observed with the slow scan described above. Similar short-branch dendrites result if the temperature distribution in the molten zone is such that its trailing edge has a concave curvature causing the expanding dendrites to impinge upon each other. The appearance of such films is quite similar to those recrystallized by radiant heat.



Figure 1a. (a) Vacuum Chamber Containing Electron Beam Gun and Specimen Holder for Electron Beam Recrystallization of InSb Films. (b) Electron Beam Assembly.

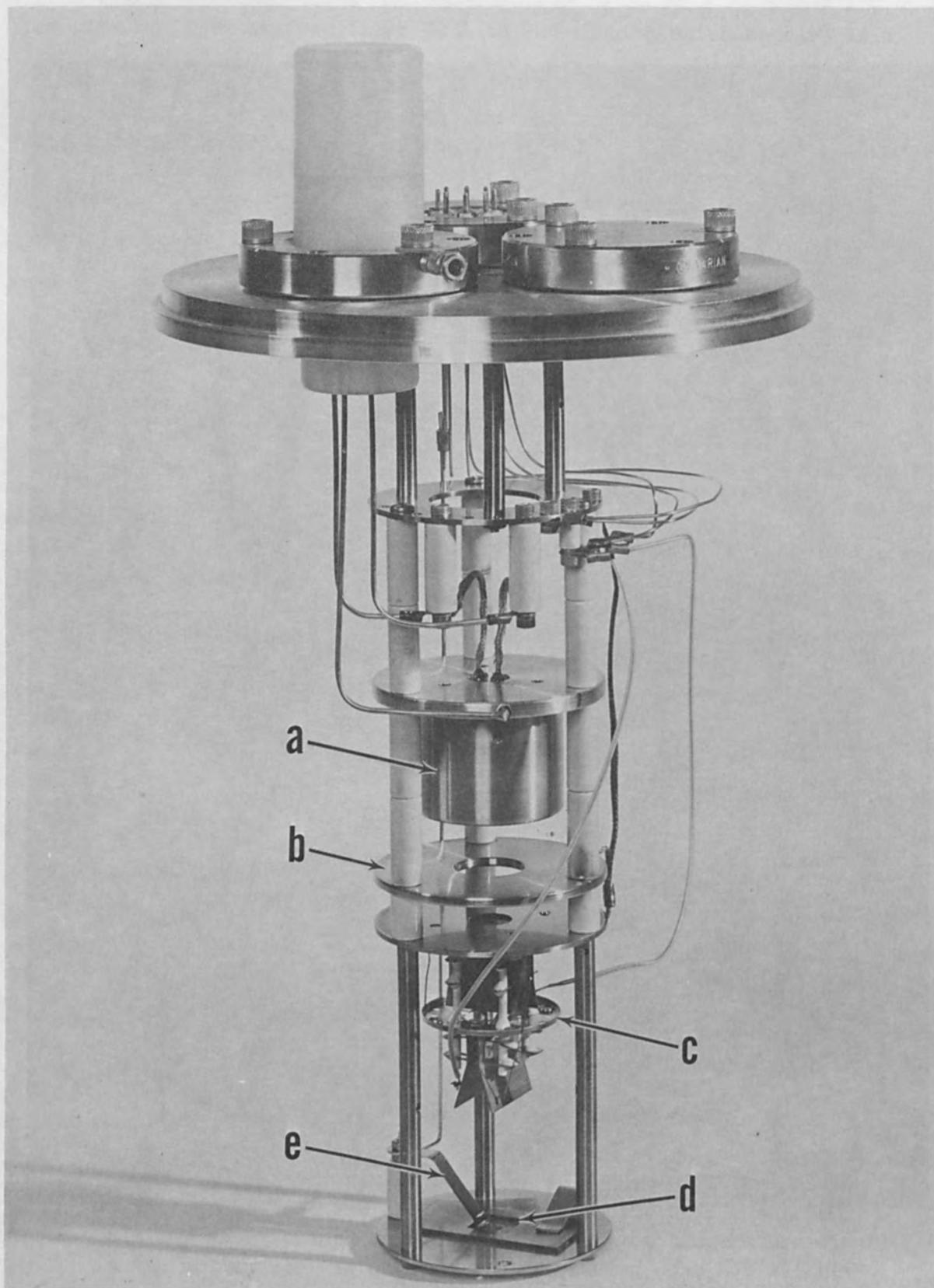


Figure 1b. Cathode Located Within Wehnelt Cylinder (a); Anode (b); Electrostatic Deflection System (c); Film Specimen (d); Grounding Contact (e).

The In_2O_3 film is of great importance to the success of the recrystallization process. It provides sufficient constraining forces to prevent agglomeration of the molten InSb film. An unoxidized InSb film, when heated in an oxygen-free atmosphere to its melting point, will form small hemispherical globules which grow at the expense of the remainder of the film producing isolated mounds. The oxide layer allows large areas of film to melt and remain continuous under the oxide "lid". In addition, the oxide inhibits the evaporation of Sb from the molten InSb film and thus aids in maintaining the desired film constitution.

Figure 2a illustrates the appearance of a film which had been previously recrystallized by radiant heat and thereafter subjected to a second recrystallization process by means of the electron beam. The region scanned by the electron beam stands out clearly. The dendrites within it are much larger than those of the standard recrystallization process and they exhibit a well-defined anisotropy compared to the thermally recrystallized region in which the dendrites are arranged in an essentially random pattern.

Figure 2b shows the obverse of the same film photographed through the glass substrate. The boundaries of the dendrites may be seen more clearly. They are delineated by the filaments of metallic In which are segregated out and trapped between the dendrites during their growth. Their role in determining the electrical properties of such films will be described in the next section.

On a macroscopic scale, the films are polycrystalline. Their Debye-Sherrer x-ray spectra indicate a preferential InSb (111) orientation of the dendrites parallel to the film and a weak In (101) peak. Figure 3 shows an electron diffraction pattern obtained by transmission through a dendrite. The (111) orientation is clearly apparent. The dendrites are single crystals, or twinned single crystals of stoichiometric InSb.

Thus far, we have been unable to grow single crystal films of InSb by manipulating the scanning process. The problems of controlling thermal gradients in the film and the tendency toward constitutional supercooling have prevented us from crystallizing entities larger than a few mm^2 . In addition, there is evidence that the width of a crystal grown as a dendrite is self-limiting and thus large area crystals could be expected only if growth were non-dendritic. Nevertheless, the electrical properties of the inhomogeneous and anisotropic films shown in Fig. 2 are of intrinsic, as well as practical, interest.

IV. GALVANOMAGNETIC PROPERTIES

The transverse magnetoresistance coefficient, $(\Delta R/R_0)$, of a

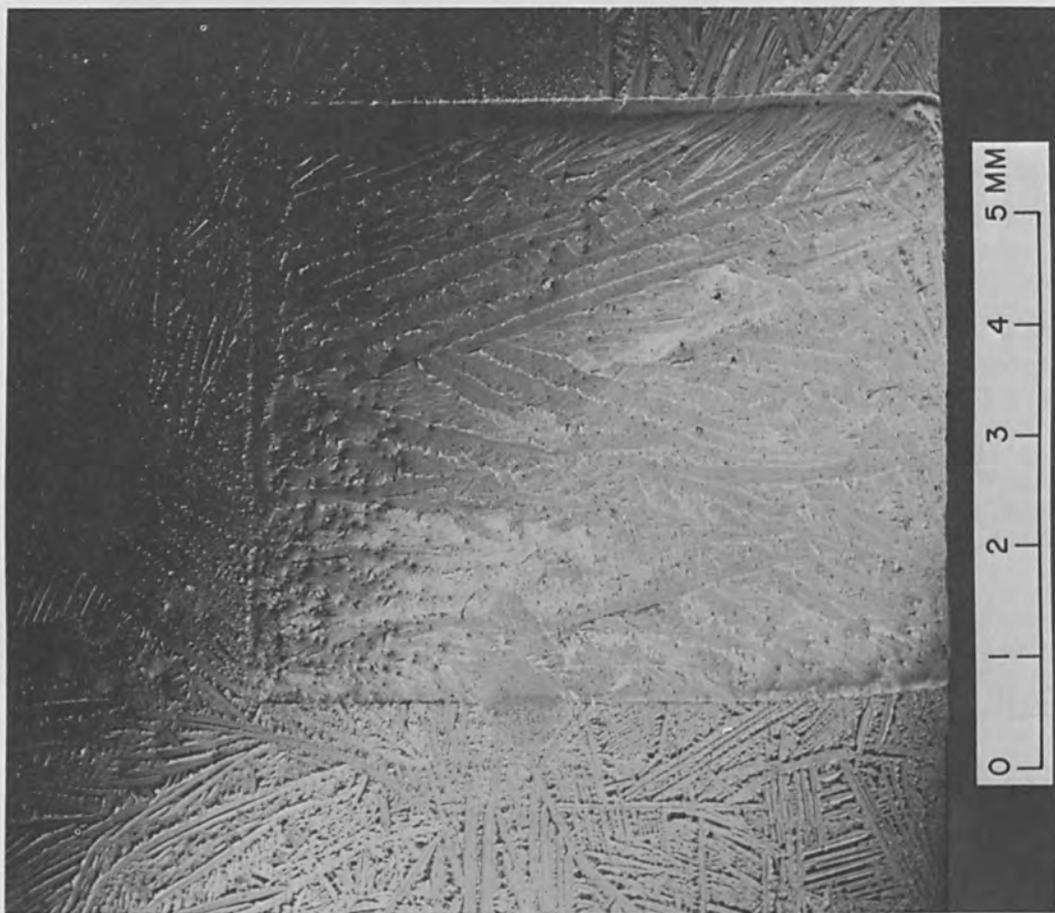


Figure 2a. (a) Recrystallized InSb Film. Randomly Oriented Dendrites Produced During Initial Synthesis and Crystallization by Radiant Heat. Central Region was Subsequently Recrystallized by Sweeping an Electron Beam Generated Molten Line from Bottom to Top.

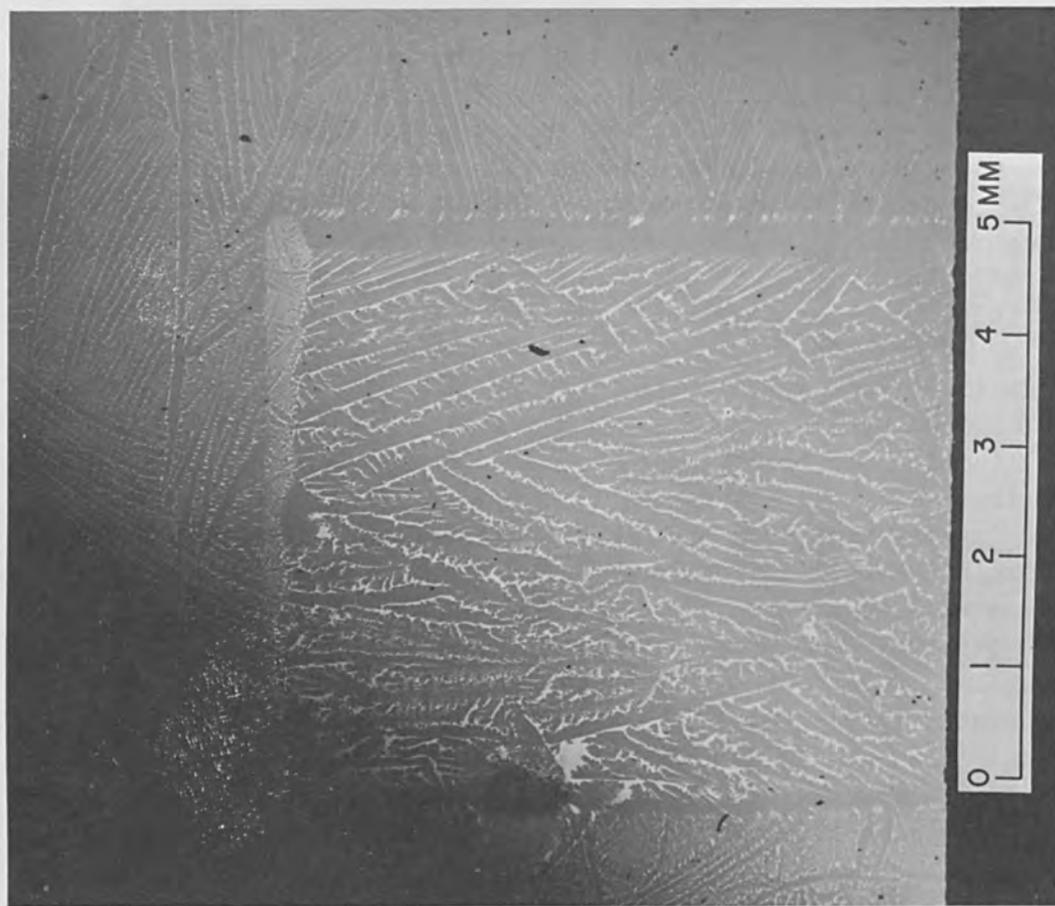


Figure 2b. Observe View of Same Film as Figure 2a, Taken Through Glass Substrate. Improved Ordering and Size of Dendrites are Apparent As Delineated by Metallic In Filaments.



Figure 3. Selected-area Transmission Electron Diffraction Pattern of Recrystallized InSb Film Showing $\{111\}$ Orientation in Plane of Film. Film Thickness is 0.14 Micron.

representative film is shown, as a function of the magnetic field H , in Fig. 4. The magnetoresistance is evidently larger when the current through the film is perpendicular rather than parallel to the dendrites.

Qualitatively, the curves are the same. Both have a quadratic dependence of $(\Delta R/R_0)$ on H for $H < 1$ kOe as shown in Fig. 5 and both have an essentially linear dependence on H in higher fields.

The anisotropy of $(\Delta R/R_0)$ vs. H is due to the difference in the electrostatic short-circuit of the metallic In filaments acting on the dendrites in which a Hall field is generated by an applied transverse magnetic field.

Only a fraction of the film volume contains the parallel oriented dendrites, the rest consists of dendrites with a random orientation. The magnetoresistance coefficient of the film may be expressed in large magnetic fields¹⁰ as:

$$\frac{\Delta R}{R_0} = \left(\frac{\ell}{w}\right) [\mu_n H + f \left(\frac{\ell}{w}\right)] - 1 \quad (1)$$

where μ_n is the isotropic electron mobility, H is the applied magnetic field, ΔR is the change in resistance, R_0 is the resistance in zero magnetic field, and $f(\frac{\ell}{w})$ is a geometrical factor which expresses the effective electrostatic shorting of the metallic filaments at the dendrite boundaries in terms of the width w and length ℓ of the dendrites. This equation applies for the current density vector perpendicular to the long axis ℓ of the dendrite. If the current density vector is parallel to ℓ then w and ℓ must be interchanged. The derivative of eq. (1) with respect to H is:

$$\frac{d(\Delta R/R_0)}{dH} = \left(\frac{\ell}{w}\right) \mu_n \quad (2)$$

If it is assumed that the dendrites are parallel to each other in the entire film volume, then on applying the current perpendicular or parallel to the dendrite axes we can express the respective $(\Delta R/R_0)$ vs. H slopes as:

$$S_{\perp} = \frac{d(\Delta R/R_0)_{\perp}}{dH} = \left(\frac{\ell}{w}\right) \mu_n$$

$$S_{\parallel} = \frac{d(\Delta R/R_0)_{\parallel}}{dH} = \left(\frac{w}{\ell}\right) \mu_n \quad (3)$$

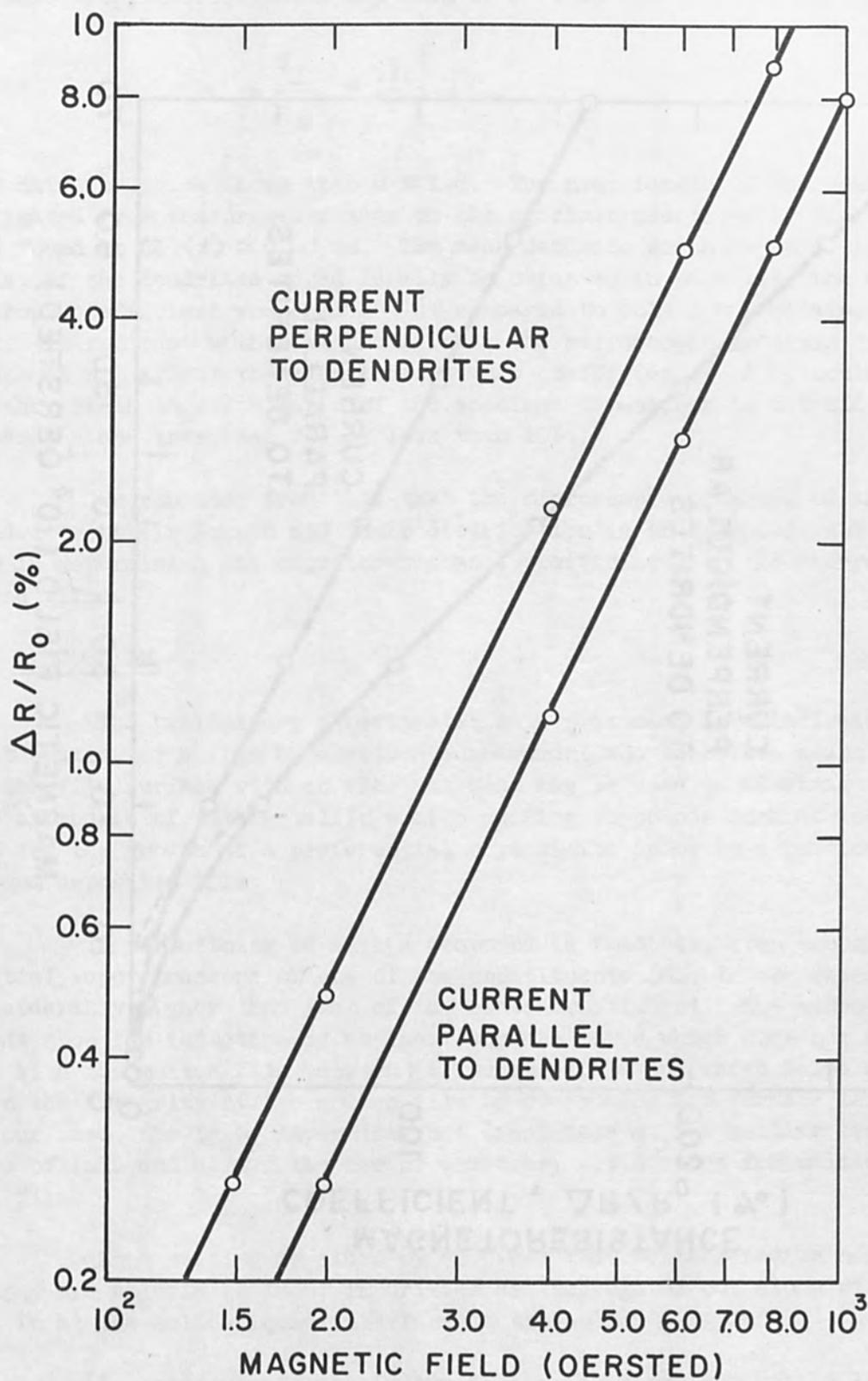


Figure 4. Magnetoresistance Coefficient as a Function of Magnetic Field for an InSb Film. Dendrites were Grown by Electron Beam Synthesis. The Two Curves were Obtained with the Current Density Vector Oriented Respectively Parallel and Perpendicular to the Long Axis of the Dendrites.

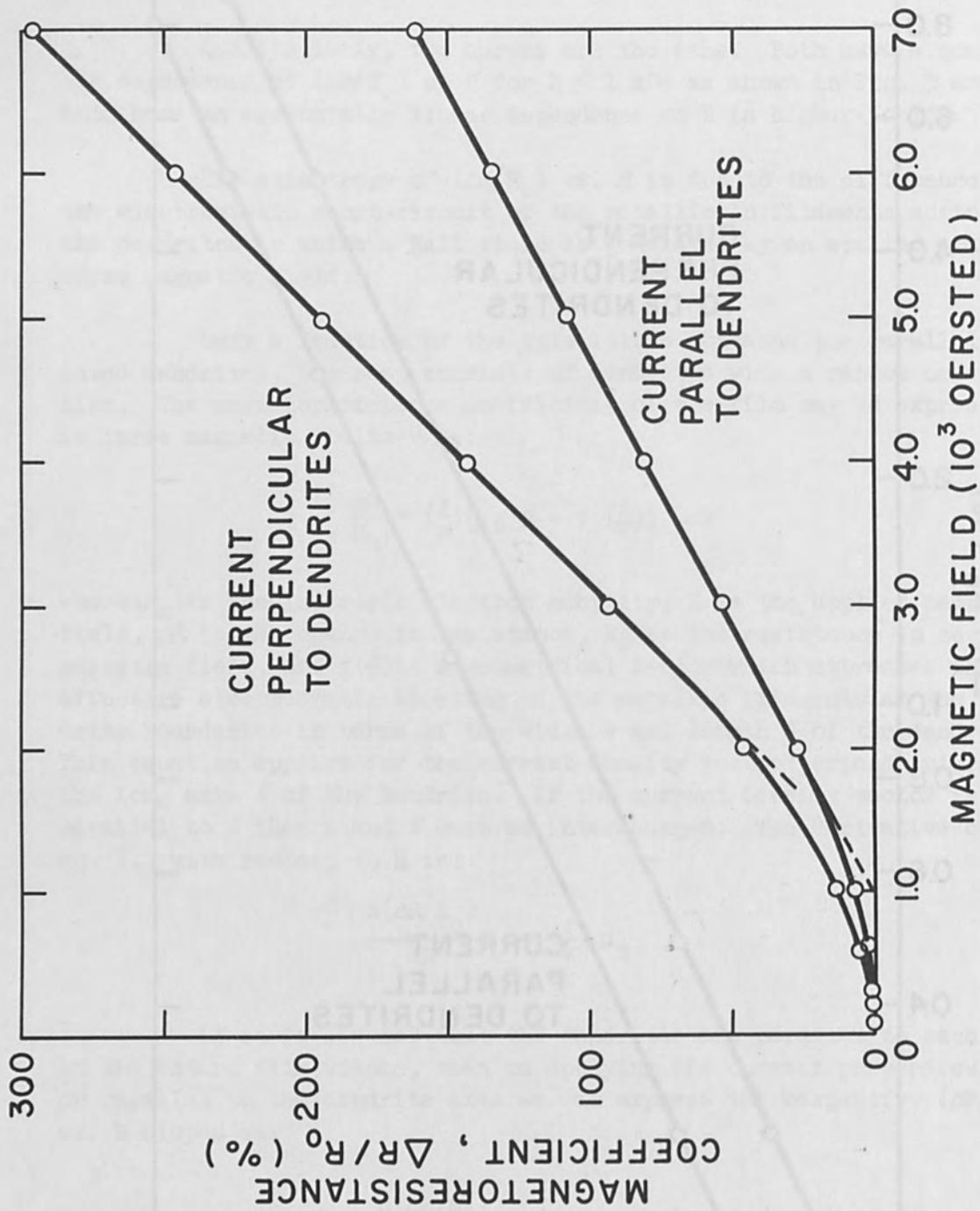


Figure 5. Magnetoresistance Coefficient for Both Parallel and Perpendicular Orientations is Shown to Have a Quadratic Dependence on the Magnetic Field for $H \leq 1$ kOe.

An anisotropy coefficient α may then be defined as:

$$\alpha = \frac{S_{\perp}}{S_{\parallel}} = \left(\frac{l}{w}\right)^2 \quad (4)$$

The data in Fig. 4 shows that $\alpha \approx 1.8$. The mean length of the dendrites estimated from measurements made on the specimen described by Fig. 4 was found to be $\langle l \rangle \approx 0.08$ cm. The mean dendrite width $\langle w \rangle \approx 0.01$ cm; thus, if the dendrites could ideally be oriented in parallel, the anisotropy coefficient would be $\alpha \approx 64$ compared to only $\alpha \approx 2$ obtained thus far. In contrast with homogenous InSb, the macroscopic specimen dimensions do not affect the magnetoresistance coefficient.³ A reduction of the length to width ratio of the specimen dimensions to 0.5 did not alter S_{\perp} and increased S_{\parallel} by less than 10%.

We conclude from this that the microscopic ordering of the dendrites, their length and their distribution is the preponderant factor in determining the magnetoresistance coefficients of the recrystallized films.

V. DISCUSSION

The preliminary experimental data presented here indicates that heating of a film by electron bombardment and selective scanning of the film surface with an electron beam may be used to advantage for the synthesis of intermetallic semiconducting compounds such as InSb and for the growth of a preferential crystalline order in a previously vacuum deposited film.

Zone refining of such a compound is feasible, even though the partial vapor pressure of one of the constituents (Sb, in our case) is considerably higher than that of the other constituent. The method depends upon the induction of an inert surface layer which does not interact with the molten film between it and substrate and which helps maintain the integrity of the molten film by overcoming its surface tension. In our case, the In_2O_3 layer does not dissociate at the melting temperature of InSb and allows the use of arbitrary width zones for melting the film.

Zone melting is aided by the fact that neither indium nor antimony are soluble in InSb; impurities are segregated out along with the In at the solid-liquid interface of the solidifying InSb.

In consequence, the indium antimonide dendrites have a low impurity concentration, of the order of 10^{16} cm^{-3} , and the electron mobility is not degraded by the impurity concentration. Using the Hall

effect, the electron mobility in isolated dendrites from recrystallized InSb films has been determined to be as high as 5×10^4 cm²/v-sec.

Since the electron mobility is the most significant material figure of merit for a material which is to be used for magnetoresistor applications, electron-beam zone refining offers a decided advantage for the fabrication of high resistance films with a high magnetoresistance coefficient required for galvanomagnetic device applications.

REFERENCES

1. "Compound Semiconductors", edited by R. K. Willardson and H. L. Goering, Vol. 1, Section VI, Rheinhold Publishing Co., New York 1962.
2. H. H. Wieder and A. R. Clawson, Solid-State Electron. 8, 467 (1965).
3. H. H. Wieder and N. M. Davis, Solid-State Electron. 8, 605 (1965).
4. H. H. Wieder, Solid-State Commun. 3, 159 (1965).
5. G. B. Gilbert, T. O. Poehler and C. F. Miller, J. Appl. Phys. 32, 1597 (1961).
6. O. A. Weinreich and G. Dermit, J. Appl. Phys. 34, 225 (1963).
7. S. Namba, Crystallization of Vacuum Evaporated Germanium Films by the Electron Beam Zone-Melting Process, Proceedings of the Electron and Laser Beam Symposium, March 31-April 2, 1965, p. 305, Pennsylvania State University.
8. J. Maserjian, Solid-State Electron. 6, 477 (1963).
9. E. B. Bas and F. Gaydon, Z. Anegew, Phys. XI, 370 (1959).
10. H. J. Lippmann and F. Kuhrt, Z. Naturforsch. 13, 462 (1958).

Al₂O₃ SINGLE CRYSTAL GROWTH
BY ELECTRON BEAM FLOAT-ZONE MELTING

by

Walter Class and Harvey R. Nesor
Materials Research Corporation
Orangeburg, New York

ABSTRACT

The electron beam float-zone process has been successfully applied to the processing of the dielectric compound Al₂O₃ to yield high quality sapphire crystals. The technique makes use of an electron accelerating grid to produce the molten zone, and an electron beam after-heater to reduce thermal gradients in the vicinity of the growing crystal interface to suppress thermal shock-induced crystal flaws.

I. INTRODUCTION

The application of electron beam float-zone refinement to metals to yield purification as well as single crystal growth is a well-established technique. Its application to inorganic, non-metallic compounds, however, has been sparingly investigated, primarily due to the insulating nature of these compounds and their instability in the vacuum environment required to sustain an electron beam. The present paper deals with the methods, results and advantages of the electron beam float-zone technique for the preparation of α -Al₂O₃ single crystals.

The basic electron beam process makes use of an electron gun system which traverses a vertically held rod at controlled speeds. The portion of the rod within the electron gun is heated by electron bombardment until a molten zone is formed. As the gun moves, the material is converted to a single crystal by means of unidirectional solidification from the melt, similar to crystal growth by the Czochralski or Bridgman methods. The molten zone is supported by its own surface tension which eliminates the need for a crucible required by the other techniques.

The advantages of the electron beam float-zone technique of crystal growth can be summarized as follows:

a) It does not require use of a crucible, which eliminates what is frequently a major source of contamination.

b) Crystal growth and purification are carried out in a single operation.

c) The small melt volume and excellent transfer of energy to the workpiece inherent to the process combine to yield a highly efficient melting process. Typically, less than 1 KW of beam power will permit the formation of a molten zone which is suitable for crystal growth.

d) The well-defined, unidirectional thermal gradient under which crystal growth occurs facilitates seeding and eliminates the problem of habit growth which can lead to platelike or needlelike crystals.

The non-metal α - Al_2O_3 , commonly known as sapphire or corundum, has been successfully processed by a modified electron beam technique (1,2). However, the problems involved in the growth of metal crystals are magnified in the case of sapphire growth. Since sapphire is a non-conductor, it cannot be zone melted using a work-accelerated gun as is used for metals. Its surface tension (690 ergs/cm²) is low compared to metals, so that poor mechanical support of the floating zone results. Its poor thermal conductivity and thermal shock resistance require more careful control of thermal gradients and fluctuations to yield good crystal growth. Consequently, the technique of electron beam float-zone growth as applied to sapphire crystals must be carefully controlled with respect to crystal growth parameters. Modifications of the standard process used for the growth of metal crystals are therefore necessary in order to fully realize the advantages of the process.

II. APPARATUS AND EXPERIMENTAL TECHNIQUES

A modified MRC Model EBZ-6000 electron beam zone refining apparatus was used for the present work. The standard model, shown schematically in Figure 1, employs a work-accelerated electron gun to traverse a vertically held specimen rod at scanning rates between 1"/hr. and 9"/hr. This type of apparatus has been described in detail elsewhere (3). In brief, it consists of an electron gun assembly mounted to a lead screw driven carriage with the entire assembly contained in an evacuated bell jar. A 5 KV high voltage source and a filament power source supply power to the work-accelerated electron gun. Electrons are emitted by a heated tungsten filament and accelerated toward the workpiece maintained at a high positive potential. The filament is enclosed in a molybdenum shield which serves to electrostatically focus the electron beam, thereby limiting the molten zone length.

Since Al_2O_3 is a non-conductor, it cannot be heated utilizing this type of electron gun. Details of the electron gun assembly employed for Al_2O_3 are shown in Figure 2. The work-accelerated gun was con-

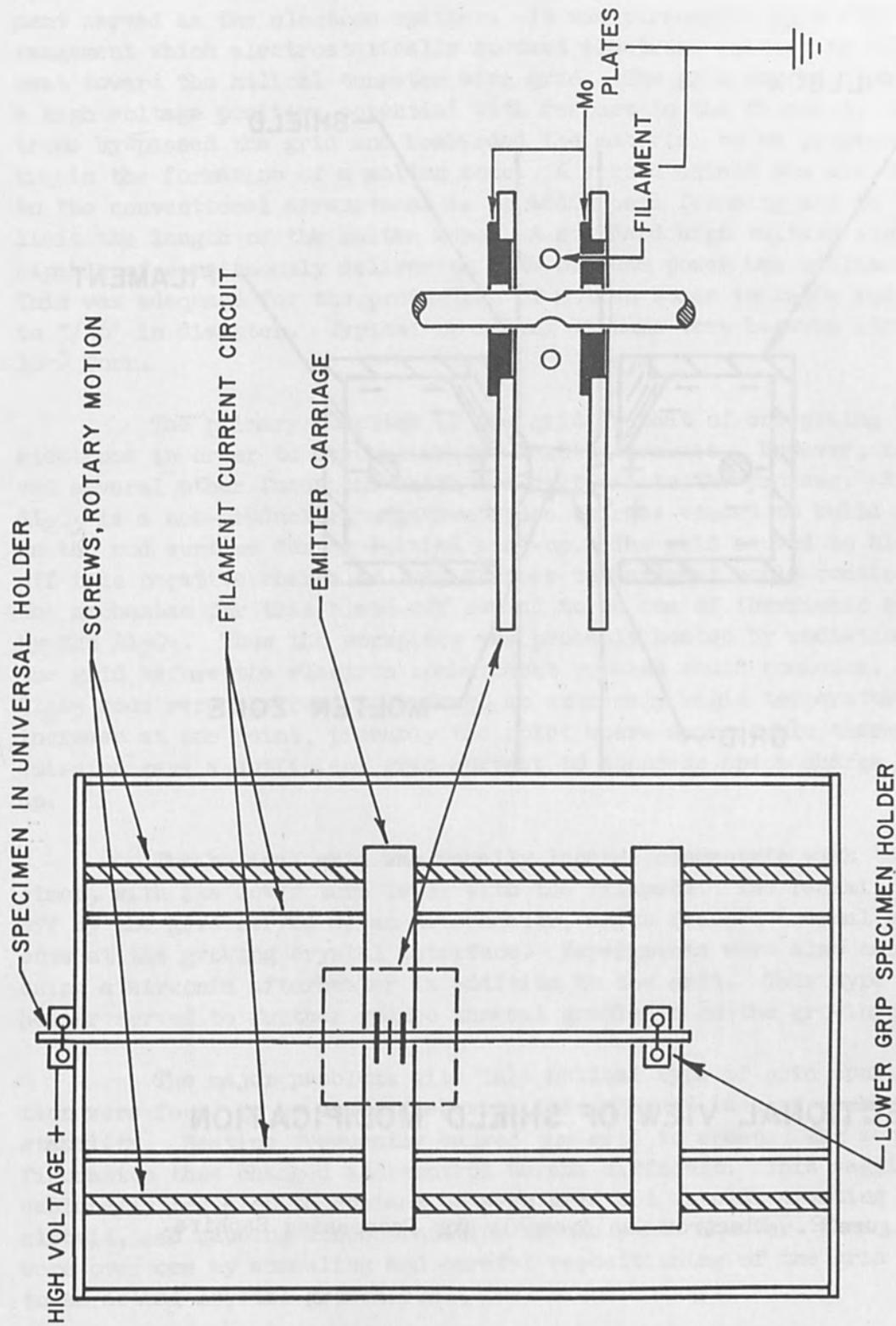
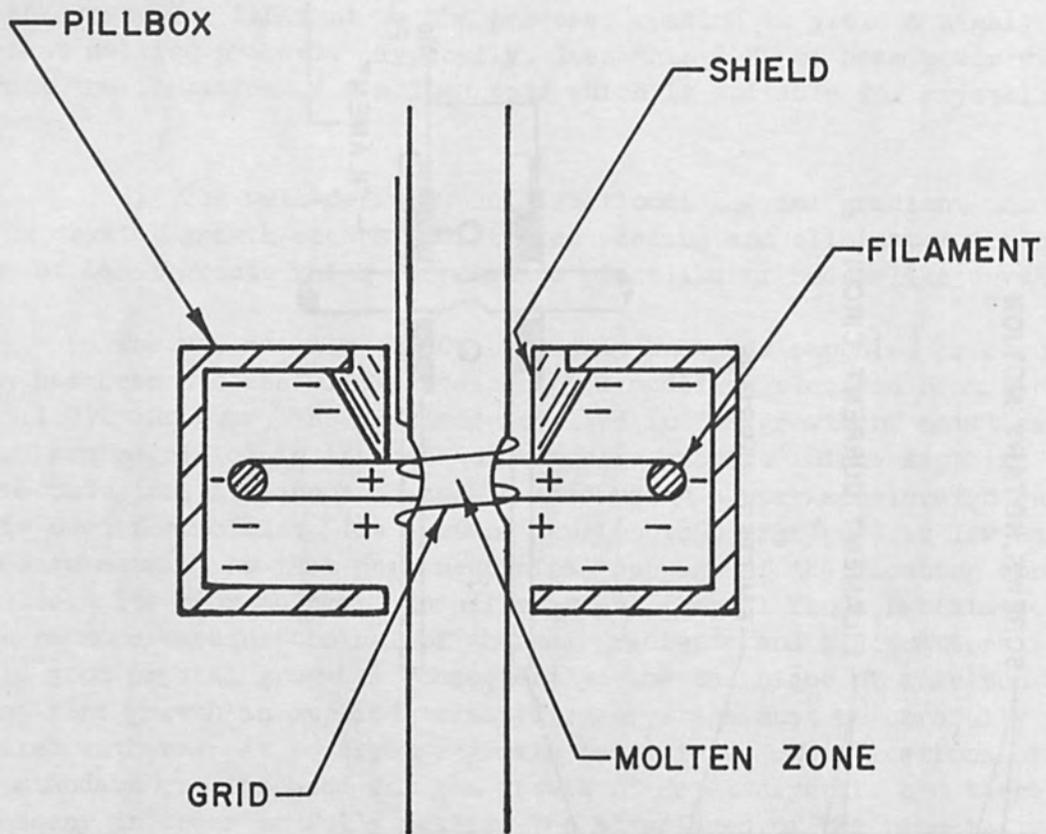


Figure 1. Standard Electron Beam Zone Melting Apparatus.



SECTIONAL VIEW OF SHIELD MODIFICATION

Figure 2. Electron Gun Assembly for Processing Sapphire.

verted to a self-accelerated type utilizing the grid technique described by Neumann and Huggins (1). A resistance heated annular tungsten filament served as the electron emitter. It was surrounded by a shield arrangement which electrostatically focused electrons emitted by the filament toward the helical tungsten wire grid. The grid was maintained at a high voltage positive potential with respect to the filament. Electrons by-passed the grid and bombarded the material to be processed, resulting in the formation of a molten zone. A second shield was also added to the conventional arrangement as an additional focusing aid to further limit the length of the molten zone. A standard high voltage supply, capable of continuously delivering 2 KW of beam power was utilized. This was adequate for the production of molten zones in Al_2O_3 rods up to 5/16" in diameter. Typical operating vacuums were between 10^{-4} and 10^{-5} Torr.

The primary function of the grid is that of energizing the electrons in order to accomplish the melting process. However, it served several other functions which are critical to the process. Since Al_2O_3 is a non-conductor, negative space charges tended to build up on the rod surface during initial heat-up. The grid served to bleed-off this negative charge so that further bombardment could continue. The mechanism for this bleed-off seemed to be one of thermionic emission by the Al_2O_3 . Thus the workpiece was probably heated by radiation from the grid before the electron bombardment process could commence. The Al_2O_3 rods were observed to undergo an extremely rapid temperature increase at one point, probably the point where appreciable thermionic emission gave a sufficient grid current to suppress space charge build-up.

The helical grid was usually located concentric with the specimen, with its lower turn level with the filament. The radiation given off by the grid served as an afterheater, which reduced thermal gradients at the growing crystal interface. Experiments were also conducted using a zirconia afterheater in addition to the grid. This type of afterheater served to further reduce thermal gradients on the growing crystal.

The major problems with this helical type of grid configuration were found to be associated with its inherent lack of mechanical stability. Heating frequently caused the grid to creep. The zone configuration thus changed and control became difficult. This sagging occasionally led also to contact between grid and shield, creating a short circuit, and causing discontinuation of the run. However, these problems were overcome by annealing and careful repositioning of the grid prior to an actual crystal growth run.

Figure 3 shows the actual scanner assembly used for sapphire crystal growth. Scan speeds for the electron gun assembly were reduced

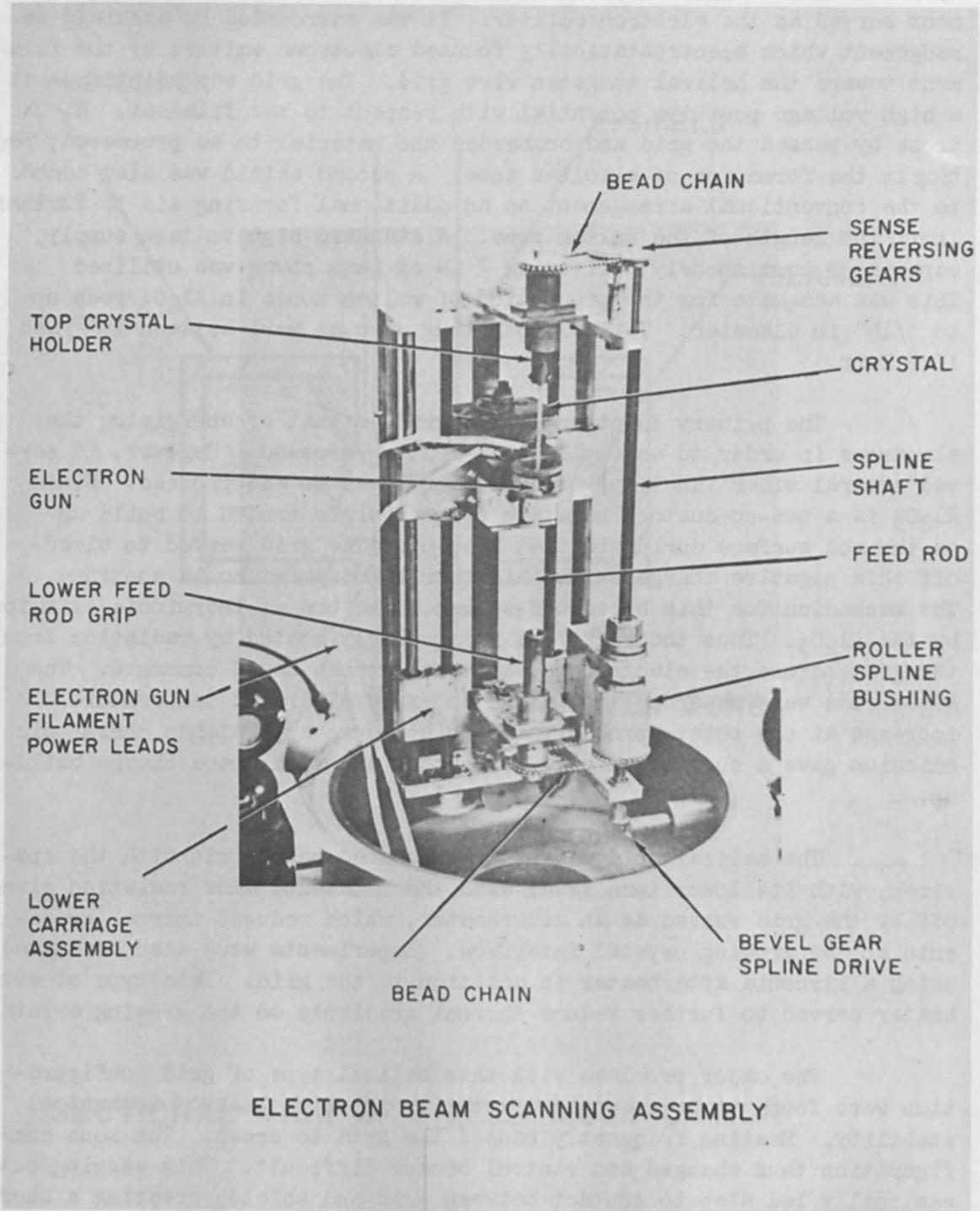


Figure 3. Modified Electron Beam Zone Melting Apparatus.

to a range of $1\frac{1}{2}$ " / hr. to 4" / hr. A seed crystal and feed rod were mounted in holders and rotated by means of a bead chain-spline drive system. The spline assembly provided opposite sense rotations of top and bottom rods and unrestricted vertical motion of the feed rod without interrupting rotation, a situation which was desirable for control of the growing crystal diameter.

A typical crystal growth run was begun by slowly bringing the seed crystal up to melting temperature. Once a molten pendant was created on the end of the seed crystal, the lower carriage was raised until the feed rod welded to the seed crystal. Power was then increased until the zone became fully molten, and rotation was then begun. Downward scanning at a rate of 2" / hr. was used in most cases since it was observed that upward scanning, normally used for metals, led to instability of the molten zone. When crystal growth was completed, the molten zone was ruptured by pulling the lower rod downward. This was done to prevent fracture of the crystal by the propagation of cracks originating from the polycrystalline feed rod.

Crystal orientations were determined by optical and x-ray techniques. Perfection was studied optically by examination for Tyndall scattering centers, and microscopically by observing dislocation etch pit densities. The etch pit procedure used has been described by Alford and Stephens (4), and Janowski and Conrad (5). This procedure consisted of mechanically and chemically polishing a crystal face which had been cut along the (0001) basal plane. The etch pits were then revealed on this face by etching in $K_2S_2O_7$ at $675^\circ C$.

III. RESULTS OF CRYSTAL GROWTH

Sapphire crystals having diameters of $1/8$ ", $3/16$ " and $1/4$ " were grown utilizing the previously described modified electron beam technique. The largest crystals grown to date, free of Tyndall scattering centers, have been approximately 5" in length by $3/16$ " in diameter. A typical array of such crystals is shown in Figure 4. The distortion of the background reference lines is an optical effect. These as-grown crystals are all free of macroscopic Tyndall scattering centers which have been a problem associated with the development of this technique. Crystals which exhibited such scattering had the appearance depicted in Figure 5. As is evident from this figure, the scattering centers are not uniformly distributed, but rather are collected into a sequence of bands. Microscopic examination has revealed that these bands are a dense collection of bubbles which are presumably entrapped during a rapid, non-steady state advance of the growing crystal interface. Consequently, scanning rates below 2" / hr. were employed. By comparison, metals are typically scanned at speeds up to 10" / hr.

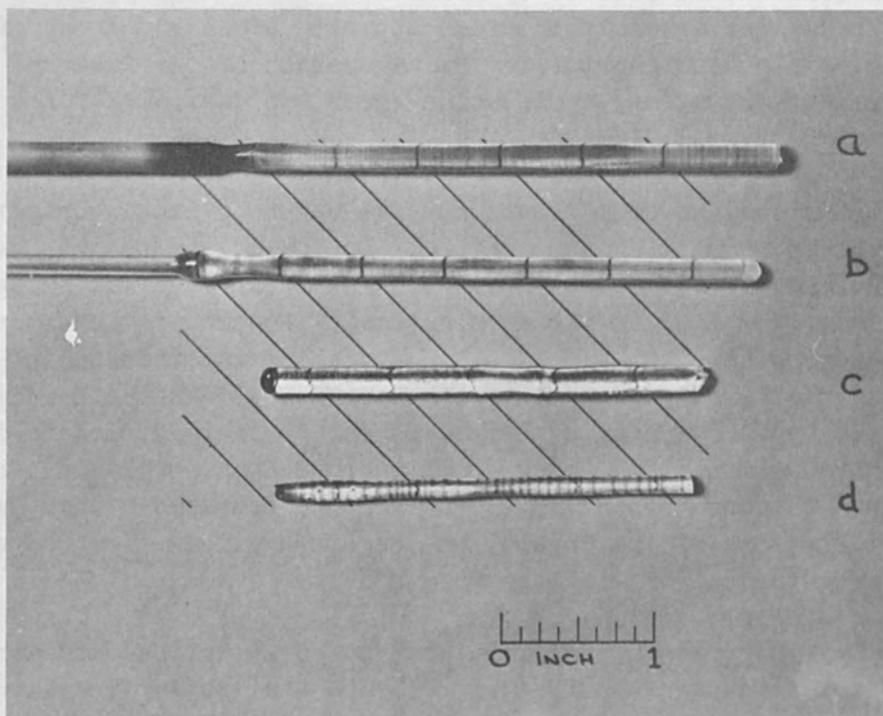


Figure 4. Sapphire Crystals Produced by the Electron Beam Technique.

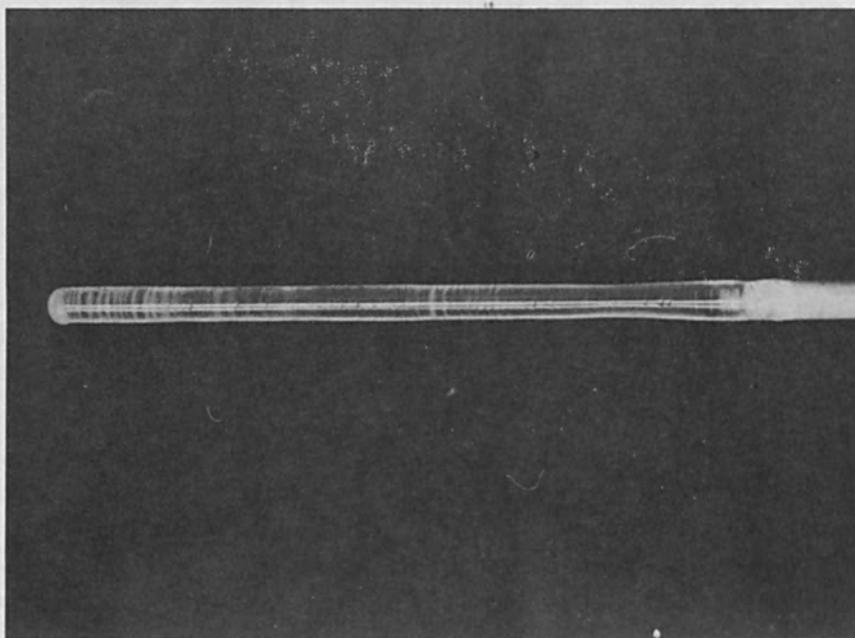


Figure 5. Sapphire Crystal Containing Tyndall Scattering Centers.

An etch pit examination of the crystals grown by this technique has revealed a surprisingly low concentration of dislocations in those regions which are free of the Tyndall scattering bands. This is revealed in Figures 6 and 7 which were taken from a crystal banded near the point of seeding but which became free of defects as the growth progressed. Figure 6 reveals the dislocation etch pit density of $7.5 \times 10^4/\text{cm}^2$ typical of a banded region. Figure 7, on the other hand, reveals a density estimated at approximately $10^2/\text{cm}^2$ which was observed in the crystal portion that was free of the scattering bands. It is further noted that the good crystal regions are also free of substructure. All measurements made to date used as-grown crystals which were not subjected to any annealing treatments subsequent to the crystal growing operation. These results compare favorably with flame-fusion crystals which exhibit substructure, require a post-anneal and typically exhibit dislocation densities of approximately $10^6/\text{cm}^2$ (4,5).

Optical and x-ray examinations of the crystals grown by the electron beam technique have revealed a preferred growth orientation in which the crystallographic c-axis makes an angle of $\sim 90^\circ$ with the crystal rod axis. Such an orientation almost invariably develops when the seed crystal is replaced by a sintered polycrystalline rod in the experimental setup. Crystals "c" and "d" of Figure 4 are the result of such a random seeding operation and are observed to produce additional distortion of the background reference lines. This additional distortion is due to the presence of lateral basal facets which develop in 90° crystals.

IV. DISCUSSION

Since most metal crystals are grown by scanning in an upward direction, this was first attempted with sapphire. However, it became apparent that the stability of the molten zone was influenced by the shape of the growing crystal. Figure 8a shows a stable zone formed when scanning upward, where the crystal diameter, D_c , is equal to the diameter of the molten zone, D_m , just above it. The onset of instability is shown in Figure 8b. The molten material flows downward due to gravity, creating a bulge at the growing crystal interface. The crystal then tends to grow with an increasing diameter as it assumes the contour of the molten zone adjacent to it. As this condition progresses, the molten zone eventually becomes entirely pinched off (Figure 8c).

The conditions for a stable molten zone could more easily be achieved by scanning downward. The configuration here is shown in Figure 9a. This results in a crystal diameter equal to that of the molten zone, and also equal to that of the feed rod, D_f . Crystal diameter could further be controlled by lowering the feed rod (Figure 9b) to decrease diameter, and raising the feed rod (Figure 9c) to increase diameter.

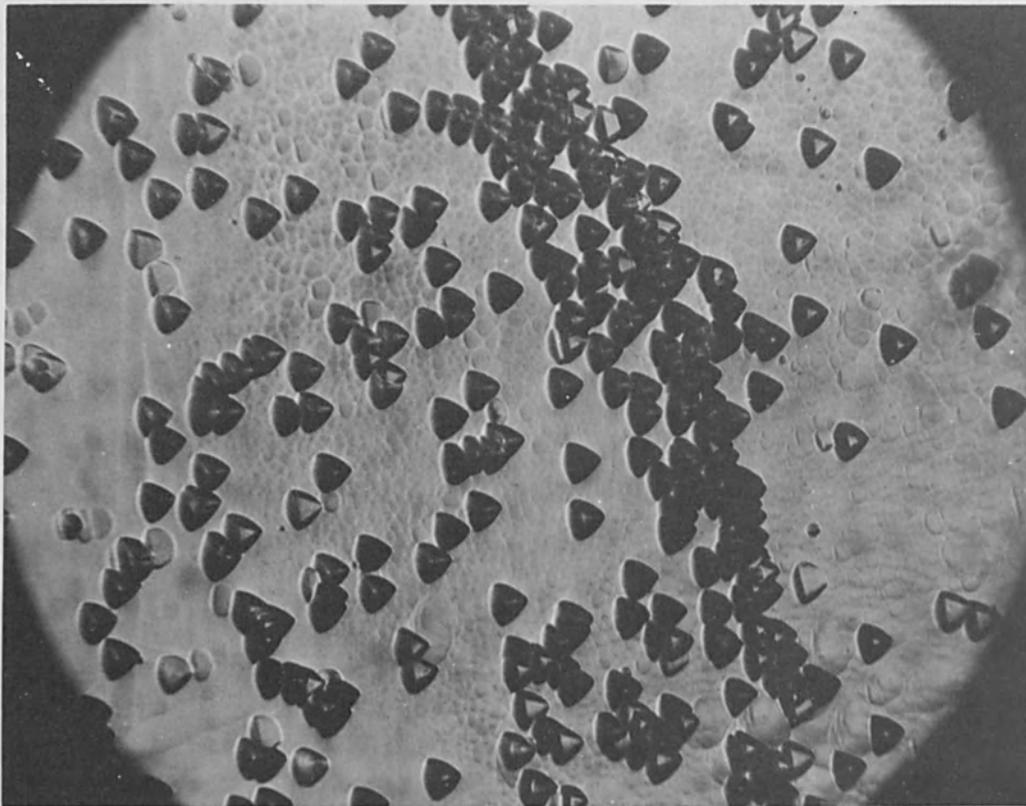


Figure 6. Dislocation Etch Pits in a Region Containing a Tyndall Scattering Center.

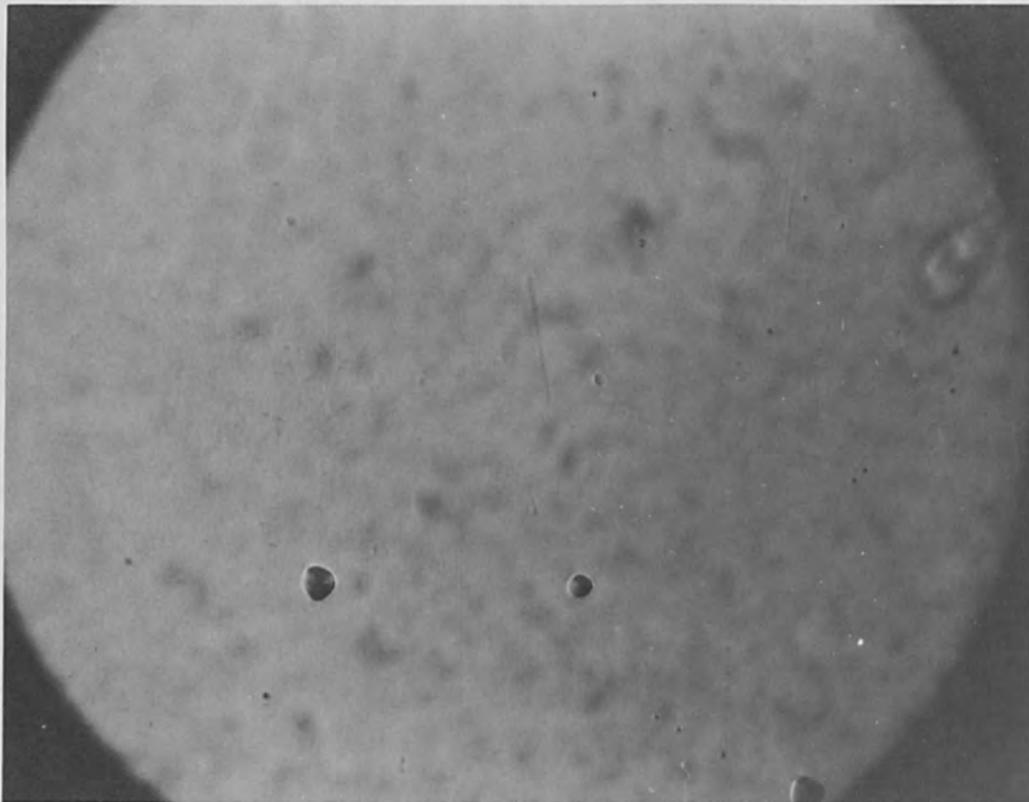
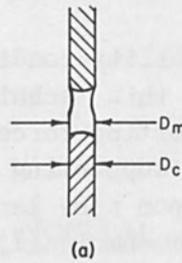


Figure 7. Dislocation Etch Pits in a Region Free of Tyndall Scattering Centers.

DIRECTION OF SCAN = ↑

STABLE



UNSTABLE

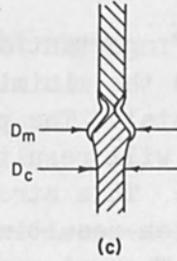
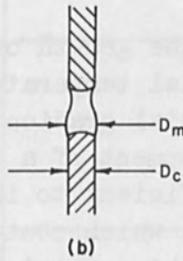


Figure 8. Zone Geometry and Instability - Upward Scanning Direction.

DIRECTION OF SCAN = ↓

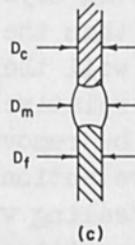
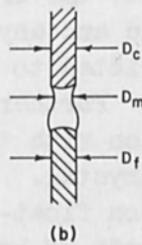
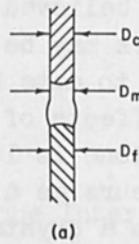


Figure 9. Zone Geometries - Downward Scanning Direction.

No instabilities were observed in these cases. In fact, since a larger, more thermally massive zone was produced by downward scanning, an additional degree of stability was introduced to the crystal growth process.

One additional instability condition limits the maximum diameter that can be processed by this technique. This results from the fact that the molten zone supporting force, being surface tension, will eventually become incapable of supporting the mass of the molten zone. At one point then, depending upon zone length, diameter, surface tension and density, the zone becomes mechanically unstable. Assuming a zone height equal to its diameter, a critical diameter equal to approximately 1" was calculated from the stability condition derived by Heywang (6). On this basis, fairly large crystals may be grown before this type of instability sets in.

One very important condition for the growth of low dislocation content crystals is the minimization of radial temperature gradients in the growing crystal. The presence of radial gradients combined with thermal expansion will result in the development of a state of stress within the crystal. This stress can be sufficient to induce dislocation motion and generation resulting in a crystal which contains a large dislocation density. Thermal gradients along the crystal length will not have this effect because here the crystal is free to undergo thermal expansion without constraint. It is evident therefore that the sharp gradient along the crystal length which occurs in the electron beam process is not deleterious to good crystal growth. However, radial gradients are also present which would be expected to give rise to dislocation multiplication. It is believed that the low dislocation densities observed in the present work may be attributed to three factors. Firstly, the diameters produced to date have been relatively small, thereby minimizing the deleterious effects of any radial gradients. Secondly, it is observed that the molten zone is longer at the surface than at the center. This phenomenon occurs as a natural consequence of the electron bombardment. It results in a crystal growth interface which is convex with respect to the growing crystal, as is shown schematically in Figure 10a. This feature implies that the isotherms in the crystal also possess a similar curvature. Consequently, if one takes a diametric section through the crystal at any point, the crystal center will be colder and harder than the crystal surface and any dislocation motion near the molten zone will therefore be restricted to the crystal surface, leaving the interior relatively undisturbed. Furthermore, dislocations near the surface can be removed by interaction with the surface thereby minimizing their retention in the grown crystal. The results of Clausen and Rutter (7), dealing with the induction float-zone refinement of NiO, confirm this supposition. Here, the depth of heating obtained with induction, coupled with the poor thermal conductivity of NiO, combined to give a concave growth interface, as in Figure 10b, resulting in

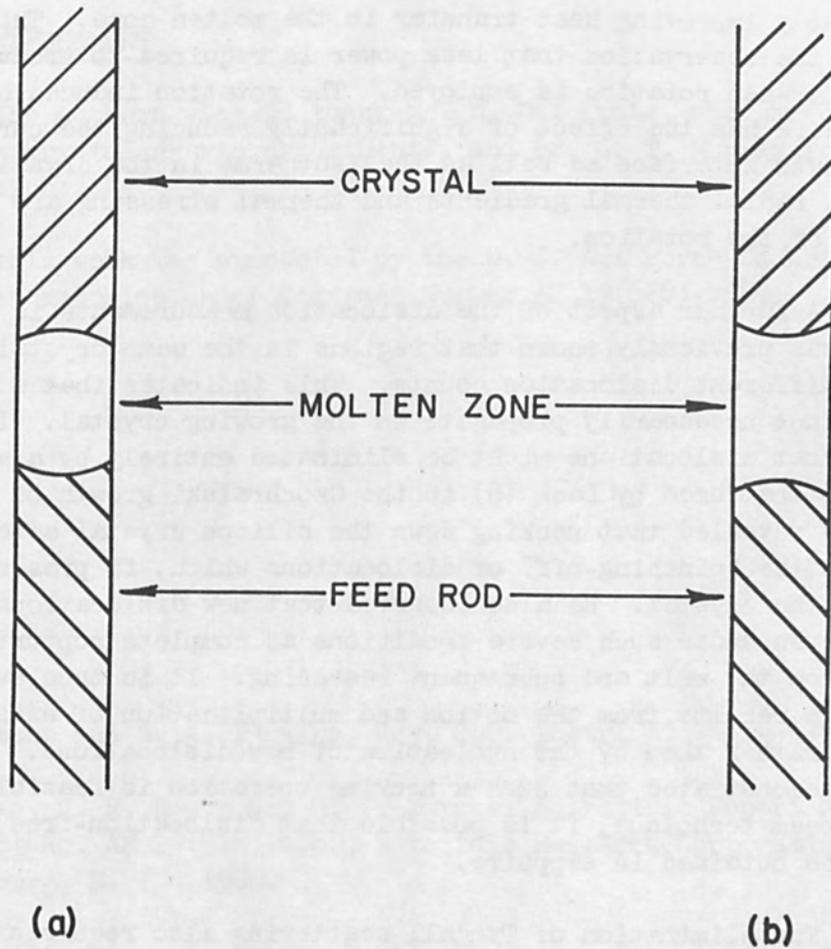


Figure 10. Zone Interface Geometries.

rather poor crystal perfection.

A third reason for the low dislocation density observed in electron beam sapphire is associated with the rotation employed during crystal growth. This rotation has the effect of stirring and thereby significantly improving heat transfer in the molten zone. This is evident from the observation that less power is required to produce a fully molten zone when rotation is employed. The rotation induced heat transfer therefore has the effect of significantly reducing the curvature of the growth interface as well as the isotherms in the crystal. Consequently, radial thermal gradients and thermal stressing are reduced by virtue of the rotation.

A further aspect of the dislocation measurements is of interest. It was previously shown that regions in the same crystal can have markedly different dislocation counts. This indicates that a dislocation will not necessarily propagate in the growing crystal. It is then possible that dislocations might be eliminated entirely by a method similar to that used by Dash (8) in the Czochralski growth of silicon. That study revealed that necking down the silicon crystal after seeding led to the "pinching-off" of dislocations which, if present, would grow with the crystal. He also observed that new dislocations would not form even under such severe conditions as complete rupturing of the crystal from the melt and subsequent reseedling. It is thus evident that most damage results from the motion and multiplication of existing dislocations rather than by the nucleation of new dislocations. Since it has been demonstrated that such a necking operation is feasible with the electron beam technique, it is possible that dislocation-free growth may also be obtained in sapphire.

The elimination of Tyndall scattering also requires careful control of the uniformity of both scanning and beam power to the molten zone. Non-uniformities in either of these quantities can result in severe perturbations in crystal growth rate because of the small thermal inertia of the molten zone. Such perturbations will result in crystal entrapment of the volatile constituents in the molten zone thereby generating a band of Tyndall scattering centers.

V. CONCLUSION

It has been demonstrated that the modified electron beam float-zone technique of crystal growth can produce Al_2O_3 (sapphire) single crystals. The principal modifications required are:

- a) The use of an electron accelerating grid, and
- b) Rotation of the molten zone during crystal growth.

Sapphires over 5" long have been grown in diameters up to 3/16", and diameters up to 1/4" have been processed. These crystals are free of Tyndall scattering centers and possess a high degree of internal perfection.

VI. ACKNOWLEDGEMENT

The authors wish to thank Mr. N. Del Frate, who performed most of the crystal growth experiments, and Dr. G. T. Murray for helpful discussions.

This work was sponsored by the U. S. Air Force Cambridge Research Laboratories under Contract Number AF 19(628)-4089.

REFERENCES

1. L. Neumann and R. A. Huggins, Rev. Sci. Instr. 33, 433-434 (1962).
2. W. Class, H. R. Nesor and G. T. Murray, Scientific Report No. 3, Contract No. AF 19(628)-4089, Materials Research Corporation, Orangeburg, N. Y., 1966.
3. S. Hurwitt and M. A. Adams, Metals Progress, Feb. 1963.
4. W. J. Alford and D. L. Stephens, J. Am. Cer. Soc. 46, 193 (1963).
5. K. Janowski and H. Conrad, Report No. TDO-169(3240-01)TN-2, Contract No. AF 04(695)-169, Aerospace Corp., El Segundo, California, July 1963.
6. W. Heywang, Z. Naturforsch, 11, 238 (1956).
7. F. M. Clausen and T. W. Rutter, G. E. Research Laboratory Report No. 64-RL-3710M, "Preparation Techniques for the Growth of Single Crystals of Non-Metallic Materials," August 1964.
8. W. C. Dash, "Growth and Perfection of Crystals," J. Wiley and Sons, 1958, p. 361.

DIFFRACTION-LIMITED SPOT SCANNING

by

Samuel Bousky
Ampex Corporation
Redwood City, California

ABSTRACT

Data record and readout systems utilizing diffraction-limited spots of laser generated light are dependent upon precise spot and line scan methods. Ideal linear scan is accomplished by four geometric configurations within three basic scanning techniques. Practical system capabilities for focused spot scanning by mechanical rotation at bandwidths to 100 MHz, line rates of 10 kHz, and line lengths to 20,000 spot diameters are indicated.

I. INTRODUCTION

Discovery of the principle of optical amplification through stimulated emission and its practical application in lasers provides, for the first time, a source of coherent light of high power, narrow beam, and monochromaticity, and offers the possibility for building optical scanning and recording systems with greatly improved capabilities. The art has evolved to a point where evaluation of these possibilities in terms of spot size, dynamic range, bandwidth, equipment size, and utility indicates that high-performance laser scanning is both feasible and practical.

Generation of laser light is a new and revolutionary concept. Because of this, it is natural to consider methods of laser scanning or beam deflection techniques which also invoke new and revolutionary concepts. Many such concepts have appeared in the literature and several have already been demonstrated in the laboratory. Notable among these are electrically induced changes of refractive index in electro-optical materials, sonically induced refractive index changes in gases or solids, switching between the ordinary and extraordinary ray in birefringent materials, interaction of light with sonic waves or light waves, interference control, etc. Such methods generally are considered in terms of small angular deflections up to 100 spot diameters. Deflections of the order of 10,000 spot diameters do not seem feasible within the foreseeable future by such techniques. Moreover, such methods are primarily applicable to deflection of light beams rather than to focused light spots.

In high resolution data recording or readout systems it is desirable to employ spots of light as small as a few microns in size. Such tiny spots can only be formed by a focusing optical system and even then the size is limited by the diffraction of light itself. This means that a spot-forming lens of diffraction-limited (or near diffraction-limited) quality is required for formation of the scanning spot. Any system for data recording or readout must include this aspect.

Consideration of techniques applicable to the deflection of diffraction-limited spots of 2 to 10 microns diameter, over line lengths of 5,000 to 20,000 spot diameters, at line rates of 1,000 to 10,000 lines per second, and at bandwidths of 10 to 100 megahertz, leads to the conclusion that only rotational mechanical scanning methods are fully applicable at the present state of the art. The performance of such mechanical scanning systems is greatly enhanced through the use of air lubricated rotor bearings which permit low-friction, ultra-high speed operation, at very high axial stability.

II. BEAM SCANNING GEOMETRY

Before considering actual scanning methods, it may be well to first examine the geometry of mechanical beam deflection. For simplicity, this examination is limited to reflective methods.

We start with the consideration of an ideally collimated beam of light reflected from a plane mirror surface, and assume that the mirror rotates about an axis through the intersection of beam centerline and mirror. Consideration of the characteristics of surfaces generated by the rotating reflected beam may be formulated into five group classes. First, consider that with respect to the incident beam the rotational axis may be either normal (i.e., a right angle), inclined, or coincident. Next, consider that the rotational axis cannot usefully be normal to the plane of the mirror since this would result in the reflected beam returning back upon itself without scanning. Thus, the rotational axis can only be inclined to, or lie within, the mirror plane. However, the mirror plane may not contain the incident beam and the rotational axis, since this would result in the incident and reflected beams always lying in the mirror plane without resultant scanning. Hence, we need consider only the five cases as listed in Table I.

Figures 1 and 2 show the reflection surface generated for the incident beam normal to the rotational axis. When the mirror plane contains this axis (Figure 1), then the reflection surface generated is a plane. When the mirror plane is inclined (Figure 2), then the reflection surface is a cone. Note, however, that the axis of the cone is tilted. It is inclined with respect to both the rotational axis and the incident beam. To scan successive straight lines in Figure 1, the film

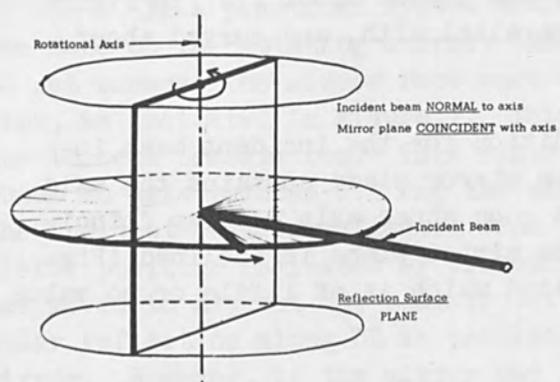


Figure 1.

Scan Geometry for Axis Normal to Incident Beam

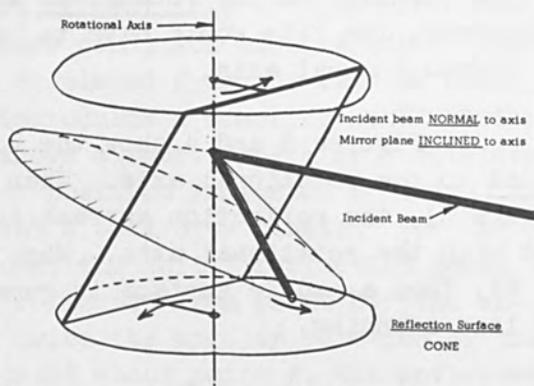


Figure 2.

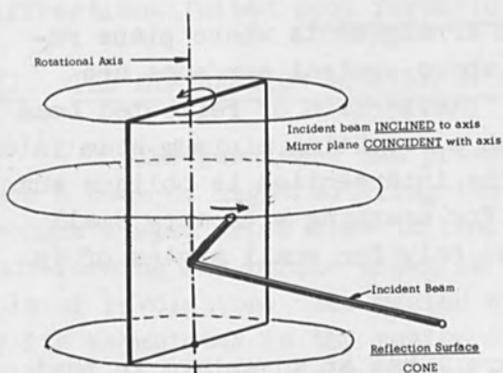


Figure 3.

Scan Geometry for Axis Inclined to Incident Beam

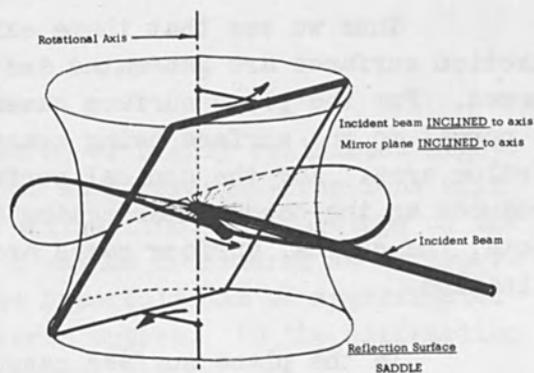


Figure 4.

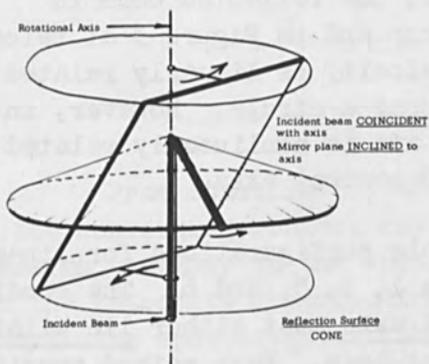


Figure 5.

Scan Geometry for Axis Coincident with Incident Beam

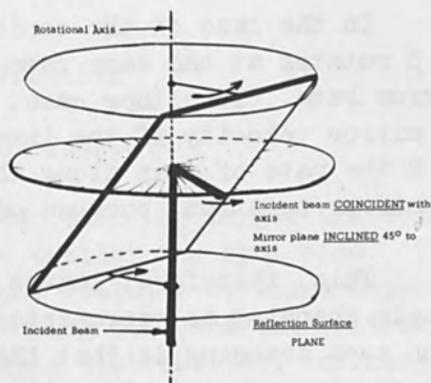


Figure 6.

must be parallel to the rotational axis and curved about it. In Figure 2, however, the film would have to be parallel with, and curved about, the tipped conical axis.

Figures 3 and 4 show the condition for the incident beam in-
clined to the rotational axis. When the mirror plane contains the axis (Figure 3), the reflection surface is a cone whose axis is also coincident with the rotational axis. When the mirror plane is inclined (Figure 4), then a saddle surface is generated which is of little or no value for line scanning.

Figures 5 and 6 show conditions for the incident beam coinci-
dent with the rotational axis. With the mirror plane inclined (Figure 5), the reflection surface is a cone with the axis coincident with the rotational axis. Figure 6 shows the special case where the inclined mirror plane is at 45° to the axis. For this condition the reflection surface is a plane.

Thus we see that there exist two arrangements where plane reflection surfaces are generated and three where conical surfaces are formed. For the plane surface cases, the intersection of reflected beam is normal to the surface being scanned with a resulting minimum beam intersection area. For the conical surfaces, the intersection is oblique and produces an increased intersecting area. For scanning with very small spots, the conical surface cases are useful only for small angles of inclination.

In the plane surface cases, Figure 1 has an advantage in that the reflected beam rotates at twice the mirror rate, but a disadvantage in that the incident and reflected beams are in each other's way during a small angular span. Figure 6 has the advantage that the beams do not interfere, but the disadvantage that mirror and reflected beam rotate at the same rate. In both cases the scan velocity is linearly related to the rotational velocity.

In the case of the conical surfaces, the reflected beam in Figure 5 rotates at the same rate as the mirror and in Figure 3 at twice the mirror rate. In either case, the scan velocity is linearly related to the mirror velocity if the intersection forms a circle. However, in Figure 2 the rate of scan along the circular arc is nonlinearly related to the mirror rotation, because of the tilted conical axis.

This, therefore, leaves four possible configurations for linear wide-angle scanning as represented by Figures 1, 3, 5, and 6. The condition for such scanning is that the rotational axis must either lie within the mirror plane or coincide with the incident beam. Each method provides one conical and one planar reflection surface.

In a practical system where it is desirable to have more than one face on the rotating mirror, the above condition of coincidence cannot be met because the mirror face must be displaced from the axis of rotation, as indicated in Figure 7. This introduces a scanning nonlinearity due to beam translation. This figure shows a four-sided mirror rotating about an axis through C. For the mirror position indicated by solid lines, the incident beam is reflected from point A back upon itself. For the mirror position indicated by broken lines, the initial reflection point has moved to A', but the beam is being reflected from point D. The angular reflection along DE is precisely twice the angular rotation of the mirror. However, if the mirror had rotated about point A, the reflected beam would have been along AB. Thus, it should be apparent that, while the angular rotation is correct, incorrect scanning is introduced because of noncoincident axes.

We have examined the mechanical rotational methods of effecting line scanning of light beams. Let us now examine the requirements of diffraction-limited spot formation.

III. THE DIFFRACTION-LIMITED SPOT

The gas laser can provide a source of highly collimated light. Such a beam of light entering the pupil of an aberration-free lens will produce a spot whose size is limited by diffraction. The envelope of the spot-forming ray bundle shown in Figure 8 may be considered as an hyperbola of revolution. The region where the hyperbola can be approximated by its asymptotes is the region of geometric optics. In the diffraction optics region, the minimum spot size is given by

$$d = k\lambda f = \frac{k\lambda}{2 \tan \alpha} \quad (1)$$

where

d = spot size at geometric focus

λ = wavelength of light

f = relative aperture, f /number

α = half angle of conical bundle.

From equation (1) above it is seen that the larger the angle α for the converging conical ray bundle, the smaller the spot size. In order to assign a number to the spot size, we must evaluate k , which depends upon the intensity distribution of the entering beam and the criterion for defining spot diameter. For the Gaussian distribution, defining both beam and spot diameters at the $1/e^2$ points (13.5%), the value of k equals $4/\pi$ or 1.27. Defining for the Airy disc in terms of the line spread

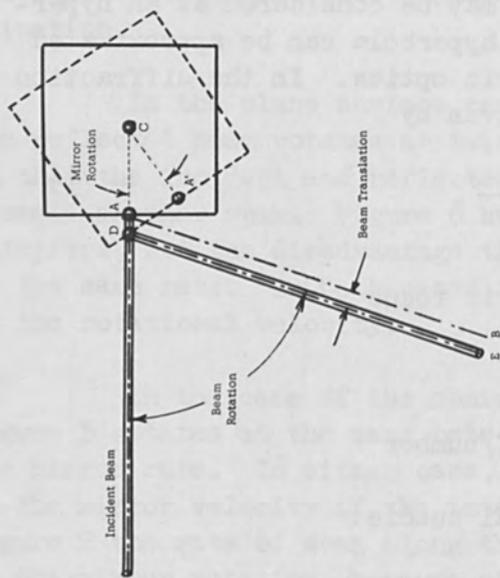


Figure 7. Beam Translation

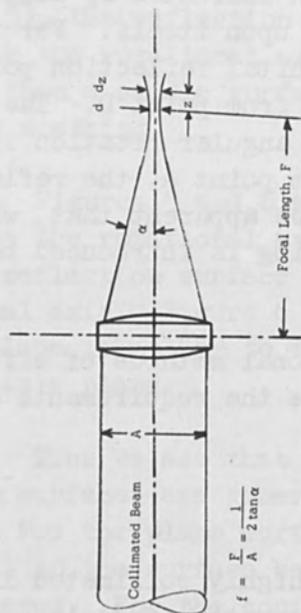


Figure 8. Spot Forming Schematic

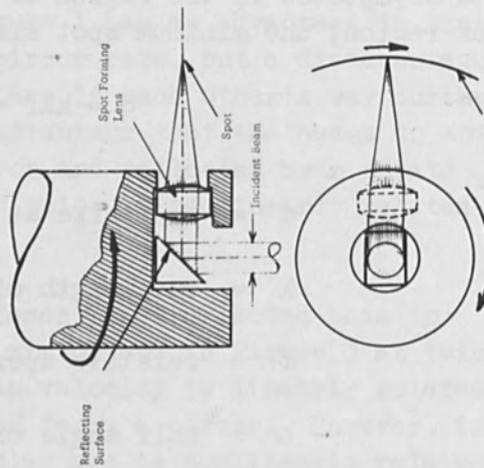


Figure 9. Pupil Rotation Principle

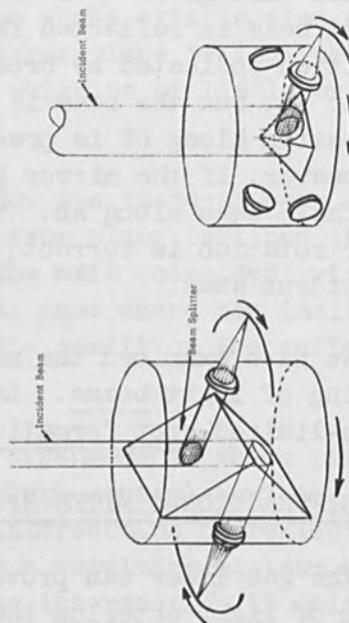


Figure 10. Pupil Rotation Methods

function, at the 10% to 90% points, the value of k is derived as 1.33. This is not the diameter of the first dark ring. k is obtained by solving the transcendental relationship:

$$\frac{1}{\pi} \int_0^{\pi/2} \left[J_0^2 \left(\frac{\pi d \sec \theta}{2\lambda f} \right) + J_1^2 \left(\frac{\pi d \sec \theta}{2\lambda f} \right) \right] d\theta = \frac{1}{10} . \quad (2)$$

In practice, lenses are not completely aberration free. This and other ordinarily minor effects increase the spot size as calculated above. Therefore, in designing a system, one employs a design f/number, f' , which is somewhat smaller than that calculated above. This margin may range from 10% to 35% depending on the system quality. The ratio of focal length F to entering beam diameter A is then fixed by the design f/number:

$$f' = \frac{F}{A} . \quad (3)$$

The magnitude of the focal length F is constrained by the required scan line length and the scanning method utilized. This leaves only the parameter A for arbitrary definition of the spot size in a scanning system. There are practical constraints in this direction as well, since the beam diameter determines the sizes of the scanning mirror and optical components of the system. As an example of practical magnitudes of beam diameter A , Table II lists such values for a 4.5-inch scan line length for which, in one specific system, the focal length would be about 3 inches.

Table II indicates that for the blue Argon laser line, the beam diameter would be about 1/2 inch while for the Neon-Helium laser line it would be about 5/8 inch. Such large beam diameters make system design exceedingly difficult.

The spot size determines the depth of focus or the accuracy of positioning the focal surface. For a Gaussian distribution*,

$$z = \frac{\pi d^2}{4\lambda} \sqrt{\left(\frac{d_z}{d}\right)^2 - 1} \quad (4)$$

where

z = the distance from geometric focus

d = the spot diameter at geometric focus

d_z = spot diameter at z

*Deschamps and Mast, "Beam Tracing and Applications", Symposium on Quasi-Optics, Polytechnic Institute of Brooklyn, June 1964.

SCANNING GEOMETRY

<u>CASE</u>	<u>ROTATIONAL AXIS WITH RESPECT TO INCIDENT BEAM</u>	<u>ROTATIONAL AXIS WITH RESPECT TO MIRROR PLANE</u>
I	NORMAL	COINCIDENT
II	NORMAL	INCLINED
III	INCLINED	COINCIDENT
IV	INCLINED	INCLINED
V	COINCIDENT	INCLINED

TABLE I

SPOT FORMING PARAMETERS

for 5 μ spot and 3" F.L.

	<u>WAVELENGTH</u>
4880 Å	6328 Å
Calculated f/no.	7.7
Design f/no.	5.7-7.0
Beam Diameter (A)	0.43-0.53

TABLE II

REFLEX CATADIOPTRIC SCANNER PARAMETERS

Line Length	4.5 inches
Spot Size	8 microns
Focal Length	3.11 inches
Beam Diameter	0.345 inches
Design f/number	9
Motor Speed	15000 rpm
Line Rate	2000 per sec.
Spots per line	14300
Bandwidth	14 megahertz
Deadtime	8 percent

TABLE III

λ = wavelength of light

In equation (4) the spot diameter is defined at the $1/e^2$ points. Considering the depth of focus as the distance at which the spot changes by 5%, equation (4) yields a value for z just under 10 microns for a 5-micron spot in 6328 Å light or about 12-1/2 microns in 4880 Å light. These dimensions are in the neighborhood of only 1/2 mil. One can see that even the film emulsion thickness will have a significant effect.

With this background in beam and spot characteristics, we will now examine methods of scanning.

IV. SCANNING METHODS

The mechanical rotational scanning methods may be grouped into three basic classes in terms of the region in the optical spot forming system where deflection is introduced. These are:

1. Pupil
2. Exit bundle
3. Entrant bundle.

The first method indicates that scanning is produced by rotation of the lens pupil itself. How this principle is accomplished is shown in Figure 9. The incident beam enters on-axis. The focused spot sweeps out a circular arc. Figure 10 shows two methods of scanning by this principle; both employ the geometry of Case V (Figure 5 and 6). This method can provide linear scans and long scan lines. It is, however, limited to relatively low rotational rates in order to maintain lens location and focal position alignments during rotation.

The principle of exit bundle scanning is shown in Figure 11. Here scanning is effected by introducing a reflecting surface in the exit ray bundle between lens and focused spot. Figure 12 shows two methods of scanning by this principle, one employing the geometry of Case III (Figure 3) and the other that of Case V (Figure 6). This principle is useful for relatively large f /numbers and short scan lines. Long scan lines introduce appreciable scan nonlinearity and shift of focal position due to translation errors (Figure 7). In addition, unless dual beams are employed, excessive deadtime occurs as the corners between successive mirror faces traverse through the exit bundle.

The principle of entrant bundle scanning is indicated in Figure 13. Here the ray bundle entering the spot-forming lens is rotated about an axis through the lens such that the beam always enters the lens

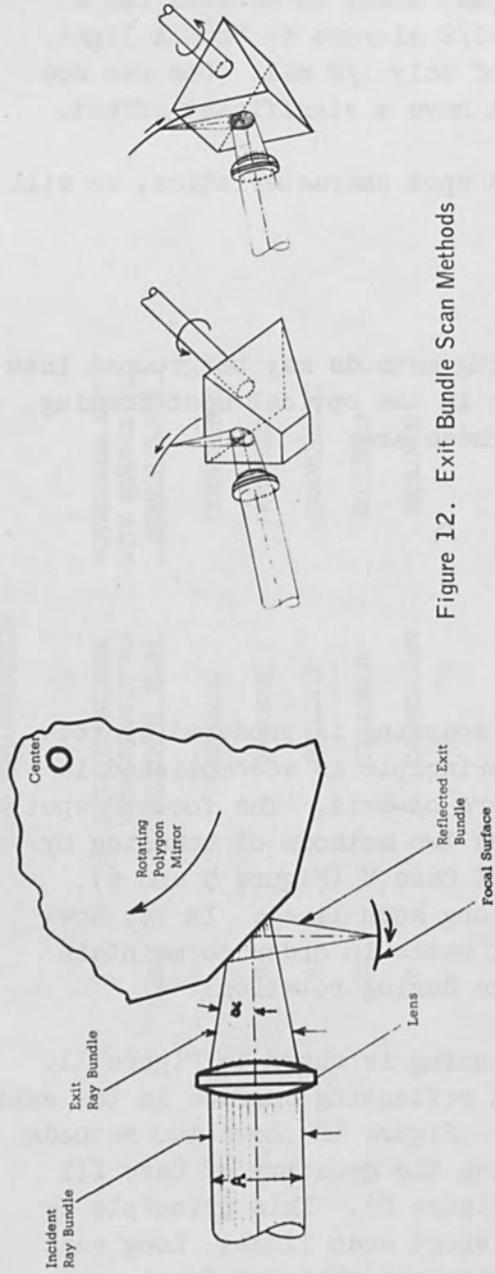


Figure 11. Exit Bundle Scan Principle

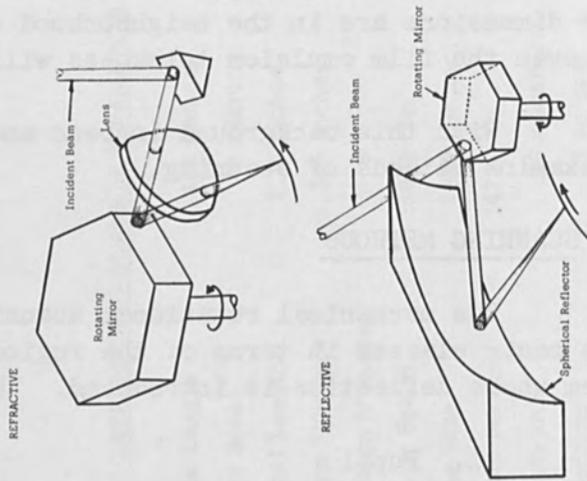


Figure 12. Exit Bundle Scan Methods

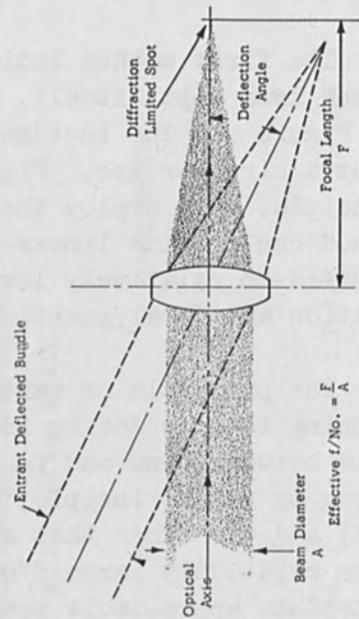


Figure 13. Entrant Bundle Scan Principle

Figure 14. Entrant Bundle Scan Methods

pupil. Two methods of accomplishing such scanning are shown in Figure 14. The simple refractive method utilizes two passes through a lens element so that the beam deflection effectively occurs within the lens at its nodal center. The simple refractive method utilizes an off-axis spherical mirror as a lens. Beam deflection occurs at an axis through the center of curvature of the mirror. For collimated light, the focal surface is halfway between the center and the spherical surface. Spherical aberration in such systems is readily corrected by the addition of a concentric meniscus refractor. Such a combination forms a catadioptric system, an interesting version of which is shown in Figure 15.

The refractive element of Figure 15 is designed so that one of its surfaces is one focal length away from the reflector. By placing a narrow reflective strip at this surface, collimated light entering the system comes to a focus at this strip where it is reflected back to the reflector and emerges from the system as a collimated beam parallel to the entering beam. If the entering beam is deflected through an arc at an axis through the center of the system, the emerging beam will cross this axis for all scan angles. For a lens placed on the axis of the emerging beam, the beam will always enter the lens pupil. Figure 16 shows such an arrangement, which is termed a catadioptric reflex scanner. The system shown is based on the geometry of Case I (Figure 1); however, it may equally well utilize Case III or Case V geometry.

The catadioptric reflex scanner is useful for small diffraction-limited spots, relatively long scan lines, and high scanning rates. By employing a special concentric lens design as shown in Figure 17, diffraction-limited performance is attainable over scan angles as great as 120° . Scan nonlinearities normally introduced by translations of the scanning beam are corrected in this system as shown in Figure 18. Beam A entering a lens pupil will come to focus at A'. Deflected beam B will come to focus at B'. If, in addition to the angular deflection, beam C is also translated as it enters the lens pupil, it will also come to focus at B', thus introducing no translation error. The catadioptric reflex scanner is capable of the highest scan linearity at high scan rates. This is important for small scan spots since it aids in avoiding variations in exposure and spot size if the scanner is to be part of an optical recording system.

A system such as that shown in Figure 16 was constructed. It utilizes two collimated beams directed at 90° toward a four-sided rotating polygon mirror. Parameters of the system for Neon-Helium laser light at 6328 \AA are listed in Table III. For Argon laser light at 4880 \AA , the spot size would be reduced to 6 microns, the spots per line would increase to 19,000, and the bandwidth to 18 megahertz. The motor was actually operated at speeds up to 30,000 rpm, which corresponds to line rates of 4,000 per second and which, of course, doubles the bandwidth.

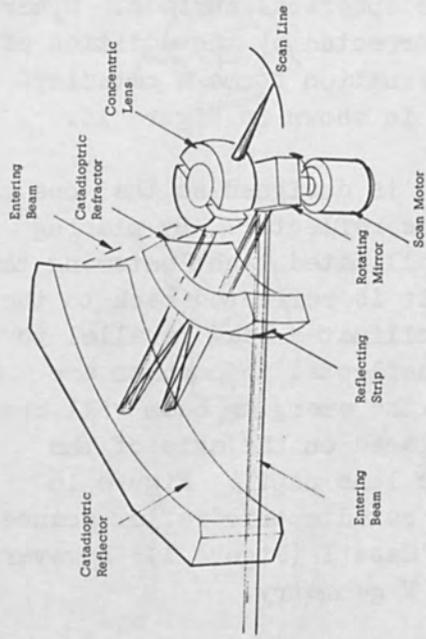


Figure 16. Catadioptric Reflex Scanner

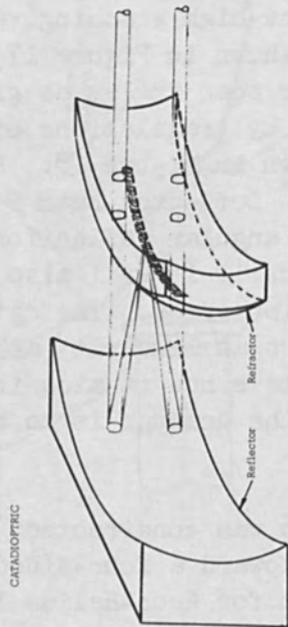


Figure 15. Catadioptric Reflex

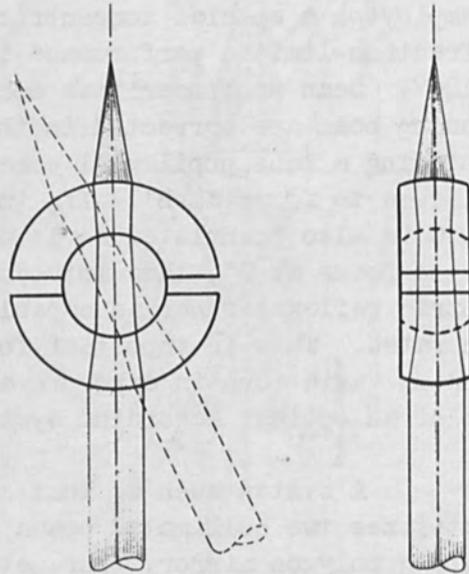


Figure 17. Concentric Lens

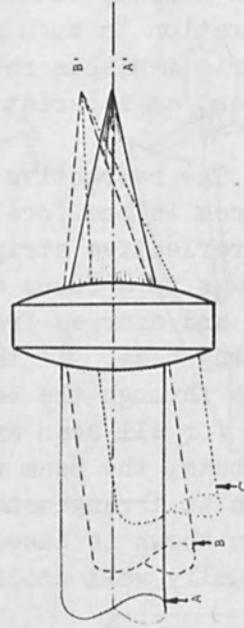


Figure 18. Translation Correction Principle

The optical design details for the catadioptric scanner and concentric lens are covered by E. K. Thorburn who performed the optical work¹.

V. ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance and support of David Power of the Air Force Avionics Laboratory under whose direction the catadioptric scanning system was developed, built and demonstrated.

The optical design details for the scanning system are described in a report by E. K. Jordan who performed the optical work.

V. ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance and support of David L. Taylor of the Air Force Avionics Laboratory upon whose ideas the parabolic scanning system was developed. His assistance is gratefully acknowledged.



Figure 14. Parabolic Scanning System



Figure 15. Cylindrical Scanning System



Figure 16. Curved Scanning System



Figure 17. Circular Scanning System



DIFFRACTION LIMITED CONCENTRIC OPTICS*

by

E. K. Thorburn
Optical Instruments Corporation
Buena Park, California

ABSTRACT

In this day of the high speed computer with automatic optimizing programs, many optical designers are prone to forget that there are some simple approaches to the solution of optical design problems. One of these simple concepts is that of the concentric optical system. Some novel applications of this principle to achieve diffraction-limited performance for flying spot recording and readout are discussed, along with possible applications to a unique infrared optical system.

I. INTRODUCTION

Several of the criteria for the design of a laser beam scanner have been reviewed in the previous paper¹. The scanning systems which were discussed in that paper will be reviewed here from the standpoint of the problems confronting the optical designer and the solutions which have been achieved.

The specific requirements of the recorder system which was to be designed were:

1. The scan angle shall be as large as possible.
2. The system shall produce a diffraction limited point image (Airy disc) over the length of the entire scan line, at a speed of $f/7.0$ or faster.
3. The dead time between successive scans shall be minimized.
4. Provision shall be made for controlled non-linearization of the scan both along the scan path and at right angles to it.

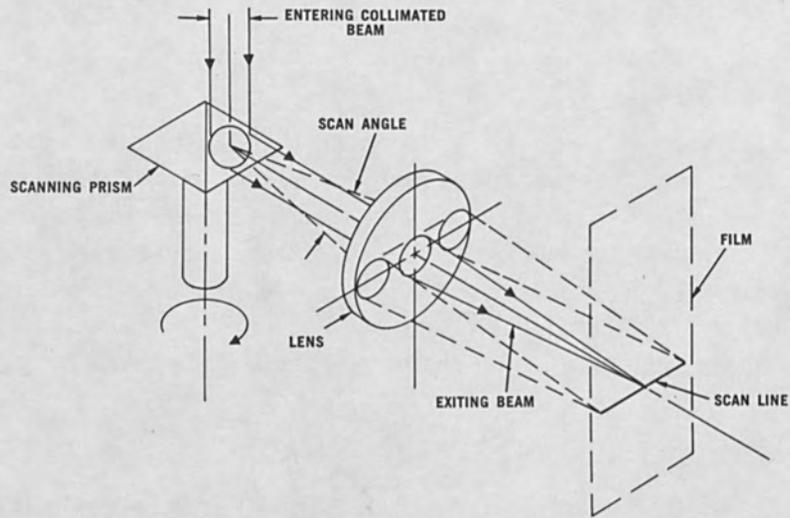
II. DESIGN CONCEPTS

In the past several line scanners were designed which used the concept shown in Figure 1. The optical design of such a system for

* The work covered in this report was supported in part by the U. S. Air Force Avionics Laboratory, Wright-Patterson Air Force Base.

SCANNER REQUIRING COMPLEX OPTICS

FIGURE 1



SIMPLE SCANNING RECORDERS

FIGURE 2a.

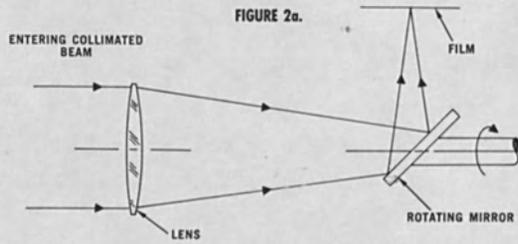
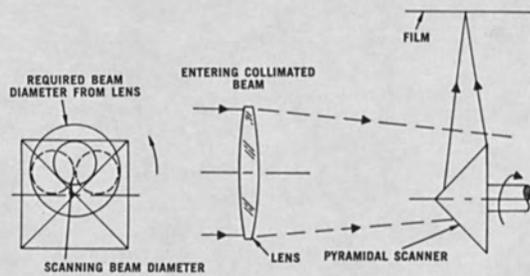


FIGURE 2b.



diffraction-limited performance is extremely difficult. In the first place, the entrance pupil is automatically located at the scanning prism. The design of a pupil-in-front lens is difficult even using modern optimizing programs on a high speed computer such as the IBM 7094. The design becomes very sophisticated, and very difficult to build, if diffraction-limited performance is truly to be achieved. The scan length is limited to a rather modest angle by the maximum obtainable lens diameters.

The lens designs for what Bousky has called the "exit bundle scanner" system, as shown in Figures 2a and 2b, is straightforward. Air spaced doublets can be used up to speeds of approximately $f/3.5$ or faster. By aspherizing one surface slightly, the speed can be made faster than $f/2.5$. Considerably higher speeds can be obtained by adding one or two additional elements. The required f -number of the lens is essentially $1/2$ that of the beam being scanned; however, as is illustrated in Figure 2b, the specified $f/7.0$ beam speed would require an $f/3.5$ lens. This system was ruled out by considerations other than optical image quality.

III. THE CONCENTRIC SCANNER

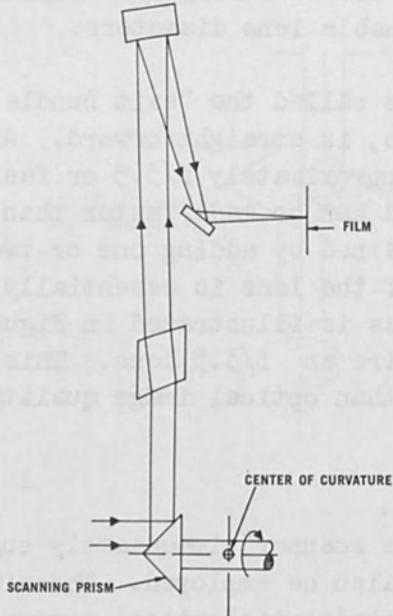
The curved field of the "exit bundle scanner" immediately suggests that a concentric optical system might also be employed. The simplest such system is the "Bouwers-Maksutov" catadioptric optical system as shown in Figure 3. It has the ideal characteristic of having the pupil in front (at the common center of the reflecting and refracting surfaces) and it can be made diffraction-limited at relatively small f -numbers. Normally, if we were to require such a system to have a circular field of view of 90° , we would find that the entire aperture is obscured by the field itself. However, as the prescribed field is a single line, we are able to use an off-center position of the pupil and thus keep the field from obscuring the aperture. A flat mirror must be used to reflect the beam onto the film, and as a result the off-axis angle becomes large.

A more accessible position for the focal surface may be achieved by employing a concentric relay lens as shown in Figure 4. The off-axis angle becomes smaller in this case, with a consequent reduction in aberrations. The entrance pupil, as determined by the scanning prism, is located adjacent to the center of curvature. The pupil is reimaged at unit magnification by the Bouwers-Maksutov system at a point located symmetrically across the center of curvature. It is at this pupil image that the concentric relay lens is located.

A revision of this design is shown in Figure 5 in which the image formed by the Bouwers-Maksutov system is focused onto an aluminized strip on the second surface of the corrector. The beam is then reflected back through the system, being recollimated in the process. The concentric relay lens need only be half as fast as the one shown in Figure 4.

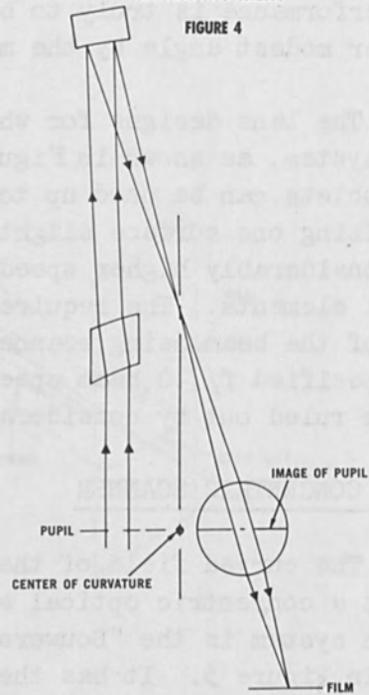
BOUWERS-MAKSUTOV SCANNER

FIGURE 3



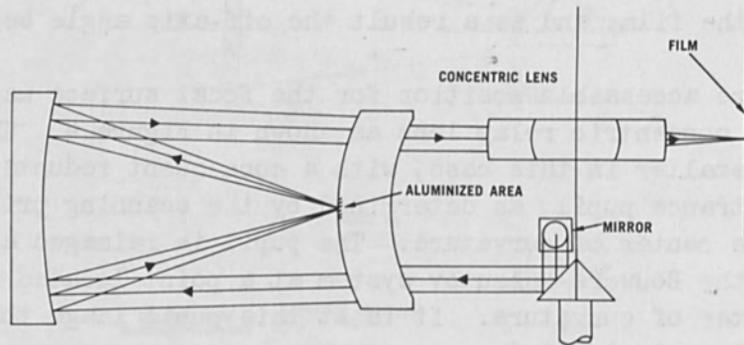
BOUWERS-MAKSUTOV SCANNER OPTICS WITH CONCENTRIC REFRACTING RELAY

FIGURE 4



COLLIMATED BEAM SCANNER

FIGURE 5



The focal length of the catadioptric portion is unrelated to that of the total system and may be chosen to give optimum performance. Because the catadioptric system is a virtual "Z" configuration, there is no coma. The less objectionable spherical aberration, and the astigmatism introduced by the beam being off-axis remain, however.

A more complex arrangement as shown in Figure 6 provides for non-linearization. In order to keep the non-linearizing device small, it should also be located in the pupil. The system in Figure 6 accomplishes this by supplying an additional pupil image at the point where the "non-linearizing mirror" is located.

IV. READOUT OPTICS

The discussion to this point has been limited to recording systems. The design of Figure 5 can be readily converted to a readout system by the addition of a spherical mirror behind the film to be read. The center of curvature of this mirror is located adjacent to the pupil of the lens and forms an image of the pupil onto the cathode of the photomultiplier (a well-known requirement for scanner systems).

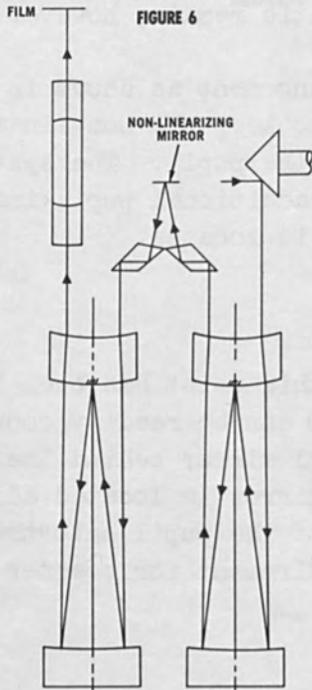
V. CONCENTRIC REFRACTORS

Concentric refractors are among the oldest wide angle photographic optical systems. In 1859 T. Sutton used a water-filled concentric glass sphere for wide angle photography. A concentric panoramic lens has also been described by Baker as a World War II development.² Such lenses can be made diffraction-limited in speeds up to approximately $f/3.0$. At a speed of $f/5.0$ or slower, the design is relatively insensitive to the ratio of indices between the low index center and the high index outer shell. For example, in an $f/5.0$ lens with an inner index of 1.5, the zonal spherical aberration is reduced only 10 per cent when the index in the outer shell is increased from 1.6 to 2.0. When we consider that a 3.1 inch focal length design with the 1.6 index outer shell had less than $1/150$ of a wavelength of spherical aberration, we see that this 10 per cent is truly insignificant.

VI. AN INFRARED CONCENTRIC DESIGN

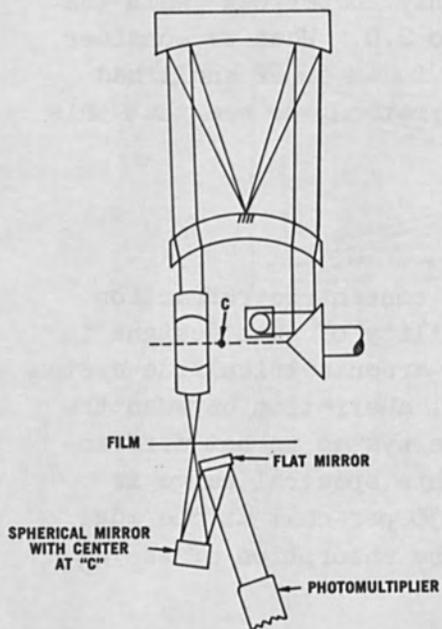
The remarkable good performance of the concentric refraction system led to the investigations of the applicability of such designs in the infrared. Figure 8 shows an $f/1.0$ germanium-arsenic trisulfide system which is achromatized and corrected for spherical aberration between the wavelengths of 2.5 and 4.0 microns. Although the system is not diffraction-limited, the theoretical resolution over this spectral range is 0.8 milliradians. The relative illumination is 50 per cent at the edge of 120° field of view. Considerations such as the absorption of As_2S_3

DOUBLE SCANNER WITH NON-LINEARIZING MIRROR



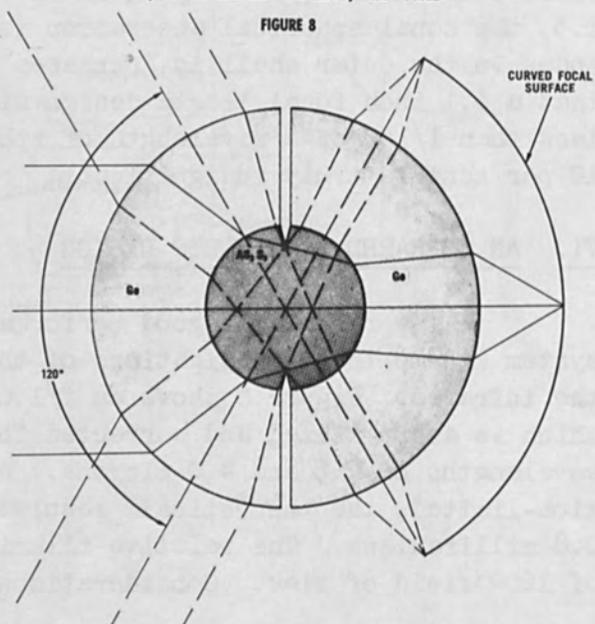
CONCENTRIC READOUT SYSTEM

FIGURE 7



OPTICAL CONFIGURATION f/1.0 IR CONCENTRIC LENS, 120° FIELD

FIGURE 8



in the 2.7 micron region rules out the practical application of such a lens to Ir imagery. It appears, however, that this can be remedied by the substitution of Irtran 4 in place of As_2S_3 . Cementing "in the dark" with the high index ($n = 2.3$) "cements" now used for immersed detectors is probably the most formidable barrier to its use. If these problems can be solved by use of alternative materials, perhaps a practical infrared application can be found for such systems.

VII. CONCLUSIONS

High performance optical systems can be designed for specific applications using only concentric surfaces. In the specific case of laser scanner systems the performance can be truly diffraction-limited. A system similar to that shown in Figure 5 was designed and fabricated. The design aberrations for the system as shown were $1/25\lambda$. Practical problems associated with introduction of multiple beams onto a flat faced scanning prism resulted in a degradation of the design performance to $1/10\lambda$.

REFERENCES

1. S. Bousky, "Diffraction-Limited Spot Scanning by a Laser Beam".
2. J. Baker and J. Chandler, "Summary Technical Report of NDRC, Div. 16, V.1., "Optical Instruments", p. 40.

in the 2.5 micron region unless the practical application of such a
 lens is infeasible. It appears, however, that this can be remedied by
 the substitution of InGaAs in place of Ag₂S. Constant "in the dark"
 with the high index in 2.5 microns, but not the increased absorption
 which would be a serious problem. If these problems
 can be solved by use of alternative materials, perhaps a practical ultra-
 red application can be found for such systems.

VII. CONCLUSIONS

High performance optical systems can be designed for specific
 applications using only conventional materials. In the specific case of
 laser scanner systems the performance can be easily diffusion-limited.
 A system similar to that shown in Figure 3 was designed and fabricated.
 The design operations for the system as shown were VLSI. Practical pro-
 cesses associated with fabrication of multiple beams onto a flat glass
 scanning prism resulted in a degradation of the design performance to



FIGURE 1. Schematic diagram of the optical system.

REFERENCES

1. J. E. Bennett, "Diffraction-limited Spot Scanning by Laser Beam",
2. J. E. Bennett and J. E. Bennett, "Summary Technical Report of RSD, Div. 16,
3. V. L. "Optical Instrumentation", p. 40.

THE PRODUCTION AND USE OF MONOENERGETIC ELECTRON BEAMS

by

J. Arol Simpson
National Bureau of Standards
Washington, D. C.

ABSTRACT

The generation of dense electron beams of energy half-width below 0.1 eV presents special problems. Among these are generation of tightly collimated beams of very low energy, space-charge effects in deflectors and anomalous energy spreads. A discussion of one successful attack on these problems will be given together with examples of the uses of such beams.

I. INTRODUCTION

In most electron beam devices the least important parameter is the exact kinetic energy of the electron beam. It is fortunate that this is the case since electron energy is one of the most difficult parameters to measure to high precision. There exist unknown contact potentials between the cathode and other parts of the system which cause the electron energy to differ by a few volts from the anode potential, and also introduce errors into retarding potential measurements. Exact kinetic energy measurements require either geometrically precise deflection analyzers or calibration against known atomic excitations. Even then, due to the fact that the almost universally used source of electrons is a thermionic cathode, the energy remains undefined to within the energy distribution of the electrons leaving the cathode.

II. BEAM ENERGY SPREAD

This distribution is the well-known Maxwellian whose full-width at half-maximum $\Delta E_{1/2}$ is given, in electron volts, by

$$\Delta E_{1/2} = 2.54 \frac{T}{11,600} .$$

T is the temperature in degrees Kelvin. Hence for the usual cathode materials, $\Delta E_{1/2}$ varies between about 0.25 and 0.5 eV. As discussed in another paper at this conference¹ the beams actually generated have still wider distributions. For purposes where precise electron energies must be known it is necessary to reduce this spread or monochromatize the beam.

Schemes have been proposed to accomplish this end by means of both klystron bunching or cooling techniques. None of these have been successful, and hence we will confine our attention to means of filtering out of the beam all but those electrons whose energies lie in a narrow band. This operation implies a serious loss of current, but a simple calculation for a tungsten cathode will show that 100 μ A of current in a one mm spot should be available with a full-width at half-maximum of 20 millielectron volts. I will admit at the start that no current of this order of magnitude with this energy spread has yet been achieved. The design problem in attempting to achieve such performance arises through the interaction of three different restricting relationships between the angle, the energy and the cross sectional area of an electron beam in free space. The first of these restrictions is the well-known Helmholtz-Lagrange Equation, which has been discussed extensively², which in the simplest form may be written:

$$E_1 \Delta \Omega_1 \Delta A_1 = E_2 \Delta \Omega_2 \Delta A_2 .$$

This law which might be subtitled the conservation of electron brightness or "Richtstrahlwert" relates the energy, E, convergence angle Ω and area of a beam at any plane, 1, in an electron optical system to that at any other plane, 2. There is one plane in the electron optical system which is not under the designer's control, and that is the cathode. If the conditions at the cathode are substituted for the conditions at plane 1 we arrive at the Langmuir Equation³ which states that the maximum obtainable current density J_{\max} is related to the cathode current density by:

$$J_{\max} = J_{\text{cathode}} \left(1 + \frac{E_2}{kT} \right) \sin^2 \theta_2$$

where the solid angle $\Delta \Omega_2$ has been expressed in terms of the convergence angle, θ , k is the Boltzmann constant, and T the cathode temperature. This relationship between the energy, the angle and the current density is of a fundamental nature arising from thermodynamic considerations and will not be violated at any point throughout the system.

In addition to this relationship we have the relationship given by the well-known⁴ space-charge-limited current in which the maximum current at the crossover point in the space-charge-limited beam is given by

$$I_{\max} = 38.5 E^{3/2} \theta^2 .$$

It must be realized at the onset, that this equation derived from the laminar flow space-charge theory cannot be correct because in laminar flow at the waist of the space-charge-limited beam the convergence angles are zero and hence if the beam arose from a cathode of finite brightness

whose temperature is above absolute zero this equation itself is in violation of the more fundamental Helmholtz-Lagrange relationship. However, in practice the electron trajectories do cross the axis and control of the angular divergence may be obtained by placing a small real image satisfying the Helmholtz-Lagrange relation at the space-charge waist. The space-charge spreading then becomes essentially a perturbation on this. The details of this analysis may be found in standard texts on electron optics.⁵ Two facts about the space-charge limitation must not be overlooked. The first is that the limitation is on current not current density and by appropriate physical scaling any desired density may be obtained. The second is that contrary to "common sense" for a device of fixed size, because of the difference in energy exponent, as the energy is raised the space-charge restriction becomes dominant over that of Helmholtz-Lagrange. At low energies space charge is not the limitation; all devices are brightness limited.

III. OTHER CONSIDERATIONS

In the cathode region another problem arises. The brightness of the cathode which must be put into the Langmuir Equation is the brightness of the cathode under the space charge conditions which exist in front of the cathode, hence one cannot merely substitute the maximum available brightness available from a tungsten source, for example, into the Langmuir Equation and obtain the correct answer unless one insures that the gun design is such that the cathode is operating under emission-limited rather than the space-charge-limited conditions. In most high current guns this is not the case. The result of this reduction of cathode brightness by space charge in the case of the well-known Pierce³ gun is that no such gun will operate according to the usual design figures if the anode voltage is less than 300 volts.⁶ Many of the uses of monoenergetic electron beams require electron beams of energies below this figure. In order to achieve these high density, low energy beams it is therefore necessary to operate the electron gun with an intermediate anode at potential considerably above that of the final beam.⁷

If one examines the common methods⁸ of sorting charged particles according to their energy either by electrostatic or magnetic deflection one soon realizes that even if the required beam is at an energy in excess of 300 volts one must, in order to keep the apparatus of convenient size, operate the electron sorter at an energy far below this value. This necessity arises out of the fact that the resolution of all known energy sorters can (to first approximation) be written as:

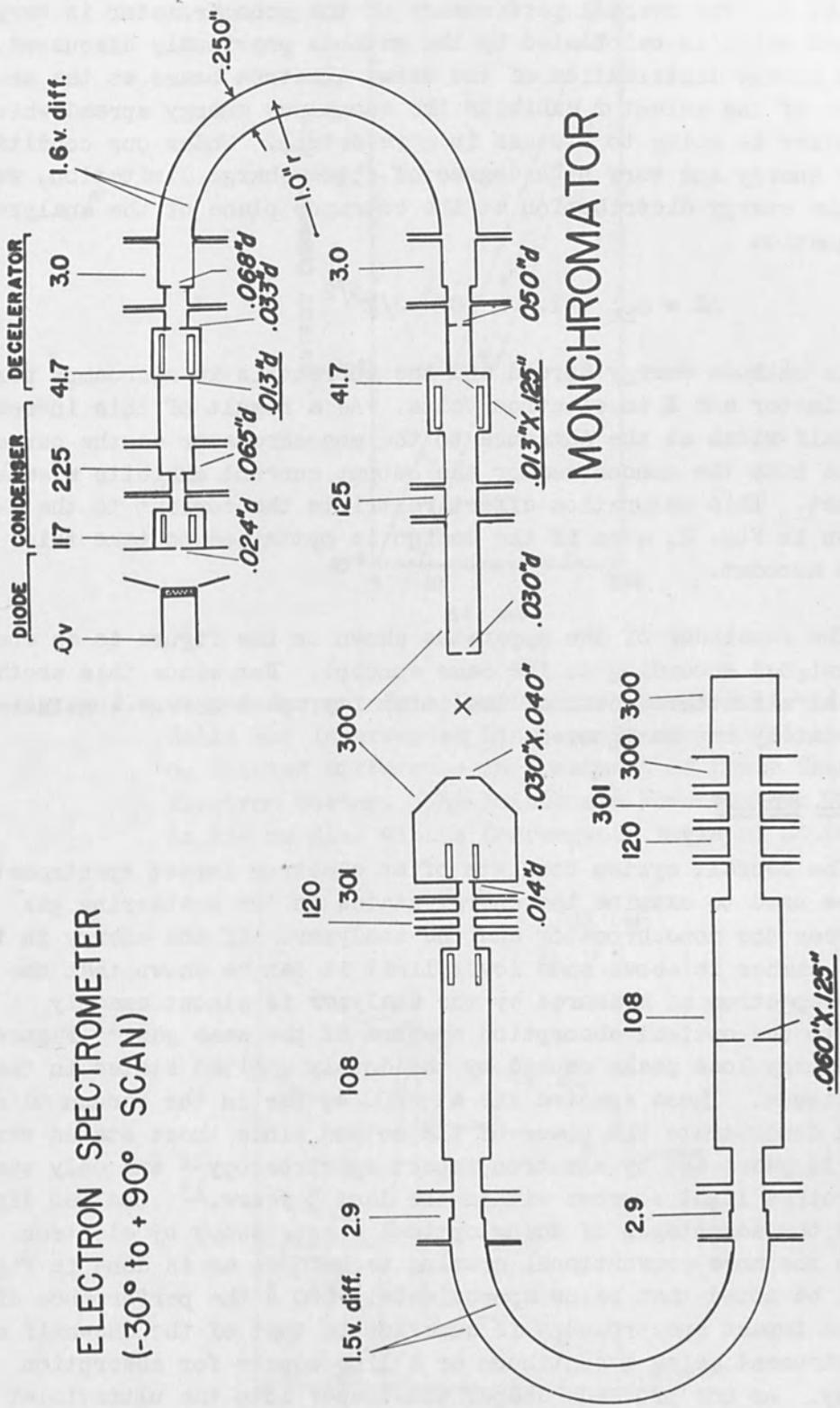
$$\frac{\Delta E}{E} = A \frac{S}{R} + B\theta^2,$$

where $\frac{\Delta E}{E}$ is the base resolution of the analyzer, S the entrance and exit

widths, R the radius of curvature within the deflector and θ the angle of convergence at the entrance slit. This equation not only requires for low absolute value of ΔE , low values of E , but also is still third restrictive relationship between I , the beam area, S^2 , and the convergence angle θ . With these facts in mind the design process is straightforward if somewhat tedious. The size of the monochromator and its type once chosen fixes R , A , and B the required resolution then fixes ΔE . The designer then must maximize I by choice of E , S and θ , subject to the space charge and brightness restrictions. Figure 1 shows the scheme by which the Electron Physics Section of the National Bureau of Standards has built a monochromator which satisfies these requirements. The cathode is an indirectly heated dispenser type operating under space-charge-limited conditions in a plane parallel diode structure. Typically the operating voltage of the diode is 100 volts. The anode hole, 0.65 mm in diameter, passes a current of $\sim 20 \mu\text{A}$. Two lenses formed by the first anode, second anode and the intermediate element second anode images this on a 13 mil aperture at an energy of approximately 20 eV. These lenses, as all the other lenses in the system, are electrostatic tube lenses based on the empirically derived designs of Spangenberg and Field.⁹ The second aperture serves as a field stop in the optical sense limiting the dimensions of the beams in two directions. A further tube lens operating at a deceleration ratio of 10:1 forms an image of this aperture in the entrance plane of the analyzer. In order to conserve length the angle stops are in the low energy section. At a voltage of two volts, this image has a diameter of 1/2 mm and a convergence angle of 0.07 radian. At this point the beam of $\sim 5 \times 10^{-7}$ amps is fully space-charge-limited with a micropervance of 38.5. This scheme of keeping the real physical aperture at a higher voltage has a number of incidental advantages such as reducing scatter into the analyzer and stopping field distortion at the beginning of the analyzer.

IV. ENERGY SELECTOR

The energy selector consists of a concentric spherical deflector operating according to the design of Purcell. Because of the presence of appreciable space charge the behavior of the beam in the selector is not known in detail. The model which we adopted was to split the focusing power of the selector into two lenses which were assumed to be one fourth of the distance around the sphere. Under these conditions and with the space-charge-limited beam one can follow the profile of the beam by means of the universal space charge equation¹⁰ and find that despite the spreading, the resolution of the analyzer should not be degraded. Experiments prove that this is the case. From the selector exit plane the beam is reaccelerated through a 10/1 lens to the exit slit at which point the energy selection is made. The beam is then further accelerated through a series of lenses to the required final energy. This device gives at the 0.75 mm entrance to what is called the scattering chamber a beam whose current varies with half-width as



ELECTRON SPECTROMETER
 (-30° to +90° SCAN)

ANALYZER

Figure 1. Schematic Drawing of NBS Electron Spectrometer. Typical Operating Potentials are Shown Above the Drawing While Aperture Dimensions in Inches are Given Below.

shown in Fig. 3. The overall performance of the monochromator is very close to that which is calculated by the methods previously discussed. However the energy distribution of the dense electron beams at the entrance plane of the selector exhibits the anomalous energy spread which Professor Ulmer is going to discuss in more detail. Under our conditions of very low energy and very high degree of space-charge limitation, we find that the energy distribution at the entrance plane of the analyzer obeys an equation

$$\Delta E = \Delta_{EK} + 1.4 \times 10^{-3} J/E^{3/2}$$

where Δ_{EK} is cathode energy spread and the current is in microamps per square centimeter and E in electron volts. As a result of this increase in energy half-width at the entrance to the monochromator as the current is increased into the monochromator the output current exhibits a saturation effect. This saturation effect restricts the current to the values shown in Fig. 2, even if the design is optimized to take this effect into account.

The remainder of the apparatus shown on the figure is an energy analyzer designed according to the same concept. But since this section is looking at a scattered beam of low intensity space-charge considerations fortunately may be ignored.

V. OVERALL SYSTEM

The overall system consists of an electron impact spectrometer which may be used to examine the energy states in the scattering gas placed between the monochromator and the analyzer. If the energy in the scattering chamber is above some lower limit it can be shown that the energy loss spectrum as measured by the analyzer is almost exactly equivalent to the optical absorption spectra of the same gas.¹¹ Figure 3 shows the energy loss peaks caused by the doubly excited states in the helium continuum. These spectra lie at ~ 200 Å, far in the vacuum ultraviolet, and demonstrate the power of the method since these states were discovered 30 years ago by electron impact spectroscopy¹² and only seen with synchrotron light sources within the last 5 years.¹³ One can directly compare the advantages of doing optical spectroscopy by electron impact with the more conventional grating techniques as is done in Fig. 4. It will be noted that below approximately 600 Å the performance of the electron impact spectroscopy is superior to that of the one-half meter grating instrument using a continuum or a line source for absorption spectroscopy. As one proceeds deeper and deeper into the ultraviolet the advantage of the electron spectrometer increases. The electron spectrometer moreover has the advantage of much smaller size and considerable increase in convenience. In order to change the wavelength

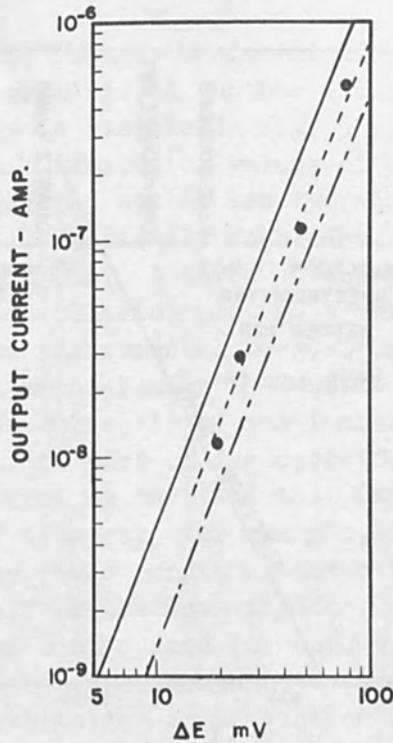


Figure 2. Output Current As a Function of Energy Half-Width. The Solid and Interrupted Line are Predicted Values Depending on Diluted Difference in Treatment of Space Charge Within Electron Sorter. The Points are Measurements. The Beam is 1/2 mm dia. With a Convergence Angle of ~ 0.007 Rad.

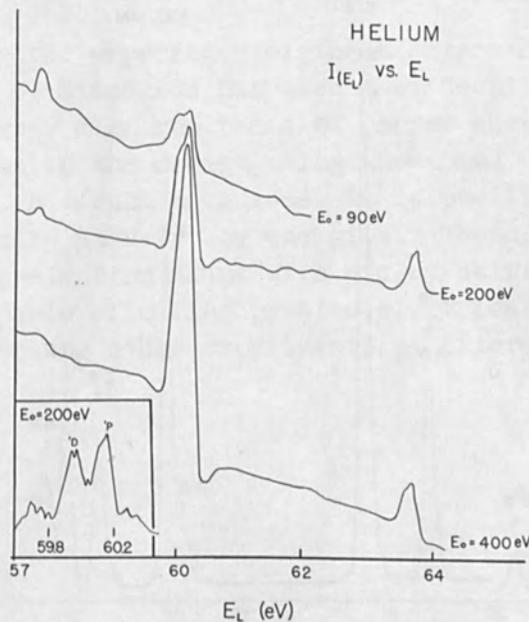


Figure 3. Energy Loss Spectrum of Doubly Excited He. Note that the 1D State Which Is Optically Forbidden Increases in Prominence and Impact Energy E_0 Is Lowered to 90 eV.

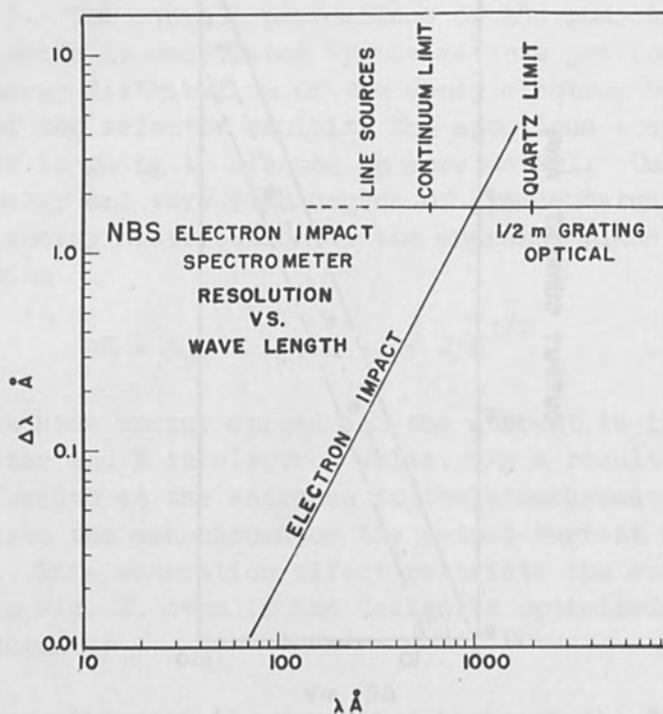


Figure 4. Comparison Between Photon Absorption and Electron Impact Spectroscopy. The Easily Obtained "Working" Resolution of a 1/2 Meter Photon Instrument Using the Indicated Light Sources Is Used As A Basis of Comparison With the Same Quantity for the Electron Impact Spectrometer.

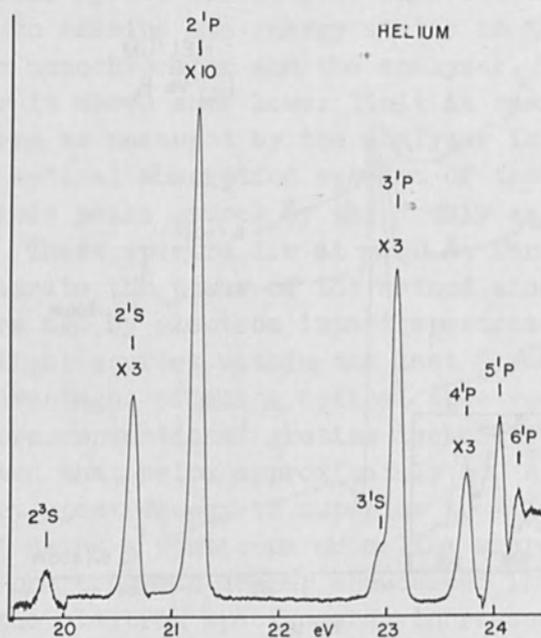


Figure 5. The Absorption Spectra of He Measured by Electron Impact. Note the Prominence of the Optically Forbidden 3^1S and 3^1P States. Resolution ~ 0.07 eV.

scale one needs merely to change an electrical potential rather than any mechanical movement of grating. A further example of the power of electron impact spectroscopy is visible on Fig. 5 where the lowest states of helium are displayed. Excitation values of the triplet S and triplet P states are seen. These two states are typical of many excited states of atoms and are so-called optically forbidden, hence their excitation cannot be achieved by photons. A third advantage is more subtle in that due to the nature of the collision process a very narrow line in a spectrum, which is below the instrumental resolution, is recorded in correct intensity relation to a much broader line while in the optical case this average is not made correctly by the instrument and must be made by proper techniques on the part of the operator. The fact that this is not always done is shown by the fact that fairly recent data on the ultraviolet spectrum of hydrogen, for example, shows errors of almost an order of magnitude in the strength of these transitions due to improper averaging on behalf of the investigator. Of course these monoenergetic electron beams can be used for unique scattering type experiments instead of aping the experiments of optical spectroscopy. for example shows the excitation cross section for the first (n equals two) states in He as a function of electron energy just above a threshold of the excitation.¹⁴ The existence of the many irregularities demonstrates the existence of hitherto unsuspected negative ion states. A considerable amount of work has been done in the past year or so and whole new classes¹⁵ of negative ion states have been discovered by electron impact.

VI. CONCLUSIONS

In summary the experience of the Electron Physics Section of the National Bureau of Standards has been that despite the bad reputation of extremely low energy electron beams of narrow energy spread, if sufficient care is taken in the design using known and understood lenses and meticulous care in vacuum technique, it is possible to do sophisticated electron optics at very low energies. These techniques have been used to develop electron beams with energy half-widths of less than 0.01 eV and capable of making precise electrical measurements on atomic states and perform other experiments of interest to Physics.

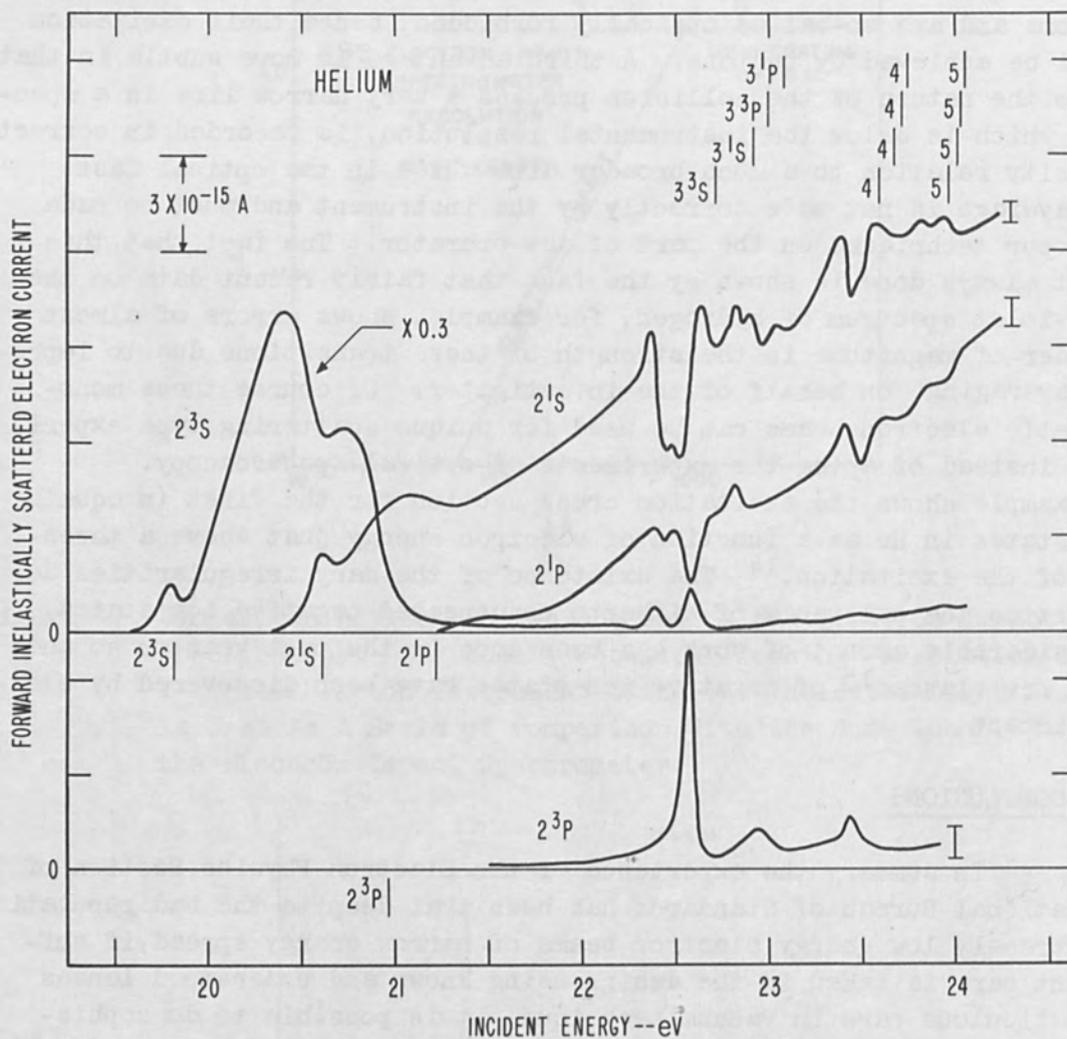


Figure 6. Excitation Functions for the Lowest States in He As Measured by NBS Electron Spectrometer. The Structure is Due to Interference Phenomena Between the Direct Excitations and Negative Ion States. The Energy Resolution is ~ 0.06 eV.

REFERENCES

1. K. Ulmer and B. Zimmermann, following paper.
2. P. A. Sturrock, Static and Dynamic Electron Optics (Cambridge University Press, Cambridge, England, 1955) p. 93, it. seq.
3. J. R. Pierce, Theory and Design of Electron Beams, second edition (Van Nostrand Book Co., New York, 1954).
4. J. R. Pierce, op. cit. Chap. 8.
5. W. Glaser, Grundlagen der Electronenoptik (Springer-Verlag, Vienna, 1952).
6. J. A. Simpson and C. E. Kuyatt, J. Res. NBS 67C, 279 (1963).
7. J. A. Simpson and C. E. Kuyatt, Rev. Sci. Instr. 34, 265, 1963.
8. K. T. Bainbridge, Experimental Nuclear Physics, E. Segre, Editor, Vol. 1, p. 571 (John Wiley, New York) (1953).
9. K. R. Spangenberg and L. M. Field, Elec. Commun. 21, 194, 1963.
10. K. R. Spangenberg, Vacuum Tubes, p. 465 (McGraw-Hill Book Co., Inc., New York, 1948).
11. E. N. Lassetre, et al., J. Chem. Phys. 40, 1208 et. seq. (1965).
12. R. Whiddington and H. Priestley, Proc. Roy. Soc. London A145, 462 (1932).
13. R. P. Madden and K. Codling, Phys. Rev. Letters 10, 516 (1963).
14. G. Chamberlain and H.G.M. Heidemann, Phys. Rev. Letters 15, 337 (1965).
15. C. E. Kuyatt, et. al., Phys. Rev. 138, A385 (1965).

1. A. Unter and B. Zimmermann, *Zeitschrift für Naturforschung*, **19**, 111 (1964).
2. F. A. Schwarz, *Static and Dynamic Properties of Solids* (Gordon and Breach, London, 1962), p. 111.
3. J. N. Stroh, *Theory and Design of Elastic Media*, second edition (Van Nostrand Reinhold, New York, 1957).
4. J. F. Johnson, *et al.*, *Phys. Rev.*, **137**, 1021 (1954).
5. W. Diller, *Verhandlungen der Naturforschenden Versammlung in Basel*, **1957**.
6. J. A. Sauer and U. E. Parrott, *J. Appl. Phys.*, **34**, 202 (1963).
7. J. A. Sauer and C. E. Kline, *Rev. Sci. Instrum.*, **34**, 202 (1963).
8. E. T. Baird, *Experimental Molecular Physics*, **1**, 1 (1951).
9. K. H. Spangenberg, *Phys. Rev.*, **137**, 1021 (1954).
10. E. H. Spangenberg, *Vacuum Tubes*, p. 463 (McGraw-Hill Book Co., New York, 1958).
11. E. H. Spangenberg, *et al.*, *J. Appl. Phys.*, **34**, 202 (1963).
12. H. W. Melville and E. T. Baird, *Proc. Roy. Soc. (London A)*, **162** (1937).
13. E. H. Spangenberg and K. G. ... *Phys. Rev.*, **137**, 1021 (1954).
14. G. ... *Phys. Rev.*, **137**, 1021 (1954).
15. C. N. ... *Phys. Rev.*, **137**, 1021 (1954).

MECHANISM OF ENERGY RELAXATION IN ELECTRON BEAMS

By

Kurt Ulmer and Bodo W. Zimmermann
Physical Institute
T. H. Karlsruhe, Germany

ABSTRACT

In electron beams with high charge density, anomalous shiftings and broadenings as well as symmetrizations of the energy distribution (Boersch 1954) are observed. Continuing former work, where these phenomena were ascribed to a process which changes the energy distribution to a new equilibrium distribution, a corresponding mechanism is considered in greater detail and an apparent paradox is resolved.

I. PROBLEM AND EXPERIMENTAL RESULTS

A considerable time has passed since Richardson in 1908 concluded from his measurements with electron beams that the velocity distribution of thermionically emitted electrons is a half-Maxwellian distribution dependent upon cathode temperature and that therefore the velocity distribution of the electrons immediately outside of an emission plane looks in the velocity space as pictured in Figure 1.

v_{ℓ} is the velocity component perpendicular to the emission plane, v_t is a transverse component. The surfaces of constant population density are half-spheres, centered at the velocity zero. The graph must be imagined to be rotationally symmetrical about the v_{ℓ} - axis.

Now, it is obvious that such a distribution is no thermodynamic equilibrium distribution, for this is isotropic with respect to the average velocity. Therefore, there must occur a relaxation process to establish equilibrium. This process and some of the physical consequences thereof are the theme of my talk. Before we go deeper into the mechanism of this process, however, let us first consider the simple experimental apparatus used by us to determine the distribution of the longitudinal velocity component. This is shown in Figure 2.

The electrons are emitted by a cathode and accelerated by the anode potential. The current is controlled by a grid G. After having drifted through the anode space the electrons enter a retarding field energy analyzer. There the current density j reaching the electrode F is measured as a function of the retarding field potential U_F .

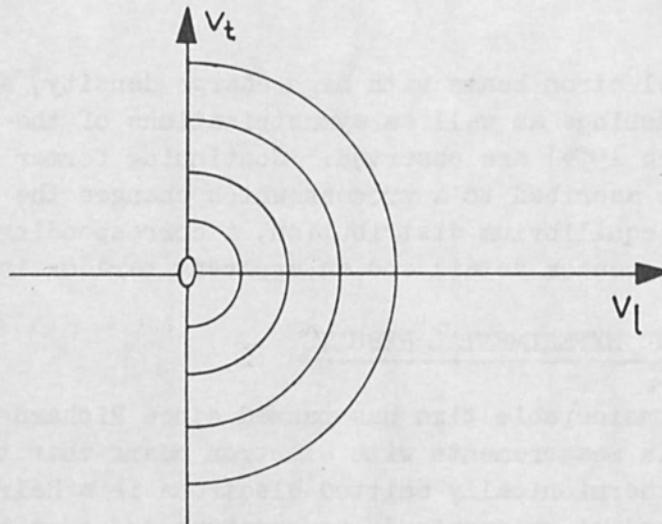


Figure 1.

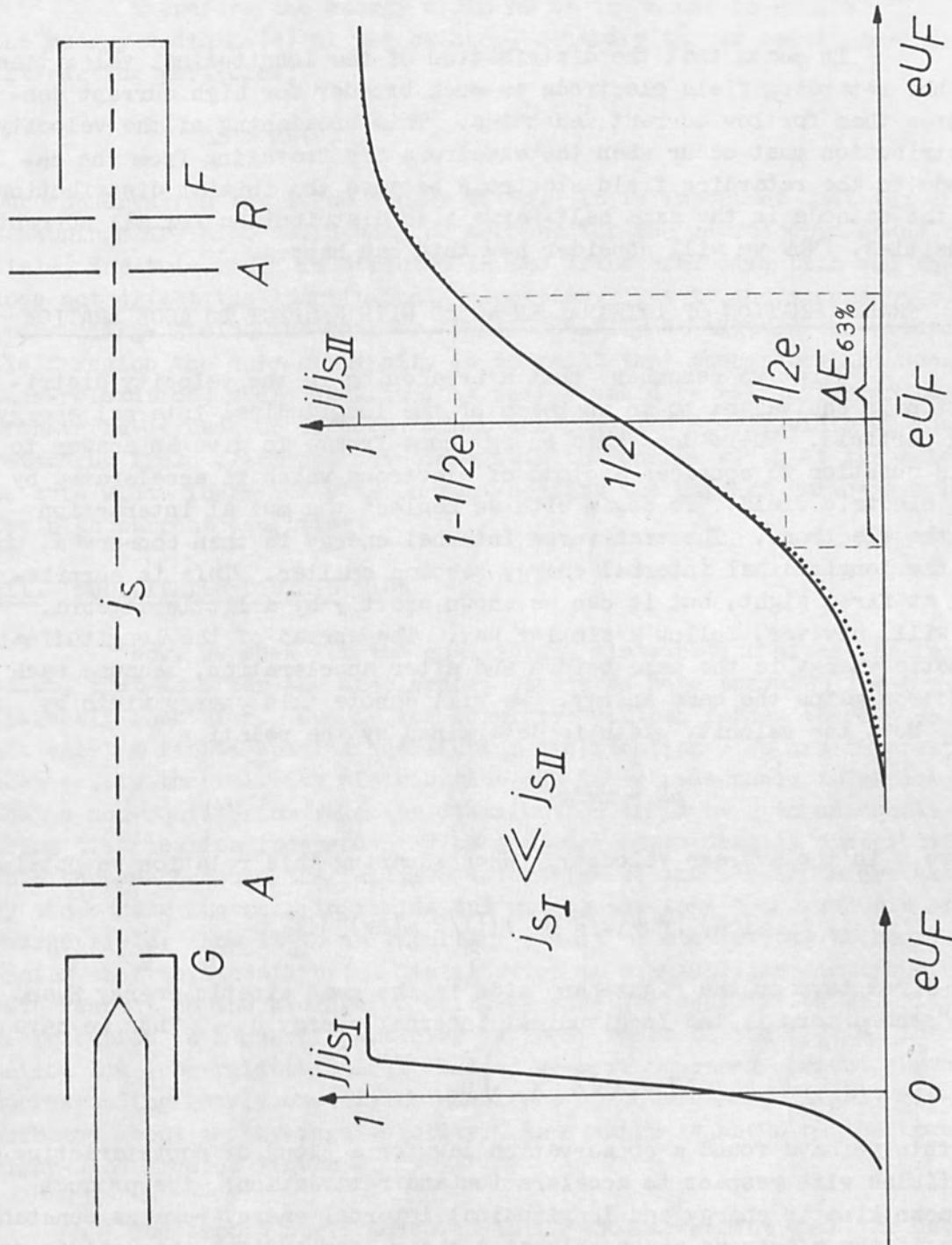


Figure 2.

On the lower half of the figure you see the experimental results for two different beam current densities j_{sI} and j_{sII} . For the small current density j_{sI} one obtains the left curve; for the high current density j_{sII} one obtains the smooth curve shown to the right. It can be seen immediately that the current curve for the high current density is much broader and more symmetrical than the curve for the low current density. What does this mean for the velocity distribution of the electrons?

It means that the distribution of the longitudinal velocities at the retarding field electrode is much broader for high current densities than for low current densities. This broadening of the velocity distribution must occur when the electrons are traveling from the cathode to the retarding field electrode because the initial distribution at the cathode is the same half-Maxwellian distribution for all current densities. Now we will consider how this can happen.

II. TRANSFORMATION OF INTERNAL ENERGIES WITH RESPECT TO ACCELERATION

First we remember that a broadening in the velocity distribution is equivalent to an increase of the longitudinal internal energy per particle. Where does this energy come from? To give an answer to this question we consider a cloud of electrons which is accelerated by the electric field. To begin with we neglect the mutual interaction of the electrons. The transverse internal energy is then conserved, whereas the longitudinal internal energy becomes smaller. This is surprising at first sight, but it can be shown exactly by a little algebra. We will, however, follow a simpler way. The spread of the longitudinal kinetic energy is the same before and after acceleration, because each electron gains the same energy. We will denote this energy width by ΔE . Now, the velocity width is determined by the relation

$$\Delta E = mv\Delta v,$$

where v is the average velocity. When squaring this relation we obtain

$$(\Delta E)^2 = 4(1/2 mv^2) [(1/2 m(\Delta v)^2)].$$

The first term on the right-hand side is the mean kinetic energy E and the second term is the longitudinal internal energy u_l . Thus we have

$$\Delta E = 2 \sqrt{E} \times \sqrt{u_l}.$$

By this we have found a conservation law for a cloud of noninteracting particles with respect to acceleration and retardation: the product of mean kinetic energy and longitudinal internal energy remains constant. Thus if the electrons are accelerated their longitudinal internal energy

decreases. As their transverse internal energy remains constant we have a deviation from thermodynamic equipartition. This deviation increases with acceleration potential. If the electron density is sufficiently high the mutual interaction of the electrons is no longer negligible and there occurs a relaxation process in which transverse internal energy is transferred to the longitudinal degree of freedom. This relaxation occurs simultaneously with acceleration.

Therefore the energy width ΔE at the anode is greater than the energy width $\Delta E(o)$ at the cathode, contrary to our result for non-interacting particles:

$$\Delta E > \Delta E(o)$$

Our conservation law is no longer valid. It is important that the above reasonings are also valid when we account for the macroscopic space charge field, for this field is directed in the transverse direction and thus does not affect the longitudinal velocities. Due to the macroscopic space charge field, the beam broadens spatially so that in the retarding field region the current density is so small that there does not occur back-relaxation, where longitudinal energy would be retransferred to the transverse degrees of freedom. Thus the observed energy width at the retarding field electrode coincides with the energy width at the anode. As this width increases with current density the current curves are broad for high current densities.

III. EQUILIBRIUM DISTRIBUTION

Now, we seek for the equilibrium distribution after the relaxation, that is after the accelerated electrons have traveled for a sufficiently long time. Due to the repulsive Coulomb forces there does not exist a finite spatial equilibrium distribution. We are interested, however, in the velocity distribution and it is reasonable to expect that a non-equilibrium velocity distribution turns over to an equilibrium distribution independently of the macroscopic density distribution. This happens because the relative velocities of the electrons are altered by the strong microscopic fields and not by the smooth macroscopic space charge field. Now it is an important result of statistical thermodynamics that this equilibrium distribution is a Maxwellian distribution with respect to the average velocity because the equilibrium velocity distribution in a gas of particles is independent of the special interaction law (non-relativistic). Indeed we have measured current curves corresponding nearly exactly to electrons which are Maxwellianly distributed about the average velocity. An example is shown on the lower right-hand side of Figure 2.

The smooth curve measured is in good agreement with the dashed

one calculated for a drifting full Maxwellian distribution of electrons with an adjusted temperature.

We conclude therefore, that a longitudinal temperature has been established. The question now is, whether the relaxation has already come to an end or whether it is only interrupted by the retarding field electrode. To decide what is correct we have to estimate the greatest longitudinal temperature which can be reached.

IV. MAXIMUM LONGITUDINAL TEMPERATURE AND MAXIMUM ENERGY WIDTH ΔE

In Figure 3 we consider the relaxation in a drifting volume element in the beam when there has not occurred any relaxation in the accelerating region. With the aid of the collision-less Boltzmann equation, it can be shown that the velocity distribution of the accelerated electrons is as pictured on the left-hand side. The surfaces of constant population density are spherical segments. The symbols w_\parallel and w_\perp denote the velocity components with respect to the average velocity. Now, as time goes on, the distribution will finally assume the isotropic form on the right-hand side. For times between, the surfaces of constant population density will look like the graph in the middle of the figure. When a longitudinal temperature has already been established these surfaces are ellipsoids. This is the case in our experiments as will be seen later. We now return to our problem of estimating the maximal longitudinal temperature. We neglect the spatial beam broadening due to the macroscopic space charge field. To avoid mathematical formalism we make use of the following intuitive conception about the moving electrons after acceleration. When electrons are leaving the anode, we lock them up in separate boxes moving with the average velocity. The total energy of such a box then remains constant and, since the boxes are closed systems, the momentum must also be conserved. Thus both the kinetic energy of the center of mass of the electrons within a box and the internal energy of the electrons remain constant. Now, the internal energy before relaxation is $2 \times 1/2kT$ because the longitudinal degree of freedom contains no internal energy. The isotropic temperature after relaxation may be denoted by θ and thus we have for the final internal energy per particle $3/2k\theta$. Because of conservation of internal energy we have

$$kT = 3/2k\theta$$

and therefore $\theta = 2/3 T$.

The same final temperature θ is obtained using hydrodynamic formulas which take into consideration the fact that a hypothetical volume element is not a closed system, but has to perform work against thermodynamic pressure forces. This work is delivered by the kinetic energy of the volume, however, and thus does not effect the internal energy. One obtains only

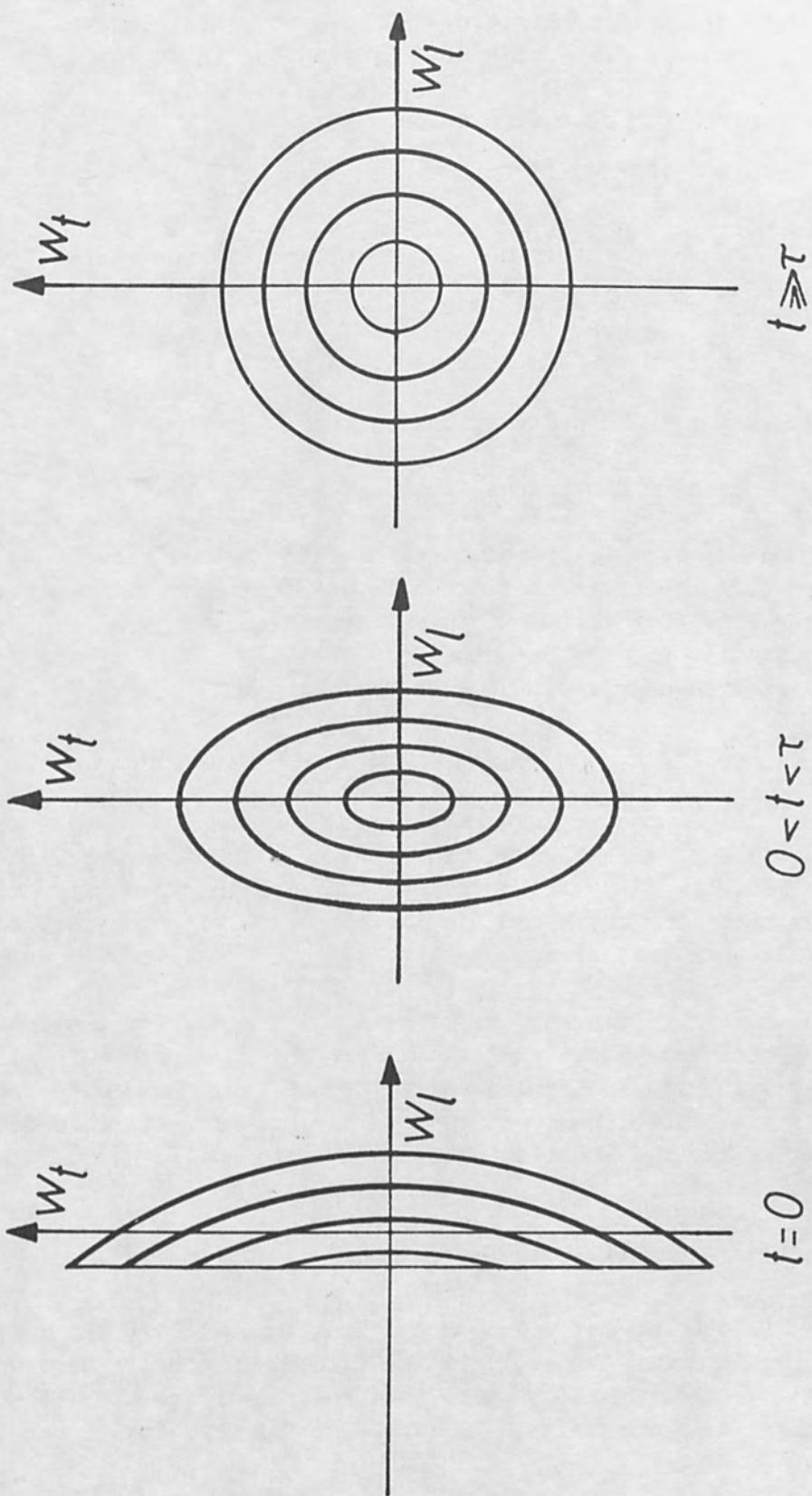


Figure 3.

a decrease of the average velocity with increasing distance from the anode and this decrease is negligible with regard to the measured energy width ΔE . (The hydrodynamic analysis is carried out in the appendix.)

Above we have derived for ΔE the relation

$$\Delta E = 2 \sqrt{eU} \times \sqrt{u_\ell},$$

where e is the electronic charge and U the accelerating potential. When relaxation has completely finished u_ℓ assumes its greatest value

$$(u_\ell)_m = 1/2 k\theta = 1/3 kT$$

so that ΔE is limited in both directions:

$$\Delta E(o) \leq \Delta E \leq 2 \sqrt{eU} \times \sqrt{1/3 kT}.$$

We have obtained experimentally about 10% of this maximal value. As these measured current curves correspond to drifting Maxwellian distributions, we conclude that there exists a longitudinal temperature long before the relaxation process has finished. We will now treat the dependence of the measured energy width ΔE on the current density.

V. RELAXATION LAW

For this purpose we must investigate the temporal development of the energy width of electrons within a volume element moving with the local average velocity $v(x)$. It turns out that for this quantity there exists a relaxation time which can then be connected with the particle density and thus with the current density.

Because one is wont to assume a relaxation law for internal energies, it is somewhat surprising that there exists a relaxation time for the energy width ΔE , and consequently for the width Δv of the longitudinal velocity distribution, that is true for much longer times than a relaxation time for the longitudinal internal energy $1/2 m(\Delta v)^2$. We have to assume that the relaxation occurs according to

$$\frac{d}{dt} (\Delta v(x) - \Delta v_0(x, T)) = \frac{1}{\tau(x)} (\Delta v(x) - \Delta v_\infty(x, T)).$$

Here $\Delta v_\infty(x, T)$ is the maximal width of the longitudinal velocity distribution after relaxation. $\Delta v_0(x, T)$ is the velocity width if there would be no relaxation. The relaxation time τ depends on the particle density and thus on the space coordinate. The differential operator $\frac{d}{dt}$ can be

expressed by

$$\frac{dx}{dt} \cdot \frac{d}{dx} = v(x) \frac{d}{dx}$$

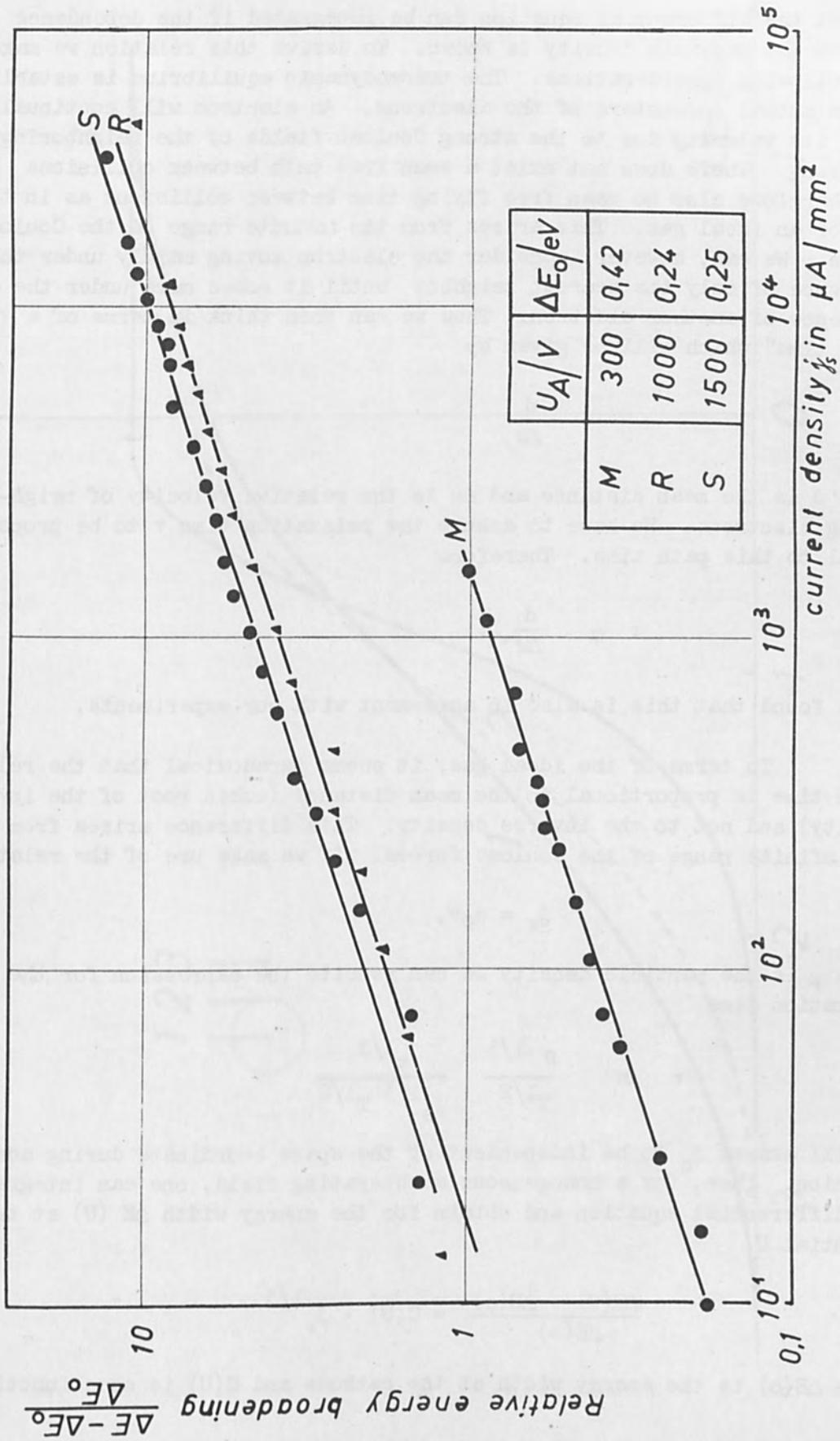


Figure 4.

so that the differential equation can be integrated if the dependence of τ on the particle density is known. To derive this relation we make the following considerations. The thermodynamic equilibrium is established by the mutual encounters of the electrons. An electron will continually alter its velocity due to the strong Coulomb fields of the neighboring electrons. There does not exist a mean free path between collisions and therefore also no mean free flying time between collisions as in the case of an ideal gas. This arises from the infinite range of the Coulomb forces. We can, however, consider the electron moving mainly under the influence of only its nearest neighbor until it comes more under the influence of another electron. Thus we can then think in terms of a "path time" which will be given by

$$\frac{d}{\Delta w}$$

where d is the mean distance and Δw is the relative velocity of neighboring electrons. We have to assume the relaxation time τ to be proportional to this path time. Therefore

$$\tau \propto \frac{d}{\Delta w} .$$

It is found that this is also in agreement with our experiments.

In terms of the ideal gas, it seems paradoxical that the relaxation time is proportional to the mean distance (cubic root of the inverse density) and not to the inverse density. This difference arises from the infinite range of the Coulomb forces. If we make use of the relation

$$j_s = e\rho v,$$

where ρ is the particle density we can rewrite the expression for the relaxation time

$$\tau \propto \frac{\rho^{-1/3}}{T^{1/2}} \frac{v^{1/3}}{j_s^{1/3} T^{1/2}}$$

We will assume j_s to be independent of the space coordinate during acceleration. Then, for a homogeneous accelerating field, one can integrate the differential equation and obtain for the energy width $\Delta E(U)$ at the potential U

$$\frac{\Delta E(U) - \Delta E(o)}{\Delta E(o)} = C(U) \cdot j_s^{1/3}$$

where $\Delta E(o)$ is the energy width at the cathode and $C(U)$ is some function

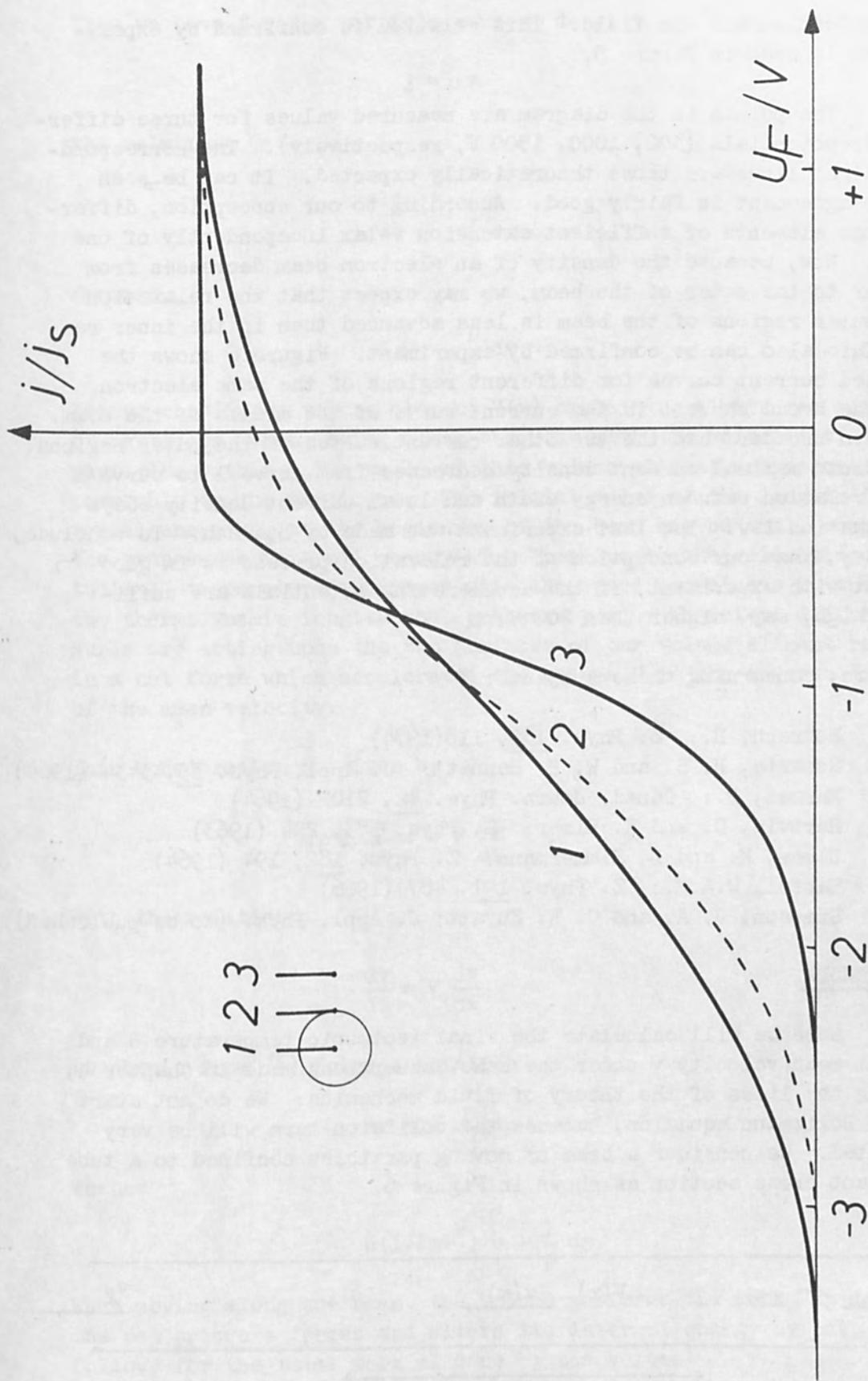


Figure 5.

of the potential and the field. This relation is confirmed by experiments, as is seen in Figure 5.

The points in the diagram are measured values for three different anode potentials (300, 1000, 1500 V, respectively). The corresponding straight lines are those theoretically expected. It can be seen that the agreement is fairly good. According to our conception, different volume elements of sufficient extension relax independently of one another. Now, because the density of an electron beam decreases from the inner to the outer of the beam, we may expect that the relaxation in the outer regions of the beam is less advanced than in the inner region. This also can be confirmed by experiment. Figure 6 shows the normalized current curves for different regions of the same electron beam. The broad curve 1 is the current curve of the middle of the beam. It is much broader than the two other current curves of the outer regions. The absolute maximal current density decreases from curve 1 to curve 3 and the relation between energy width and local current density obeys our relaxation law. The last experiment was made by Boersch. To conclude, we may say, that our conception of the relaxation process is in good agreement with experiment, if the accelerating potentials are sufficiently high, say, higher than 200 V.

Literature, concerning these subjects:

- Boersch, H.: Z. Phys. 139, 115(1954)
 Roberts, H. S. and W. H. Bennett: J. Appl. Phys. 35, 3434 (1964)
 Marmet, P.: Canad. Journ. Phys. 42, 2102 (1964)
 Hartwig, D. and K. Ulmer: Z. Phys. 173, 294 (1963)
 Ulmer, K. and B. Zimmermann: Z. Phys. 182, 194 (1964)
 Hartel, W.A.M.: Z. Phys. 191, 487 (1966)
 Simpson, J. A. and C. E. Kuyatt: J. Appl. Phys. (to be published)

VI. APPENDIX

Here we will calculate the final isotropic temperature θ and the final mean velocity v under the same assumptions made in Chapter 4, but along the lines of the theory of fluid mechanics. We do not start with the Boltzmann equation, because the collision term will be very complicated. We consider a beam of moving particles confined to a tube of constant cross section as shown in Figure 6.

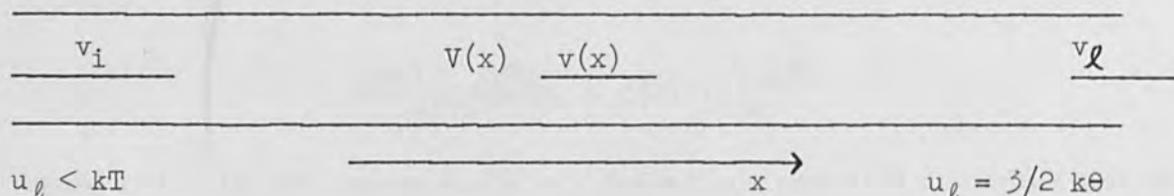


Fig. 6

Thus we have for the particle current density

$$j = \rho v \quad (1)$$

the relation

$$\frac{dj}{dx} = 0, \quad (2)$$

whence

$$d\rho = -\frac{\rho}{v} dv. \quad (3)$$

Let us consider a volume element $V(x)$ containing always the same number of particles, when moving along the beam with the velocity $v(x)$ of the particles within. The walls of this volume element are assumed to be permeable to the particles. The initial value of the longitudinal internal energy u_{\parallel} at the beginning of the beam may be small compared with the transverse internal energy kT . Then due to the relaxation, the longitudinal temperature increases with the space coordinate x , and so does the thermodynamic longitudinal pressure $p(x)$. Therefore different pressures are acting upon the two surfaces of our volume element resulting in a net force which accelerates the volume V in the inverse direction of the mean velocity.

This acceleration is given by

$$m\rho V \frac{dv}{dt} = -V \frac{dp}{dx}. \quad (4)$$

With the relation

$$\frac{dv}{dt} = v \frac{dv}{dx} \quad (5)$$

we obtain from (4) Euler's equation

$$dv = -\frac{1}{mj} dp \quad (6)$$

whence

$$d(1/2mv^2) = -\frac{1}{\rho} dp. \quad (7)$$

When moving along the beam, the volume performs the work Vdp against the net pressure forces and alters its internal energy by pdV , whence follows for the total work dA done by the volume

$$dA = d(pV). \quad (8)$$

Thus, the work done per particle is

$$da = d(p/\rho). \quad (9)$$

For sufficiently high mean velocity v heat conduction is negligible, so that the principle of energy conservation demands

$$d(1/2 mv^2) + du + da = 0, \quad (10)$$

where u is the total internal energy per particle.

Inserting (7) in (10) we obtain

$$du = \frac{p}{\rho^2} d\rho, \quad (11)$$

which can be rewritten with the aid of equation (3) as

$$du = -\frac{p}{j} dv. \quad (12)$$

Making use of Euler's equation (6) we obtain

$$du = \frac{1}{2mj^2} dp^2. \quad (13)$$

We can integrate this relation and obtain

$$u_f - u_i = \frac{1}{2mj^2} (p_f^2 - p_i^2). \quad (14)$$

The initial internal energy u_i is given by

$$u_i = kT, \quad (15)$$

whereas the final internal energy is

$$u_f = \frac{3}{2} k\theta. \quad (16)$$

We may assume that the initial longitudinal pressure p_i is negligible compared with the final isotropic pressure p_f .

Thus we have

$$\frac{3}{2} k\theta - kT = \frac{1}{2mv_f^2} (p_f/\rho_f)^2. \quad (17)$$

Together with the integrated Euler's equation (6)

$$v_f - v_i = -\frac{1}{mv_f} (p_f/\rho_f), \quad (18)$$

and the equation of state for the electrons we could calculate the interesting quantities v_f and θ . When we assume, however, the equation of state for an ideal gas

$$p_f/\rho_f = k\theta, \quad (19)$$

we obtain the approximate solutions

$$\theta \approx \frac{2}{3} T \quad (20a)$$

$$\frac{1}{2} m v_f^2 \approx \frac{1}{2} m v_i^2 - k\theta. \quad (20b)$$

The first result was also obtained by our intuitive conception in the main text. The second result is different from our simple considerations and states the fact that the hypothetical volume elements which have to be considered are not closed systems. The result (20b), however, does not modify the conclusions in the main text, because the relative change of the mean kinetic energy is negligible, if this energy is sufficiently high compared with internal energy.

and the question of error for the elements we could calculate as follows
during quantities ϵ and δ when we assume, however, the equation of

$$-d(\rho/\rho_0) = \dots$$

of latitude at longitude λ and latitude ϕ is given by the following
approximate solution

$$(10) \quad \dots$$

where ρ_0 is the mean density of the earth and ρ is the density at the
point in question

$$(11) \quad \dots$$

The first result was also obtained by the following consideration
of the fact that the geopotential is constant on the surface of the
earth and that the geopotential value elements which have
to be considered are not closed systems. The results (10) and (11)
are valid in the case where the relative error
of the mean kinetic energy is negligible in this respect as compared
with the kinetic energy.

$$(12) \quad \dots$$

It is also possible to calculate the relative error of the
geopotential value elements

$$(13) \quad \dots$$

where ρ_0 is the mean density of the earth and ρ is the density at the
point in question

$$(14) \quad \dots$$

where ρ_0 is the mean density of the earth and ρ is the density at the
point in question

$$(15) \quad \dots$$

where ρ_0 is the mean density of the earth and ρ is the density at the
point in question

$$(16) \quad \dots$$

where ρ_0 is the mean density of the earth and ρ is the density at the
point in question

$$(17) \quad \dots$$

TECHNIQUES FOR THE STUDY OF
SELF-FOCUSING ELECTRON STREAMS*

by

S. E. Graybill and S. V. Nablo
Ion Physics Corporation
Burlington, Massachusetts

ABSTRACT

A brief review is presented of the requirements of a relativistic electron stream such that self-magnetic (Bennett) confinement will occur. A description is given of a 4 million volt pulsed electron field emission accelerator capable of generating kiloampere electron streams under well-defined conditions. The several techniques which have been used for the investigation of these high current density electron beams are reviewed. The thin film dosimetry and beam calorimetry techniques which have been developed to provide high resolution mapping of a 17,000 ampere, 20 nanosecond stream are described and experimental data characteristic of accelerator performance are presented.

I. INTRODUCTION

Due to the extreme energy fluxes associated with self-magnetically focused relativistic electron streams drifting in a field-free region, unusual mapping techniques have been necessary to permit the experimental study of the phenomenon. The theory of such a partially neutralized stream with no applied longitudinal field was first laid down by Bennett in 1934 and has since been discussed by Bennett⁽¹⁾ and by Lawson⁽²⁾ for rather idealized conditions. It has only recently been possible to generate such streams under sufficiently well-controlled conditions in the laboratory so that detailed parametric studies of the effect can be made under true drifting conditions⁽³⁾ (as compared with those conditions in the magnetically confined arc).

For the case of a low density stream, Bennett has calculated a critical current I_c , that must be exceeded in order to achieve focusing. This current is dependent on the electron transverse energy, and on the degree of charge neutralization. It is also shown that there is a permanent loss of some components and that the stream is not subject to the "kink" instability. Based upon considerations of the electric and magnetic forces on a partially neutralized stream, Bennett has

*This work was partially supported under Contract AF08(635)-5029, Detachment 4, RTD, Eglin Air Force Base, Florida.

derived the following expression for the critical current:

$$\frac{I_c^2}{c^2} = 2 N E_t + \left[\frac{q}{\ell} \right]^2 . \quad (1)$$

The relation of equation (1) defines the critical current, in which N is the total number of particles per unit length of the beam, E_t is the average transverse kinetic energy and q/ℓ is the net charge per unit length. For the conditions to be described here in which we have studied a 2 Mev electron beam passing through gas at low pressures, the following expression is valid:

$$\frac{I_c^2}{c^2} = 2 N_e E_t + [N_e e (1 - f)]^2 . \quad (2)$$

Here we assume that the electrons liberated in ionization are immediately accelerated out of the beam by the radial E field and that the residual ion energies are negligible. N_e is the number of primary beam electrons per unit length and f is the ion:electron density ratio. Because only the radial electrical field due to the electron beam itself is present and because of the difference in electron and ion mobilities, $f \geq 1$ for all gas densities, which guarantees an observation of pure magnetic self-focusing. Equation (2) may be rewritten to express these conditions:

$$1 - \left[\frac{1}{\beta} (1 - f) \right]^2 \geq \frac{2 c E_t}{\beta e I} . \quad (3)$$

Here we again assume complete escape of secondary electrons so that $N_e = I/e\beta c$. Equation (3) then describes the relationship between f , E_t , and I essential for self-magnetic focusing of the stream.

Lawson has expressed the generalized perveance (K) of such a partially neutralized beam in the following form:

$$K = \frac{2 N_e^2 (1 - \beta^2)^{1/2}}{m_0 c^2 \beta^2} (1 - \beta^2 - f) . \quad (4)$$

One notes that as expected, this quantity is proportional to $I/V^{3/2}$ at low energies and must be negative in order for the beam to be self-focusing. The criterion of $\beta^2 + f \geq 1$ which makes K negative is just that which makes $\bar{v} \times \bar{B} > E$. Under these conditions the electrons will cross the beam axis at pinch points separated by $\lambda/2$ where b is the beam radius, where

$$\lambda = 2\pi b / (-K)^{1/2} . \quad (5)$$

This sinusoidal motion may then be expected to be characteristic of electrons in a drifting beam of constant radius and uniform current density for which one assumes no initial transverse energy and a small angle of convergence into the pinch points.

For the low density stream at low gas pressures, the ion:electron and ion:neutral ratios resulting from primary ionization may be written as:

$$f = n_i/n_e = \beta c \sigma \tau n_a \quad (6)$$

and

$$n_i/n_a = \phi \sigma \tau \quad (7)$$

where n_i , n_e and n_a are the volume densities, τ the pulse duration, σ the ionization cross section and ϕ the primary electron flux. As before, the expression of equation (6) is valid when there is complete secondary electron escape from the beam with n_i/n_a small.

II. APPARATUS

Figure 1 illustrates the basic power supply used in these studies. It consists of a 4 million volt pressure insulated Van de Graaff generator which is used to charge a 3 meter long, 280 pf, gas insulated, coaxial capacitor. This coaxial geometry forms a simple transmission line which, when end switched, produces a 20 nsec pulse. At the opposite end of the pressure vessel is the field emission diode which is shown in detail in Figure 2. This tube geometry is based upon an earlier AWRE design⁽⁴⁾ and similar configurations have since been used in pulsed accelerators at much higher tube voltages.⁽⁵⁾

As shown in Figure 2, the coaxial capacitor is switched to the termination and cathode of the field emission tube by means of a triggered spark gap which employs the insulating gas as the switching medium. The field emission cathode is housed in a capacitively graded vacuum envelope. The electrons pass through a thin titanium window into the drift tube region with only a few percent energy degradation; e.g., $\Delta E = 60$ kev at 2 Mev. The field emission tube is maintained at 10^{-5} torr while the drift region pressure can be varied from 10^{-4} torr to one atmosphere using a controlled leak from a reservoir of the desired gas species (air, in the studies to be described). The integrated system performance has been found to be sufficiently reproducible so that through precision control of the capacitor charging voltage, the variation in energy release per pulse can be held to less than 3%.

Under typical operating conditions the system is charged to

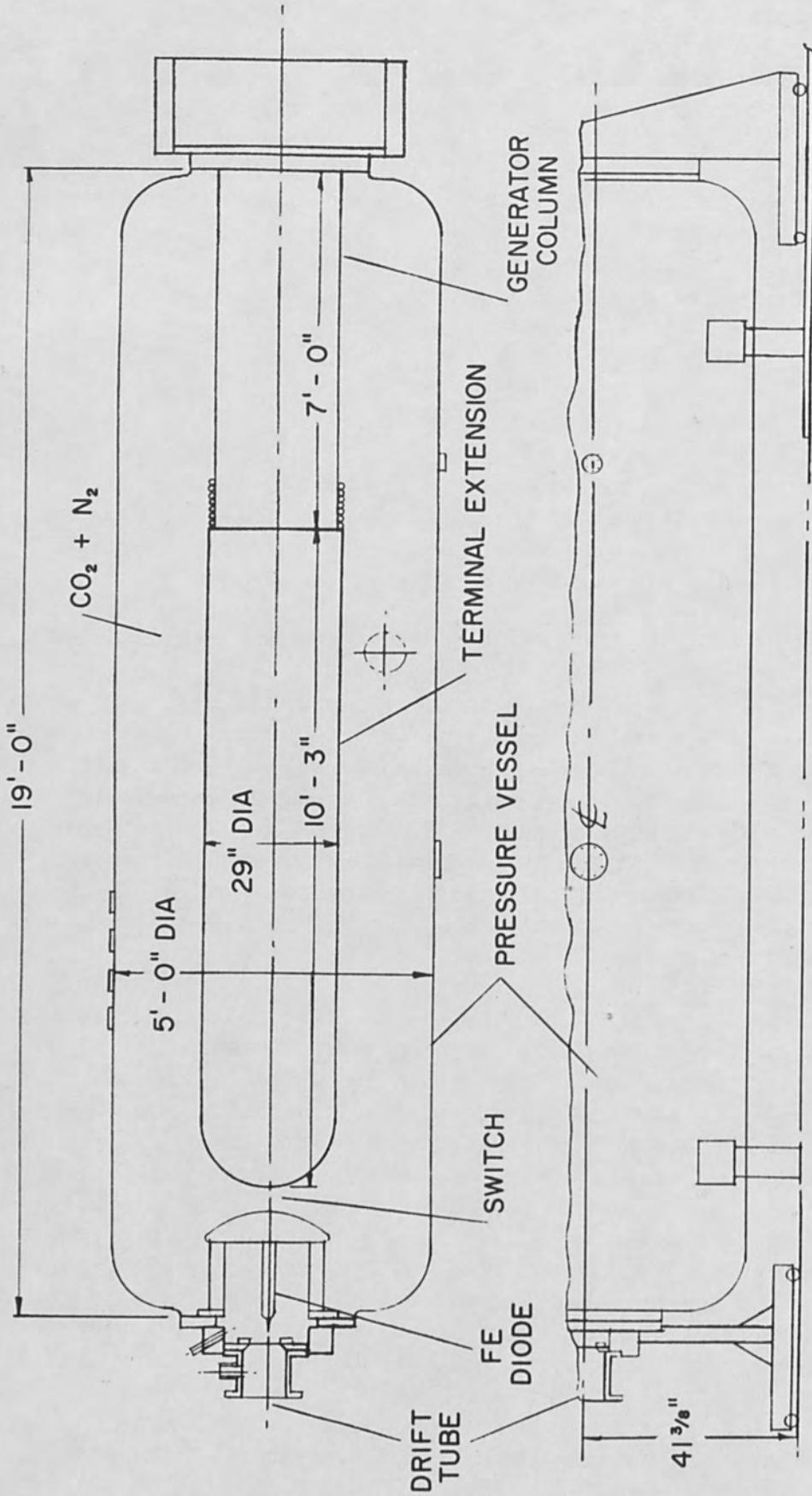


Figure 1. 4 Megavolt, 2 Kilojoule Field Emission System.

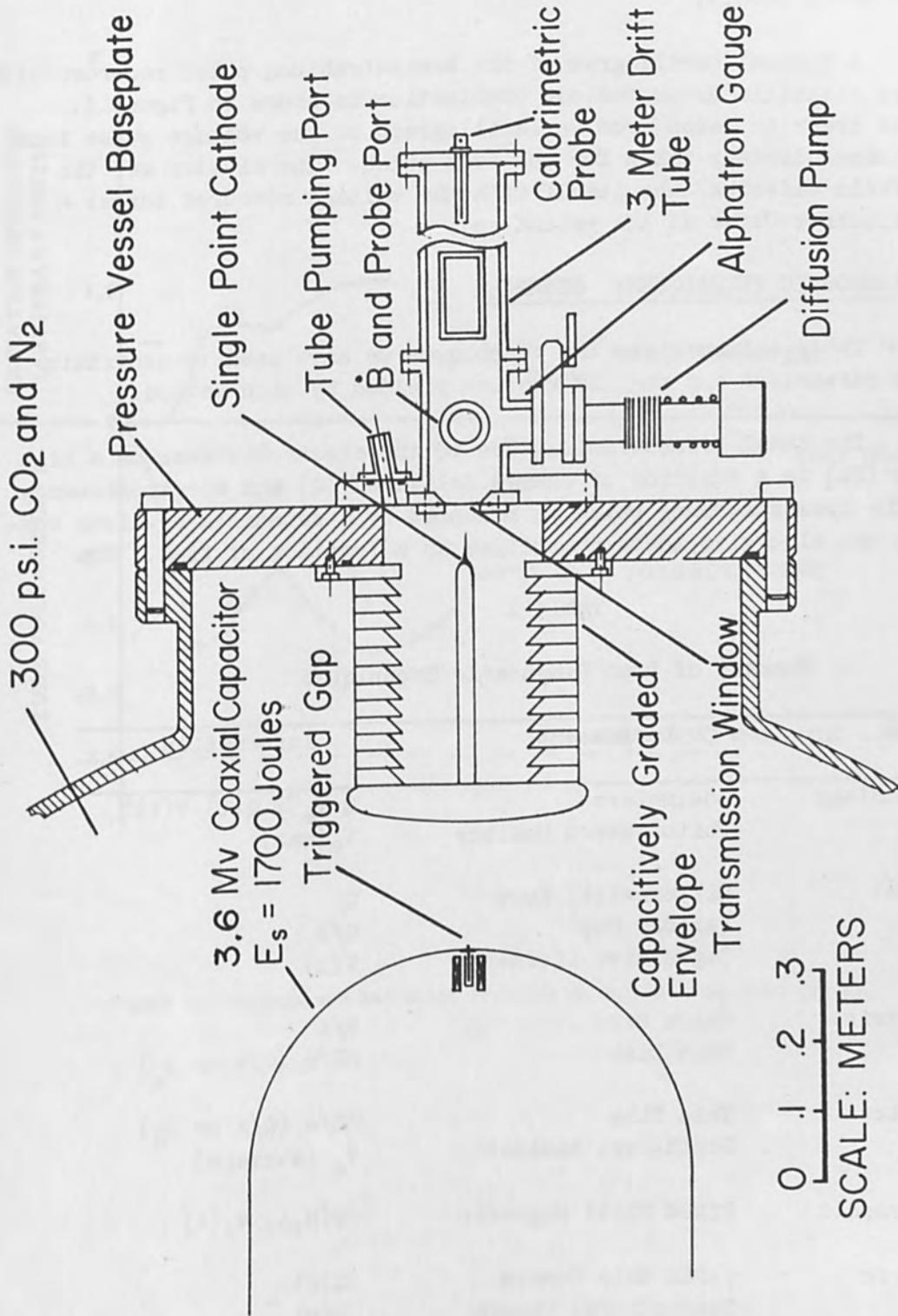


Figure 2. The Electron Field Emission Diode - Drift Tube Assembly Mounted on the 3.6 Megavolt Generator System.

3.4 Mv (V_0) so that for the selected impedance mismatch a maximum tube potential of 2.3 Mv (V_t) results. The maximum beam current is found to be 1.7×10^4 amperes with a pulse width of 20 nsec (fwhm). It is obvious that most conventional direct probing techniques are no longer applicable at these power levels.

A typical oscillogram of the bremsstrahlung pulse recorded with a plastic scintillator-photodiode combination is shown in Figure 3. The lower trace is taken from an oscillograph of the voltage pulse from a capacitance divider taken for the same pulse. The divider was the graded field emission tube itself with the voltage measured across a thin dielectric wafer at the ground end.

III. DIAGNOSTIC TECHNIQUES: GENERAL

Table 1 summarizes the techniques we have used to ascertain the beam parameters and the information yielded by each method.

The total bremsstrahlung due to the electron stream on a high Z target-(Ta) is a function of charge delivered (Q) and electron energy (V_e). The bremsstrahlung pulse as measured by a scintillator-diode combination can also give these quantities as a function of time. The

Table 1

Summary of Beam Diagnostic Techniques

Method	Instrument	Information
Bremsstrahlung	Dosimeters	$Q V_e^n$; $Q(t) V(t)^n$
	Photoneutron Monitor	$V_e(\text{max})$
Electrical	Differential Loop	Q
	Faraday Cup	Q/A
	Capacitive Divider	V(t)
Calorimetric	Thick Disc	E/A
	Thin Disc	$\Delta E/m$ (Q/A or j_e)
Dosimetric	Thin Film	$\Delta E/m$ (Q/A or j_e)
	Depth-Dose Laminate	V_e (average)
Spectrographic	Fixed Field Magnetic	$N(H\rho)$, $V_e(t)$
Photometric	γ -Pin Hole Camera	Q(r)
	Conventional Camera	E(r)

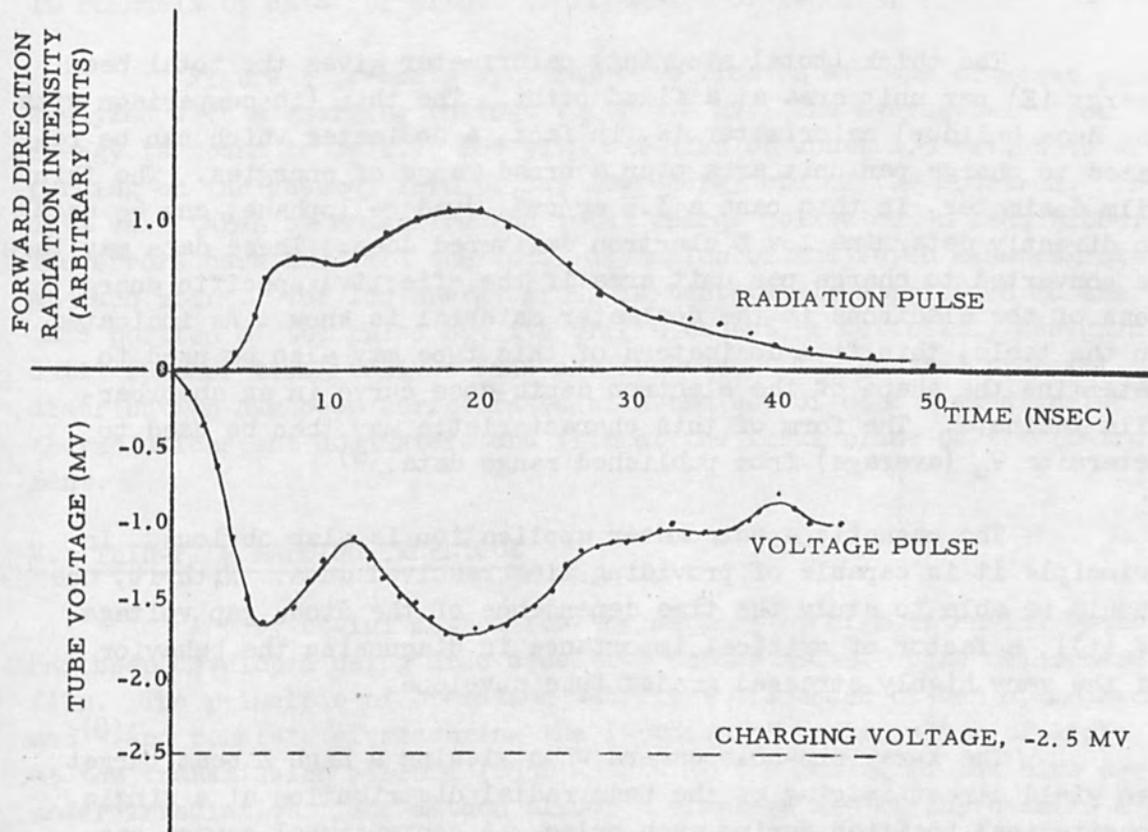


Figure 3. Voltage and Radiation Profiles on Pulsed X-Ray Tube (FX 1).

bremsstrahlung is very useful for ascertaining the maximum electron energy. This is accomplished by measuring the thresholds for the Be (γ, n) and D(γ, n) reactions, which are 1.67 and 2.23 Mev respectively, and relating these to the charging voltage of the energy storage system; i.e., an indirect calibration of V_t .

The magnetic flux loop, Faraday cup, and voltage divider yield the quantities shown in a straightforward manner.

The thick (total stopping) calorimeter gives the total beam energy (E) per unit area at a fixed point. The thin (in comparison with the dose buildup) calorimeter is, in fact, a dosimeter which can be related to charge per unit area over a broad range of energies. The thin film dosimeter, in this case a 3.5 mg/cm², dyed cellophane, can be used to directly determine low Z electron delivered dose. These data may then be converted to charge per unit area if the effective specific energy loss of the electrons in the dosimeter material is known. As indicated in the table, thin film dosimeters of this type may also be used to determine the shape of the electron depth-dose curve in an absorber-film laminate. The form of this characteristic may then be used to determine V_e (average) from published range data.⁽⁶⁾

The magnetic spectrometer application is also obvious. In principle it is capable of providing time resolved data. With it, one should be able to study the time dependence of the diode gap voltage ($V_e(t)$), a factor of critical importance in diagnosing the behavior of the very highly stressed graded tube envelope.

The X-ray pin-hole camera when viewing a high Z beam target can yield direct imaging of the beam radial distribution at a single longitudinal position during each pulse. A conventional camera can produce spectacular, but not highly useful, pictures of the recombination light after a pulse. Fast image converter camera techniques can provide useful information on beam behavior particularly during the plasma generation-neutralization period, when very fast changes in the beam envelope can occur.

IV. MAGNETIC SPECTROMETER

The momentum spectrum of the electron beam drifting in vacuum external to the accelerator has been measured using a 180° fixed field magnetic spectrometer of the type used so successfully by Kinsey, Groshev, Motz, and others as a Compton electron spectrometer to study γ spectra. The instrument shown here was designed and made available by W. R. Van Antwerp⁽⁷⁾ of the Edgewood Arsenal. The entrance collimators were in line with the beam axis with the drift tube at a sufficiently low pressure (10^{-4} torr) that the beam diverged from normal space-charge repulsion.

The entrance current density was thus reduced to $0.1 - 1 \text{ A/cm}^2$ at the 1 meter position from an entrance current density of $\sim 10^3 \text{ A/cm}^2$ at 2 Mev. The analyzed beam was normally monitored by ten charge collectors positioned along the focal plane complemented by film and LiF dosimetry. The useful energy range of the configuration was $0.100 - 3 \text{ Mev}$. A view of the one kilogauss instrument is shown in Figure 4 mounted on the 10 cm diameter evacuated drift tube. The connector header shown provided 10 channels of data for either oscillograph or recorder readout.

Figure 5 presents spectral data from an average of seven pulses analyzed with a charging voltage V_0 of 3.4 Mv. The average electron energy is found to be 1.75 Mev with a median of about 1.9 Mev, both depending on the assumed form of the low energy end of the spectrum. These data were taken by measuring the total charge collected on each probe. The errors bars indicate the total deviation of the seven measurements at each point. The low energy group is tentatively attributed to the late portion of the pulse but its origin cannot be definitely assigned until further time resolved data are recorded. The general spectral distribution has been corroborated with the use of both 1 mm diameter thermoluminescent dosimeters and film at the focal plane of the instrument.

V. THIN-FILM MAPPING TECHNIQUE

A very useful method for the mapping of high intensity beams has been developed using dose sensitive DuPont MSC-300 blue cellophane film. The principle of operation was first discussed by Henley and Richman⁽⁸⁾ and consists of measuring the increase of transmission of light at the transmission minimum (6550 Å) due to the fading of the blue dye under irradiation. The method allows a complete map of the beam in a single pulse due to the negligible energy loss of a relativistic electron in traversing the 3.5 mg/cm^2 film; e.g., $\sim 6 \text{ kev}$ at 2 Mev.

One of the limitations of the technique results from inadequate data concerning rate and pressure effects on dye response. Two calibration curves recorded in air for the film are shown in Figure 6 and demonstrate an apparent rate sensitivity of the film, a problem which had not appeared to exist at levels up to at least 10^6 r/second .

The upper calibration curve was recorded with the use of the AFCRL Co⁶⁰ facility at a fixed dose rate of 120 rads/second. The lower curve was recorded on the pulsed electron accelerator with single pulse irradiations ($\tau = 20 \times 10^{-9} \text{ secs}$) so that the rates varied from $10^{13} - 10^{15} \text{ rads per second}$. The delivered dose was calculated from the recorded temperature rise in a thin foil immediately behind the dosimeter. Readout of the film was normally accomplished with a Beckman Model DU Spectrophotometer. One comparative curve is shown in Figure 6, for the elec-

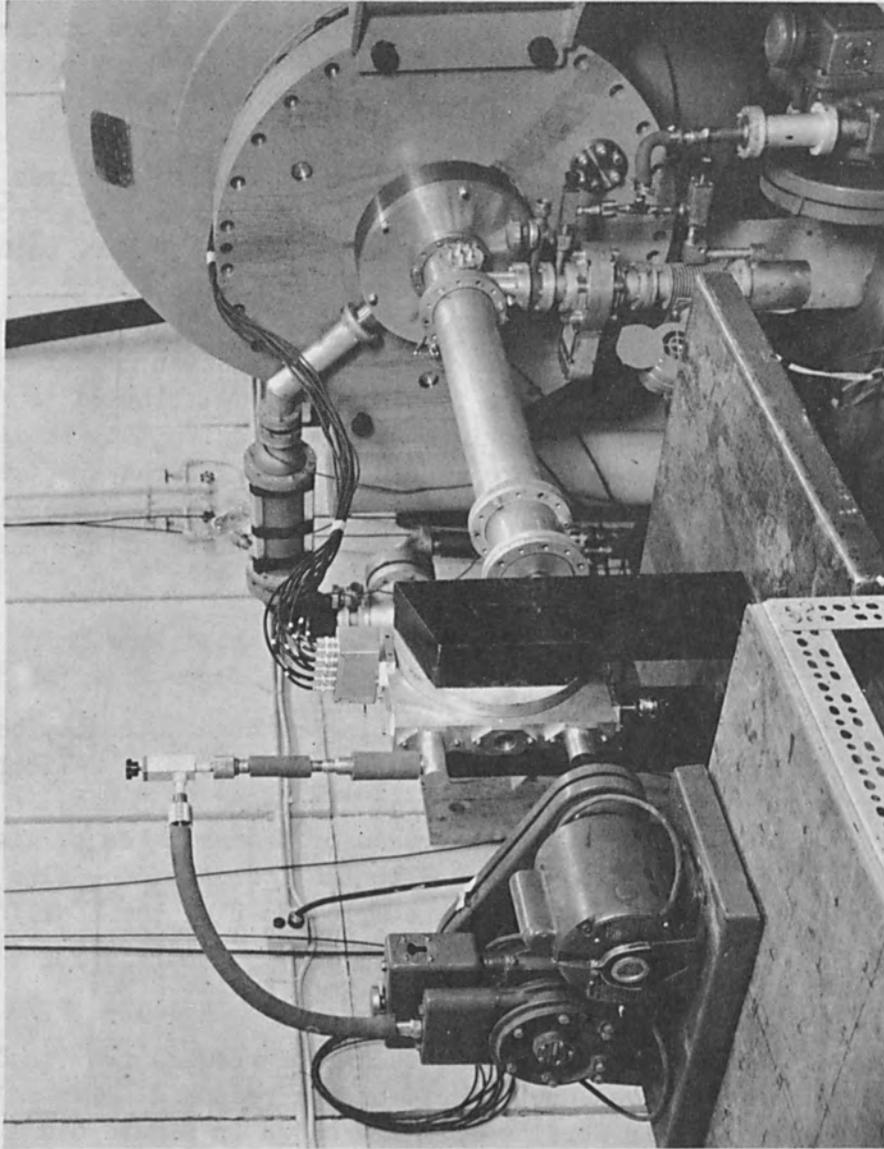


Figure 4. Experimental Geometry - Beam Spectral Analysis.

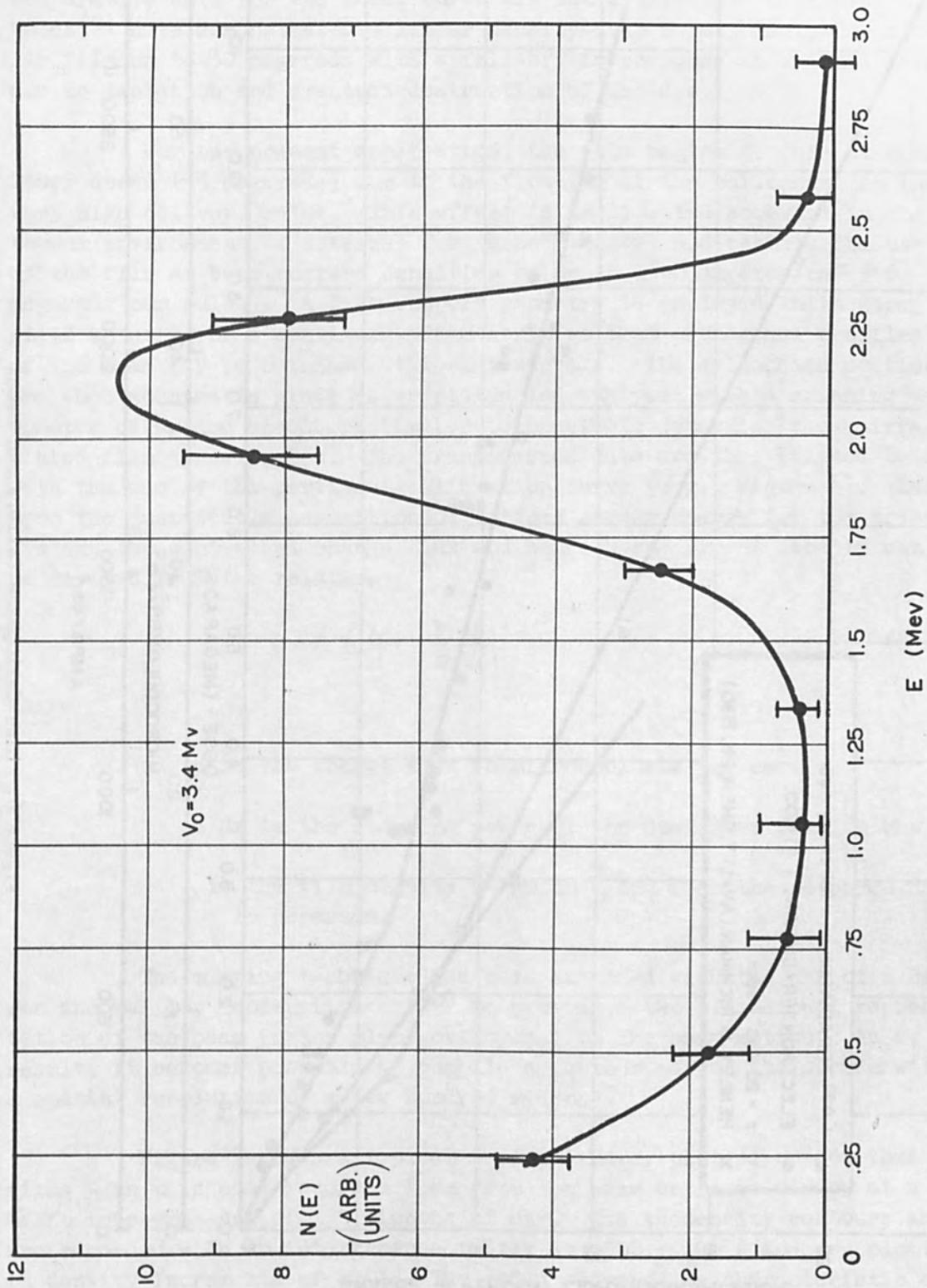


Figure 5. Pulsed Electron Beam Energy Spectrum.

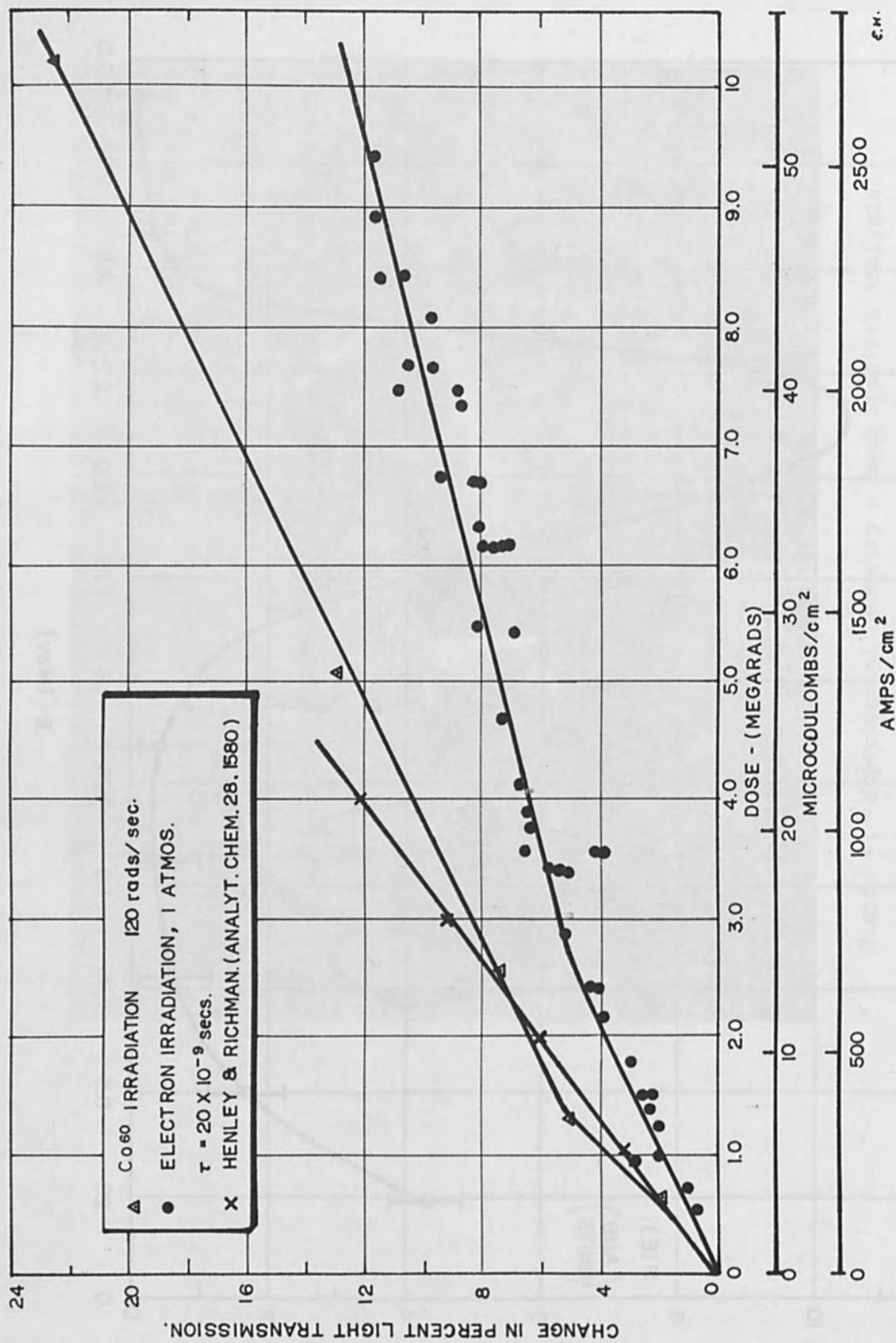


Figure 6. Blue Cellophane Calibration Data.

tron calibration of similar film at 2 Mev reported by Henley and Richman⁽⁸⁾ taken at dose rates of $>2 \times 10^5$ rads/ second. No electron data appear to have been reported at rates much above 10^6 rads/sec, hence comparative data for the lower curve are not available. Charlesby and Woods⁽⁹⁾ have demonstrated a linear density-dose relationship for similar film up to 30 megarads with a fall-off in response above this level due to depletion and eventual destruction of the dye.

For the present application, the film begins to fail at much lower doses (~ 5 megarads) due to the frosting of the cellophane at these very high delivery rates. This effect is ameliorated somewhat in the vacuum environment of interest for these studies, and permits the use of the film at beam current densities of up to 1600 amperes/cm² (~ 6 megarads per pulse). A film support geometry is employed which permits rigid mounting of a continuous film strip so that orthogonal profiles of the beam may be obtained at 1 cm intervals. The cellophane sections are then mounted in glass cover plates and analyzed with a scanning densitometer using the spectrophotometer to precisely determine the unirradiated film transmission. The transmission data are then related to dose with the use of the pertinent calibration curve (e.g., Figure 6). Based upon the justifiable assumption of uniform stream energy for the present system, the equivalent charge flux and hence beam current density can be derived from the relation:

$$\phi = (0.1 \frac{1}{\rho} \frac{dE}{dx})^{-1} D \quad (8)$$

where

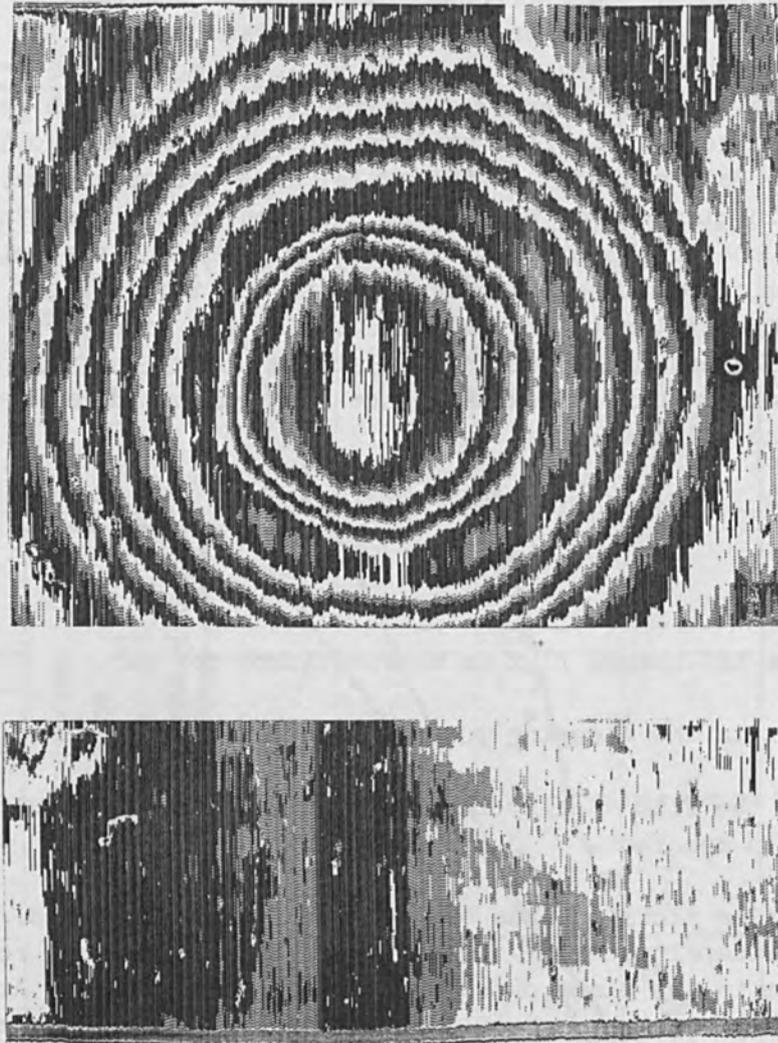
ϕ is the charge flux in microcoulombs per cm²,

dE/dx is the stopping power of the dosimeter film in Mev cm⁻¹

ρ is the film density in gm cm⁻³ and D is the measured dose in megarads.

The mapping technique has been extended with the use of a Beckman and Whitley Isodensitracer⁽¹⁰⁾ to provide a two-dimensional representation of the beam in any plane orthogonal to the beam motion. As a result, it becomes possible to compile a contour map of the stream with a spatial resolution of a few hundred microns.

Figure 7 shows a typical isodensitracer plot of one of these films - in this case recorded 1 cm from the beam entrance window at a drift tube pressure of 150 microns of air. The isodensity contours shown are recorded with an interference filter at $6500 \pm 100 \text{ \AA}$ and are plotted in density increments of 0.01. The coded sequence is characteristic of the Beckman and Whitley instrument and moves in a dash-dot-space sequence



DATE 9/15/65	RECORDING NO. Book 209, p. 33	IDENTIFICATION Shot No. 1
Spot Height 190 μ	Wedge Number D Filter	Magnification: 2 \rightarrow
Spot Width 190 μ	Δ D Increment 0.01	2 \uparrow
Objective X5	Differential Control Out	Scale: One Inch = 1.27 cm

Figure 7.

for increasing transmission. The isodensity contour corresponding to the peak current density of the entrance stream ($\sim 1000 \text{ amp/cm}^2$ at 2 Mev) is clearly shown in the figure with the fall-off in the central portion of the hollow stream. A sample of unirradiated film is shown at the left of the figure and demonstrates the small degree of transmission variation due to film inhomogeneities. For the case shown, the entrance diameter of the annulus of peak current density was 3 cm.

A direct intercomparison of the data resulting from the scanned film technique with calorimetric mapping techniques has been performed using a four-point calorimeter. Figure 8 shows two radial profiles of the hollow relativistic beam in each instance measured simultaneously by the two techniques. The calorimetrically determined points are shown as the 3 mm wide bars corresponding to the average energy flux recorded across thick calorimeters of this diameter. The upper curves in each case correspond to the dyed cellophane data reduced with the isodensitracer. The disagreement in the central portion of the profile indicates the presence of a low energy component in the core of the stream, likely due to the soft electron group recorded in the magnetic analysis studies. The data shown here were taken at 150μ and at one atmosphere in air, 4 cm from the entrance window. These comparative data emphasize the primary limitation in the use of the dosimetry technique for beam mapping: any significant variation from the assumed film stopping power can lead to significant errors in the calculated values of j_e . The importance of complementary spectral data is obvious in the interpretation of both the dosimetric and calorimetric data.

VI. GENERAL FEATURES OF THE SELF-FOCUSED STREAM

Some data demonstrating the application of these several techniques to the direct observation of self-magnetically focused streams will now be presented. The first experiment utilized the simplest energy flux monitor, namely the thick (total stopping thickness) calorimeter, to study beam confinement as a function of pressure at a fixed distance in the isolated drift tube.

Figure 9 shows data taken by thick calorimeter probes with thermocouple readouts. These data were taken at 20 cm with one calorimeter on axis and the other 8 mm off axis while the drift tube pressure was varied from 10^{-4} torr to 760 torr. The low pressure end of the curve shows the energy density rising as the critical current requirement is reduced due to space-charge neutralization. The peak current density in the stream occurs at 0.2 torr where it is calculated that $f = 1$ in approximately 5 nsec. The current density begins to fall off where the radial escape of electrons is becoming impeded. At higher pressures where the beam generated plasma density is 2-3 orders of magnitude higher than the beam electron density we apparently get an effective shielding

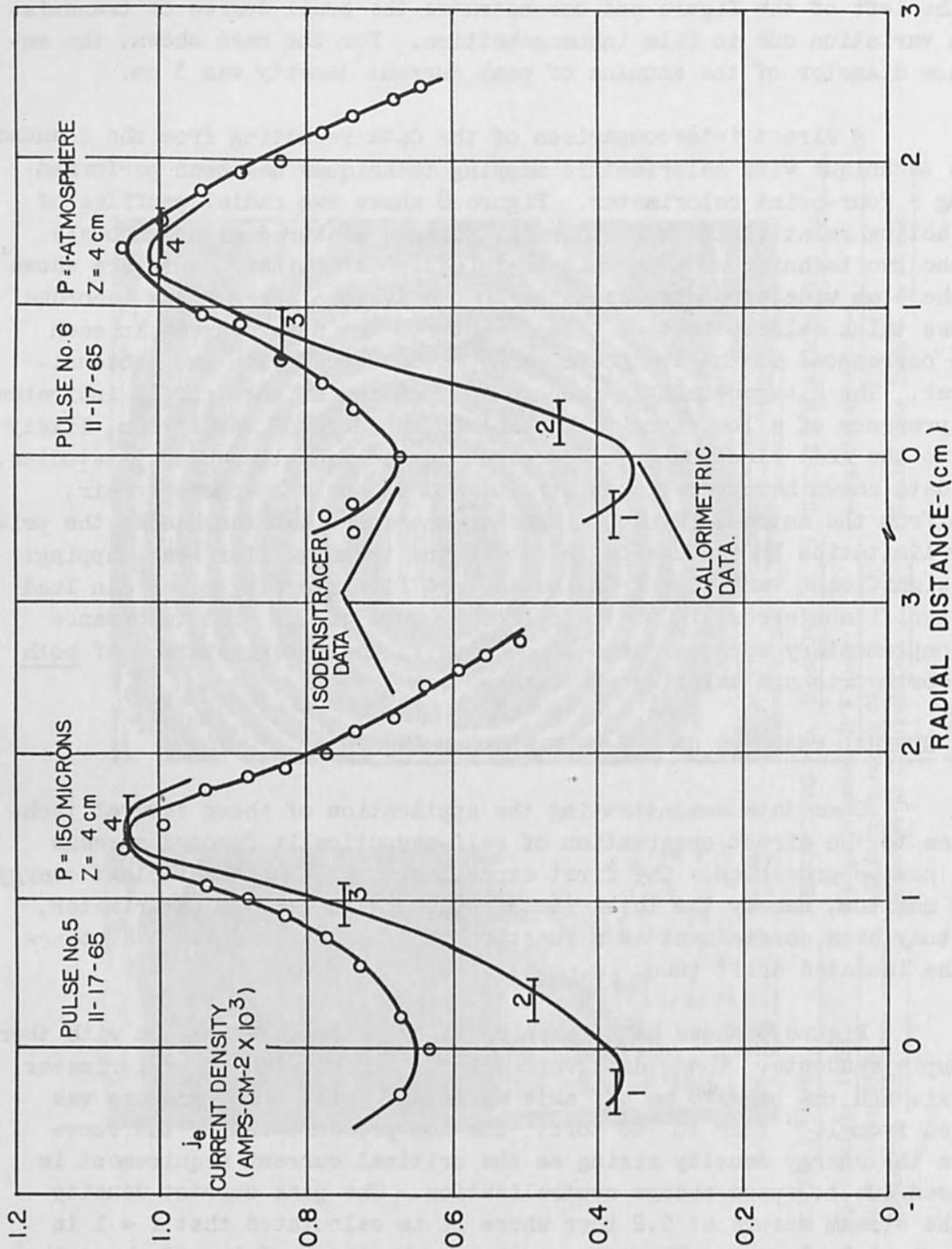


Figure 8. Cellophane Dosimeter - Calorimeter Intercomparison.

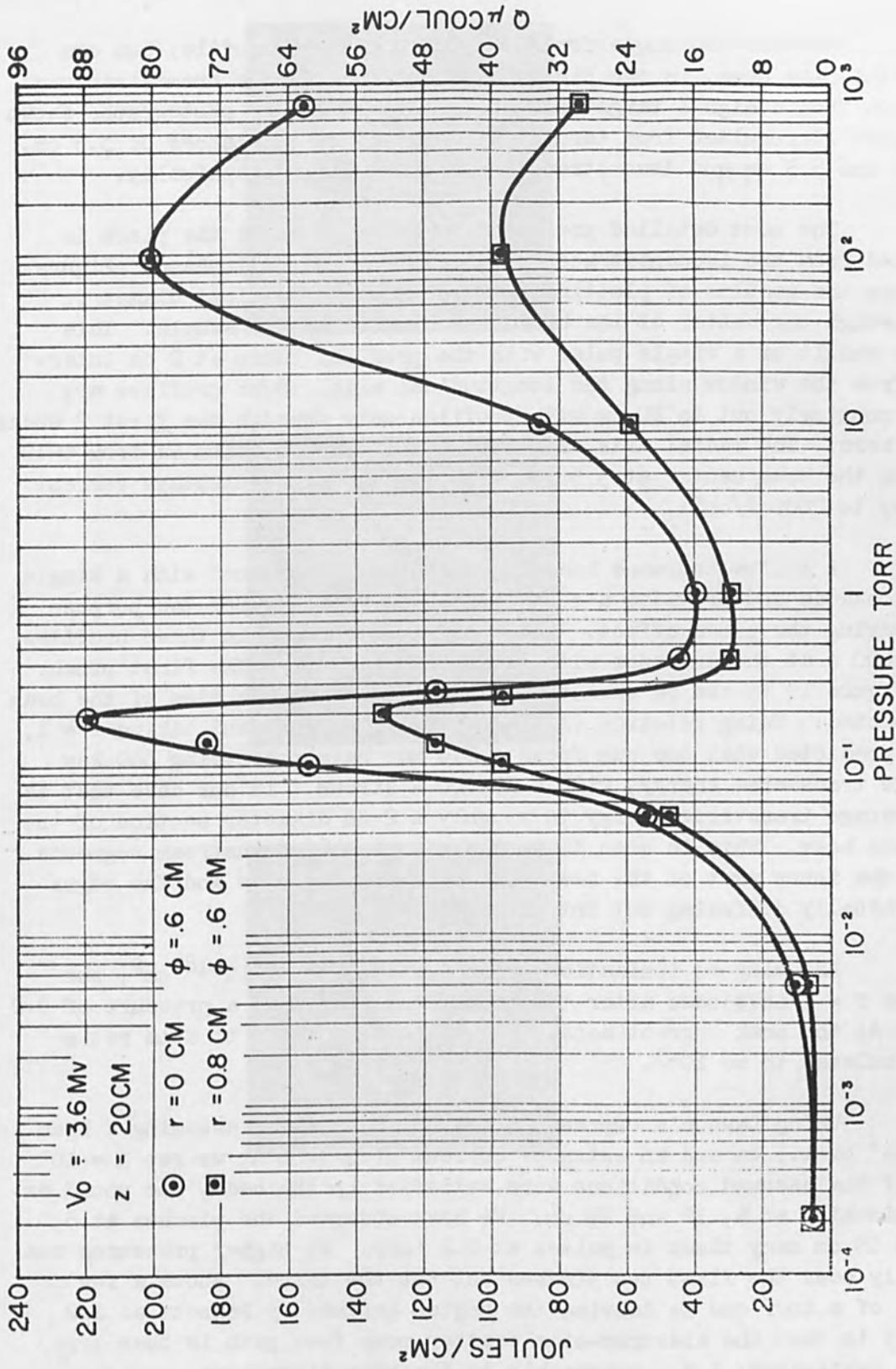


Figure 9. The Pressure Dependence of the Beam Energy Flux Monitored Calorimetrically 20 cm from the Entrance Window.

of the beam electrons from each other and self-focusing is again evident.

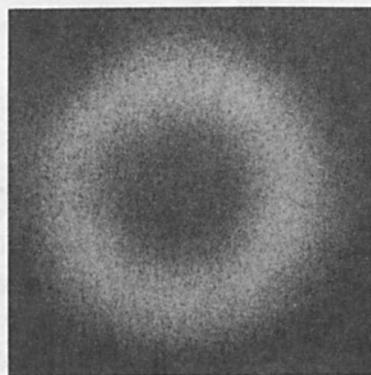
Another technique for measurement of beam profile, but one that stops the beam, is the pin hole photography of the bremsstrahlung emission from a high Z target placed in the beam. The photographs shown in Figure 10 resulted from targets at longitudinal distances of 3.5 cm, 4.5 cm and 5.5 cm and demonstrate the Bennett pinch graphically.

The most detailed presentation of the form of the pinch is obtained from the isodensitracer plots of the blue cellophane. Figure 11 shows the results of plotting the dose along a vertical diametric cut through the center of the beam for each cellophane sample. This is the result of a single pulse with the profiles taken at 2 cm intervals from the window along the longitudinal axis. Such profiles are taken routinely out to 20 cm with profiles only through the first 2 nodes shown here. The radial axis as shown in this presentation is from 0-12 cm with the beam center at 5.5 cm. The vertical axis displays current density to 2000 A/cm².

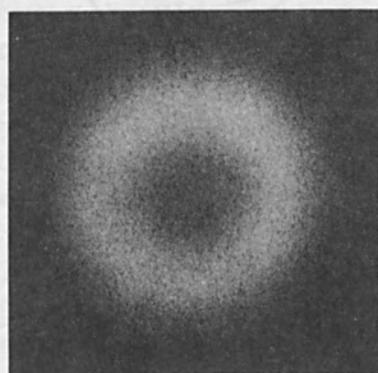
A hollow entrance beam is consistently obtained with a single point cathode and provides a convenient beam geometry for the purpose of studying the pinch effect. The integrated current in these profiles is 17,000 A at the entrance with some reduction after the first pinch. From Figure 10 it can be seen that the main loss is the edge of the beam as expected. Using relation (1) for critical current and taking $f = 1$, it is predicted that one can focus a 2.0 Mev beam possessing 300 kev average transverse energy, with a 15,000 A stream. In our case that is the average transverse energy in roughly a 2 cm diameter section of the entrance beam. This is seen to be the case in the downstream segments where the inner part of the beam converges and diverges and the edges are gradually diffusing out into a broad beam.

Assuming an ionization cross section⁽¹¹⁾ of 10^{-18} cm², one obtains $f = 1$ in 5 nsec after the beam front passes at a pressure of 0.2 torr. At the peak current densities obtained, the ion to atom ratio is calculated to be 10^{-3} .

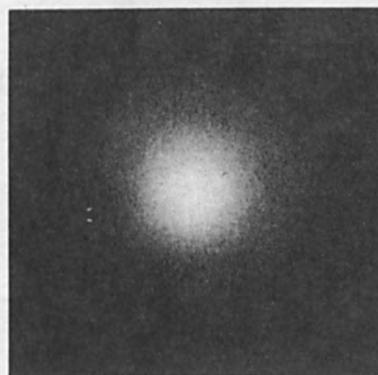
Using Lawson's expression (equation (5)) and assuming a beam "radius" of 1.5 cm and an entrance current of 17,000 A, we get $\lambda = 16$ cm. If the assumed conditions were satisfied by the beam, one would expect pinching at 4, 12 and 20 cm. We have observed the pinches at 5, 13 and 19 cm many times in pulses at 0.2 torr. At higher pressures one normally sees the first two pinches but not the third. Above a few tenths of a torr one is leaving the region defined by Bennett as low density in that the electron-atom scatter mean free path is then less than a centimeter; i.e., comparable to the beam dimensions.



(1)



(2)



(3)

Figure 10. Bremsstrahlung Emission from Tantalum Target Placed at (1) 3.5 cm, (2) 4.5 cm, (3) 5.5 cm.

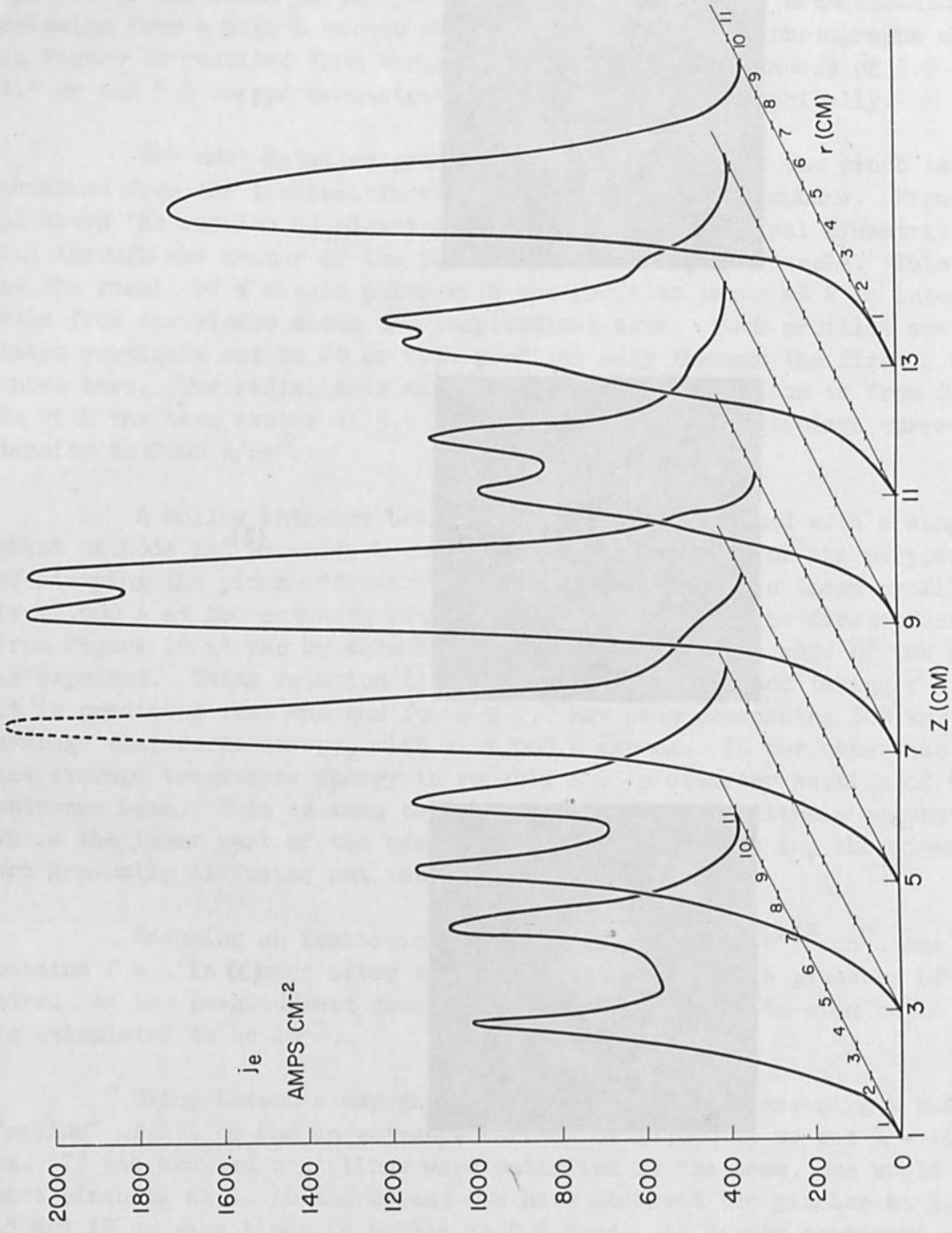


Figure 11. Drifting Electron Beam in Vertical Profile.

VII. CONCLUSIONS

The pulsed relativistic beam provides a unique tool to study the Bennett pinch unobscured by other effects. The pulse length removes the need to consider recombination and ion drift effects. The high current density causes the pinch points to fall at convenient distances and overcomes divergent entrance conditions. The high beam energy makes energy losses negligible over long paths so that a true drifting beam can be studied. It is evident that the effect permits the reproducible generation of very high electron delivered energy fluxes and the techniques described above are being used to determine the intrinsic limitations of such magnetically self-focused streams.

VIII. ACKNOWLEDGEMENTS

The authors acknowledge the constructive comments and assistance of Dr. A. Guenther and Dr. P. Moore of the Air Force Weapons Laboratory, Albuquerque, New Mexico. The many productive discussions held with Mr. J. C. Martin of AWRE, Aldermaston, England, have helped expedite these studies. The assistance of Mr. Fred Parsons of Tech Ops, Burlington, Massachusetts, in utilizing the Isodensitracer was essential to development of the high resolution scanning technique. The co-operation of Mr. W. R. Van Antwerp of the Edgewood Arsenal, Baltimore, Maryland, in supplying the fixed field spectrometer and in offering many helpful suggestions for its use, is gratefully acknowledged. Finally, the contributions of Mr. W. C. Beggs, Mr. J. P. Shannon and Dr. A. S. Denholm, and the capable assistance of Mr. G. Ames and Mr. G. Goosney in the execution of these experiments, have been essential to the program.

REFERENCES

1. Bennett, W. H., Phys. Rev., "Magnetically Self-Focusing Streams", 98, 1584, (1955); 45, 890, (1934).
2. Lawson, J. D., Journal of Electronics and Control 3, 587, (1957); 5, 146, (1958).
3. Graybill, S. E. and Nablo, S. V., Applied Physics Letters 8, 18, 1966.
4. Smith, I. D., "Pulse Breakdown of Insulation Surfaces on a Poor Vacuum", Proc. Conf. Insulation of High Voltages in Vacuum, 261, MIT, Boston, Massachusetts, October 1964.
5. Denholm, A. S., et al, "The Feasibility of Designing High Voltage X-Ray Tubes ...", AFWL-TR-65-62, Ion Physics Corporation, Burlington, Massachusetts, August 1965.
6. Birkhoff, R. D., "The Passage of Fast Electrons Through Matter", Hand. der. Physik 34, 129, Springer-Verlag, Berlin, 1958.
7. Van Antwerp, W. R., An Initial Gamma-Ray Spectrograph, U. S. Army Chemical Corps, Nuclear Defense Lab., Army Chemical Center, Maryland, February 1962.
8. Henley, E. J. and Richman, D., Analyt. Chem. 28, 1580, (1956).
9. Charlesby, A. and Woods, R. J., "Radiation Dosimetry with Dyed Cellophane", Int. Jour. App. Radn. and Isotop. 14, 413, 1963.
10. Miller, C. S., Parsons, F. G. and Kofsky, I. L., "Simplified Two-Dimensional Microdensitometry", Nature 202, No. 4938, 1196, 1964.
11. Barnett, C. F., et al "Atomic and Molecular Collision Cross Sections of Interest in Controlled Thermonuclear Research", ORNL-3113, Oak Ridge National Laboratories, Oak Ridge, Tennessee, May 1961.

THE QUEST FOR SUPERDENSE ELECTRON BEAMS

by

W. E. Waters
Varian Associates
Palo Alto, California

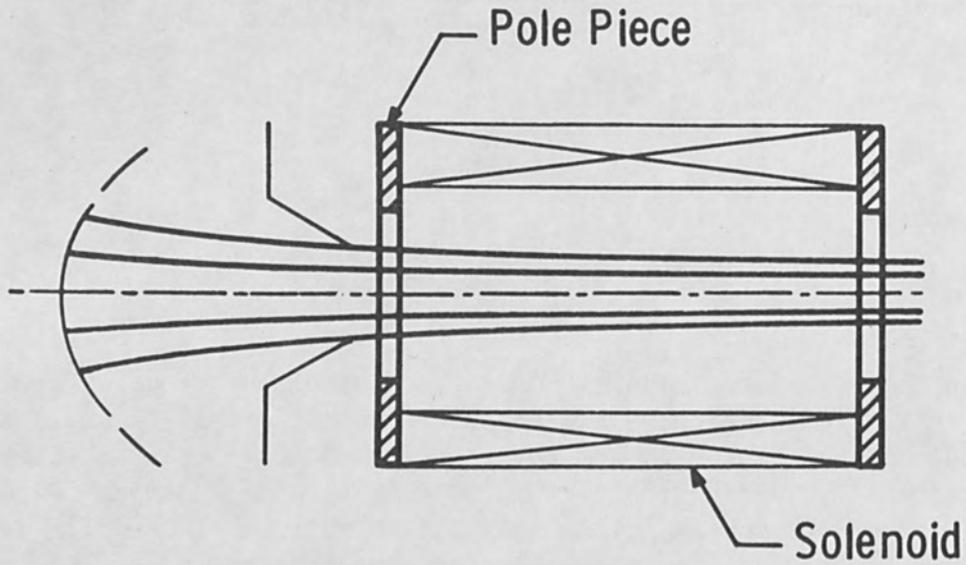
ABSTRACT

In the theories of the steady flow of electrons, focused by a uniform magnetic field, which have been published to date it is shown that the electron plasma frequency ω_p of the unrippled focused beam is related to the electronic cyclotron frequency ω_h of the magnetic focusing field by the relation $\omega_p = k \omega_h$; one finds that k lies between $1/2$ and 1 . We have inquired into the possibility that nature might permit operation of an axially symmetric system such that ω_p can exceed ω_h either somewhere, or throughout, the beam. The results of this inquiry are negative and we believe, definitive; this paper presents our method of exploring the question.

A beam is herein defined to be superdense if, neglecting rippling, thermal velocities and relativistic effects, the plasma frequency exceeds the cyclotron frequency somewhere in the beam. It is possible to derive a rigorous mathematical expression which relates the relative radial positions (i.e., the radial order) of two adjacent lamina of the focused beam to the relative radial positions of those same two lamina at the cathode surface, under the assumption that the beam is superdense in the region being considered. It is then possible to prove that the radial order of two adjacent lamina in the region of superdensity is the opposite of the radial order produced by the same focusing system, but operating at a total current level small enough to permit space-charge forces to be neglected (i.e., the "one-electron-at-a-time" radial order). Arguing heuristically, we believe this state of affairs implies inherent instability of a superdense beam, for otherwise the gradual increase of beam space charge from zero would at some point require adjacent lamina to reverse their relative positions in the focused beam, and we find it extremely difficult to imagine an electron optical arrangement which permits this to occur.

I. INTRODUCTION

A familiar and widely used electron beam system for linear microwave tubes is the so-called "Brillouin-focused" arrangement¹, depicted schematically in Figure 1. In this system ideally the cathode surface



Pierce Gun

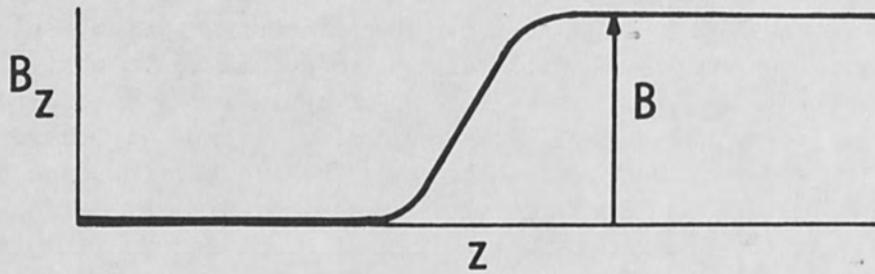


Figure 1. Schematic Arrangement of a Pierce-Gun Focusing System and the Axial Magnetic Field Variation.

is completely shielded from magnetic flux linkage.

It is characteristic of an ideal Brillouin beam, in the region of the uniform magnetic field, that:

1. Every electron in the beam rotates around the axis with an angular velocity equal to the Larmor frequency of the magnetic field, i.e., $d\theta/dt = \eta B/2$.
2. Every electron drifts in the axial direction with the same axial velocity \bar{v}_0 .
3. The total beam current \underline{I} enclosed within some radius \underline{a} is given by the relation*

$$I = \pi \epsilon_0 \eta^{3/2} B^2 \phi_0^{1/2} a^2 / \sqrt{2}, \quad (1)$$

where ϕ_0 is the potential on the beam axis.

Upon introducing certain subsidiary relations, i.e.,

$$J = \rho v = - I / \pi a^2$$

(Current Density = charge density x velocity),

$\omega_h = \eta B$, (cyclotron frequency definition) and

$\omega_p^2 = - \eta \rho / \epsilon_0$ (plasma frequency definition),

it is readily shown that Eq. (1) is equivalent to the relation

$$\omega_p = \omega_h / \sqrt{2}. \quad (2)$$

Hence, in an ideal Brillouin beam the electron stream adjusts itself to a radial distribution such that the plasma frequency is uniform and equal to $1/\sqrt{2}$ times the magnetic field cyclotron frequency. In the sheet beam equivalent of a round Brillouin beam the plasma frequency is again uniform, but exactly equal to ω_h .

Figure 2 depicts a very general axially-symmetric electron beam system. For convenience the system is divided into three general regions, namely the GUN region, immediately adjacent to the cathode, a

*Rationalized MKS units (amperes, volts, meters, webers/sq. m); $\eta = e/m$.

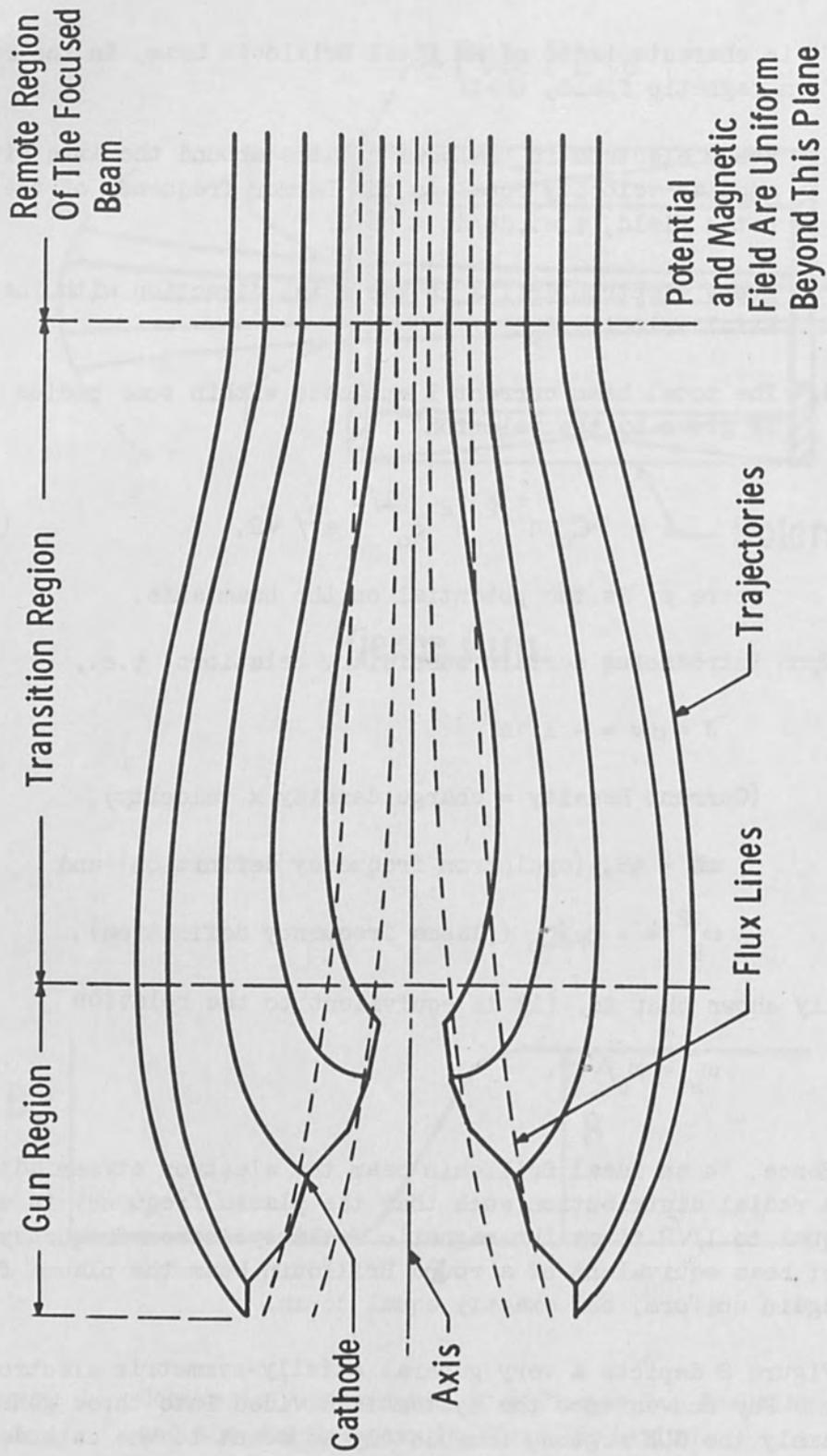


Figure 2. A General Axially-Symmetric Focusing System.

TRANSITION region, and the REMOTE region, well beyond the gun. In the remote region it is assumed that the magnetic field is uniform and axial. By contrast the only restriction placed on the magnetic field in the gun region is that the field must be axially symmetric; it is also assumed that cathode is axially symmetric and has a continuous variation of both local slope and local emission density. (A magnetic field is assumed to exist at the cathode surface. Properties of the beam arising from a magnetically shielded cathode are readily derivable, and it is easy to prove that such beams are not superdense).

II. DERIVATION OF THE CHARGE DENSITY RELATION

The remote beam is taken to possess the following characteristics:

1. Any one electron moves on a helical path centered on the axis;
2. Both the angular velocity around the axis and the axial velocity are functions of radius only;
3. The system is invariant in time.

Under these assumptions one may write several important relations pertaining in the remote beam, namely:

The radial force balance equation,

$$m r \omega^2 + e r \omega B + e E_r = 0; \quad (3)$$

the energy relation,

$$2 \eta \phi = r^2 \omega^2 + v^2; \quad (4)$$

the electric field relation,

$$E_r = - d\phi/dr; \quad (5)$$

Poisson's equation,

$$\rho = -\epsilon_0 \nabla^2 \phi. \quad (6)$$

In the above relation, ϕ is the potential, ω is the angular velocity, v the axial velocity, and all are considered to be differentiable functions of radial position.

Equation (4) may be differentiated with respect to r and the resultant substituted into Eq. (3); this process yields a first-order non-

linear differential equation relating v and ω , namely

$$v \, dv/dr + \omega_h^2 \sigma \, d [r^2 (\sigma - 1/2)] / dr = 0, \quad (7)$$

$$\sigma = \omega/\omega_h, \quad \omega_h = \eta B.$$

Equation (7) represents a necessary relation connecting v and σ , and is applicable to any axially symmetric magnetically focused beam, hollow or solid, regardless of the detailed nature of the electron gun used to generate the beam.

It was stated previously that $\omega = \omega_h/2$ in a Brillouin beam, i.e., $\sigma = 1/2$. Equation (7) then requires that $d \sigma / dr = 0$, or $\sigma = \text{constant}$, a condition which was also stated to be a property of a Brillouin beam. Although at this point of the development a proof is lacking, it is possible to show that the condition $\sigma = 1/2$ imposes a very stringent condition on the electron gun, namely that the cathode surface must be completely shielded from magnetic flux. (Refer to Eq. (10)).

Next, using the value of $d\phi/dr$ obtained from Eq. (4), and also using Eq. (7) to eliminate $v \, d\gamma/dr$, one may derive the following expression for the charge density ρ from Eq. (6):

$$\rho = -\epsilon_0 \omega_h^2 / \eta [2 \sigma (1 - \sigma) + r(1 - 2\sigma) \, d\sigma/dr] . \quad (8)$$

Upon defining the plasma frequency through the relation

$$\omega_p^2 = - \eta \rho / \epsilon_0$$

and letting $R = \omega_p^2 / \omega_h^2$, it is found that

$$R = 2\sigma (1 - \sigma) + r (1 - 2\sigma) \, d\sigma/dr. \quad (9)$$

The superdensity condition is simply $r > 1$. (Note that in a Brillouin beam $\sigma = 1/2$, so that $R = 1/2$ from Eq. (9), or $\omega_p = \omega_h/\sqrt{2}$.) So far in the development only the remote beam has been considered. In order to progress the theory it is now necessary to establish a relation connecting a lamina of the remote beam and its corresponding cathode position.

III. RELATION OF A BEAM LAMINA TO THE CATHODE

Figure 3 illustrates how a lamina of the remote beam may be related to the corresponding region on the cathode surface. Busch's theorem² may be used to write an explicit relation for the angular velocity,

$$\sigma = 1/2 - \psi_c / 2\pi r^2 B. \quad (10)$$

In passing from a radius r to a radius $r + dr$ in the remote beam, the corresponding cathode radius and flux linkage change to $r_c + dr_c$ and $\psi_c + d\psi_c$, respectively; Eq. (10) therefore yields the differential

$$d\sigma = [2 \psi_c dr/r^3 - d\psi_c/r^2] / 2\pi B, \quad (11)$$

which may be inserted directly into Eq. (9) to yield

$$R - 1 = 2\sigma(\sigma - 1) - (1 - 2\sigma) dP/2 r dr, \quad (12)$$

where

$$P = \psi_c / \pi B.$$

IV. REQUIREMENTS OF THE SUPER DENSITY CONDITION

It is important to note that ψ_c/B , and hence P , may be either positive or negative, depending upon whether the axial magnetic field threading the cathode has the same or the opposite direction as the field in the remote beam. If $P > 0$ (no field reversal) one may show that σ lies between 0 and 1, and if $P < 0$, $\sigma > 1/2$. From Eq. (10) it is found that

$$r = [P/(1 - 2\sigma)]^{1/2}, \quad (13)$$

and the preceding remarks guarantee that r will be real regardless of the sign of P .

The condition of superdensity then requires, from Eq. (12), that $R - 1 > 0$; upon inserting Eq. (13) into Eq. (12) and recalling the remarks just preceding Eq. (13), superdensity then requires that

$$\frac{-d(P)^{1/2}}{dr_c} \cdot \frac{dr_c}{dr} > \frac{2\sigma(1 - \sigma)}{(1 - 2\sigma)^{3/2}} \geq 0 \quad \text{if } P > 0, \quad (14a)$$

$$\frac{-d(-P)^{1/2}}{dr_c} \cdot \frac{dr_c}{dr} > \frac{2\sigma(1 - \sigma)}{(2\sigma - 1)^{3/2}} \geq 0 \quad \text{if } P < 0. \quad (14b)$$

Our remarks are now to be confined to the class $P > 0$; the reader may easily verify the same conclusions when $P < 0$.

Noting that $dP^{1/2} = 1/2P^{-1/2} dP$, Eq. (14a) requires that

$$-\frac{dP}{dr_c} \cdot \frac{dr_c}{dr} > 0, \quad (15)$$

so that if dP/dr_c is positive, dr_c/dr must be negative, and conversely. It is worthwhile to explain explicitly the nature of these derivatives.

dP/dr_c is $1/\pi B$ times the rate of change of magnetic flux linkage as one passes from a given point on the cathode to a nearby point having, in general, an incrementally different radius and an incrementally different flux linkage (see Fig. 3).

Corresponding to the above shift in cathode position is a shift in radius in the remote beam; an electron leaving the cathode from a point having radius r_c is focused to a radius r in the remote beam, and another electron leaving the cathode at radius $r_c + dr_c$ is focused to a radius $r + dr$ in the remote beam. The derivative of interest is simply the ratio of the two corresponding increments, dr_c and dr .

Figures 4a and 4b illustrate conditions under which P may be negative or positive, respectively, whereas Figures 5a-d illustrate conditions which result in positive or negative values of dr_c/dr .

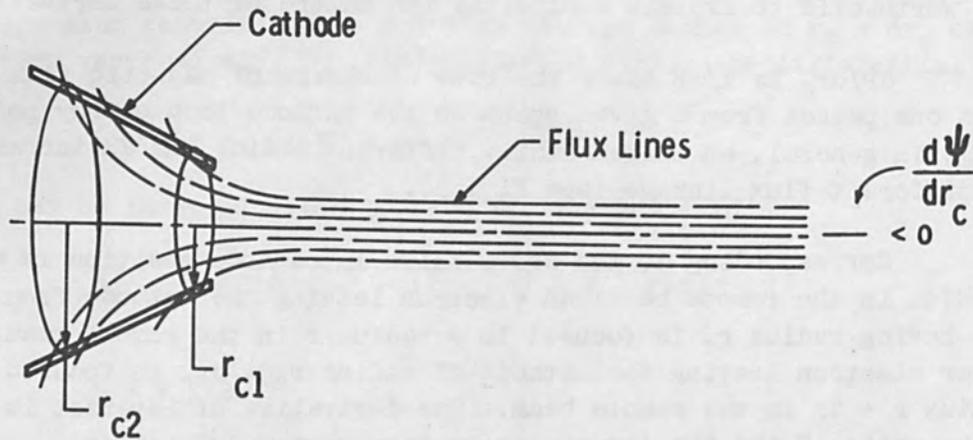
Let us now consider the system when the current is so small that space-charge forces are ignorable. Then the electric field in Eq. (3) vanishes, and combination of Eqs. (3) and (10) lead to $r^2 = \psi c/\pi B$, which may be differentiated to yield

$$\frac{dP}{dr_c} \cdot \frac{dr_c}{dr} = 2 r, \quad (16)$$

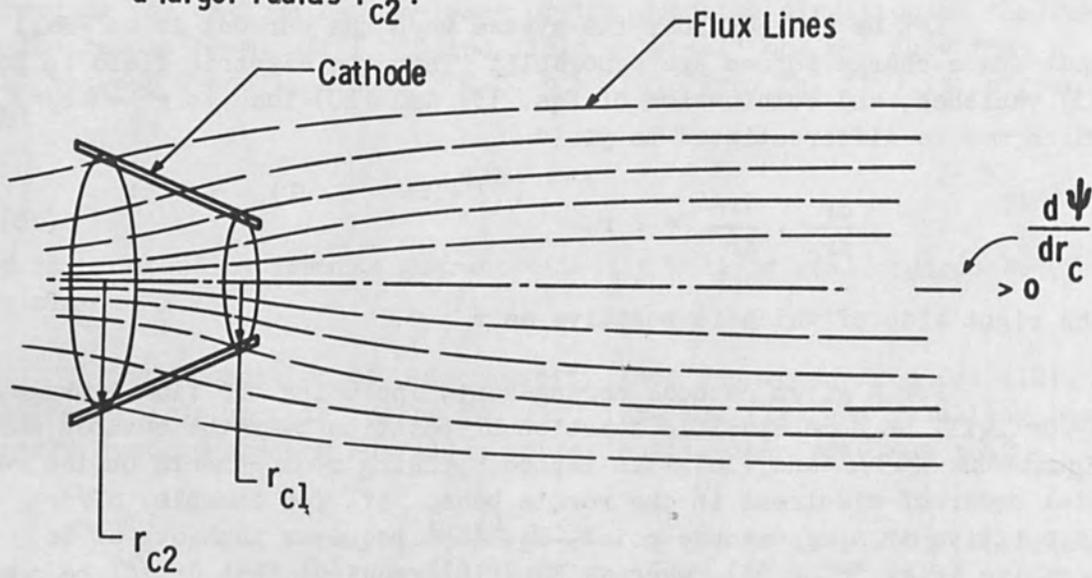
the right side of which is positive as $r > 0$.

For a given cathode arrangement, including the flux linkage, dP/dr will be some specific function of position over the cathode surface. Equations (15), and (16) then impose opposing requirements on the radial order of electrons in the remote beam. If, for example, dP/dr_c is positive at some cathode point, Eq. (15) requires that dr_c/dr be negative (Figs. 5b or 5d), whereas Eq. (16) requires that dr_c/dr be positive.

Thus, the condition of superdensity demands that the radial order of electrons in the focused beam, when the current is finite, must be the opposite of the radial order when the current vanishes. We find it difficult to visualize a system which would permit such a reversal of radial order and still be stable. Equation (16) is a mathematical statement of the fact that, in the absence of current and in the absence of rippling in the remote beam each electron moves along a flux line, thus cutting no flux. Any deviation from a flux line results in restoring forces acting to accelerate the electron back toward that flux line. Since the total ensemble of flux lines constitutes a well-ordered non-

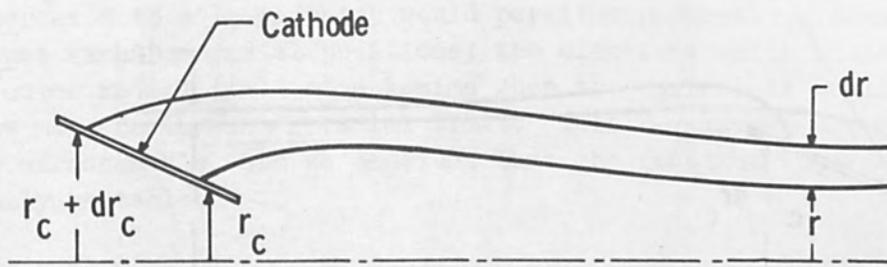


The flux linkage $d\psi/ar_c$ is negative if the flux expands outward from inside the cathode. A circle of radius r_{c1} encloses more total flux than a second circle having a larger radius r_{c2} .



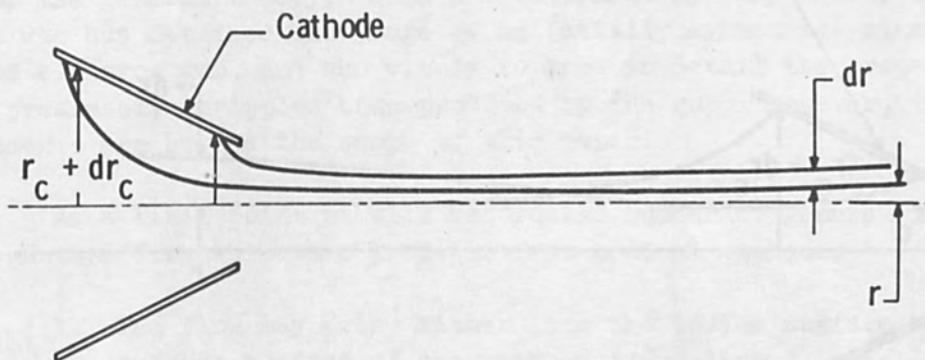
The flux linkage is positive if the flux converges from outside the cathode. In this case the smaller circle of radius r_{c1} encloses less flux than a larger circle of radius r_{c2} .

Figure 4. Illustrating Two Distinct Types of Flux Linkage.



Illustrating how $dr_c/dr > 0$ in a laminar beam.

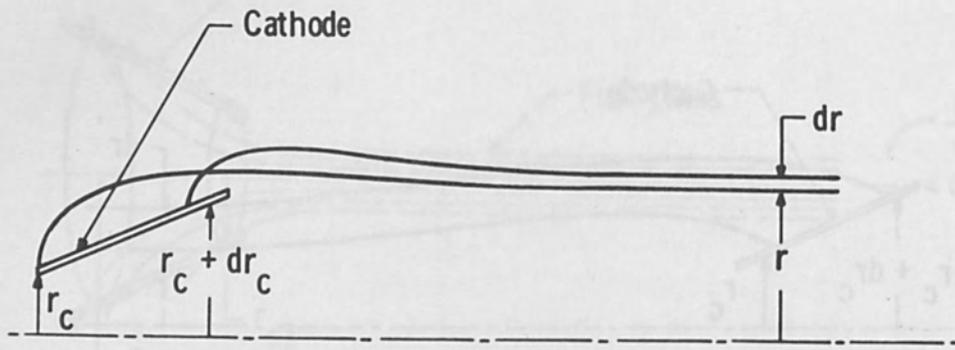
Figure 5a



Illustrating how $dr_c/dr < 0$ in a laminar beam

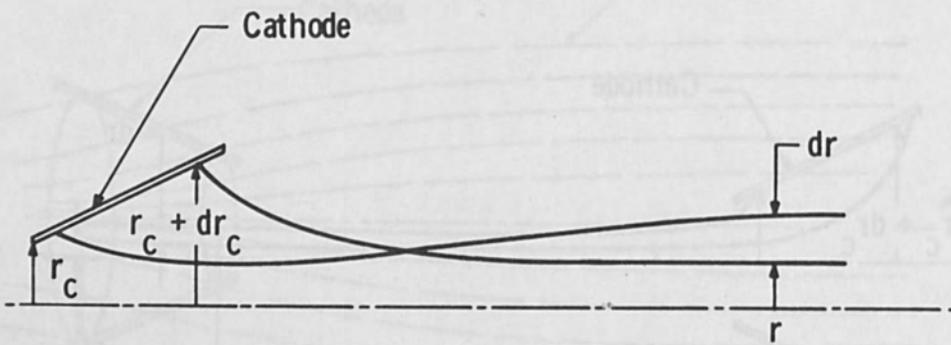
Figure 5b

Figure 5. Showing How dr_c/dr May be Positive and Negative in Laminar and Non-Laminar Segments.



Illustrating how $dr_c/dr > 0$ in a nonlaminar beam.

Figure 5c



Illustrating how $dr_c/dr < 0$ in a nonlaminar beam.

Figure 5d

intersecting set (in a vacuum), then presumably the electron trajectories in the absence of space charge would do the same.

Equation (15) on the other hand requires that when the current has been increased to a level which would permit superdensity, somehow electrons must exchange radial positions; the electrons which would move on the outer radial limit of a lamima when the current is negligible must now move on the inner radial limit. This occurrence seems to us totally unreasonable, and we conclude that the onset of superdensity is inherently unstable.

V. CONCLUSIONS

The analysis presented in this paper is very useful for establishing completely general relations in round beams. In general one must also give consideration to the variations of both the local slope and local emission density over the cathode surface. Once these quantities are known, a second first-order, non-linear differential equation relating σ and γ may be established, and detailed properties of the remote beam may be calculated by simultaneously solving this equation and Eq. (7). All possible round beams, hollow or solid, then become special cases of the general theory. Such a development is very useful to the analyst who has detailed knowledge of an (axially symmetric) magnetically immersed electron gun, and who wishes to know in detail the properties of the presumably unrippled beam produced by the gun. However, these developments are beyond the scope of this paper.

As a final point we will record the number of general types of space charge flow which are possible from conical cathodes.

1. The flow may arise either from the inside surface or the outside surface of the cathode (two classes).
2. The flow direction may be generally from the large end of the cathode toward the small end, or conversely (two classes).
3. The magnetic field may converge toward the axis of the cathode (positive flux linkage) or diverge away from the axis (negative flux linkage) (two classes).
4. There may or may not occur a magnetic field reversal between the cathode and the remote beam (two classes).

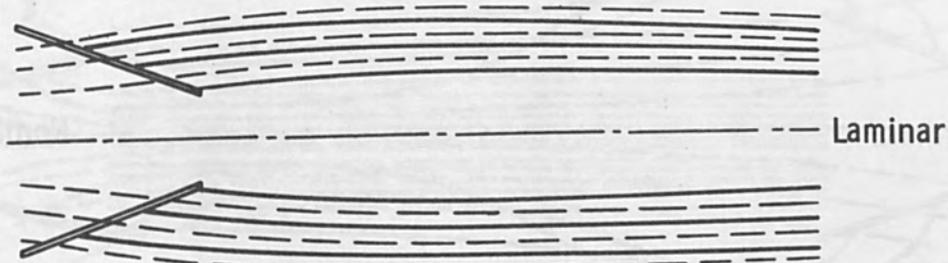
Thus, there are two choices for each of four mutually exclusive flow characteristics, giving a total of sixteen general classes of flow patterns.

Figures 6a through 6h which follow illustrate the general character of the eight flow patterns not involving a field reversal; trajectories have been sketched on the assumption of negligible space charge.

REFERENCES

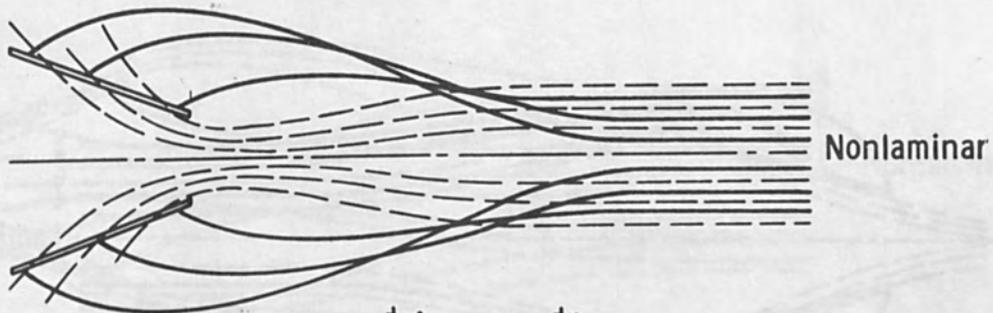
1. Pierce, J. R., "Theory and Design of Electron Beams", 1st ed., D. Van Nostrand, 1949.
2. Op. cit., Sect. 4.4.

—— Trajectory
- - - Flux



Large-toward-small $\frac{d\psi_c}{dr_c} > 0$ $\frac{dr_c}{dr} > 0$

Figure 6a



Large-toward-small $\frac{d\psi_c}{dr_c} > 0$ $\frac{dr_c}{dr} > 0$

Figure 6b

Figure 6. Eight General Types of Flow Patterns from Conical Cathodes.

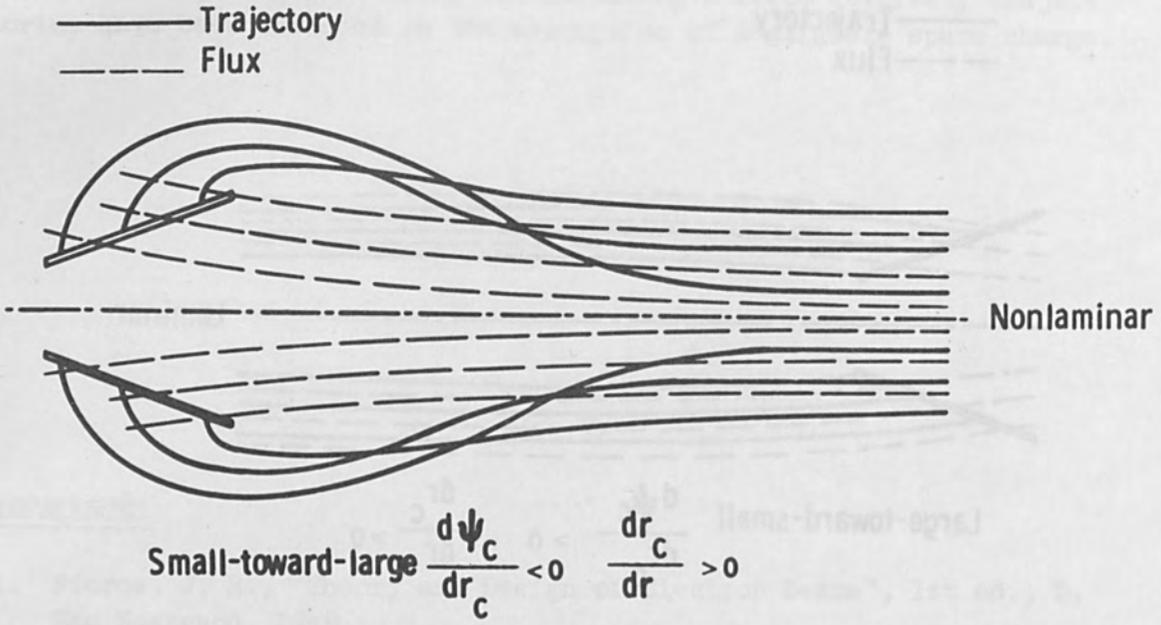


Figure 6c

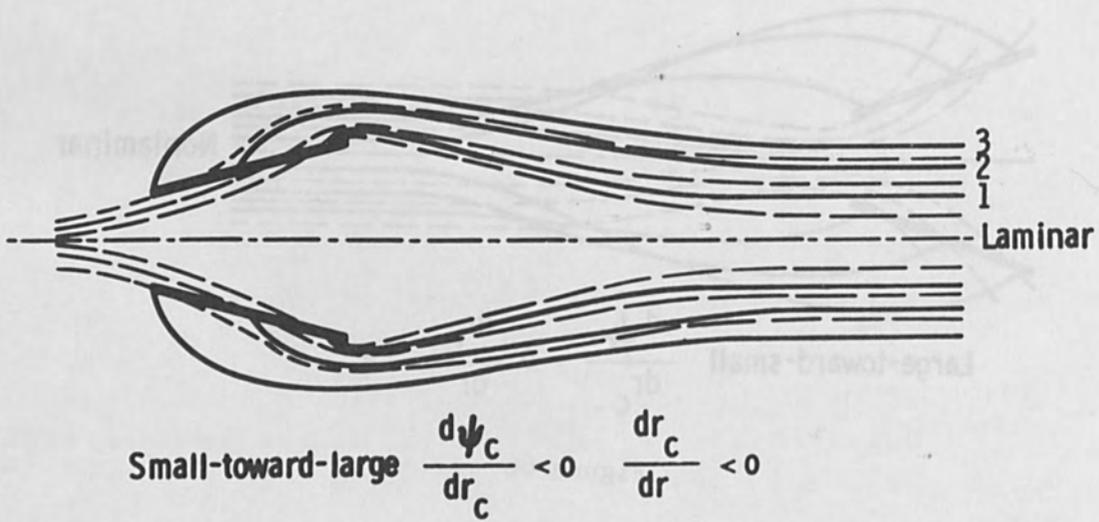


Figure 6d

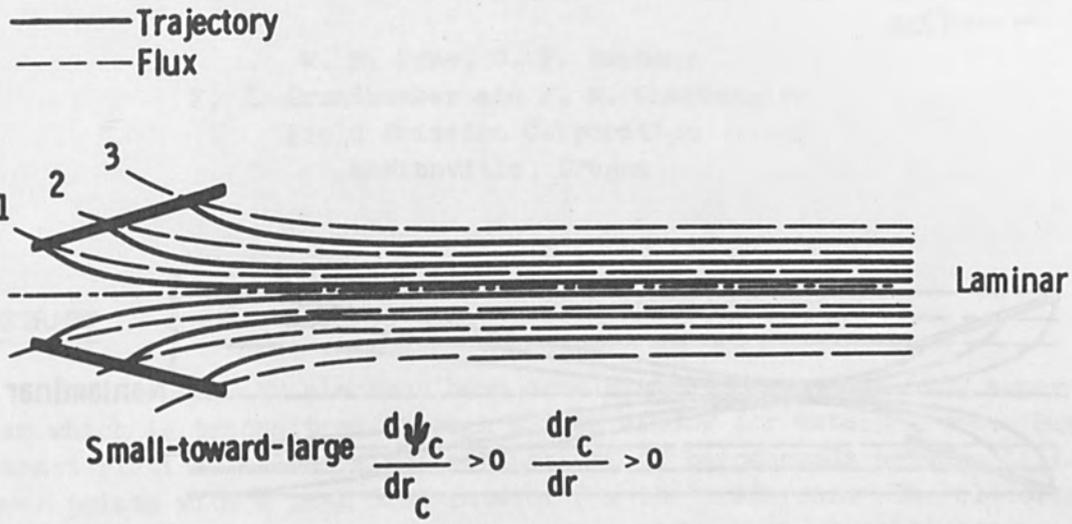


Figure 6e

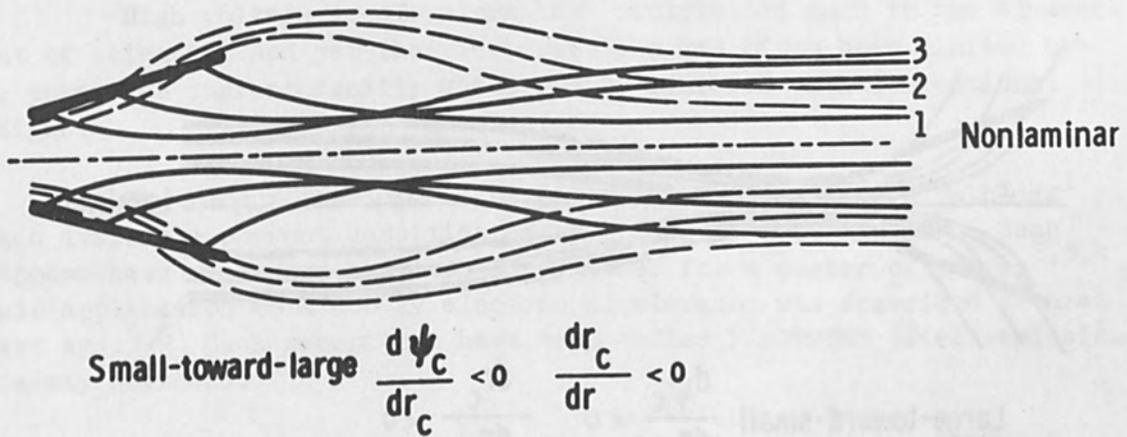


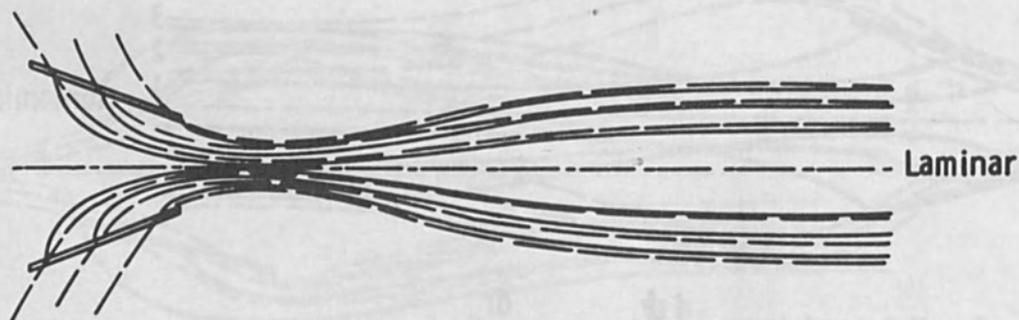
Figure 6f

— Trajectory
- - - Flux



Large-toward-small $\frac{d\psi_c}{dr_c} > 0$ $\frac{dr_c}{dr} > 0$

Figure 6g



Large-toward-small $\frac{d\psi_c}{dr_c} < 0$ $\frac{dr_c}{dr} < 0$

Figure 6h

AN ELECTRON ACCELERATOR WITH A PEAK BEAM POWER OF 10^{10} WATTS

by

W. P. Dyke, J. P. Barbour
F. J. Grundhauser and F. M. Charbonnier
Field Emission Corporation
McMinnville, Oregon

ABSTRACT

A new 2 Mv electron beam accelerator generates a 5000 ampere beam which is transmitted through a thin window for external use. Beam characteristics include a pulse length of 30 nanoseconds between half-power points with a peak dose rate of 2×10^{14} rads/sec. The electron beam penetration and energy density are continuously variable, giving flexibility and control in studies of the effect of the sudden deposition of energy in materials. The peak output is 10^{10} watts corresponding to an output power of 300 joules per pulse. In heavy materials, such as tungsten, a temperature rise of 500°K can be obtained in less than 40 nanoseconds. The output variation is less than 3 percent rms from pulse to pulse.

I. INTRODUCTION

High voltage accelerators have contributed much to the advancement of science. And yet their contribution has often been limited by the power and current density which can be achieved with conventional designs.

The recent development of practical field emission cathodes¹ makes available current densities of the order of 1000 amp/cm². Such cathodes have been used in flash X-ray tubes for a number of years.^{2,3,4} Their application to a 600 kv electron accelerator was described several years ago.^{5,6} Such generators have been called FEBETRONS (field-emission-beta-ray devices).

The present paper reports the extrapolation of the FEBETRON principle to an electron beam current of 10,000 amperes at 2.5 Mv. The electron beam is transmitted through a thin window for external use. An energy density of 100 joules/cm² is delivered in a pulse length of 30 nanoseconds. Dose per pulse and dose rate are of the order of 10^7 rads/pulse and 10^{14} rads/sec respectively.

The key features of the generator will be described briefly

and several applications will be noted.

II. ELECTRON BEAM TUBE

The FEBETRON tube is a diode in which electrons are emitted from a cold field emission cathode, accelerated in vacuum and then transmitted through a thin metal window (Figure 1). Small size is a characteristic of such structures. Typically, the electron discharge can be generated, focused and dissipated in a volume about the size of a golf ball at a peak power of 10^{10} watts (Figure 2).

In order to do quantitative research the beam characteristics (dose, wave shape, etc.) must be highly reproducible from pulse to pulse. In order to minimize the time spent on dosimetry and calibration it is desirable for the output to be identical from tube to tube. It has been possible to achieve both objectives by use of sealed-off envelopes with very high vacuum and a multiple needle cathode configuration. High vacuum avoids several mechanisms through which gas can affect the generator output in an uncontrollable manner: the Bennett pinch⁷ which alters beam cross section and hence current; stray conduction along tube walls due to gas desorbed under electron bombardment; electrode alteration by ion bombardment, etc. The multiple needle cathode configuration both insures that current is in phase with voltage and distributes current density evenly and reproducibly over a well-defined area. By use of these features dose is controlled with less than an rms variation of 3 percent (Figure 3). In fact the latter includes the errors in both the manual adjustment of charging voltage and in dosimetry. The residual dose variation is believed to be of the order of 1 percent.

III. PULSER

The high voltage is supplied by a Marx-surge pulser in which capacitors are charged in parallel at a relatively low voltage and then discharged in series to increase voltage. Switching is done with spark gaps.

The Marx-surge circuit has a number of basic advantages:

1. The high voltage exists for the shortest possible time -- about a factor of 100 less than in pulse charged systems and about 100 million less than in systems which use a dc charging voltage that is equal to or higher than the output pulse. The shorter the time, the greater the insulation strength. The latter can be traded off in such practical advantages as lower gas pressure (30 psi freon in the present case), smaller structures and reliability.

2. Switching is done at low voltage (30 kv) and at low energy

MODEL 545 TUBE

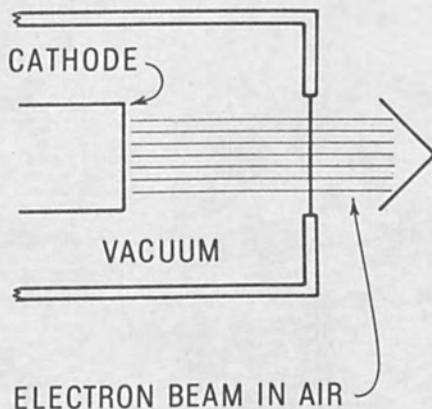


Figure 1. Schematic Representation of the Tube Used for Generating the High Current Beam.

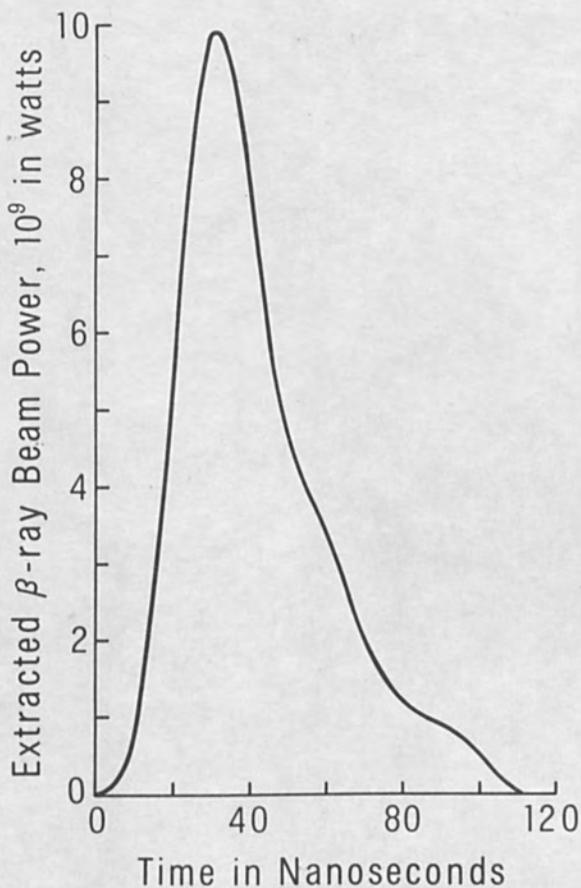


Figure 2. Beam Power is Shown as a Function of Time; the Data were Obtained from Oscillograms of the Current and Voltage Pulses.

Reproducibility of Electron Dose vs Depth Curves
Aluminum Absorber 2" from Tube Window
Cellophane Dosimetry Run of Ten Consecutive Pulses

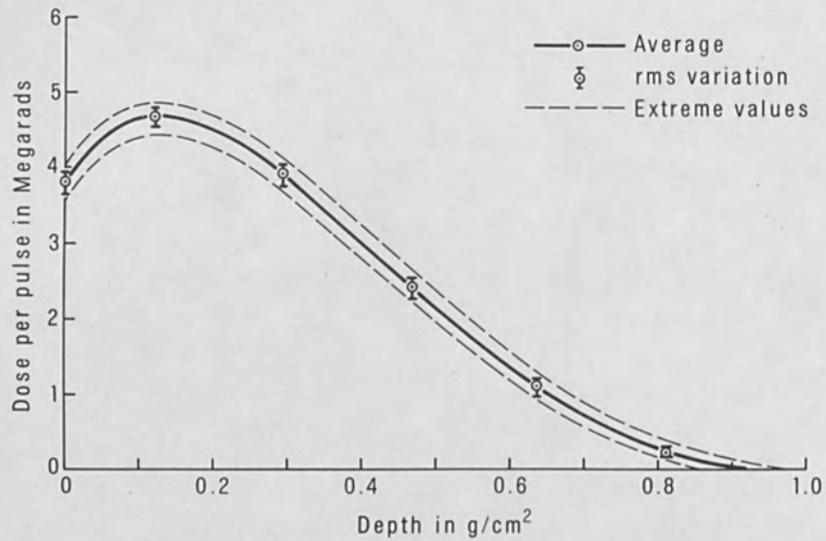


Figure 3. The Energy Deposition in Aluminum as Measured with Cellophane Dosimetry.

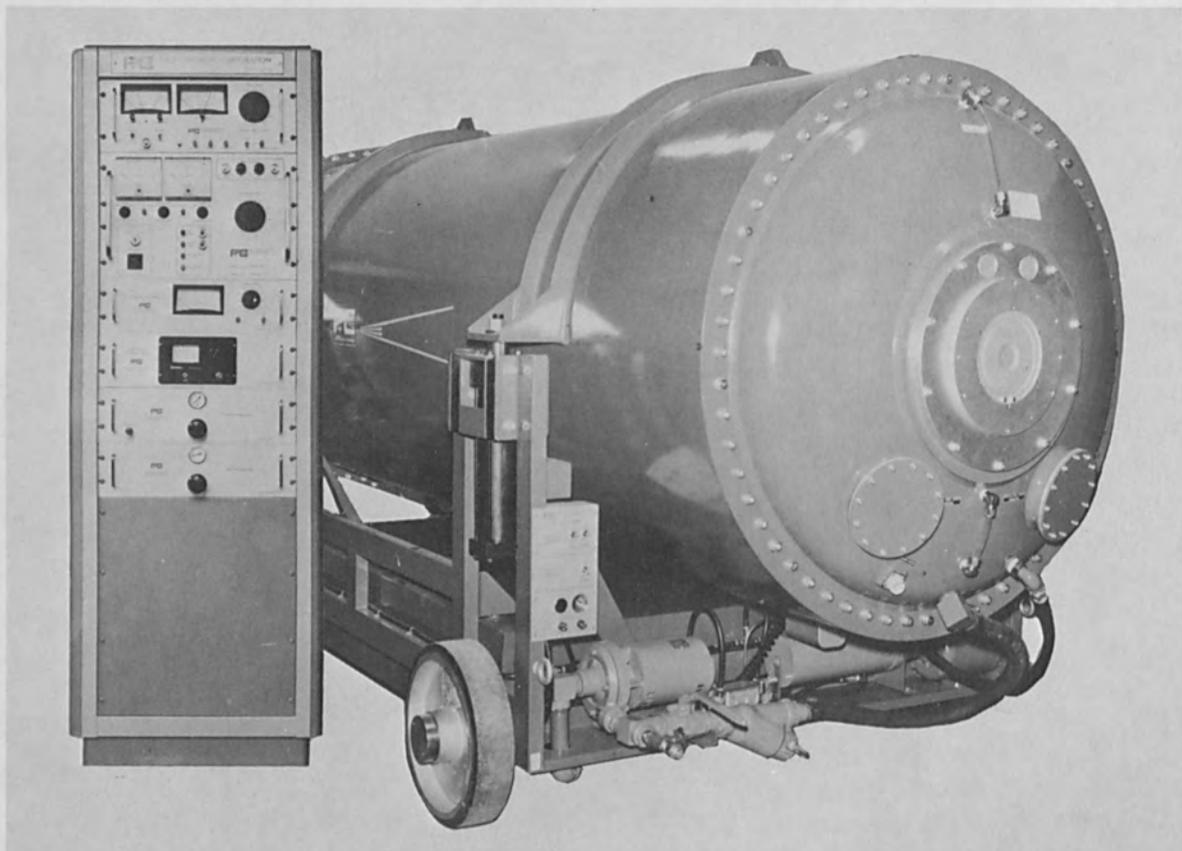


Figure 4. Complete 2 Mv System.

per switch. Thus it is possible to control the firing time of the pulser within 5 nanoseconds in order to synchronize the radiation pulse with the event under study. Also the gaps are stable, have long life and operate in a convenient atmosphere (nitrogen at 12 to 75 psi).

3. The capacitive storage isolates the generator's electrical performance from short-term line voltage fluctuations.

4. The voltage is evenly distributed along the length of the generator by use of modular construction.

Several disadvantages of earlier Marx-surge generators have been overcome with design improvements:

a. The discharge is enclosed in a pressurized metal container in order both to reduce noise (to the order of a millivolt on a 70-ohm cable at the tube face) and to isolate the electrical behavior from such atmospheric influences as dust, humidity, pressure, etc.

b. The discharge is made coaxial in order to optimize wave shape.

c. Artificial lines are used instead of simple capacitive storage in order to provide a voltage wave that is more nearly square (maximizes efficiency and minimizes size for a given output energy).

d. The simultaneous firing of multiple gaps has been made straightforward by use of improved gap designs with ultraviolet coupling. The resulting generator is shown in Figure 4. It is a simple extrapolation of the lower voltage machines which have been used for several years in flash X-ray machines.

IV. PERFORMANCE CHARACTERISTICS

The electron beam may be used directly for experimental purposes. Alternatively a tungsten target can be used to intercept the beam and generate X-rays. If the tungsten target is sufficiently thick to stop electrons the output radiation will then be in the form of X-rays only. A dose of 6 megarads is obtained with the electron beam at a distance of 2 inches from the tube window. This corresponds to a peak dose rate of 2×10^{14} rads/sec. A dose of 6000 rads is obtained at the surface of the tungsten target when the system is used to generate X-rays.

The dose obtained with the electron beam was determined by measuring the change in optical transmittance of a blue cellophane using techniques and calibrations described in the literature.^{8,9,10} The beam

current density was also measured with an apertured Faraday cup and the dose calculated from known values of dE/dx , the rate at which electrons lose energy in matter. The latter data indicate that the doses measured with cellophane dosimetry are conservative, and actual values are somewhat greater than those quoted. The variation of dose with depth of penetration was determined by interspersing cellophane dosimeter discs with sections of aluminum. The results are shown in Figure 3, and similar results have been obtained using other metals for the absorbing material. Data were obtained on 10 consecutive pulses and average values are shown along with the rms variation and the extreme values. The output voltage can be varied by merely adjusting the dc charging voltage, and Figure 5 shows data obtained at three different voltages. Both the dose and penetration depth increase with voltage, and the highest curve corresponds to a peak voltage of 2.1 Mv.

The high dose rates are extremely useful in the study of radiation effects on electronic components and circuits, as well as with biological effects and radiation chemistry. In some applications, it is necessary to have the dose more uniform throughout the depth of the specimen. When using a low atomic number absorber a high atomic number material may be suitably positioned so that backscattered electrons build up the dose. This is illustrated in Figure 6 where the dashed curve shows the dose versus depth relationship for monoenergetic electrons in beryllium. A tungsten plate was placed behind 0.62 gram/cm^2 of beryllium and the dose was measured in both the tungsten and beryllium. The solid curve shows the experimental data. The surface dose is built up because of the lower energy electrons which are accelerated during the rise and fall time of the pulse voltage. The dose is built up in front of the tungsten plate by the large number of electrons backscattered from the high atomic number material. The dose remains uniform within ± 5 percent for a beryllium thickness of 0.6 g/cm^2 . This technique would be especially useful in radiation chemistry where the irradiated samples are normally composed primarily of low atomic number materials.

The beam energy density has been measured on axis as a function of distance from the window. The measurements were made at atmospheric pressure and are shown in Figure 7. The solid curve shows values derived from the dose versus depth curves in aluminum taken at various distances from the window. Since by definition one megarad represents an energy absorption of 10 joules per gram, the beam energy density in joules/cm^2 is equal to 10 times the area under the dose depth curve, with the dose in megarads and the depth in gram/cm^2 . The electrical data were obtained using a Faraday cup with a small aperture to measure the beam current density and its time variation. The energy was then obtained with the product of the average electron energy and the beam charge density. Again the values measured with cellophane dosimetry are 30 percent lower than values obtained with direct electrical measurements.

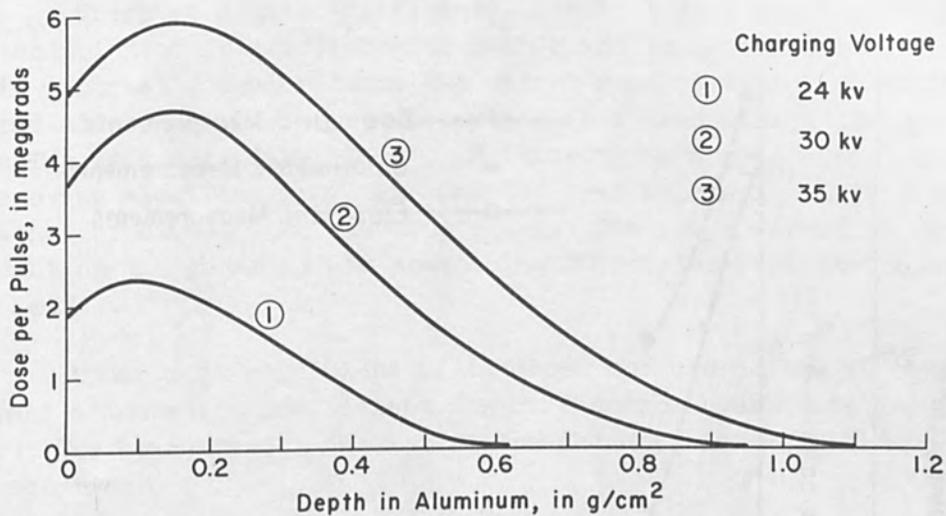


Figure 5. Data Obtained at Different Voltages Showing How Dose Rate and Penetration Depth Vary.

Highly Uniform Dose vs Depth, for Low Z Absorber Backed by High Z Reflector

- ① Be absorber, 0.62 g/cm² thick, backed by W plate Febetron at full power, absorber 2" away from tube window
- ② Monoenergetic electrons in solid Beryllium

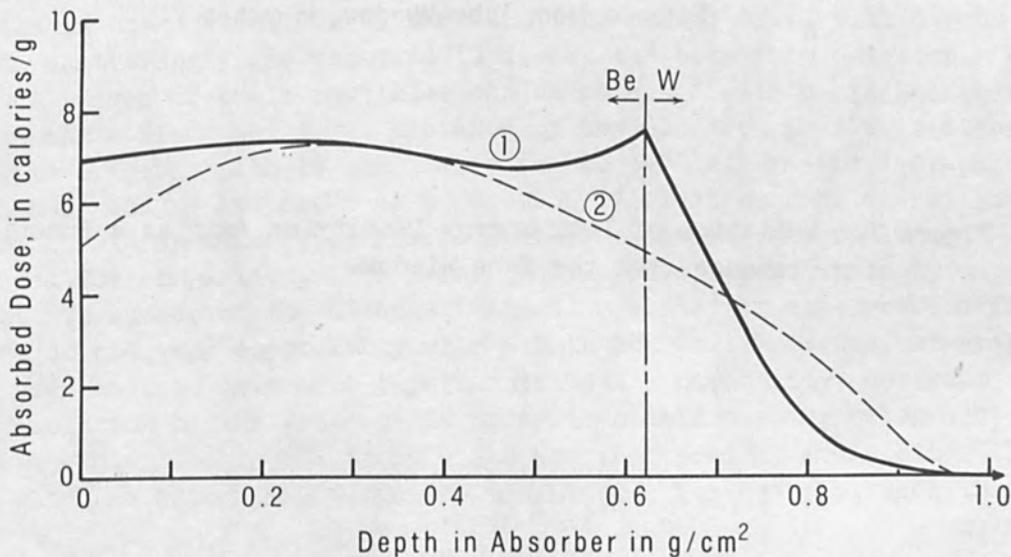


Figure 6. Curves Showing the Effect of Electron Backscatter on the Deposition of Energy in Beryllium.

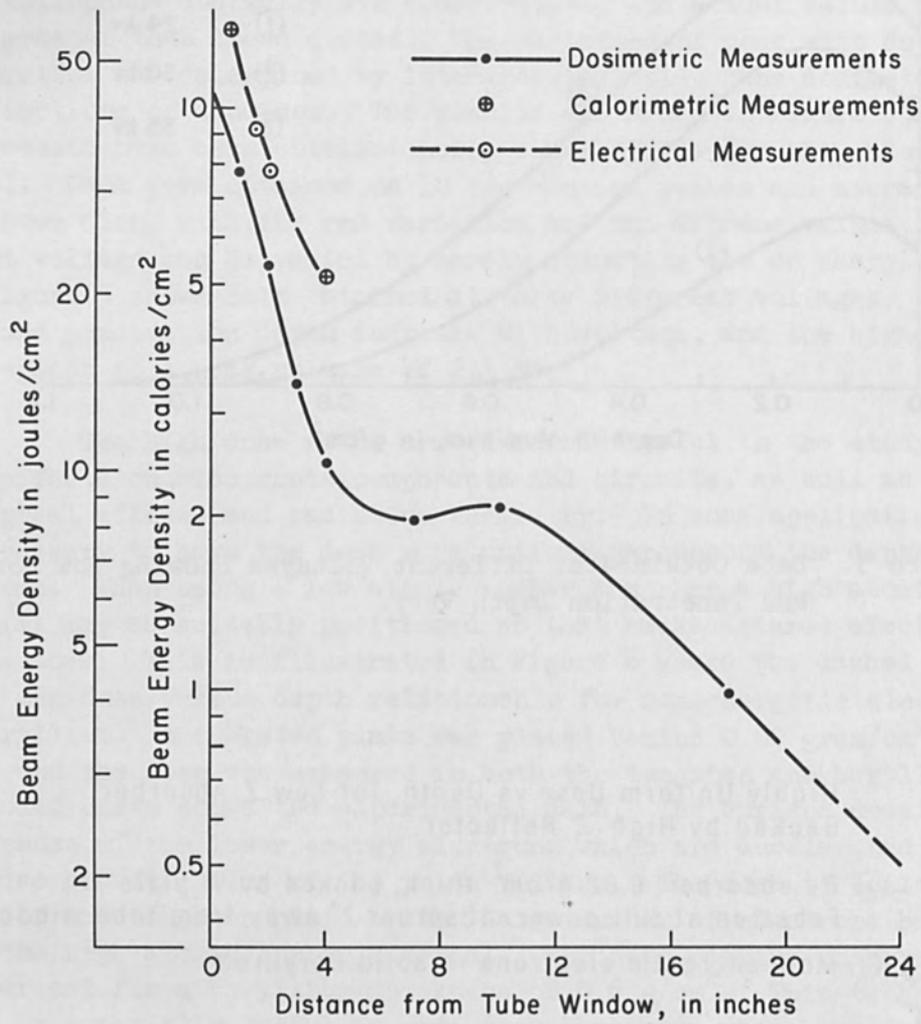


Figure 7. Variation of Beam Energy Density on Axis as a Function of Distance from the Tube Window.

resolution can be obtained with objects as small as 10 microns.

There is a basic difference between X rays and electrons from the standpoint of interaction with matter and image formation. Photographic contrast occurs because the object removes a portion of the incident radiation. With X rays, adsorption is a predominant process and scattering is a secondary effect. Scattering is a predominant mechanism for removing electrons from the directed beam unless the object thickness approaches or exceeds the electron range. The image formed is thus dependent on the geometry and scattering coefficients of the objects to be imaged.

Under most conditions of interest one can define an equivalent mass attenuation coefficient for an electron beam traveling through matter. The beam current density at the detector is described by the usual equation

$$J(x) = J_0 e^{-\alpha \rho x}$$

where J_0 is the incident current density without a scattering medium, α is the attenuation coefficient, ρ is the density of the medium and x is the distance traveled in the medium. There are two cases which are important experimentally and which can be treated with available electron scattering data. First, the case of small particles for which values of α_I are appropriate, and second extended dilute media like gas for which values of α_J apply.

Values of α_I have been obtained from published measurements on thin objects, and from the multiple scattering theory Moliere (as summarized by Bethe and Ashkin)^{11,12} for thicker objects. Schumacher¹³ has summarized both the experimental results and the type of apparatus required for accurate measurements. Since scattering is the dominant process in the attenuation of the beam, the values of α_I will depend on the scattering angle required to direct the beam from detector. Well-defined images of small particles can be obtained with a minimal scattering angle since the total distance of the electron must be scattered from its original path is comparable to the particle radius. Graphs of both α_I and α_J are shown as a function of electron beam energy in Figure 8. X-ray adsorption coefficients for several materials are included for comparison. The X-ray coefficients depend strongly on the atomic number of the attenuating medium while the electron coefficients do not vary appreciably with atomic number. Numerous betagraphs have been made to determine experimentally the capability, contrast and resolution of the system. In order to obtain a sharp betagraphic image the blur due to electrons which are scattered by media other than that which is being imaged must be minimized. For example, when the

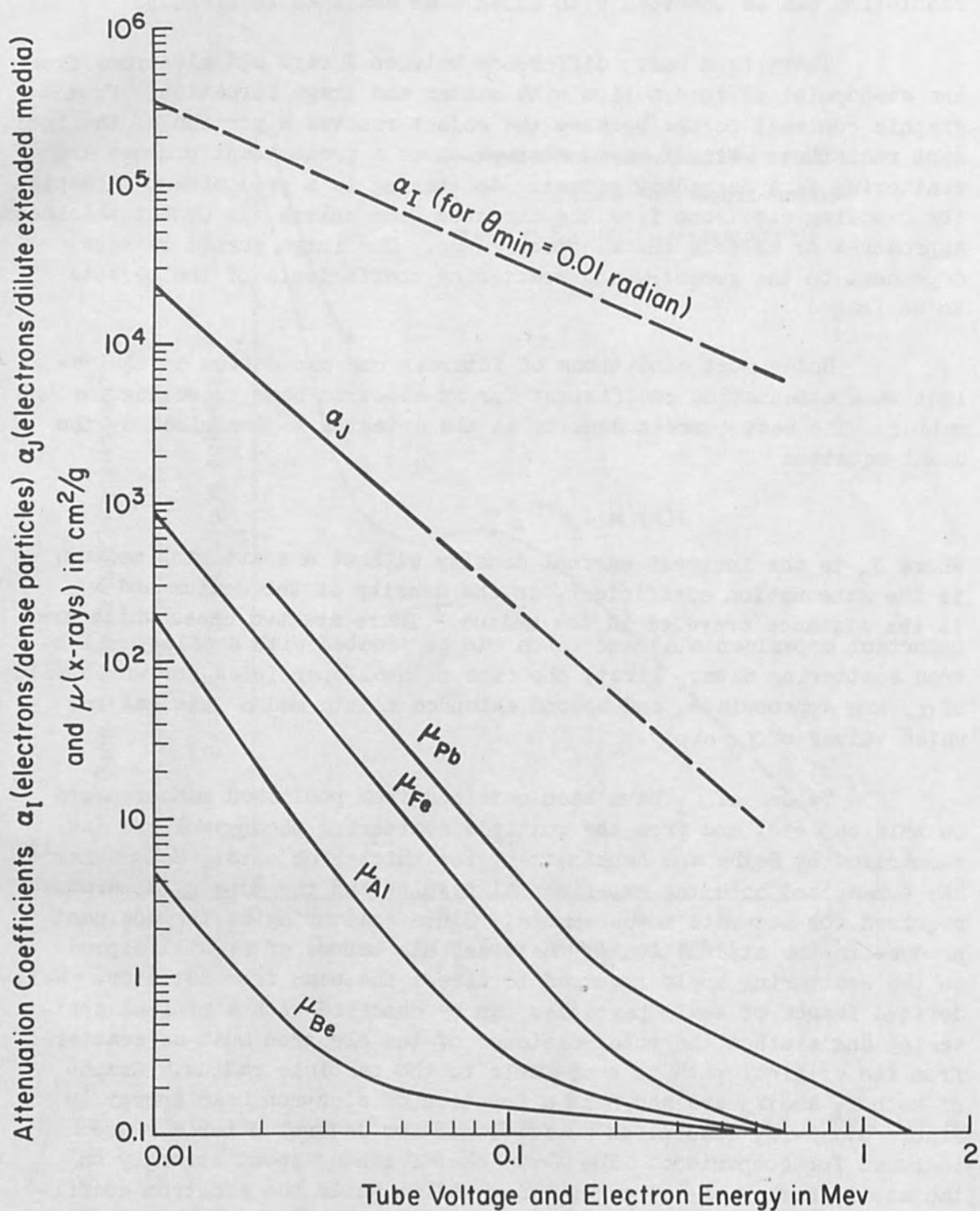


Figure 8. Graphs Showing Electron Beam Attenuation Coefficients α_I for Thin Particles and α_J for Extended, Tenuous Media and x-ray Absorption Coefficient μ for Several Materials as a Function of Tube Voltage and Electron Beam Energy in Mev.

In heavy materials, such as tungsten, the beam will cause a temperature rise of 1200°C in less than 40 nanoseconds. The temperature rise is thus 3×10^{10} degrees per second. The beam can thus be used to generate thermal shock waves and spallation of material. Preliminary data have been obtained using an external magnetic field to reduce further the beam diameter after transmission through the window. A substantial increase in energy density has been achieved and it is expected that power densities in excess of 10^{10} watts/cm² will be obtained. The system is thus suitable for the study of thermal damage and spallation with extremely short pulses.

V. BETAGRAPHY

The electron beams obtained from either the 600 kv or 2 Mv systems have sufficiently high current densities (1000 amps/cm²) so that high resolution electron shadowgraphs can be obtained.

Small objects of low atomic number material often do not absorb sufficient X-rays to yield a high contrast radiographic image. Since electrons interact strongly with matter, very small amounts of material will significantly attenuate the primary electron beam, and thus be detected on the image which the beam forms on a photographic film. The term betagraphy has been coined to describe this method of image formation. By simultaneous use of both X-rays and an electron beam originating from the same location, observation can be extended to a wide range of object sizes and densities. X-rays will elucidate the more dense portions of the subject, and the electron beam will sharply image the low density portions such as small particles, thin films and gas clouds.

There are several conditions which must be satisfied to obtain high resolution betagraphs of rapidly changing events. The three main factors which contribute limitations to the lateral resolution are:

- (1) diameter of the radiation source
- (2) motion blur caused by object movement during exposure
- (3) intrinsic resolution of the film

The effective source size can be controlled by placing a limiting aperture in the beam. The systems have sufficient current density so that a source size of 0.5 millimeters gives adequate film exposure. Thus penumbra contributed by the finite source size is in general negligible. The short exposure time of 20 nanoseconds also minimizes motion blur. The electron energy is efficiently used in exposing film so that even with the small source size and short pulse length described above a high resolution film, such as Kodak Type R, can be used. Thus, adequate

object is a small particle in air, the air may contribute scattered electrons. In principle, such electrons could be removed by special grids which are designed to collect scattered electrons. As an alternative the pressure can be reduced in the experimental chamber to minimize scattering collisions between electrons air molecules. For example, in the case of the small glass beads (0.3 to 3 mil diameter) shown in Figure 9 the pressure was reduced to 10^{-2} torr. The glass beads were in free fall at the time the exposure was made.

Both X-rays and electrons can be obtained from the arrangement shown schematically in Figure 10. This allows essentially identical positions for both of the radiation sources and their corresponding photographic films. Electrons were stopped in the first film where the beta-graphic exposure was made. X rays passed through that film without exposing it (because of slow exposure index, Kodak Type R) and proceeded to the second film where the radiographic exposure was made (because of the high exposure index needed, Kodak Royal Blue film was used together with phosphor intensifying screens).

The ballistic object shown in Figure 11 was arranged to provide a wide range of objects, sizes and density, wood chips to lead fragments, and to hide part of the event in an optically opaque region in the form of a hollow aluminum cylinder. The betagraph Figure 12a elucidates the less dense portions: wood W, wood chips WC, vapor caused by the wood impact VW, and vapor, presumably lead, released at projectile impact VL. The radiograph, Figure 12b records the more dense portions: 30:06 bullet B, lead disc L, lead fragments F and aluminum cylinder A. A composite frame of the two films can be easily made because of their common positions. It records particles with a size range of 1000 and a density range from wood vapor to metallic lead.



Figure 9. Betagray showing small glass beads (7 to 70 microns) in free fall 1 cm from photographic film; 0.5 mm electron source size; 600 kv accelerating potential; 45 cm source-to-film; Kodak Type R film.

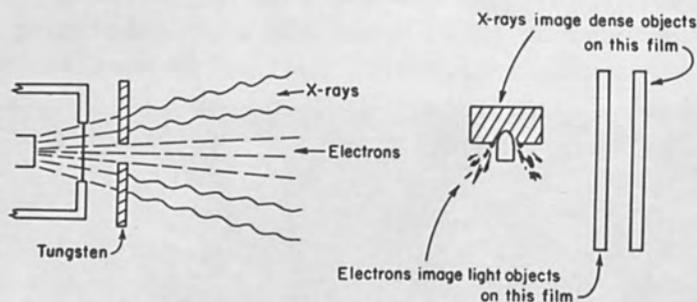


Figure 10. Drawing showing arrangement used for simultaneous radiographs and betagraphs; a portion of the electron beam illuminates the object, another portion forms x-rays at the external target; the first film records the electron shadowgraph (x-rays pass through this film making no exposure because of its slow speed); the second film records the radiograph (it receives no betagraphic exposure since electrons are stopped in the first film). Typically the first film was Kodak Type R (no screens) and the second was Kodak Royal Blue with Industrial phosphor screens.

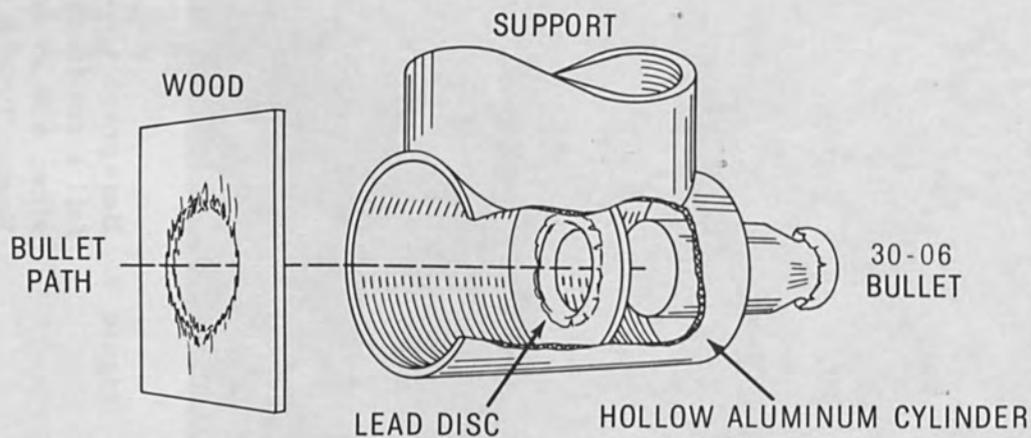


Figure 11. Drawing of ballistic object used in simultaneous radiographs and betagraphs; a 30-06 bullet penetrated first a wood then a lead disc target; the latter was "hidden" in a hollow aluminum cylinder.

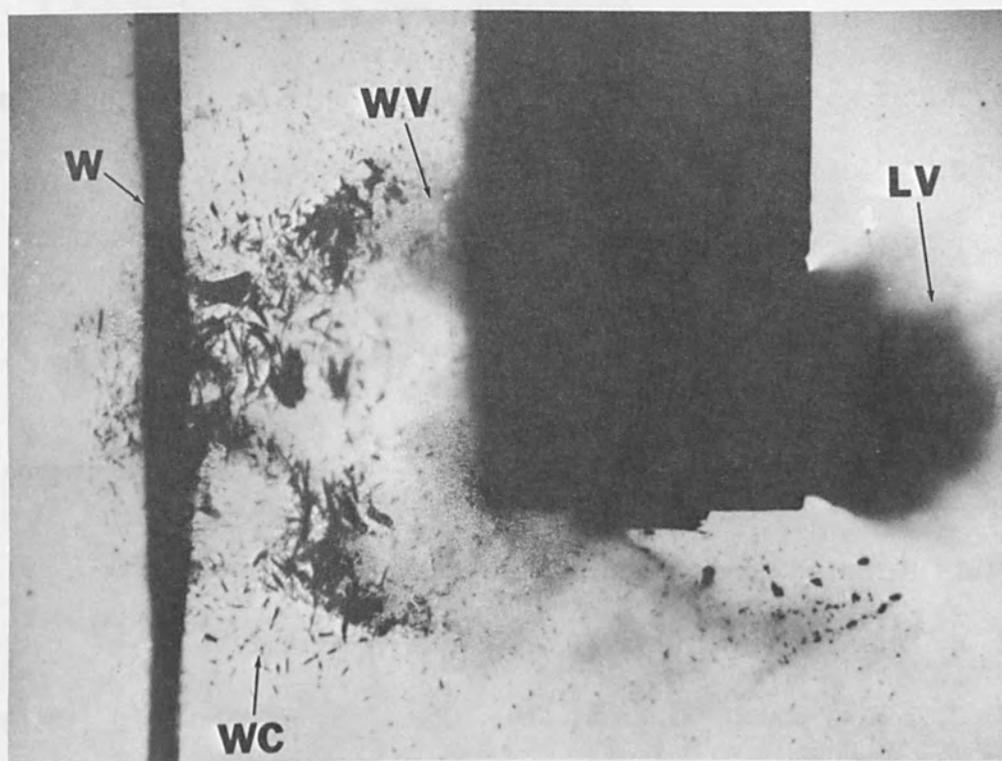


Figure 12a. Betagraph of event showing wood (W), wood chips (WC), vapor (WV) caused by wood impact and vapor (LV), presumably lead, released by projectile impact.

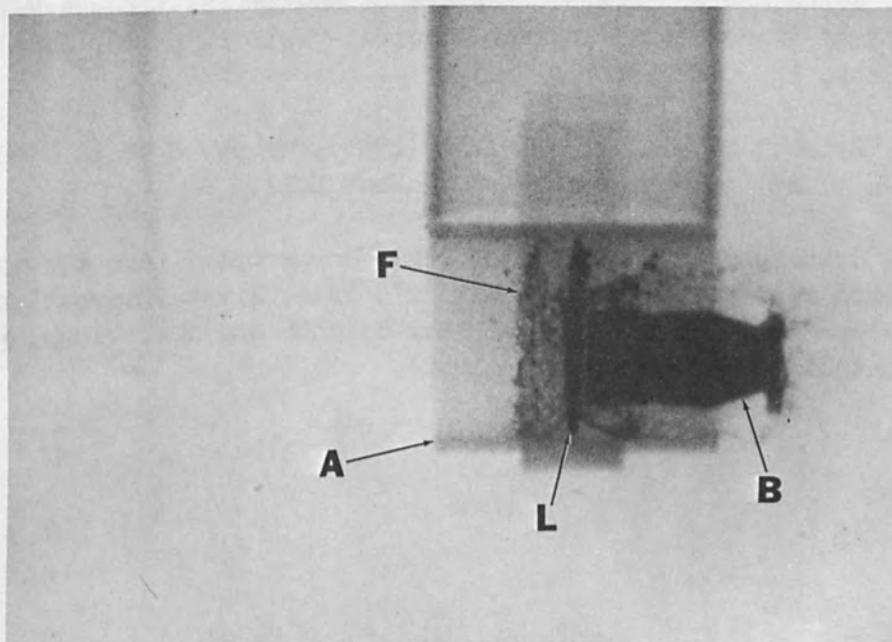


Figure 12b. Radiograph of same event showing 30-06 bullet (B), lead disc (L), lead fragments (F) and aluminum cylinder (A); this radiograph records the ballistic event shown in Figure 10 and was taken simultaneously with the betagraph shown in Figure 12a.

REFERENCES

1. W. P. Dyke, "Advances in Field Emission", Sci. Am., 210, 108 (1964).
2. F. J. Grundhauser, et al., J. SMPTE 70, 435, (1961).
3. W. P. Dyke, et al., J. SMPTE 71, 82, 1962.
4. F. J. Grundhauser, et al., "A New Two Million Volt Flash X-Ray Machine" 7th International Congress on High-Speed Photography, Zurich, 12-18 Sept '65.
5. F. J. Grundhauser, et al., "International Symposium on Aerospace Nuclear Propulsion", Las Vegas, Nevada, 1961.
6. Field Emission Corporation U. S. Patent No. 3173006.
7. W. H. Bennett, Phys. Rev. 45, 890, 1934.
8. E. J. Henley and D. Richman, Anal. Chem. 28, pp 1580-1582 (1946).
9. Goldblith and Mateles, Nucleonics 16, p 102 (1958).
10. A. Charlesby and R. J. Woods, Intl. J. of Appl. Rad. and Isotopes 14, 414 (1963).
11. H. A. Bethe and J. Ashkin: Passage of Radiations through Matter, pt. II of Experimental Nuclear Physics. Vol. I, p 166, John Wiley and Sons, Inc., New York (1952).
12. H. A. Bethe: Phys. Rev. 89, 1256 (1953), or pp. 8-38 to 8-40 of the Am. Inst. of Phys. Handbook, McGraw Hill.
13. B. W. Schumacher: "A review of the (Macroscopic) Laws for the Electron Penetration Through Matter", Proc. First International Conference on Electron and Ion Beam Science and Technology, published by John Wiley and Sons, Inc. (1965).

