



# RCA Review

*RC*

**September 1975**

**Volume 36 No. 3**

RCARCI 36(3) 381-651(1975)

RCA Review, published quarterly in March, June, September and December by RCA Research and Engineering, RCA Corporation, Princeton, New Jersey 08540. Entered as second class matter July 3, 1950 under the Act of March 3, 1879. Second-class postage paid at Princeton, New Jersey, and at additional mailing offices. Effective Jan. 1, 1971, subscription rates as follows: United States and Canada: one year \$6.00, two years \$10.50, three years \$13.50; in other countries, one year \$6.40, two years \$11.30, three years \$14.70. Single copies (except for special issues) up to five years old \$3.00.

## Contents

- 383** Introduction  
Richard Williams
- 385** From Camera Tubes to Solid-State Sensors  
Paul K. Weimer
- 408** An Approximate Model of the Beam-Blocking Contact in a PbO Vidicon  
Alvin M. Goodman
- 425** Semiconductor Heterojunction Vidicons  
C. R. Wronski
- 444** Simplicity in Theory: An Anecdotal Account of Current Injection in Solids—  
In Tribute to Albert Rose  
Murray A. Lampert
- 467** Photoconductivity in Amorphous Semiconductors  
Richard H. Bube
- 485** Speed of Response of Photocurrents in CdSe  
H. Kiess and B. Binggeli
- 499** On the Concept of "Lifetimes" in Photoconductors  
F. Stöckmann
- 508** Oxygen Chemisorption and Photodesorption Processes on ZnO Surfaces  
D. Eger, Y. Goldstein, and A. Many
- 531** A. Criterion for the Onset of Hot Electron Transport  
Richard S. Crandall
- 542** Structure and Energy of the Interface Between an Insulator and an  
Electrolyte  
Richard Williams
- 551** Solar-Cell Capacitance  
A. R. Moore
- 566** Basic Concepts of Charge-Coupled Devices  
W. F. Kosonocky and J. E. Carnes
- 594** Tunneling Conductivity in Granular Metal Films  
B. Abeles
- 621** Pattern Meaningfulness and Pattern Detectability  
Joseph J. Mezrich
- 632** An X-Ray-Sensitive Fiber-Optic Intensifier Screen for Topography  
R. W. Smith
- 638** Biography of Albert Rose
- 640** Technical Papers Published
- 643** Patents
- 646** Authors

## **RCA Corporation**

**Robert W. Sarnoff** Chairman of the Board and Chief Executive Officer  
**A. L. Conrad** President and Chief Operating Officer

## **Editorial Advisory Board**

**Chairman, J. A. Rajchman** RCA Laboratories

**A. A. Ahmed** Solid State Division  
**E. D. Becken** RCA Global Communications  
**G. D. Cody** RCA Laboratories  
**D. M. Cottler** Government and Commercial Systems  
**N. L. Gordon** RCA Laboratories  
**G. B. Herzog** RCA Laboratories  
**J. Hillier** RCA Research and Engineering  
**E. O. Johnson** RCA Research Laboratories, Inc., Tokyo  
**J. Kurshan** RCA Laboratories  
**C. H. Lane** Picture Tube Division  
**D. S. McCoy** Consumer Electronics  
**K. H. Powers** RCA Laboratories  
**R. E. Quinn** RCA Laboratories  
**P. Rappaport** RCA Laboratories  
**J. H. Scott, Jr.** RCA Laboratories  
**L. A. Sholliff** International Licensing  
**T. O. Stanley** RCA Laboratories  
**F. Sterzer** RCA Laboratories  
**J. J. Tietjen** RCA Laboratories  
**W. M. Webster** RCA Laboratories

**Secretary, Charles C. Foster** RCA Laboratories

**Editor** **Ralph F. Ciafone**

## **Associate Editors**

**W. A. Chisholm** RCA Limited (Canada)  
**J. E. Steoger** RCA Service Company  
**W. O. Hadlock** RCA Research and Engineering  
**D. R. Higgs** Missile and Surface Radar Division  
**W. A. Howard** National Broadcasting Company  
**C. Hoyt** Consumer Electronics  
**E. McElwee** Solid-State Division  
**C. A. Meyer** Distributor and Special Products Div.  
**M. G. Pletz** Government and Commercial Systems  
**C. W. Sall** RCA Laboratories  
**I. M. Seideman** Astro-Electronics Division  
**W. S. Sepich** Commercial Communications Systems Division

## Introduction

Television cameras have extended human vision throughout the world and into the depths of space beyond. Though its social impact is more recent, the development of electronic vision has engaged scientists and engineers over most of the past half century. In the process, they have contributed knowledge to many fields. Some of our most fundamental understanding of photoconductivity arose from analyses of the performance of the vidicon. A vigorous effort continues, to extend the capabilities of electronic vision and to improve our understanding of its properties.

Those who have worked in this field have felt the profound influence of Albert Rose and of his research in vision, electronics, and solid-state physics. His many published works tell only part of the story. To an extraordinary degree he has put his time and effort into helping his colleagues and guiding their work. For them, he has been a constant source of good advice, encouragement, and technical judgment, as well as a warm personal friend. This collection of articles, by some of the many people who have worked with him over the years, attests to the continuing activity in the field. It is dedicated to him on the occasion of his retirement after almost forty years of association with RCA Laboratories.

**Richard Williams**



# From Camera Tubes to Solid-State Sensors

**Paul K. Weimer**

RCA Laboratories, Princeton, N.J. 08540

**Abstract**—Electronic television has from its beginning depended on electron beams for picking up and reproducing the image. It now appears that self-scanned semi-conductive devices will be able to carry out many of the functions of camera and storage tubes. Research at RCA has played a significant role in the development of a large family of imaging tubes and their more recent solid-state counterparts. The present paper reviews the basic problems of image scanning and compares the solutions obtained by means of tubes and semiconductors.

## 1. Introduction

The operational requirements for image pickup are so demanding that the overall performance of a sophisticated television system is often limited by the characteristics of its camera tube. The use of television in astronomical research, space exploration, and low-light-level viewing has usually needed higher levels of camera performance than was available in commercial devices. Even though beam-scanned image tubes have been investigated continuously for more than fifty years, improvements are still being sought in image pickup and storage tubes.

During the past decade a parallel effort has been aimed at developing self-scanned solid-state image sensors that could substitute for camera tubes in many applications. Reductions in size, cost, and complexity were the immediate objectives, but the possibilities for improved performance and for entirely new applications of television

Name of Tube	Photocathode	Storage Target	Electric Beam	Video Signal Output	Ref.
 IMAGE DISSECTOR	Continuous transparent photoemissive cathode	None	No electron gun	Electrons from each picture element passed into electron multiplier	9
 ICONOSCOPE	Photoemissive target with insulated elements deposited on mica sheet backed by conducting signal plate		High velocity electrostatic focus	Video signal induced in target signal plate	10
 ORTHICON	Same as iconoscope except target signal plate is transparent		Low velocity magnetic focus. Electrostatic deflection	Video signal induced in target signal plate	11
 IMAGE ORTHICON	Continuous transparent photoemissive cathode	Thin semiconducting glass target with fine mesh screen to collect secondaries	Low velocity magnetic focus and deflection	Total modulated return beam passed into electron multiplier	20
 IMAGE ISOCON	Same as image orthicon	Same as image orthicon	Low velocity magnetic focus and deflection with scattered electron separator	Scattered fraction of modulated return beam passed into electron multiplier	21

Fig. 1—Five types of photoemissive camera tubes

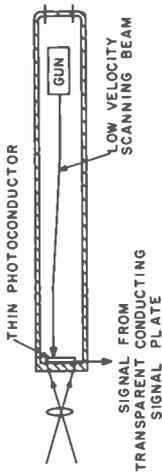
have also excited much interest. Although the basic problems of signal-to-noise ratio, resolution and uniformity were as demanding as ever, entirely new approaches were required for their solution. The recent advances in "charge-coupled"<sup>1</sup> and "charge-injection" devices<sup>2</sup> have lent much encouragement to the belief that solid-state sensors will soon play a significant role. Already these sensors are finding unique applications for which they are particularly suited. However, the most advanced solid-state sensors developed to date do not match the best camera tubes in sensitivity or resolution, and further improvements in sensors will require a sustained effort in both sensor design and in silicon technology.

Solid-state analog memory devices, analogous to beam-scanned storage tubes, should also emerge as a by-product of the work on solid-state sensors. Nondestructive readout of analog storage elements has already been demonstrated in signal correlators<sup>3</sup> and in charge-injection image sensors.<sup>4</sup> The possible application of analog memories to advanced types of computers may also stimulate work in this area.

The objective of the present paper is to examine the beam-scanned and solid-state approaches to imaging devices by comparing the solutions to the problems of sensitivity, signal-to-noise ratio, resolution, and spurious signals found for each approach. It will be assumed that those unfamiliar with the operation of these devices will be able to follow the discussion by reference to Figs. 1-5. More detailed discussion of camera tubes and solid-state sensors are given in review papers<sup>5,6</sup> and in the original references listed in the figures. A discussion of image pickup in its broadest terms will be found in Rose's early paper,<sup>7</sup> "Television Pickup Tubes and the Problem of Vision," and in his recent book<sup>8</sup> entitled *Vision—Human and Electronic*.

## 2. Charge Storage in Camera Tubes and Image Sensors

One of the earliest principles learned in the design of television camera tubes was that useful operating sensitivity would require the continuous gathering and storage of picture information throughout the entire  $\frac{1}{30}$  sec frame time. The human eye has storage ( $\sim 0.2$  sec), as do all camera devices except the Farnsworth image dissector. The image dissector,<sup>9</sup> which contained a high-gain signal multiplier behind the scanning aperture, would have been a nearly ideal pickup device (with large signals and very low noise in the dark) if it had been able to utilize all the photoelectrons generated at each element *between* successive scans. Without storage, it became the least sensitive of the photoemissive tubes listed in Fig. 1.



Name of Tube	Photoconductive Material	Target Construction	General Characteristics	Ref.
Vidicon	Sb <sub>2</sub> S <sub>3</sub>	Thin porous photoconductor on transparent conducting signal plate	Good sensitivity, perceptible lag at low light levels, $\gamma = 0.7$ , high resolution, easy to manufacture	12
Lead Oxide Vidicon	PbO	Approximately same as the vidicon	Higher sensitivity, reduced lag, $\gamma = 1.0$ , good resolution, difficult to manufacture	48
Silicon Vidicon	Si	Array of p-type photodiodes on thin n-type silicon wafer	Excellent sensitivity, low lag, impervious to burned-in images, resolution potentially limited by photodiode structure, costly to manufacture	15
Cadmium Selenide Vidicon	CdSe (+ ZnS or AS <sub>2</sub> S <sub>3</sub> )	Continuous double-layer	Currently under development as competitor to lead-oxide vidicons	14
Miscellaneous Experimental Vidicons	Amorphous selenium, granular metal barriers, etc.	Hetero-junctions, multi-layers, etc.	Promising approaches, but usually deficient in some major characteristic	50 56 57 58

Fig. 2—Photoconductive camera tubes

The first camera tube employing storage was the iconoscope, invented by Zworykin<sup>10</sup> over fifty years ago. In this tube the light from the scene was imaged upon an insulated photosensitive area called the "mosaic." An array of capacitors with one plate in common served to integrate the charges produced by photoemission of electrons from the mosaic. The increase in signal produced by full storage is given by the ratio of frame time to element time—a factor of about 250,000 in the usual television system. If the iconoscope had had *efficient* storage it would have been far more sensitive than the image dissector, instead of only a few times, as was actually the case. A major loss in storage efficiency occurred because of the self-discharging effect of secondary electrons from the high velocity scanning beam raining back on the mosaic, or "target" as it was called later. This problem was eliminated by Rose and Iams<sup>11</sup> with their development of the orthicon. This tube utilized a low-velocity (i.e., low-energy) beam whose secondary-emission ratio at the target was less than unity. Such a beam required a decelerating electric field in front of the target thereby eliminating the rain of secondary electrons that had discharged the target prematurely.

The video signal was generated capacitively in the common signal plate of the target when the low-energy beam replaced the electrons that had been ejected from each element by photoemission. Low-velocity scanning, which permitted 100% efficiency in storage and discharge, was a major advance in electron-beam technology that was subsequently adopted in virtually all camera and storage tubes. However, the sensitivity of the orthicon was still relatively low by modern standards because of the low responsivity of its insulating photoemissive target. A very much higher yield of electrons per photon was achieved later with photoconductive targets such as used in the vidicon.

The advantage of the vidicon<sup>12</sup> over earlier attempts at photoconductive pickup was largely derived from its efficient charge storage. This was achieved by combining the efficiency of low-velocity scanning with the selection of photoconductive materials whose dark relaxation time exceeded the television frame time (1/30 sec) (see Fig. 2). By this means, it was found that insulating photoconductors could be used whose photoconductivity had not even been detected in prior tests. Low-energy scanning of the photoconductors also reduced (but did not completely eliminate) the effects of spurious conductivity induced in the target by the scanning beam. It may be mentioned parenthetically that the high photoconductive gain in well-known photoconductors, such as cadmium selenide, offered an alternative (but less desirable) method of light integration by *excitation* storage rath-

er than by *charge storage*. However, it has not proved feasible to obtain charge storage simultaneously with photoconductive gains significantly greater than unity.<sup>13</sup> Photoconductive tubes now in use do not exceed unity gain even when a material such as cadmium selenide<sup>14</sup> is used.

The *silicon vidicon*<sup>15</sup> achieved an important advance in image pickup by demonstrating that the high resistivity required for television frame storage could also be obtained in silicon with reverse-biased photodiode arrays. The reduced trap levels permitted higher responsivities (up to one charge per photon) and improved response times (less lag in moving images). Equally important, the results provided further incentive for the development of self-scanned devices combining the integrated-circuit capabilities of silicon with its photosensitive and charge storage properties.

Although the silicon photodiode arrays had good responsivity and good uniformity, the small signal output from these elements introduced noise problems when they were first incorporated into self-scanned sensor arrays. Another approach, which was later abandoned for reasons of nonuniformity, made use of phototransistors<sup>16</sup> which provided both signal gain and charge storage. The most promising solid-state sensors for television today utilize reverse-biased MOS capacitors<sup>1,2</sup> instead of diffused photodiodes<sup>17</sup> as the photosensitive elements. The picture charge is accumulated at the interface between the reverse-biased semiconductor and the oxide layer without requiring a diffused photodiode at each element. While the responsivity of such elements is no better than that of diffused photodiodes, they offer advantages in ease of fabrication and in the readout of signals. The low signal levels at each element are tolerable provided the scanning noise can be kept small. The reduction of scanning noise in self-scanned sensors is a major problem which has been greatly alleviated by the newly developed charge-transfer methods.<sup>18,19</sup> However, as seen in the next section, solid-state sensors have not yet equalled the low-light-level performance of the most sensitive camera tubes.

### 3. Signal-to-Noise Problems in Imaging

The noise sources in a television imaging device can be divided into three categories: photon fluctuation noise in the light image, scanning noise introduced by the device, and circuit noise from the output amplifier. The objective of the device designer is to make the last two sources inconsequential compared to the first.

One reason that the iconoscope and the orthicon required very high scene illumination was because the unamplified video output signal

was obtained directly from the target "signal plate" which coupled to all elemental capacitors. The resultant signal-to-noise ratio was limited by the noise threshold of the input stage of the video amplifier. ( $\sim 2.0 \times 10^{-9}$  A for a pass band of 5 MHz.) The noise level of the amplifier was increased by the high shunt capacitance of the target and output lead to ground (10–20 pF).

The image orthicon<sup>20</sup> incorporated three features that accounted for its greatly increased sensitivity over the orthicon:

- (1) The separation of the photoemission and storage functions by means of a separate photocathode and target permitted a more sensitive type of photocathode to be used.
- (2) The acceleration of electrons from the photocathode to the target gave a secondary emission gain of 3–5 times in stored charge at the target.
- (3) The signal was derived from an electron multiplier that amplified the fraction of the beam that was not required to discharge the target.

The return-beam signal was equal and opposite in polarity to the signal generated at the target, but it had the advantage that it could be amplified by a substantially noiseless multiplier to such a level that amplifier input noise was of no consequence. The electron multiplier introduced a problem of uniformity, however, since the return beam retained some scanning motion that could detect spatial variations in gain of the first stage.

The image orthicon, which became the work horse of broadcast television in the 1950's, was for many years the most sensitive camera tube available. It was recognized very early, however, that the signal-to-noise ratio in the image orthicon was limited by the shot noise in the return beam, which unfortunately was a maximum in the dark portions of the picture where the noise was most visible. A substantial reduction in noise at very low light levels was obtained in the image isocon,<sup>21</sup> a modification of the image orthicon in which only the scattered portion of the returning electrons was allowed to enter the multiplier. Since electrons were scattered by the target only in the lighted areas, the polarity of the output signal was opposite to that of the image orthicon and the percentage modulation of the signal entering the multiplier was much higher.

An entirely different approach to improving the signal-to-noise ratio of camera tubes was to introduce a large gain in the image charge prior to storage. (See Fig. 3.) By this means the video signal produced at the target could be made large enough to completely dominate the amplifier noise. Under such conditions a simple low-velocity scanning beam without a signal multiplier is adequate. The

Name of Tube	Description of Intensifier System	Total Gain Prior to Storage	Ref.
SEC TUBE (SECONDARY-EMISSION CONDUCTIVITY TARGET)	High energy electrons from photoemissive cathode strike porous KCL target producing enhanced charge pattern by internal secondary emission within KCL. Low-velocity scanning beam generates signal at the target.	200 at 8 kV (from photocathode to target)	22
SIT TUBE (SILICON INTENSIFIER TARGET)	High energy electrons from photoemissive cathode strike thin silicon vidicon target producing bombardment-induced conductivity in silicon target. Low-velocity scanning beam generates video signal at the target.	2000 at 8 kV (from photocathode to target)	23
I-ISOCON (IMAGE INTENSIFIER FOR IMAGE ISOCON)	Image intensifier tube coupled to image isocon tube by fiber optic face plates provides a light gain of approximately 40 times	$40 \times 3$ (120 total)	24
I-SEC (IMAGE INTENSIFIER FOR SEC TUBE)	Image intensifier tube coupled to SEC tube by fiber optic face plates provides a light gain of approximately 40 times	$40 \times 200$ (8000 total)	24
I-SIT (IMAGE INTENSIFIER FOR SIT TUBE)	Image intensifier coupled to a SIT tube by fiber optic face plate provides a light gain of approximately 40 times	$40 \times 2000$ (80,000 total)	24

Fig. 3—Five types of low-light-level camera tubes employing image intensification prior to storage

SEC ("secondary-emission-conductivity") tube<sup>22</sup> employed internal secondary emission in a porous KCl target to obtain charge gains of several hundred times. The target capacitance was made sufficiently great that the large signal levels did not cause saturation. A second type of image-intensifier tube uses bombardment-induced conductivity in a thinned silicon vidicon target.<sup>23</sup> This tube, which is known as the SIT tube (silicon-intensifier-target) gives charge gains up to 2000 times. Such tubes are highly sensitive but do not quite reach the goal of detecting individual photoelectrons. Their sensitivity can be increased still more by adding another stage of image intensification. Separate light-intensifier image tubes have been coupled to the SIT, SEC, and image-isocon tubes by use of fiber-optic faceplates to produce useful signals under overcast starlight illumination.<sup>24</sup> Fig. 6 shows how closely the SIT and the intensifier "I-SIT" tubes approach the theoretical limits of an ideal device for that type of photocathode. The only way of improving their operating sensitivity further would be to use a more efficient photocathode.

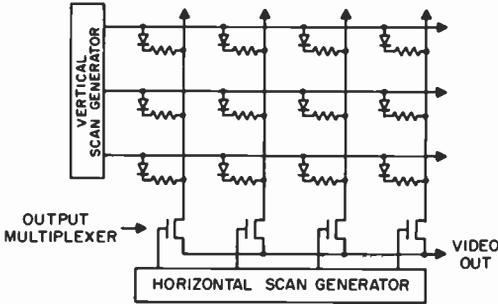
The restrictions imposed on self-scanned sensors have made the achievement of adequate signal-to-noise ratios at low light levels a major problem. Since the objective has been to produce a low-cost, solid-state device, the addition of a conventional vacuum-tube image-intensifier stage to a self-scanned solid-state sensor,<sup>25</sup> although improving its sensitivity, would not satisfy that objective. The simplest solution would be to use a method of solid-state scanning that does not introduce excessive scanning noise. The problem of obtaining adequate sensitivity in a solid-state sensor is eased slightly by the fact that in silicon the average quantum yield for visible light is several times larger than that obtained in the photoemissive cathodes used in intensifier tubes.

The first types of self-scanned sensors to be investigated consisted of arrays of photoconductive<sup>26</sup> or phototransistor elements<sup>16,27</sup> connected to perpendicular *XY* address strips (Fig. 4.). Scanning was accomplished by applying scanning pulses to the strips while taking the video signal from the strips or from the common substrate. The pulse generators, which selected the proper rows or columns, were digital shift registers or decoders formed on the periphery of the array.

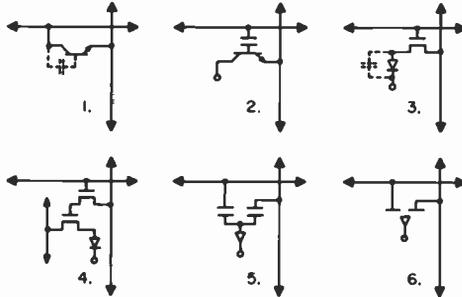
The scanning of *XY* sensors in this manner produces two types of noise: (1) fixed pattern noise arising from nonuniform scan pulses or imperfections in the structure, and (2) fluctuation noise in the output amplifier, which is enhanced by the relatively large capacitances shunting the address strips or output bus. The output capacitance is particularly high when the entire substrate is used to collect the signal from all elements. Its limits on signal-to-noise ratio can be toler-

ated only at high light levels or in relatively small arrays such as the  $50 \times 50$  element sensors now on the market.<sup>28</sup>

In large sensors (such as the early  $256 \times 256$  element photoconductive sensors,<sup>26</sup> or the recent  $244 \times 248$  charge-injection sensors,<sup>4</sup>) the preference has been to transfer the signals from an entire row simultaneously into the column buses where the signal is held temporarily



A. DIGITALLY SCANNED XY SENSOR WITH PHOTOCONDUCTIVE ELEMENTS



B. SILICON SENSOR ELEMENTS FOR XY SCANNING

Fig. 4—Solid-state image sensors with XY address strips. A. Typical digital readout circuit with external multiplexer (photoconductive array, Ref. [26]). B. Various types of silicon XY sensor elements: (1) phototransistor (Ref. [16]); (2) phototransistor (Ref. [27]); (3) photodiode (Ref. [17]); (4) photodiode (Ref. [28]); (5) charge injection photodiode (Ref. [26]); and (6) charge-injection MOS capacitors (Ref. [2]).

until it can be discharged sequentially into a horizontal output bus. The scanning noise can be minimized by keeping the capacitance of the columns and the output bus as small as possible, and by integrating the output amplifier on the same chip with the sensor. Since the rms noise associated with resetting the excess capacitance varies as the square root of the capacitance,<sup>29</sup> the threshold noise level for XY

scanning of a large array could be 10–50 times the noise of discharging a single element.

The development of image sensors incorporating internal charge-transfer registers (see Fig. 5) offers an opportunity to reduce the scanning noise to approximately that of discharging a single element. A charge-coupled sensor usually consists of a vertical array of CCD registers that are illuminated either through transparent gates or from the reverse side through the thinned substrate. The picture charge is allowed to build up under the reverse-biased MOS capaci-

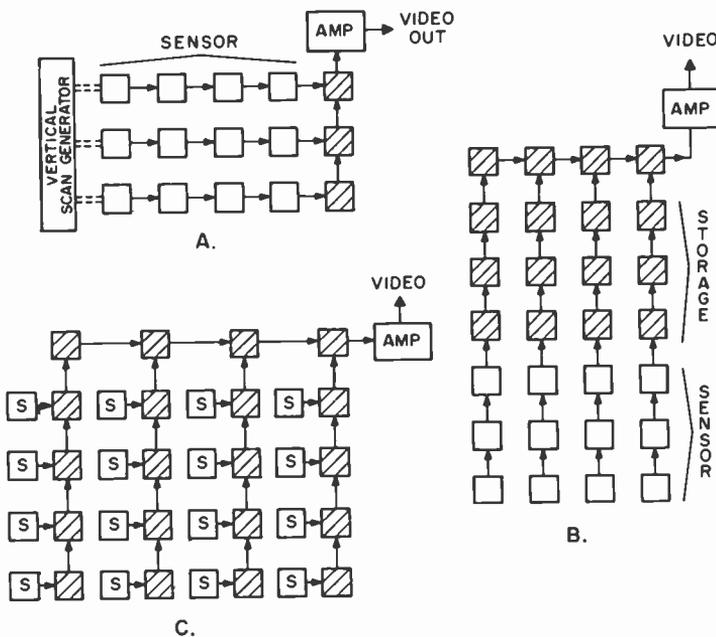


Fig. 5—Solid-state image sensors employing charge-transfer registers for scanning. A. Horizontal line-transfer system used in early 32 X 44 element bucket brigade sensor (Ref. [32]). B. Vertical frame-transfer system used in most CCD sensors (Ref. [1]). C. CCD sensor with nonilluminated registers and interleaved sensor elements (Ref. [47]).

tors in the sensor area during the normal field period while the transfer through these registers is stopped. All registers are then activated during the vertical retrace period causing the charge pattern to be transferred intact from the sensor area into a temporary storage area. During the next scanning field, the stored charges continue to advance in parallel toward the top of the array where each row is trans-

ferred out in sequence at the horizontal rate by the charge-coupled output register.

The favorable noise characteristics of the CCD sensor arise from the fact that the capacitance of each stage of the registers leading from the sensor element to the output amplifier on the same chip is no larger than the capacitance of the sensor element itself. Since the noise introduced by the transfer process itself can be kept relatively small, the CCD design appears to offer the possibility of producing a low-light-level sensor suitable for starlight illumination without requiring image intensification such as has been necessary with tubes. To meet this objective, however, all available light should be utilized in the silicon and all sources of scanning noise in the output must be kept to a minimum. Fig. 6 compares the low-light performance of sev-

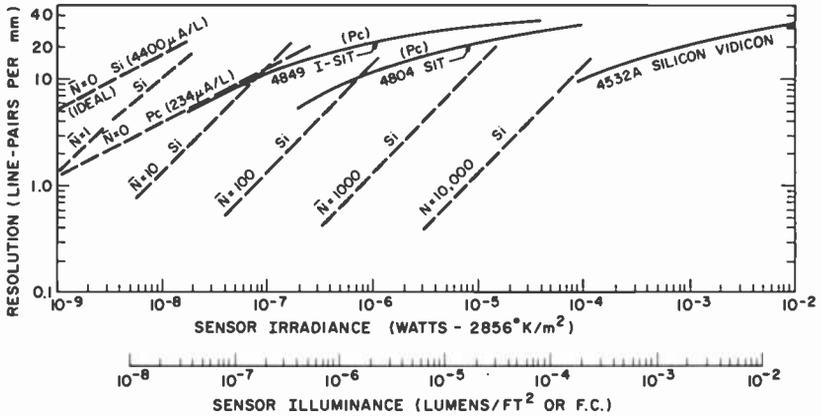


Fig. 6—Predicted low-light-level performance of silicon solid-state sensors (broken curves) compared with three existing camera tubes (the silicon vidicon, the SIT tube, and the I-SIT combination). The parameter  $\bar{N}$  represents the total extraneous scanning noise of the sensor expressed in electrons per integration period  $t$ . (The value of  $t$  is assumed to be 0.2 sec, to match the storage time of the eye.) (Refs. [6], [30]).

eral existing camera tubes with the predicted performance of solid-state sensors for different values of noise fluctuation at the output.<sup>30,6</sup> The responsivity of the solid-state sensor has been assumed to be 4400  $\mu\text{A}/\text{lumen}$ , or 87  $\text{mA}/\text{watt}$ , of illumination from a tungsten lamp at 2856°K. This responsivity corresponds to that of a thinned back-side-illuminated silicon vidicon target with unfiltered tungsten light. It is noted that the solid-state sensor could approach the low-light performance of the I-SIT tube provided its total scanning-noise

fluctuations  $\bar{N}$  are kept below 10 electrons per integration period. (The integration period is assumed to be 0.2 sec to match the integration time of the eye.) The total scanning noise in the CCD sensor includes the transfer and trapping noise, the shot noise in the background currents, and reset noise in the on-chip amplifier. The suppression of noise fluctuations to the required low level has been predicted for cooled, buried channel CCD sensors employing distributed floating-gate output amplifiers.<sup>31</sup> However, the above curves do not include fixed-pattern noise arising from scanning voltage pickup or from spatial variations in sensor dark current. A full comparison of solid-state sensors with intensifier camera tubes for very low light level applications will have to include other factors such as highlight resolution, freedom from spurious signals, convenience, and cost. The degree of inconvenience introduced by the need to cool the silicon target for low-light-level operation has not been evaluated.

In solid-state camera applications, in which the light levels are large enough that cooling would not be required the signal-to-noise advantages of an internally scanned CCD are not as significant. The recent advances in the *XY* scanning circuits<sup>2</sup> used for the charge-injection sensors have reduced both the fluctuation noise and the fixed-pattern noise which plagued earlier *XY* sensors. Although the reset noise is 10–20 times higher than for a CCD owing to the capacitance of the address strips, the total  $\bar{N}$  scanning noise with an on-chip amplifier could still be less than that of a typical silicon vidicon, which is considered a relatively sensitive tube. The more open-gate structures permitted by the use of *XY* address strips (instead of internal registers) may permit more effective introduction of light from the gate side without requiring the more expensive thinning process necessary for back-side illumination. Alternative methods of scanning *XY* arrays by use of external charge-transfer registers<sup>32</sup> also offer the possibility of reducing the noise below that now obtained with an external horizontal multiplexer.

#### 4. Resolution in Camera Tubes and Sensors

Although the resolution of image pickup devices at very low light levels is limited by noise, their resolution at high lights is affected by entirely different considerations. In a solid-state sensor the highest possible resolution is limited by the total numbers of picture elements in the array. In a camera tube such as the vidicon, whose target contains no structure, much higher limiting resolutions are possible. In either device the resulting modulation transfer function falls off with increasing spatial frequency in the image for a variety of causes, some

of which are common to both. Failure of charge to be transferred without loss from one element to the next in a CCD sensor causes the MTF to drop off in a amount dependent upon the product of the total number of transfers and inefficiency of transfer.<sup>1</sup> A similar effect can occur in both camera tubes and sensors if surface leakage between elements causes the picture charge to spread before scanning. Electron optics and beam spot size can limit the resolution of tubes, but their resolving power is still far ahead of present day sensors.

While sensors with more elements have been produced experimentally, the largest solid-state sensor commercially available today is the frame-transfer CCD sensor called the "SID."<sup>25</sup> In this device, the

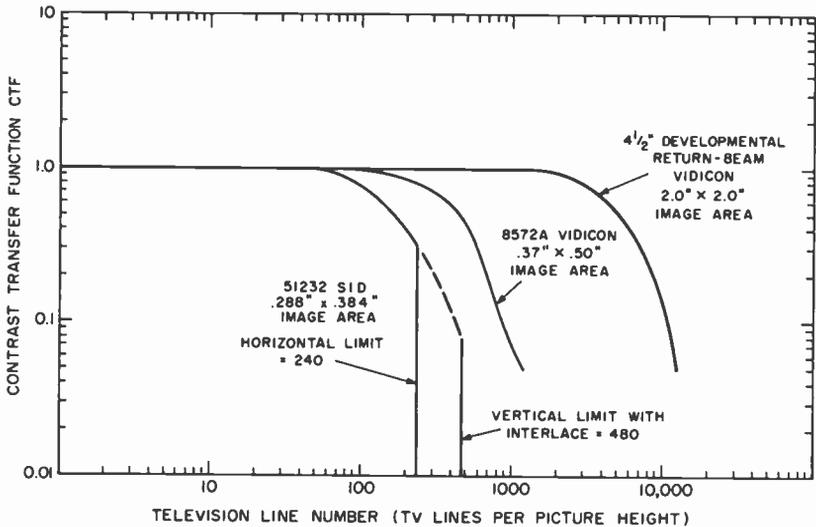


Fig. 7—Resolution transfer function of the 512 × 320 element CCD sensor compared with several types of vidicons.

CCD registers in the sensor area are illuminated directly, and the resulting charges are transferred to the output register, as shown in Fig. 5(b). The SID sensor contains 256 × 320 picture elements in each field, but the element centers can be shifted vertically by one-half element on successive fields to give an interlaced<sup>33</sup> picture of 512 × 320 elements. The center-to-center spacing of the elements is 0.6 mil (0.015 mm) in the vertical direction and 1.2 mils (0.030 mm) in the horizontal with a picture format of 7.31 × 9.75 mm. Overall picture quality is about comparable to the small  $\frac{2}{3}$ -inch-diameter vidicons.

Fig. 7 compares the square wave response (CTF) for the SID with that of two different sizes of vidicons. Although the large  $4\frac{1}{2}$ -inch-

diameter experimental return-beam vidicon<sup>34</sup> has a resolving power of 10,000 TV lines, the large capacitance of the ASOS (antimony sulfide-oxysulfide) target produces excessive image smear for moving scenes. The complete freedom of the CCD sensor from such lag effects is an important advantage for this type of sensor.

## 5. Uniformity and Freedom from Spurious Signals

The problem of uniformity and freedom from spurious signals in the reproduced picture is of such importance that it affects the whole design of the sensing device. The requirements are actually more severe than in the human eye itself, where the brain and eye motion can compensate for nonuniform response in the retina. In a sensor the maximum signal from each picture element must match that of its neighbors to within a few percent to escape visible imperfections in the transmitted picture. Such perfection can be achieved more easily (and at lower cost) if the device design avoids the use of critical processes wherever possible.

An unavoidable requirement is that the responsivity of the photosensitive material be uniform as well as high. The photoemissive and secondary electron emissive surfaces used in the early camera tubes were initially quite nonuniform, but were brought under control after extensive development. The silicon vidicon had the technology of the entire semiconductor industry to draw on, but the requirements on uniformity and low dark current<sup>35</sup> in the photodiode array were much more severe than for conventional integrated circuits. Satisfactory picture quality was finally achieved after many years of development. Early attempts at using phototransistor arrays<sup>16</sup> for image sensors showed objectional spatial variations in the transistor gain ( $\beta$ ) too large for high-quality imaging.

In spite of the uniformity problems accompanying image intensification, the low quantum yield of photoemitters required that intensification be used in most photoemissive camera tubes. Good picture quality in the image orthicon depended upon a uniform secondary emission ratio over the back side of the glass target and over the first stage of its signal multiplier. This tube is capable of excellent performance, despite its complexity, but it has remained a relatively expensive tube to produce. Other forms of intensifiers that have achieved acceptable uniformity for low-light-level applications include the secondary-emission-conductivity (SEC) target and the bombardment-induced-conductivity silicon SIT target. The channel-type secondary-emissive intensifiers<sup>36</sup> require the work function and other surface properties to be uniform over extensive areas in the tube.

Another type of signal nonuniformity found in beam-scanned tubes was a center-to-edge variation in signal or background. In the iconoscope this was caused by an inherent redistribution of secondary electrons raining back on the target. In the early image orthicons and vidicons, a dark border was caused by incorrect electron optics that prevented the proper landing of the scanning beam at the edges of the target. The effect has now been corrected. At the present time, camera tubes of many types are capable of the highest quality television pictures. Only the very simplest tubes, however, such as the photoconductive vidicons, have achieved good quality at relatively low cost.

In solid-state sensors, the uniformity problem has yet to be controlled as satisfactorily as it has been in camera tubes. Although some types of nonuniformity have been eliminated, others have been introduced. In addition to the usual requirements on uniform responsivity, there is the need to avoid fixed pattern noise caused by geometrical imperfection in the structure or by the clock pulses required for scanning. The CCD sensor has proven to be excellent in providing signals relatively free of fixed pattern noise arising from the charge transfer process. However, the size and complexity of the arrays and the tight tolerances on imperfections requires careful control in manufacture. The  $512 \times 320$  element sensor covers a silicon chip 12.7 by 19.05 mm in area, and has 10  $\mu\text{m}$  center-to-center spacing of electrodes over the entire active area. The quality of the masks required for producing this device is far more demanding than usually needed for other large integrated circuits, such as random access digital memories.

The  $XY$  addressing schemes that were developed prior to charge coupling are being utilized in the recent forms of charge-injection sensors. Such sensors offer potential advantages as well as disadvantages over the frame-transfer CCD sensor discussed above. A critical problem of operating  $XY$  sensors has been to avoid fixed pattern striations in the picture produced by the multiplexed scanning process. Variations in size and shape of the scan pulses produced by the digital scan generator lead to variations in the signal that are difficult to remove. The proposed use of external charge transfer registers for readout of  $XY$  arrays would be aimed at removing this source of nonuniformity.<sup>32</sup>

Another type of spurious signal can arise in image pickup devices when the light level in local areas of the picture considerably exceeds that required for charging (or discharging) the elemental capacitors. A possible consequence of such overload is that the extra charge spills over into the adjacent elements, causing the bright area to grow and

mask picture detail in the neighboring areas. This spreading of the signal, sometimes called "blooming," has been a problem in both camera tubes and solid-state sensors. It was particularly severe in the early orthicon, where the scanned surface of the entire target could become positively charged, causing the signal to be cut off until the target surface was manually reset to ground. Blooming was never a problem in the iconoscope or in the image orthicon, where low-velocity secondary electrons returning to the target would preserve contrast in the picture even though the light overload exceeded full storage by several orders of magnitude. However, in the image orthicon an isolated bright area would be surrounded by a dark halo that could obscure neighboring detail.

The blooming problem is a minor one in the  $Sb_2S_3$  vidicons partly because of the less-than-unity transfer characteristic ( $\gamma$ ) of the photoconductor and partly because of the effective discharge of the porous photoconductor by the low-velocity beam. Lateral spreading of the charge in the photoconductor is minimized by the thinness of the layer and by the high field across it. Blooming occurs more prominently in the silicon vidicon because of the long diffusion lengths of carriers in silicon. Various techniques have been developed for confining the charge spreading<sup>37</sup> or draining off the excess charge.<sup>35</sup>

In charge-coupled devices, blooming<sup>38</sup> initially appeared to be more serious than in the silicon vidicon owing to the channeling of excess charge along the register. However, a mode of operation has been found<sup>25</sup> for the frame transfer sensor whereby the spreading of charge down the register can be greatly reduced. This consists in holding two of the three sets of gates in accumulation during the light-integration period, thereby establishing a barrier at each element. Blooming in the  $XY$ -addressed charge-injection sensors appears to be easily controlled, at least in samples in which the center-to-center spacing of the elements is several times larger than in the present CCD sensors.

A less conspicuous type of spurious signal arises in the frame transfer CCD sensor as a result of illumination of the sensor registers during the vertical retrace period while the charge pattern is being transferred out of the sensor area. A brightly illuminated element can cause a slight smearing of signal over the entire column. A similar degradation could occur in an  $XY$  addressed sensor if the column buses were photosensitive.

## 6. Analog Memory Devices Derived from Camera Tubes and Sensors

Camera tubes which employ storage for light integration were recognized very early as a possible form of memory<sup>39</sup> for digital or analog

signals. A whole family of storage tubes were soon developed which were used primarily for analog storage of television or radar signals.<sup>40-42</sup> The most recent of these tubes is the "silicon storage tube"<sup>43</sup> which can hold a single television frame for hours while the signal is being read out "nondestructively", and for days without continuous readout. In this tube the charges are stored on an array of insulated electrodes whose voltage controls the landing of a low-velocity beam on an adjacent target electrode.

Solid-state image sensors also have their counterparts in semiconductor memories. Photodiode arrays with XY address strips and one or two MOS transistors at each element are closely analogous to certain types of random-access memories employing charge storage. The major difference is that the memory chip must provide address circuits for both reading and writing signals at random addresses instead of for simply reading sequentially as in a sensor. Since signal charges leak away in a few seconds in semiconductor memories at room temperature, longer storage times require continual refreshing of the signals by reading and re-writing. Digital signals can be refreshed indefinitely in this manner, but analog signals are difficult to maintain at a constant level. Cooling of a silicon memory or sensor would allow analog signals to be stored for hours but the signal could not be read more than a few times unless some form of nondestructive readout is devised. Various types of nondestructive readout schemes have been proposed for reading an analog memory element without discharge. Burke and Michon<sup>4</sup> have described a method of operating the double-capacitor charge-injection sensor with nondestructive readout. They have also demonstrated that a cooled charge-injection sensor can be used to integrate charge at very low light levels for periods up to three hours while continually reading the signal nondestructively.

The addition of an input register to a frame-transfer CCD sensor allows it to be used as an SPS (series-parallel-series) digital or analog delay line.<sup>44</sup> The signal can be read nondestructively by means of an MOS transistor whose gate is connected to a potential well near the end of the output register. Digital signals can be refreshed and stored indefinitely by recycling the output signal back into the input register. Analog signals can also be re-read<sup>45</sup> a limited number of times by re-cycling, but exact refreshing of the signal is difficult and the signal soon becomes degraded by transfer losses or lack of linearity in the input-output circuits. A nonvolatile analog store is conceptually possible using MNOS transistors at each element<sup>46</sup> but suitable methods of reading and writing must be found. Analog information can, of course, be digitalized and stored in a digital memory, but additional

memory elements are required, and analog-to-digital converters must be provided.

## 7. Discussion

After more than ten years research and development in various laboratories, solid-state cameras suitable for standard television have now reached the market. Fig. 8 shows two cameras manufactured by the RCA Solid State Division which utilize the  $512 \times 320$  element CCD

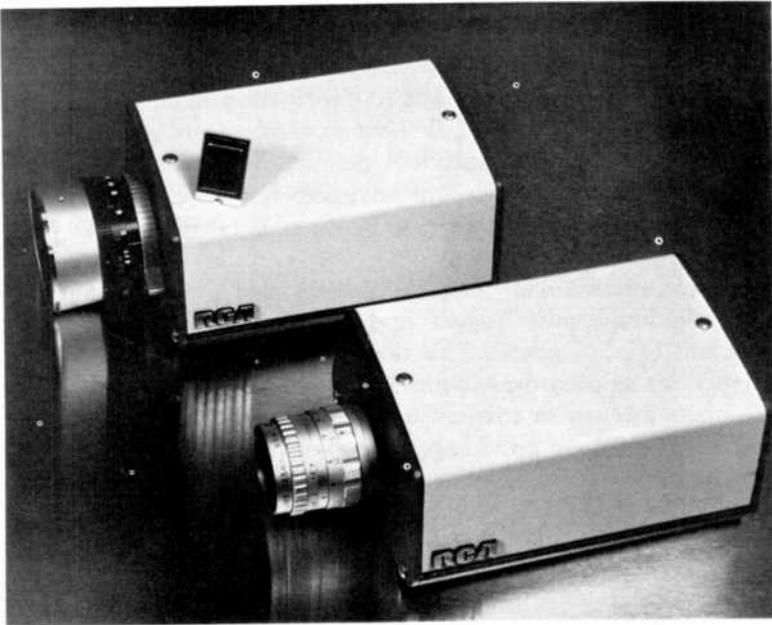


Fig. 8—Two solid-state cameras employing the RCA  $512 \times 320$  element CCD sensor. A sample of the SID sensor itself is shown lying on top of one of the cameras.

sensor described earlier. They provide a picture quality comparable to that of the  $\frac{2}{3}$ -inch vidicon with advantages that include small size, precise image geometry, freedom from lag and microphonics, and resistance to image burn. The next largest commercial sensor, produced by General Electric Co., is a  $244 \times 188$  element charge-injection sensor (CID) with two close-coupled MOS capacitors at each element. This type of sensor has  $XY$  address strips and peripheral digital shift registers for scanning.

The choice between CCD sensors with internal charge transfer reg-

isters, and *XY*-addressed arrays with external registers, is still an open one. Each will continue to be used for those applications that it serves best. Small  $100 \times 100$  element sensors with the photoelements interleaved between non-illuminated CCD registers<sup>47</sup> are being marketed for surveillance purposes. Still smaller *XY* photodiode arrays<sup>28</sup> with  $50 \times 50$  elements and many types of single-line sensors are being sold for such applications as process control and character recognition devices.

Advances in camera tubes have continued to be made during the period over which the solid-state sensors were developed. The principal areas of concern have been to find improved photoconductors for the vidicon, and to extend its capabilities for color and infrared pickup. The lead-oxide vidicon<sup>48</sup> gives higher sensitivity and less lag than the antimony sulfide vidicon, but its fabrication is more costly. The lead oxide tube has replaced the image orthicon in broadcast cameras, but porous antimony sulfide<sup>49</sup> has continued to be used in most general purpose vidicons. Recent advances have been reported with cadmium selenide<sup>14</sup> and other new types of heterojunction<sup>50</sup> targets for vidicons.

The silicon photodiode<sup>14</sup> vidicon target is also a fairly recent development, yielding a more rugged and more sensitive tube for special applications (e.g., in space). This target (which is thin) also functions effectively for bombardment-induced conductivity in the SIT and I-SIT image-intensifier tubes. Such tubes have yielded the highest signal-to-noise ratio of any image pickup devices at very low light levels.<sup>24</sup>

Although considerable progress has been made in improving the performance of single-tube color cameras, the color cameras<sup>51</sup> used for broadcasting employ three lead oxide vidicons. This tube has displaced the image orthicon because of its greater ease of registry, its linear transfer characteristic, and its smaller size. Clearly, color cameras could be built with three solid-state sensors provided their resolution, sensitivity, and other characteristics were suitable. The geometrical accuracy of the image and the smaller size of the solid-state sensor would be particularly advantageous here.

Solid-state sensors also offer promise for the ultimate development of single-sensor color cameras with built-in elemental filters, provided the number of picture elements can be increased to fit the desired standards of the system. The color signals can be generated as three simultaneous signals or as a single color-encoded signal. In the latter case the accurate scan rate of the solid-state sensor eliminates the need for generating a reference carrier as is required in some color-encoded single-tube color cameras.<sup>52</sup>

The search for improved cameras for thermal imaging has resulted in the pyroelectric vidicon.<sup>53</sup> Various types of CCD imagers with infrared-sensitive sensor elements are currently being investigated.<sup>54</sup>

The solid-state image sensor development has now reached the stage where an assessment of its future role should be possible. There can be no doubt that solid-state scanning, either by means of XY address strips or by internal charge transfer, can provide a useful tool for image sensing, analog storage, and signal processing. Although the image sensors now available do not fully meet the resolution requirements of broadcast television, the prospects are good that sensors suitable for broadcast purposes will eventually be produced at competitive prices. The extent to which sensors will be able to compete with the highly refined tubes now used for special applications will depend upon the additional development each receives. A long-term objective of the image sensor work at RCA Laboratories has been to produce a low cost camera which could be used in new consumer applications, such as electronic photography.<sup>55</sup> Fulfillment of this objective will require a sustained effort along an approach that demands least of its fabrication technology.

### Acknowledgments

The writer is indebted to Albert Rose, whose contributions to imaging have spanned the years from early camera tubes to image sensors, for his continuing friendship and valued advice on many subjects.

### References:

- <sup>1</sup> C. H. Secuin, D. A. Sealer, W. J. Bertram, Jr., M. F. Tompsett, R. R. Buckley, T. A. Shankoff and W. J. McNamara, "A Charge-Coupled Area Image Sensor and Frame Store," *IEEE Trans. Electron Dev.*, ED-20, p. 244 (1973).
- <sup>2</sup> G. J. Michon and H. K. Burke, "Charge Injection Imaging," Digest of Technical Papers, 1973 IEEE Solid State Circuits Conf., p. 138, Philadelphia, Pa.
- <sup>3</sup> J. J. Tiemann, W. E. Engeler, R. D. Baertsch, and D. M. Brown, "Intracell Charge-Transfer Structures for Signal Processing," *IEEE Trans. Electron Dev.*, ED-21, p. 300 (1974).
- <sup>4</sup> G. J. Michon, H. K. Burke, and D. M. Brown, "Recent Developments in C.I.D. Imaging," paper presented at the Symposium on Charge-Coupled Technology for Scientific and Imaging Applications, Jet Propulsion Laboratory, Pasadena, Calif., March 1975 (Conf. Proceedings Special Publication 4321).
- <sup>5</sup> P. K. Weimer, "Television Camera Tubes: A Research Review," *Advances in Electronics and Electron Physics*, 13, p. 387 (1960).
- <sup>6</sup> P. K. Weimer, "Image Sensors for Solid State Cameras," *Advances in Electronics and Electron Physics*, 37, p. 181 (1975).
- <sup>7</sup> A. Rose, "Television Pickup Tubes and the Problem of Vision," *Advances in Electronics*, Vol. 1, p. 131 (1948).
- <sup>8</sup> A. Rose, *Vision, Human and Electronic*, Plenum Press, New York (1973).
- <sup>9</sup> P. T. Farnsworth, "Television by Electron Image Scanning," *J. Franklin Institute*, 218, p. 411 (1934).

- <sup>10</sup> V. K. Zworykin, "The Iconoscope," *Proc. I.R.E.*, **22**, p. 16 (1934).
- <sup>11</sup> A. Rose and H. Iams, "Television Pickup Tubes Using Low-Velocity Electron Beam Scanning," *Proc. I.R.E.*, **27**, p. 547 (1939).
- <sup>12</sup> P. K. Weimer, S. V. Forgue, and R. R. Goodrich, "The Vidicon Photoconductive Camera Tube," *Electronics*, **23**, p. 70 (1950); *RCA Review*, **12**, p. 306 (1951).
- <sup>13</sup> A. Rose, "Maximum Performance of Photoconductors," *Helvetica Physica Acta*, **30**, p. 242 (1957).
- <sup>14</sup> K. Shimizu, O. Yoshida, S. Aihara, and Y. Kiuchi, "Characteristics of the New Camera Tube with a CdSe Photoconductive Target," *Advances in Electronics and Electron Physics*, **33A**, p. 293 (1972).
- <sup>15</sup> M. H. Crowell, T. M. Buck, E. F. Labuda, J. V. Dalton, and E. J. Walsh, "The Silicon-Diode-Array Target," *Bell System Tech. J.*, **46**, p. 491 (1967).
- <sup>16</sup> G. Strull, W. F. List, E. L. Erwin, and D. Farnsworth, "Solid State Array Cameras," *Appl. Optics*, **11**, p. 1032 (1972).
- <sup>17</sup> R. H. Dyck and G. P. Weckler, "Integrated Arrays of Silicon Photodetectors for Imaging Sensing," *IEEE Trans. Electron Dev.*, **ED-15**, p. 196 (1968).
- <sup>18</sup> W. S. Boyle and G. E. Smith, "Charge-Coupled Semiconductor Devices," *Bell System Tech. J.*, **49**, p. 587 (1970).
- <sup>19</sup> F. L. J. Sangster and K. Teer, "Bucket-Brigade Electronics—New Possibilities for Delay, Time Axis Conversion and Scanning," *IEEE J. Solid State Circuits*, **SC-4**, p. 131 (1969).
- <sup>20</sup> A. Rose, P. K. Weimer and H. B. Law, "The Image Orthicon Sensitive Television Pickup Tube," *Proc. I.R.E.*, **34**, p. 424 (1946).
- <sup>21</sup> P. K. Weimer, "The Image Isocon—An Experimental Television Pickup Tube Based on the Scattering of Low Velocity Electrons," *RCA Review*, **10**, No. 3, p. 366 (1949).
- <sup>22</sup> G. W. Goetze and A. B. Laponsky, "Camera Tubes Employing High-Gain Electron-Imaging Charge-Storage Targets," Chapter 11 in *Photoelectronic Imaging Devices*, Vol. 2, L. M. Biberman and S. Nudelman, eds., p. 217–251, Plenum Press, (1971).
- <sup>23</sup> R. L. Rodgers, III, G. S. Briggs, W. N. Henry, P. W. Kaseman, R. E. Simon, R. L. Van Asselt, E. Cave, F. C. Duigon, A. D. Cope, and G. A. Morton, "Silicon Intensifier Target Camera Tube," *Digest of Technical Papers, IEEE International Solid-State Circuits Conf.*, Philadelphia, Pa., p. 176, Feb. 1970.
- <sup>24</sup> R. W. Engstrom and G. A. Robinson, "Choose the Tube—for L<sup>3</sup>TV," *Electro-Optical Systems Design*, June 1971 (ST-4693).
- <sup>25</sup> R. L. Rodgers, III, "Charge-Coupled Imager for 525-line Television," paper presented at the 1974 IEEE Intercon Meeting in New York.
- <sup>26</sup> P. K. Weimer, W. S. Pike, G. Sadasiv, F. V. Shallcross, and L. Meray-Horvath, "Multi-Element Self-Scanned Mosaic Sensors," *IEEE Spectrum*, **6**, p. 52 (1969).
- <sup>27</sup> P. K. Weimer, F. V. Shallcross, and V. L. Frantz, "Phototransistor Arrays of Simplified Design," *IEEE J. Solid State Circuits*, **SC-6**, p. 135 (1971).
- <sup>28</sup> Reticon Corporation, Mountain View, California
- <sup>29</sup> J. E. Carnes and W. F. Kosonocky, "Noise Sources in Charge-Coupled Devices," *RCA Review*, **33**, p. 327 (1972).
- <sup>30</sup> S. B. Campana, "Charge-Coupled Devices for Low Light Level Imaging," *CCD Applications Conf. Proc.*, TD-274, p. 235, sponsored by the Naval Electronics Laboratory, San Diego, Calif., September 1973.
- <sup>31</sup> D. D. Wen, J. Early, C. K. Kim, and G. F. Amelio, "A Distributed Floating-Gate Amplifier in Charge-Coupled Devices," *Digest of Technical Papers, 1974 IEEE International Solid State Circuits Conf.*, p. 24, Philadelphia, Pa., Feb. 1975.
- <sup>32</sup> M. G. Kovac, W. S. Pike, F. V. Shallcross, and P. K. Weimer, "Solid State Imaging Emerges from Charge Transport," *Electronics*, **45**, p. 72 (1972).
- <sup>33</sup> C. H. Sequin, "Interfacing of Charge-Coupled Devices," *IEEE Trans. Electron Devices*, **ED-20**, p. 535 (1973).
- <sup>34</sup> Otto H. Schade, Sr., "Electron Optics and the Signal Read-out of High Definition Return-Beam Vidicon Cameras," *RCA Review*, **31**, p. 60 (1970).
- <sup>35</sup> R. L. Rodgers, III, "Beam-Scanned Silicon Targets for Camera Tubes," *IEEE Intercon Technical Paper*, New York Session 1, Paper I/1 (1973).
- <sup>36</sup> C. E. Catchpole, "The Channel Image Intensifier," Chapter 8 in *Photoelectronic Imaging Devices*, Vol. 2, L. M. Biberman and S. Nudelman, eds. Plenum Press, New York (1971).
- <sup>37</sup> B. Singer, J. Kostelec, M. Polshuk, and E. Arnold, "Theory Design and Performance of Low-Blooming Silicon Diode Array Imaging Targets," *International Electron Devices Meeting*, Wash., D. C., Dec. 1972.

- <sup>38</sup> W. F. Kosonocky, J. E. Carnes, M. G. Kovac, P. Levine, F. V. Shallcross, and R. L. Rodgers, III, "Control of Blooming in Charge-Coupled Imagers," *RCA Review*, **35**, No. 1, p. 3, (1974).
- <sup>39</sup> R. A. McConnell, "Video Storage by Secondary Emission from Simple Mosaics," *Proc. I.R.E.*, **35**, p. 1258 (1947).
- <sup>40</sup> L. Pensak, "The Graphechon—A Picture Storage Tube," *RCA Review*, **10**, p. 59-73 (1949).
- <sup>41</sup> A. S. Jensen, "The Radechon—A Barrier Grid Storage Tube," *RCA Review*, **16**, p. 197-215 (1955).
- <sup>42</sup> L. Pensak, "The Metrechon—A Half-Tone Picture Storage Tube," *RCA Review*, **15**, p. 145-162 (1954).
- <sup>43</sup> R. S. Silver and E. Luedicke, "Electronic Image Storage Utilizing a Silicon Dioxide Target," *IEEE Trans. Electron Devices*, **ED-18**, p. 229, (1971).
- <sup>44</sup> M. F. Tompsett and E. J. Zimany, Jr., "Use of Charge-Coupled Devices for Delaying Analog Signals," *IEEE J. Solid State Circuits*, **SC-8**, p. 151, (1973).
- <sup>45</sup> P. K. Weimer, W. S. Pike, F. V. Shallcross, and M. G. Kovac, "Video Processing in Charge-Transfer Image Sensors by Recycling of Signals Through the Sensor," *RCA Review*, **35**, No. 3, p. 341 (1974).
- <sup>46</sup> M. H. White, D. R. Lampe, J. L. Fagan, and D. A. Barth, "A Non-Volatile Charge-Addressed Memory (NOVCAM) Cell," Digest of Technical Papers, p. 115, 1974 Electron Devices Meeting, Wash., D. C.
- <sup>47</sup> L. Walsh and R. H. Dyck, "A New Charge-Coupled Area Imaging Device," CCD Applied Conference Proceedings, Naval Electronics Laboratory, San Diego, Calif., TD-274, p. 21 (1973).
- <sup>48</sup> E. F. De Haan, A. van der Drift, and P. P. M. Schampers, "The Plumbicon: A New Television Camera Tube," *Philips Technical Rev.*, **25**, p. 133 (1963-64).
- <sup>49</sup> S. V. Forgue, R. R. Goodrich, and A. D. Cope, "Properties of Some Photoconductors, Principally Antimony Trisulfide," *RCA Review*, **12**, p. 335 (1951).
- <sup>50</sup> C. R. Wronski, "Semiconductor Heterojunction Vidicons," *RCA Review*, **36**, p. 425 (1975) (this issue).
- <sup>51</sup> H. Breimer, W. Holrn, and S. L. Tan, "A Color Television Camera with 'Plumbicon' Camera Tubes," *Philips Tech. Review*, **28**, No. 11, p. 336 (1967).
- <sup>52</sup> D. H. Pritchard, "Stripe-Color-Encoded Single-Tube Color Television Camera Systems," *RCA Review*, **34**, p. 217 (1973).
- <sup>53</sup> R. Watton, C. Smith, B. Harper and W. M. Wreathall, "Performance of the Pyroelectric Vidicon for Thermal Imaging in the 8-14 Micron Band," *IEEE Trans. Electron Devices*, **ED-21**, p. 462 (1974).
- <sup>54</sup> E. Kohn "An Infrared-Sensitive Charge-Coupled Imager," *IEEE Trans. Electron Devices*, **ED-21**, p. 737 (1974).
- <sup>55</sup> A. Rose. Private Communication.
- <sup>56</sup> P. K. Weimer and A. D. Cope, "Photoconductivity in Amorphous Selenium," *RCA Review*, **12**, p. 314 (1951).
- <sup>57</sup> N. Goto, Y. Isozaki, K. Shidara, E. Maruyana, T. Hirai, and T. Fujita, "Saticon: A New Photoconductive Camera Tube with Se-As-Te-Target," *IEEE Trans. Electron Devices*, **ED-21**, p. 662 (1974).
- <sup>58</sup> C. R. Wronski, B. Abeles, R. E. Daniel, and Y. Arie, "Granular Metal-Semiconductor Schottky Barriers," *J.A.P.*, **45**, p. 295 (1974).

# An Approximate Model of the Beam-Blocking Contact in a PbO Vidicon\*

Alvin M. Goodman

RCA Laboratories, Princeton, N. J. 08540

**Abstract**—A simple physical model for the beam-blocking contact of the photoconductor layer in a PbO vidicon has been used to derive an approximate theory of the static dark current. The theory describes two regimes of dark-current density versus applied-voltage behavior:

(1) A drift-velocity-limited regime in which the dark-current density is limited by the drift, in the photoconductor layer, of conduction-band electrons produced by thermal excitation from traps in the beam-blocking region.

(2) A thermal-generation-limited regime in which the dark-current density is limited only by the supply of conduction-band electrons to the photoconductor layer resulting from thermal excitation in the beam-blocking region. Regime (2) will occur at a higher voltage than regime (1).

The theory relates the electrical behavior to the physical parameters of the model—thickness, trap-density distribution, and doping of the beam-blocking region, thickness of the photoconductor layer, and the electron mobility. Reasonable agreement is obtained between the theory and measurements on PbO vidicons. In addition, the theory provides an understanding of the factors involved in the optimization of the Fermi-level position in the beam-blocking region.

## 1. Introduction

The lead oxide vidicon<sup>†</sup> is a high-sensitivity, low-dark-current television camera tube which utilizes tetragonal PbO as the photoconductive target material.<sup>1-3</sup> An extensive literature exists describing many

\* A preliminary version of this material was presented at the IEEE Device Research Conference, Boulder, Colorado, June 27, 1973.

<sup>†</sup> Lead oxide vidicons are also often referred to by the registered trademark names of their manufacturers; viz., Plumbicon (N.V. Philips of Holland), Vistacon (RCA), and Leddicon (English Electric Valve).

aspects of the tube operation,<sup>2-4</sup> fabrication,<sup>5</sup> structure,<sup>6-8</sup> and physics.<sup>9-16</sup> In the conventional description of its operation,<sup>2</sup> the target is thought of as a reverse-biased p-i-n junction utilizing a sandwich structure geometry in which the i-region is an insulator and the p- and n-sections are predominantly extracting (or blocking) contacts. That is, the p- and n-regions are contacts (to the i-region) which inject very low currents into the i-region and yet can readily accept relatively large currents of photogenerated carriers coming from it. One of the most critical steps in the fabrication of the target is the formation of the p-layer on the scan-side. Although a variety of recipes is available for carrying out this step,<sup>5</sup> there seems to be little or no information published about how the p-layer functions as a beam-blocking contact. The purpose of this article is to present an approximate model of the steady-state operation of the beam-blocking region under dark (non-illuminated) conditions.

Let us consider first the PbO photoconductive layer that would exist in the vidicon if there were no beam-blocking contact. This is shown schematically in Fig. 1a. The layer is polycrystalline and composed of small tetragonal crystal platelets with dimensions of the order of  $10^{-4} \times 10^{-4}$  cm in the (001) plane and 50 to 100 Å in thickness; i.e., normal to the (001) plane.<sup>6,12</sup> The platelets have a preferential orientation with their (001) directions parallel to the substrate.<sup>6</sup> The substrate shown in Fig. 1a is a transparent conducting coating (usually SnO<sub>2</sub>) on glass. This forms a non-injecting contact to the PbO that is equivalent to a positively biased n-region.<sup>2</sup> The energy-band diagram for this case is shown in Fig. 1b.\* In this case, the scanning electron beam would enter the conduction band of the PbO layer and drift through it causing a large dark current limited only by electron trapping within the layer (space-charge-limited current) or the maximum available beam current. In the space-charge-limited-current case, the potential of the scanned surface reaches a quasi-steady-state value such that on the average, over a full scanning period, the accepted beam current just equals the current through the PbO layer.

If, however, a thin region on the scan side of the PbO layer is made

\* The justification for the use of an energy-band scheme to describe the polycrystalline layer as an equivalent homogeneous material has been given by van den Broek.<sup>12</sup> The electrical behavior of such layers is dominated by surface states near midgap in the individual platelets; in the equivalent homogeneous layer these surface states become deep bulk states. While the equivalent bulk density of these deep states has not been unambiguously determined, a detailed consideration<sup>15</sup> of the observed doping compensation and charge carrier trapping due to these deep states has led to the conclusion that the permissible product of their concentration  $N_t$  and capture cross-section  $S$  must be  $\lesssim 10^{-5}$  cm<sup>-1</sup>. This could be due to a combination of  $N_t \approx 10^{15}$ /cm<sup>-3</sup> and  $S \approx 10^{-20}$  cm<sup>2</sup>. On the other hand, experiments with metal contacts to polycrystalline layers indicate the presence of at least  $10^{13}$  states/cm<sup>2</sup> at the metal-PbO interface. If a similar density of states is present at a "free" PbO surface, then the equivalent bulk density  $N_t$  must be greater than  $10^{19}$ /cm<sup>-3</sup> with a concomitant value for trapping cross section,  $S < 10^{-24}$  cm<sup>2</sup>. Neither of these possibilities may be excluded at the present time.

p-type, it "acts as a contact hindering the entry of electrons" into the i-region.<sup>2</sup> According to Ref. [5] a method for obtaining this thin p-layer is by exposing the scan side of the layer to an oxygen-ion or oxygen-atom bombardment; the depth of the resulting layer is then from 10 to 200 Å. The oxygen atoms are presumed either to be trapped in interstitial sites or to fill oxygen vacancies (which had previously been compensated by lead vacancies) and to act as electron acceptors.

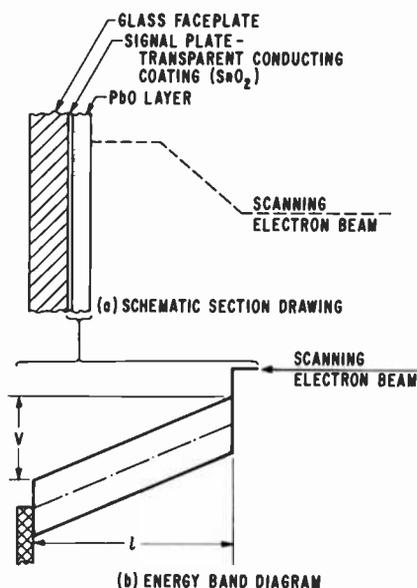


Fig. 1—PbO layer and substrate without a beam-blocking contact: (a) schematic section drawing and (b) energy-band diagram.

In Section 2, an approximate model for this beam-blocking layer and its steady-state operation under dark conditions\* is described. In Section 3, the predictions of the model are compared with experimental measurements carried out on PbO vidicons.

It is helpful in the understanding of the following sections to anticipate one important result of the present work—the recognition that the beam-blocking region (hereinafter abbreviated BBR) is not necessarily a p-type region; it may, in fact, be intrinsic or even slightly n-type. Moreover, the BBR may not even be entirely PbO; this is discussed further in Section 4.

\* When the PbO is illuminated, the behavior of the beam-blocking layer is considerably more complicated and will be considered in a future paper.

## 2. Approximate Theoretical Model

### 2.1 Formulation and Approximations

An energy-band diagram for the model is shown in Fig. 2a with an enlarged view of the BBR shown in Fig. 2b. A plane-parallel layer geometry leading to a one-dimensional description<sup>2,5,9</sup> is assumed here. The use of this geometry for the model is clearly appropriate since the width of the vidicon scanning beam ( $\sim 20 \mu\text{m}$ ) is large compared with the crystallite dimensions ( $\lesssim 1 \mu\text{m}$ ); thus, the effect of the complex crystallite structure is "averaged out" (not resolved) by the beam. The lack of well-defined band edges in the BBR (as shown schematically in Fig. 2b) indicates both the probable composite physical nature of the BBR and our inability to describe it more explicitly.

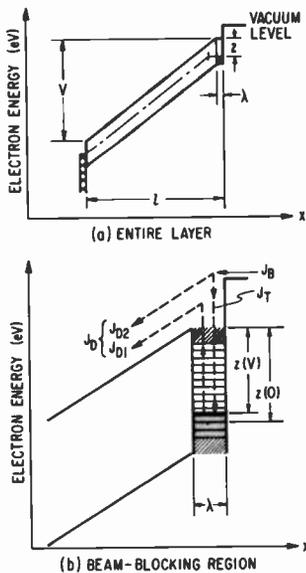


Fig. 2—Energy-band diagram of a PbO layer with a beam-blocking region: (a) entire layer and (b) beam-blocking region.

The following additional assumptions are made: (1) The BBR caused by the oxygenation process has an extent  $\lambda$  which is very much smaller than the layer thickness  $l$ . (2) Except for the BBR, the layer is intrinsic PbO. (3) The potential difference across the BBR,  $V_\lambda$ , is very much smaller than  $V$ , the potential difference across the entire layer, and in the present approximation  $V_\lambda$  is taken equal to

zero. (4) A large density of (unoccupied) electron traps exists above the Fermi level in the BBR. This may be at least partially due to the emptying of previously filled electron traps as their electrons occupied the newly created acceptor states. In addition, new electron traps may have been created during the oxygen-ion or oxygen-atom bombardment. (5) The depth,  $z$ , of the Fermi level in the BBR (measured with respect to the conduction-band edge in the intrinsic PbO at its interface with the BBR) is a function of  $V$ . When  $V = 0$ ,  $z = z(0)$  and the net charge in the BBR is zero. For non-zero  $V$  the BBR becomes negatively charged due to the filling of deep electron-trap levels, and  $z(V)$  is less than  $z(0)$ . (6) The incoming beam (current density  $J_B$ ) enters at the vacuum level; it is partially transmitted through the BBR ( $J_{D2}$ ) and partially thermalized and trapped in the BBR ( $J_T$ ). The trapped electrons may be thermally re-emitted and drifted through the layer ( $J_{D1}$ ). The total dark-current density  $J_D$  is the sum of  $J_{D1}$  and  $J_{D2}$ .<sup>\*</sup> In a "good" beam-blocking contact  $J_{D2} \ll J_{D1}$  and the lower limit on  $J_D$  is equal to  $J_{D1}$ . In what follows, we shall assume that  $J_{D2} = 0$  and  $J_D = J_{D1}$ , although it is recognized that in a "bad" beam-blocking contact,  $J_{D2}$  may not be negligible. Contributions to the dark current due to (1) hole injection from the signal plate and (2) bulk thermal generation of hole-electron pairs in the PbO are assumed to be negligible.<sup>2,9</sup>

The density of negative charge per unit area required by Gauss's law in the BBR is

$$q(V) = cV = \frac{K\epsilon_0}{l} V, \quad [1]$$

where  $c$  is the layer capacitance per unit area,  $K$  is the effective relative dielectric constant of the layer (taken as 12),<sup>5,15</sup> and  $\epsilon_0$  is the permittivity of free space. The volume density of electron traps per unit energy in the BBR at an energy  $\mathcal{E}$  below the intrinsic PbO conduction-band edge (at the intrinsic PbO-BBR interface) is taken as  $D_S(\mathcal{E})/\text{cm}^3\text{-eV}$ . The net charge per unit area in these traps (equal to  $q(V)$  in Eq. [1]) is given to a good approximation by

$$q(V) = - \int_{\mathcal{E} = z(0)}^{\mathcal{E} = z(V)} e D_S(\mathcal{E}) \lambda d\mathcal{E}, \quad [2]$$

where  $e$  is the magnitude of the electronic charge.

\* These current densities may be thought of as time-averaged values (using a frame time as the integration period). The value of  $J_B$  is a function of the BBR surface potential (or alternatively, the beam landing energy). It therefore varies with the discharge condition of the surface; i.e., how much current has previously flowed. As a trivial example, if no discharge current flows between beam scans, the surface potential does not change and no beam electrons can land;  $J_B = 0$  and  $J_{D1} = J_{D2} = 0$ . It is perhaps worthwhile to point out that  $J_{D1}$  can flow continuously during the entire frame time while the instantaneous flow of  $J_{D2}$  at a given spot on the target can occur only during the time that the beam is landing.

By differentiating Eq. [2] we obtain

$$\frac{dq(V)}{dV} = -e\lambda D_S(z) \frac{dz}{dV}. \quad [2a]$$

By combining Eqs. [1] and [2a] we obtain

$$\frac{dz}{dV} = -\frac{K\epsilon_0}{e\lambda D_S(z)l}; \quad [3]$$

i.e., the rate of decrease of the Fermi-level depth in the BBR with increasing applied voltage is inversely proportional to the density of traps per unit energy at the Fermi level. If  $D_S(\mathcal{E})$  is a constant, then,

$$\frac{dz}{dV} = -\frac{K\epsilon_0}{e\lambda D_S l}; \quad [3a]$$

the depth of the Fermi level in the BBR decreases linearly with increasing applied voltage. It is seen later that measurements carried out thus far are consistent with a considerable range of validity for Eq. [3a].

The maximum available dark-current density that may be drawn through the intrinsic layer  $J_{DM}$  is given by the unilateral random current density of electrons entering the intrinsic layer at its interface with the BBR due to thermal release\* from traps in the BBR;

$$J_{DM} = \frac{en\bar{v}}{4} = \left[ eN_c \sqrt{\frac{kT}{2\pi m}} \right] \exp \left[ -\frac{ez(V)}{kT} \right] \quad [4]$$

where  $n$  is the conduction-band electron density,  $\bar{v}$  is the average thermal velocity,  $N_c$  is the effective density of states in the conduction band,  $m$  is the electron effective mass (all in the intrinsic layer),  $k$  is Boltzmann's constant, and  $T$  is the absolute temperature. By expanding the exponent in Eq. [4] about the point  $V = 0$  and using Eq. [3a], we obtain

$$J_{DM} = \left[ eN_c \sqrt{\frac{kT}{2\pi m}} \right] \exp \left[ -\frac{ez(0)}{kT} + \frac{K\epsilon_0 V}{\lambda D_S l kT} \right]. \quad [4a]$$

We now consider the possible regimes of dark-current density versus applied voltage.

\* It is implicitly assumed that the dark current drawn from the BBR is not large enough to cause significant deviations from thermal equilibrium.

(a) *Drift-velocity-limited regime,  $J_D = J_{DDVL}$*

In this regime  $J_D$  is limited by the rate at which the electric field  $E$  in the *i*-region is capable of drifting the available electron dark current density  $J_{DM}$  away from its interface with the BBR. Thus,

$$J_{DDVL} = env_d = \frac{v_d}{\bar{v}/4} J_{DM}, \quad [5]$$

where  $v_d$  is the electron drift velocity in the *i*-region. This may be written as

$$J_{DDVL} = \left[ \frac{4\mu_n V}{\bar{v}l} \right] J_{DM}, \quad [6]$$

assuming that the field in the *i*-region is uniform and  $v_d = \mu_n V/l$ . This is equivalent to assuming insignificant electron trapping in the *i*-region; i.e., the electron transit time is much shorter than the trapping time. The maximum value of  $V$  for which Eq. [6] is valid is that for which  $J_{DDVL} = J_{DM}$ ; i.e.,

$$V_{DMS} = \bar{v}l/4\mu_n. \quad [7]$$

This is the minimum value of  $V$  required to draw the maximum available dark-current density.

(b) *Thermal-generation-limited or high-voltage regime,  $J_D = J_{DM}$*

This occurs when  $V \geq V_{DMS}$  and  $J_D$  is given by Eq. [4] or [4a].

A third possible regime, one in which the current is space-charge limited,<sup>17</sup> might also occur, at least in principle. This could happen if, contrary to our previous assumption, trapping in the *i*-region were important and if the Fermi level in the BBR were sufficiently shallow. A further discussion of space-charge-limited dark current is beyond the intended scope of this paper. It should be noted, however, that detailed measurements of space-charge-limited dark current as a function of temperature and voltage might be used to provide information about electron transport and trapping in PbO.

It is of interest to consider the limiting forms of  $J_D$  versus  $V$  that may be expected on the basis of Eq. [3a] for very small and for very large values of  $\lambda D_S$ .

(1) If  $\lambda D_S$  is small, the position of the Fermi level in the BBR will shift rapidly with applied voltage. If  $z \gtrsim E_G/2$  initially (i.e., at  $V = 0$ ), the low-voltage portion of the  $J_D$  versus  $V$  curve is likely to be in the

drift-velocity-limited regime. As  $V$  is increased, the Fermi level in the BBR will rise and  $J_D$  will rise until at some value of voltage one or more of our assumptions will be violated. For example, the charge in the i-region may become so large that it is more important than the charge in the BBR in determining the voltage across the layer, thereby invalidating Eq. [1]. Another possibility is that the extraction of the charge from the BBR may become so large that the assumption of thermal equilibrium in the BBR is invalid. The limit of  $\lambda D_S \rightarrow 0$  is equivalent to no beam-blocking contact; this would lead to a very large dark current as discussed in the introduction.

(2) If  $\lambda D_S$  is large, the position of the Fermi level in the BBR will shift only slightly with applied voltage. In the limit of very large  $\lambda D_S$ , the Fermi level shift would be negligible. In this case, if  $z \gtrsim E_G/2$  initially (i.e., at  $V = 0$ ), it would remain so as  $V$  is increased. The  $J_D$  versus  $V$  curve would then appear to be linear below  $V_{DMS}$  and constant (saturated) at higher values of  $V$ . If  $z$  were somewhat less than  $E_G/2$ , initially a space-charge-limited regime might be expected at low  $V$  followed by drift-velocity-limited and thermal-generation-limited regimes (or possibly only the latter) at increasingly higher values of  $V$ .

## 2.2 Optimization of Operating Region

In a practical vidicon, it is necessary to limit the dark-current density to some maximum value at the operating voltage. Let us assume for simplicity that the tube is operating in the *thermal-generation-limited* or *high-voltage* regime. This implies a minimum value for  $z$  according to Eq. [4]. The value of  $J_{DM}$  has been plotted in Fig. 3 on a logarithmic scale against  $z$  assuming that  $N_c = 10^{19}/\text{cm}^{-3}$ ,  $T = 300^\circ\text{K}$  and  $m/m_0$  the effective mass ratio is unity. The solid portion of the plot indicates the approximate maximum range over which current density measurements might conveniently be carried out. It is limited at the high end by available beam current and at the low end by increasing experimental difficulties and the onset of noise. Although the actual maximum  $J_D$  may vary according to the application, it is likely, for many purposes, to lie in the range  $10^{-10}$ – $10^{-9}$  A/cm<sup>2</sup>. As an example, if the maximum value of  $J_{DM} = 10^{-10}$  A/cm<sup>2</sup>, then  $z_{min} = 0.99$  eV defines one limit of a region of proper operation as shown in Fig. 3. Note that if  $J_{DM} = 10^{-9}$  A/cm<sup>2</sup>, then  $z = 0.93$  eV and the PbO adjacent to the BBR in this case would actually be slightly n-type rather than p-type.

The other limit on the position of the Fermi level in the BBR, viz.  $z_{max}$ , can be obtained from a consideration of the effect of transverse conduction (i.e., conduction along the direction of the target surface

rather than through it) upon the resolution of the tube. Proper vidicon operation requires that the scan side of the layer discharge according to the photoconduction resulting from the light pattern impinging on the tube.<sup>1</sup> Moreover, the resistivity of the layer must be great enough to maintain the resulting charge pattern with very little loss of detail for a frame time  $T_f$ . If the depth of the Fermi level in the BBR were sufficiently large, hole conduction could degrade the resolution of the charge pattern.

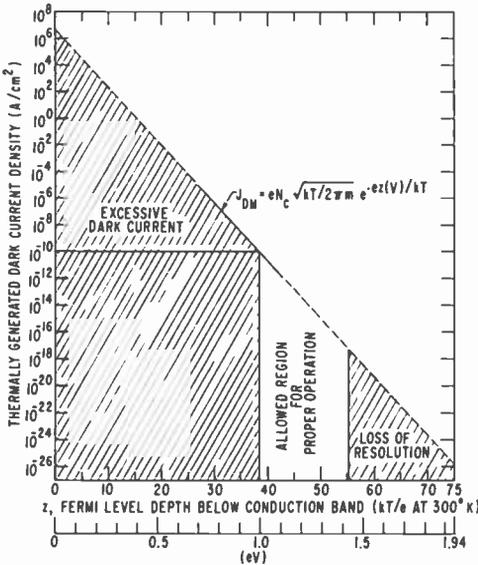


Fig. 3—Effect of the Fermi level position in a PbO vidicon beam-blocking layer.

We recognize that this hole conduction could take place in either the BBR or the adjacent nominally intrinsic PbO or both. As befits our limited knowledge of the structure and composition of the scanned surface of the target, we shall make the simplifying assumption that the transverse hole conduction occurs in a single p-region of PbO whose thickness is  $\delta$ . The following argument allows the approximate computation of  $z_{max}$ .

Let us assume that the p-region is divided into rectangular parallelepiped elements each having the dimensions  $d$  by  $d$  in the transverse directions (i.e., along the scanned-surface) and a thickness  $\delta$ . Let us also assume that in order to maintain the desired resolution, the acquired charge difference between two adjacent elements must

not diminish by more than 10% during the frame time  $T_f$ . Thus,

$$RC \gtrsim 10T_f. \quad [8]$$

where  $R$  is the effective resistance and  $C$  is the effective capacitance between two adjacent elements. To a reasonable approximation  $R \approx \rho/\delta = (\mu_p e p \delta)^{-1}$  and  $C \approx K \epsilon_0 d$ , where  $\rho$  is the volume resistivity of the p-layer,  $\mu_p$  is the hole mobility, and  $p$  is the effective hole density in the p-region.\* The dependence of  $p$  upon  $z$  is given by

$$p = N_i \exp \{-e[E_G - z(V)]/kT\}. \quad [9]$$

Combining Eqs. [8] and [9], we obtain the desired expression for  $z_{max}$ ;

$$z_{max} = E_G - \left(\frac{kT}{e}\right) \ln [10N_i e \mu_p \delta T_f / (K \epsilon_0 d)]. \quad [10]$$

For the values  $E_G = 1.94$  eV,  $T = 300^\circ\text{K}$ ,  $N_i = 10^{19}$  cm $^{-3}$ ,  $\mu_p = 1$  cm $^2$ /V-sec,  $\delta = 100$  Å,  $T_f = (1/30)$  sec,  $K = 12$ , and  $d = 10^{-3}$  cm, we obtain  $z_{max} = 1.42$  eV. This is shown in Fig. 3 as the upper limit beyond which loss of resolution occurs.

We note that  $z(V)$  decreases with increasing  $V$  and also that the portion of the i-region contributing to the transverse hole conduction will vary inversely with  $V$ ; thus the model predicts that if a loss of resolution is present in a real tube *due to transverse conduction in the p-region*, operation at a higher value of  $V$  should improve the resolution.

It should be emphasized that the limits on the allowed region of operation shown in Fig. 3 are illustrative and not definitive. In view of the approximations made and the relative uncertainty of some of the assumed parameter values, the actual limits might vary by several ( $kT/e$ ) units or more from those shown in Fig. 3.

### 3. Comparison with Experiment

#### 3.1 Description of Measurements

In order to compare the theory with experiment, the dark current  $I_D$  versus applied target voltage  $V$  was measured for 6 tubes (3 RCA and 3 Philips). The beam-blocking layer in each of the RCA tubes was formed by using an oxygen-glow discharge. The Philips tubes were obtained commercially and their fabrication details are unknown.

\* A somewhat different argument can be used to obtain an identical expression for the  $RC$  time constant of the discharge.<sup>18</sup>

The measurements were carried out using a modified GBC Model AE-50 television camera. The modifications were necessary to allow the camera to accommodate a 1.2-inch PbO vidicon rather than the 1-inch tube for which the camera was originally intended. The raster (focused beam) was adjusted to be approximately 2 cm<sup>2</sup> with the nor-

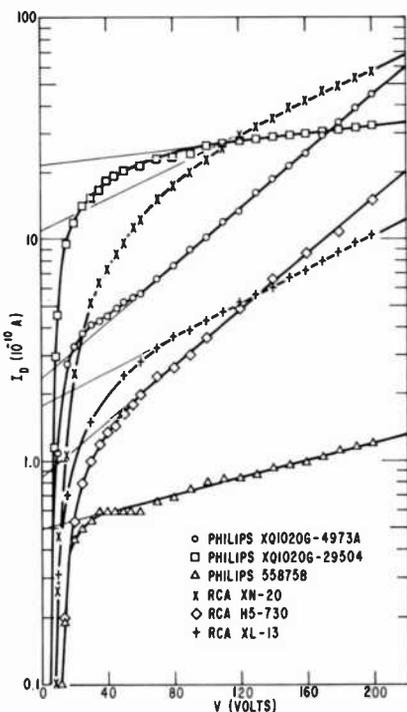


Fig. 4—Dark current versus target voltage for 6 PbO vidicons.

mal 4 × 3 aspect ratio. The vidicon signal plate lead was connected to an external power supply and a picoammeter\* for the  $I_D$  versus  $V$  measurements. In carrying out the measurements, a settling time of 5 minutes was allowed after each voltage change before measuring the dark current.

### 3.2 Results and Discussion

The dark-current versus target-voltage measurement results are plotted in Fig. 4. At very low target voltage all of the tubes showed a negative dark current (not shown in Fig. 4). The magnitude of this nega-

\* Keithley Model 417.

tive  $I_D$  at  $V = 0$  ranged from  $2 \times 10^{-11}$  to  $2 \times 10^{-9}$  A and did not seem to be correlated with the observed dark currents in the normal direction shown in Fig. 4. The value of  $V$  at which  $I_D$  changed from negative to positive was typically between 4 and 12 volts.

Three possible sources of the negative  $I_D$  are: (1) leakage paths inside the tube between the target and other electrodes ( $G_3$  and  $G_4$ ) that are normally operated at a high positive potential, typically 300 to 1000 volts; (2) electron photoemission from the PbO layer stimulated by soft x-rays generated at  $G_2$ ,  $G_3$ , and  $G_4$  by electrons from the scanning beam, and (3) positive ions (created from residual gases by the scanning beam) striking the PbO layer. Both (2) and (3) would create holes at the scan side of the PbO layer; these holes would then be transported by drift and/or diffusion through the layer toward the signal plate giving rise to the observed negative  $I_D$ . By making  $I_D$  versus  $V$  measurements with the scanning beam off, (1) was shown to be relatively unimportant; i.e., the maximum observed current due to (1) was  $<10^{-11}$  A. It was not possible to separate (2) and (3) experimentally. However, it seems reasonable to assume that the x-ray induced photoemission component of the negative  $I_D$  was approximately the same for each tube since they were operated under identical electrical conditions. The large variation in the negative  $I_D$  among the tubes was probably due to differences in the residual gas content of the individual tubes.

It can easily be seen that each of the tubes shows a more or less extended region of agreement with Eq. [4a] in the high-voltage regime.

Below this region, we might expect to find a drift-velocity-limited regime in which Eq. [7] should apply. Experimentally, we find no agreement in the case of the Philips' tubes and fair agreement in the case of RCA tubes as shown in Fig. 5. The abrupt transition between the drift-velocity-limited regime and the high-voltage regime predicted by the theory is a result of the simplifying assumption that at  $V = V_{DMS}$ , all of the available electrons are drifted away from the beam-blocking contact and that none are scattered back and retrapped. The extent to which the transition is gradual rather than abrupt is probably an indication of the actual back-scattering and retrapping that occurs. Also, although the theory assumes that the layer thickness is uniform over the scanned area, this is in fact only an approximation. Variations in thickness of actual layers are typically of the order of 10%.<sup>19</sup>

At low  $V$  ( $\leq 20$  volts), the experimentally observed  $I_D$ 's are again below the values predicted by Eq. [7]. This is likely to be due to a combination of electron trapping (ignored in the model) and the negative  $I_D$  contributions previously discussed.

We may obtain an approximate value for the electron mobility from Eq. [7a] and the experimentally determined value of  $V_{DMS}$ , as shown in Fig. 5. For the lowest value of  $V_{DMS}$ ,  $\mu_n = 135$  and for the highest,  $\mu_n = 64$  cm<sup>2</sup>/V-sec. These may be compared with the reported bulk single-crystal value<sup>20</sup> of the order of 100 cm<sup>2</sup>/V-sec.

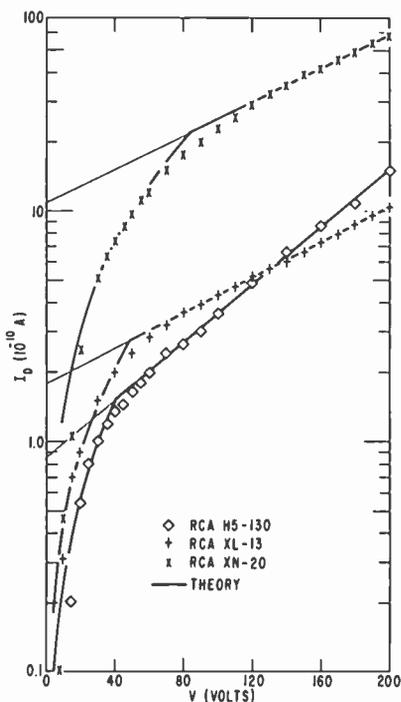


Fig. 5—Comparison of theoretical drift-limited dark current with experimental data for 3 RCA tubes.

The complete lack of agreement between the data for the Philips tubes and Eq. [7] is striking. Since no information about the fabrication method used for forming the beam-blocking layer is available, we may conjecture upon the possibility that it is different from that used for the RCA tubes and that one or more of the assumptions on which Eq. [7] is based is invalid.

An alternative explanation<sup>21</sup> for the nonagreement is based upon the existence of a photocurrent due to light from the heated cathode. This photocurrent is (in the present work) inseparable from the true dark current. The photocurrent would be expected to be largest for

two of the Philips tubes (XQ1020G-4973A and XQ1020G-29504) that have dispenser-type cathodes, which are maintained at a higher temperature than the normal oxide cathodes in the Philips 55875B and the RCA tubes. However, because the true dark-current level is so low in the 55875B, even a relatively small photocurrent component would affect the "apparent dark current" versus voltage curve.

A similar photocurrent component of the measured dark current may also be present in the RCA tubes. If present, it would tend to raise the apparent dark current; the effect would be most pronounced in the low-voltage region where the real dark current is low and less important in the high-voltage region where the real dark current is high. Further, it would have the greatest relative effect upon the tubes in which the true dark current is low. One result would be that the fitted theoretical curves in Fig. 5 would be skewed toward lower voltage and the values of  $V_{DMS}$  obtained from the curve fitting would be misleadingly low. We see that in Fig. 5 the value of  $V_{DMS}$  does tend to be lower for the tubes with lower dark current. If indeed, there is a photocurrent component in the RCA tubes, say of the order of a few times  $10^{-11}$  A, the effect on the RCA tube with the highest dark current (XN-20) would be slight and the value of  $V_{DMS}$  obtained from fitting the dark current curve for this tube would be the most accurate.

#### 4. Further Discussion, Speculation, and Summary

By using Eq. [4a] and the straight-line approximations to the experimental data shown in Fig. 4, we may obtain values of  $z(0)$  and  $\lambda D_S$  for each tube. These are shown in Table 1. We note that  $z(0)$  is less

Table 1—Beam-blocking Layer Parameters for 6 PbO Vidicons

Tube	$z(0)$ [eV]	$\lambda D_S$ [( $\text{cm}^2\text{-eV}$ ) $^{-1}$ ]
Philips XQ1020G-4973A	.986	$8.87 \times 10^{12}$
Philips XQ1020G-29504	.930	$6.52 \times 10^{13}$
Philips 55875B	1.026	$2.97 \times 10^{13}$
RCA XN-20	.947	$1.56 \times 10^{13}$
RCA H5-730	1.013	$9.05 \times 10^{12}$
RCA XL-13	.995	$1.49 \times 10^{13}$

than 0.97 eV for two of the tubes indicating that the BBR is n-type when  $V = 0$  and is therefore n-type for all values of  $V$ . For the other tubes,  $z(0)$  is greater than 0.97 eV indicating that the BBR is p-type at least at low  $V$ . The value of  $\lambda D_S$  is found to be of the order of mag-

nitude of  $2 \times 10^{13}/\text{cm}^2\text{-eV}$ . Although we have considered this as a product of the beam-blocking layer thickness and a density of states in volume-energy space, it is an equivalent density of states in surface-energy space. Considered in this way it is interesting to compare it with the energy density of surface states ( $\sim 4 \times 10^{13}/\text{cm}^2\text{-eV}$ ) found by du Chatenier and van den Broek<sup>15</sup> to be required at a metal contact to PbO in order to explain their results. The effective work function at a metal contact to PbO was found in that work to be  $\sim 0.95$  eV and relatively independent of the metal; this value is quite comparable with the  $z(0)$  values found in the present work. These comparisons indicate that a thin metal film on the scan side of the PbO layer (as suggested in Ref. [5]) would act as an effective beam-blocking layer. The film would have to be thin enough or discontinuous enough to prevent high lateral conduction and a consequent loss of resolution.

It was assumed in Sec. 2 that  $J_{D2}$  was small enough to be neglected. The relatively good agreement between the theory and the measured results confirm that this assumption is a reasonable one for "good" tubes. Although there appear to be no published studies of the structure and chemistry of the BBR produced by an oxygen-glow treatment, it seems likely that the result is a somewhat disordered material with a high trap density. The conventional description of the oxygen-glowed material is that it is PbO with a large acceptor doping.<sup>2,5</sup> It is, however, not inconceivable that the resultant material is at least partially a higher oxide of lead; e.g.,  $\text{PbO}_2$ ,  $\text{Pb}_3\text{O}_4$ , or even a non-stoichiometric oxide  $\text{PbO}_x$  where  $1 < x < 2$ . This is, of course, highly speculative. Nevertheless, the high degree of inelastic scattering and trapping in the beam-blocking region required by the model is consistent with what might be expected from a structurally and chemically disordered region.

The model used in Sec. 2 assumed a simple geometry—a BBR of uniform thickness  $\lambda$ . The actual geometry is considerably more complicated than this model. Because of the platelet structure of the layer, an incoming electron has considerably more surface area available to it. This high effective surface to volume ratio of the beam-blocking region may significantly enhance the thermalization and trapping of the incoming beam.

It is important to emphasize that we have dealt here with a schematic representation in which a complicated three-dimensional structure has been reduced to a highly simplified one-dimensional model retaining only those features that are necessary to explain the observed dark-current behavior of the PbO vidicon. As other behavioral characteristics are studied, it is anticipated that this "bare-bones"

model may require further detailing or particularization in order to conform to reality.

One of the ultimate limitations on the operating life of PbO vidicons is the failure of the beam-blocking region resulting in high dark current.<sup>2</sup> It is hoped that the model presented here will be helpful in the characterization of tubes during and after their useful operating lives and in this way contribute to an understanding of the failure mechanism.

In summary, a simple physical model for the beam-blocking contact in a PbO vidicon has been used to derive an approximate theory of the static dark current. The theory describes two regimes of dark-current density versus applied voltage:

- (1) A drift-velocity-limited regime in which the dark-current density is limited by the drift, in the photoconductor layer, of conduction-band electrons produced by thermal excitation from traps in the beam-blocking region.
- (2) A thermal-generation-limited regime in which the dark-current density is limited only by the supply of conduction-band electrons to the photoconductor layer resulting from thermal excitation in the beam-blocking region. Regime (2) will occur at a higher voltage than regime (1).

The theory relates the electrical behavior to the physical parameters of the model; thickness, trap-density distribution, and doping of the beam-blocking region, thickness of the photoconductor layer, and the electron mobility. Reasonable agreement is obtained between the theory and measurements on PbO vidicons. In addition, the theory provides an understanding of the factors involved in the optimization of the Fermi-level position in the beam-blocking region.

### Acknowledgments

It is a pleasure to acknowledge many valuable discussions with J. A. Inslee, A. R. Moore, and C. C. Wang during the course of this work. Helpful and encouraging comments by H. Kiess, M. A. Lampert, and A. Rose on various drafts of the paper are greatly appreciated. In addition, I wish to thank J. M. Breece for his help in carrying out the measurements and J. A. Inslee for providing the tubes used for those measurements.

It is appropriate here to acknowledge still another and greater debt—to Albert Rose who has served us all by his example of excellence, as a scientist, as an inventor, and as a warm, compassionate human being.

## References:

- <sup>1</sup> A description of conventional vidicon operation may be found in P. K. Weimer, S. V. Forgue, and R. R. Goodrich, "The Vidicon—Photoconductive Camera Tube," *RCA Rev.* 12, 306 (1951) or D. G. Fink, *Television Engineering*, McGraw-Hill Book Co., New York, (1960).
- <sup>2</sup> E. F. de Haan, A. van der Drift, and P. P. M. Schampers, "The 'Plumbicon', A New Television Camera Tube," *Philips Technical Review*, 25, p. 133 (1964).
- <sup>3</sup> L. Heijne, "The Lead Oxide Vidicon," *Acta Electronica*, 2, p. 124 (1957).
- <sup>4</sup> A. G. Van Doorn, "The 'Plumbicon' Compared with Other Television Camera Tubes," *Philips Technical Review*, 27, p. 1 (1966).
- <sup>5</sup> E. F. de Haan, P. P. M. Schampers, J. H. N. van Vucht, *Method of Manufacturing a Photoresponsive Device Comprising a Photoresponsive PbO Layer*, U.S. Patent No. 3372056, March 5, 1968.
- <sup>6</sup> A. van der Drift, "Texture of a Vapour-Deposited Lead-Monoxide Layer," *Philips Res. Repts.*, 21, p. 289 (1966).
- <sup>7</sup> A. van der Drift, "Evolutionary Selection, A Principle Governing Growth Orientation in Vapour-Deposited Layers," *Philips Res. Repts.*, 22, p. 267 (1967).
- <sup>8</sup> A. van der Drift, T. E. Horsman, and C. Langereis, "Investigations on the Preferred Orientations in Vapour-Deposited Lead-Monoxide Layers," *Philips Res. Repts.*, 23, p. 48 (1968).
- <sup>9</sup> L. Heijne, "Photoconductive Properties of Lead-Oxide Layers," *Philips Res. Repts.*, Suppl. 4 (1961).
- <sup>10</sup> J. van den Broek, "Contact Barriers in Red Lead Monoxide," *Philips Res. Repts.*, 20, p. 674 (1965).
- <sup>11</sup> J. van den Broek, "Optical Absorption of Red Lead Monoxide," *Philips Res. Repts.*, 22, p. 36 (1967).
- <sup>12</sup> J. van den Broek, "The Electrical Behaviour of Vapour-Deposited Lead-Monoxide Layers," *Philips Res. Repts.*, 22, p. 367 (1967).
- <sup>13</sup> J. van den Broek, "Optical Lattice Vibrations and Dielectric Constant of Tetragonal Lead Monoxide," *Philips Res. Repts.*, 24, p. 119 (1969).
- <sup>14</sup> F. J. du Chatenier, "Space-Charge-Limited Photocurrent in Vapour-Deposited Layers of Red Lead Monoxide," *Philips Res. Repts.*, 23, p. 142 (1968).
- <sup>15</sup> F. J. du Chatenier and J. van den Broek, "Electrical Properties of Vapour-Deposited Layers of Red Lead Monoxide," *Philips Res. Repts.*, 24, p. 392 (1969).
- <sup>16</sup> A. M. Goodman and A. Rose, "Double Extraction of Uniformly Generated Electron-Hole Pairs from Insulators with Non-Injecting Contacts," *J. Appl. Phys.*, 42, p. 2823 (1971).
- <sup>17</sup> M. A. Lampert and P. Mark, *Current Injection in Solids*, Part I, Academic Press, New York (1970).
- <sup>18</sup> R. H. Dean and P. M. Schwartz, "Field Profile in n-GaAs Layer Biased Above Transferred-Electron Threshold," *Solid State Elect.*, 15, p. 417 (1972).
- <sup>19</sup> P. J. Zanzucchi (private communication).
- <sup>20</sup> R. Keezer, D. Bowman, and J. Becker, "Electrical and Optical Properties of Lead Oxide Single Crystals," *J. Appl. Phys.*, 39, p. 2062 (1968).
- <sup>21</sup> J. A. Inslee (private communication).

# Semiconductor Heterojunction Vidicons

C. R. Wronski

RCA Laboratories, Princeton, N. J. 08540

**Abstract**—The introduction of the silicon vidicon demonstrated that a good-quality semiconductor can be used as a vidicon target. However, the fabrication of a structured target that consists of a large array of electrically isolated n-p homojunctions is difficult. The use of an appropriate insulator or granular metal (cermet) heterojunction offers the possible elimination of such a structure, but stringent constraints are imposed on such junctions by the camera-tube requirements. Various silicon heterojunctions have been investigated and their junction properties as well as vidicon characteristics studied. By extending some of this work to wider-bandgap semiconductors, the dark current, resolution, and sensitivity could be directly correlated with the semiconductor, the insulating film, and their interface.

## 1. Introduction

The introduction of the silicon vidicon demonstrated that a good quality n-type semiconductor target can be used in a camera tube.<sup>1</sup> The low densities of centers in the forbidden gap, the excellent free-carrier transport, and the bandgaps of such materials offer possible improvement in the sensitivity and response of photocurrents over those of the previously used wider-bandgap, insulating photoconductors.<sup>2,3</sup> However, targets that utilize n-type semiconductors with resistivities very much less than  $10^{12}$  ohm-cm require blocking contacts to the electron beam in order to have acceptably low dark currents. This has been achieved in the Si vidicon by thermalizing the electrons from the beam in a thin p-region of an n-p homojunction. These p regions are highly conducting, so that adequate TV resolu-

tion can only be obtained if an array of electrically isolated diodes, typically  $660 \times 660$ , is used. The difficulties associated with such a structured target are eliminated if the blocking contact has a lateral resistance  $\geq 10^{13}$  ohms per square.<sup>4</sup> Such high resistances have been obtained with heterojunctions formed between a semiconductor and an insulating film and more recently in semiconductor-granular-metal junctions.<sup>5</sup> The granular metals consist of a fine dispersion of metallic particles in an insulator; granular metals act like a metal for electron and hole transport, but at the same time have an electron mobility low enough to maintain the required lateral resistance.

Although work on semiconductor heterojunction vidicons has been going on for several years,<sup>6</sup> there have been no systematic studies of the dependence of vidicon characteristics on junction properties nor any discussion of the physical processes underlying their operation. It is the purpose of this paper to consider the dependence of the vidicon characteristics, i.e., of dark current, photocurrent, resolution, and lag on the junction properties and to provide an evaluation of the semiconductor-insulator and semiconductor-granular-metal vidicons.

## 2. Operation of a Semiconductor Heterojunction Vidicon

The operation of a semiconductor heterojunction target is similar to that of an Si vidicon, and is shown schematically in Fig. 1. The target

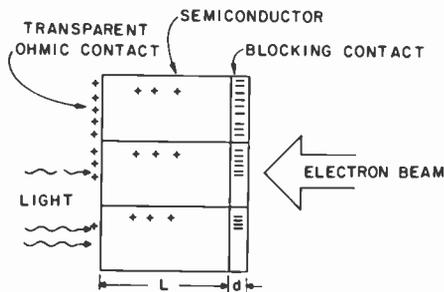


Fig. 1—Schematic representation of semiconductor heterojunction vidicon.  $L$  and  $d$  are the thicknesses of the semiconductor and the blocking contact, respectively.

consists of an n-type semiconductor of thickness  $L$ , with an ohmic contact on one side and an insulating film of thickness  $d$  on the other. A positive voltage  $V_T$  in the range  $10 < V_T < 50$  volts, is applied across the target while an electron beam scans the blocking contact

charging up the target every frame time,  $t_f = 1/30$  sec. This reverse bias on the heterojunction forms a depletion region in the semiconductor which, for optimum target performance, must extend through the entire thickness  $L$ . Light incident through the transparent ohmic contact generates holes in the semiconductor that are then driven towards the blocking contact and discharge the target to form a charge pattern of the scene. The video signal is generated by the charge that is deposited by the beam due to the voltage pattern of the scanned surface,  $V_0$ , that occurs during a frame time,  $t_f$ . This charge is  $Q_0 = C_0 V_0$ , where  $C_0$  is the capacitance of the target; the photocurrent is  $I_L = Q_0/t_f = C_0 V_0/t_f$ . For a vidicon to have a 100% readout efficiency, the beam must be able to charge the target in a time much less than  $t_f$ . This charging time depends on the effective impedance of the beam and the capacitance and thickness of the target. A target capacitance of  $\sim 1500$  pF  $\text{cm}^{-2}$  is desirable for a charging time much less than  $t_f$ , which for a dielectric constant of 10 corresponds to a fully depleted semiconductor of  $\sim 5 \times 10^{-4}$  cm thickness. For 100% quantum efficiency each incident photon must generate a hole that recombines with an electron deposited by the beam. In the absence of surface recombination, a saturated photocurrent having a quantum efficiency of 100% can be obtained for wavelengths having energies greater than the bandgap of the semiconductor. Such photocurrents can be achieved even in semiconductor targets that have a thickness greater than the diffusion length of holes if the semiconductor is completely depleted. Such operation ensures not only that there is no field-free region in which the photogenerated holes can undergo recombination but also that the capacitance of the target is at a minimum. A semiconductor having a bandgap  $\leq 1.8$  eV is capable of a uniform sensitivity across the visible spectrum ( $\lambda = 400$  to  $700$  nm). The operating voltage,  $V_T$ , must be sufficiently high to ensure this depletion, even in the presence of  $V_0$  and any trapping of the holes prior to reaching the steady state. All such holes must reach the location of the deposited beam electrons, which makes this effective location in the target important. The two types of heterojunction structures considered here both have a blocking contact formed by depositing onto the semiconductor a thin ( $\sim 1000$  Å) film with a resistivity  $> 10^{13}$  ohms per square. In such a target the beam electrons are thermalized close to the scanned surface and the resistivity of the film does not limit the target resolution.

The energy-band diagrams of two types of such targets are shown in Fig. 2. Fig. 2(a) shows a semiconductor insulator target where the beam electrons are trapped in the insulator close to the scanned surface. Also shown are the negative charges arising from charged sur-

face states at the interface and the centers in the bulk of the insulator. The semiconductor has a bandgap  $E_1$ , an ionized space-charge density  $N_+$ , and a density of hole trapping centers below midgap  $N_t$ . At the interface there is a density  $N_s$  of surface states, assumed to be localized at an energy  $E_s$ ; the discontinuity between the valence bands of the semiconductor and the insulator is  $\Delta E_v$ . The applied target voltage  $V_T$  divides into  $V_1$  and  $V_2$  across the two sides of the junction. For efficient operation  $V_1$  must be large enough to ensure depletion of the semiconductor and  $V_2$  sufficient to ensure that all the photogenerated holes reach the scanned surface. This requires

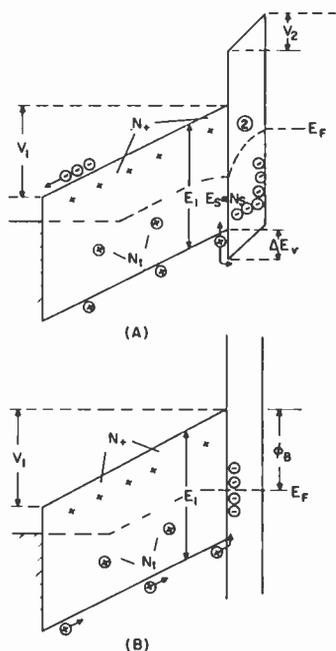


Fig. 2—Schematic representation of two heterojunction vidicon targets under reverse bias: (A) Semiconductor-insulator junction and (B) granular-metal-semiconductor junction.

that hole currents up to  $\sim 1 \mu\text{A}/\text{cm}^2$  be injected into and transported across the insulator film. The operating conditions of the vidicon target must be such that this photocurrent can be transported to the scanned surface.

Fig. 2(b) is a diagram of an n-type semiconductor-granular-metal

(cermet) Schottky barrier.<sup>5</sup> In this structure, the thermalized beam electrons reside at the junction interface where their injection into the semiconductor is limited by the barrier height  $\phi_B$ . In this case, essentially all of the target voltage  $V_T$  is across the semiconductor, and the photogenerated holes driven to the semiconductor interface recombine with the electrons stored there.

The advantages offered by the semiconductor can be fully utilized only if the vidicon characteristics are not limited by the blocking contact. It is, therefore, necessary to evaluate the contribution of both heterojunction materials to the performance of the target. This can be done by studying the heterojunction properties and vidicon characteristics of various semiconductor-insulator and semiconductor-granular-metal systems. The dependence of some of the vidicon characteristics on these junction properties is discussed below, and results with structures formed on n-type Si, GaAs, and CdSe semiconductors are presented.

### 3. Dark Currents

The vidicon dark current is due to free-carrier injection and the thermal generation of hole-electron pairs in the structure. For the ohmic contact to the n-type semiconductor, hole injection can be neglected. In the case of the semiconductor-insulator target, electron injection from the electron beam can be neglected if all the electrons are permanently trapped close to the scanned surface and cannot reach the semiconductor. In the granular metal target, although the beam electrons are thermalized close to the scanned surface, they migrate to the junction interface. The dark current is then determined by the barrier height,  $\phi_B$ , and the well-known Richardson equation.<sup>7</sup> For a uniform junction and electric fields  $\leq 10^5$  V/cm this is a saturated current which for  $\phi_B \geq 0.9$  eV is  $\leq 10^{-9}$  A/cm<sup>2</sup>.

The dark current due to hole-electron generation in the semiconductor depends on its bandgap  $E_1$  and on the densities and energies of the centers.<sup>8</sup> This current can be written

$$j_B = q\sigma_e N_c v_{th} N_g l \exp\left\{-\frac{E_1 q}{2kT}\right\}, \quad [1]$$

where  $q$  is the electronic charge,  $\sigma_e$  is the electron capture cross section of the centers,  $v_{th}$  is the carrier thermal velocity,  $N_c$  is the density of states in the conduction band,  $N_g$  is the effective density of centers at midgap,  $E_1$  is the bandgap of the semiconductor, and  $l$  is the depletion width of the semiconductor. When  $l = L$  this current becomes independent of the target voltage.

In the case of the insulator heterojunctions, current is generated by the surface states at the interface. Since the normal densities of such states,  $10^{12}$  to  $10^{13}$   $\text{cm}^{-2}$ , can exceed the total number of the generating centers in the bulk of the semiconductor target, their contribution to the dark current must be considered. This generation is analogous to that of the bulk except that the effective bandgap for electron-hole generation is  $E_1 + \Delta E_v$  and the density of centers is constant, being independent of the depletion width. For  $N_S$   $\text{cm}^{-2}$  surface states at an energy  $E_S$ , as shown in Fig. (2a), to a good approximation this generation current can be written

$$j_S = q\sigma_e v_{th} N_c N_s \exp\left\{\frac{-q\left(E_1 + \Delta E_v\right)}{kT} - \Delta E\right\}, \quad [2]$$

where  $\Delta E$  is the distance of  $E_S$  from the steady-state Fermi level located close to the midpoint of the junction barrier ( $E_1 + \Delta E_v$ ). At electric fields of  $\lesssim 10^5$  V/cm this generation current is independent of voltage.

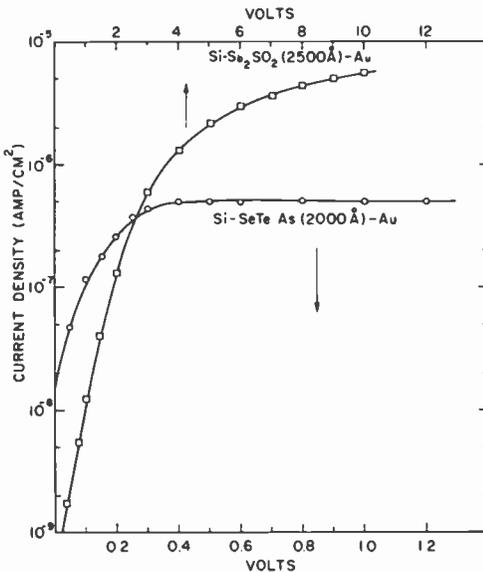


Fig. 3—Reverse bias currents of a Si-SeTeAs (2000 Å) and an Si-Sb<sub>2</sub>SO<sub>2</sub> (2500 Å) heterojunction. The voltage scale for the Si-Sb<sub>2</sub>SO<sub>2</sub> junction is ten times larger than that for the Si-SeTeAs one.

The dark currents in the semiconductor-insulator targets can be limited by the transport of holes through the insulator film or, as in the case of semiconductor p-n junctions,<sup>9</sup> at reverse biases greater

than a few  $kT/q$  they can be determined only by the carrier generation in the junction. The dark reverse bias currents were investigated for a variety of insulator heterojunctions using both evaporated electrodes and electron-beam scanning. Examples of field-limited and non-field-limited dark currents in a semiconductor-insulator junction are shown in Fig. 3. The currents are those of an Si-SeTeAs and an Si-Sb<sub>2</sub>SO<sub>2</sub> junction with gold contacts where both of these junctions were formed on single-crystal silicon wafers such as used in the silicon vidicon. It should be noted that the voltage scale for the Sb<sub>2</sub>SO<sub>2</sub> junction is an order of magnitude greater than that for the SeTeAs junction. The SeTeAs junction reverse current saturates at less than a volt and remains constant up to voltages in excess of 30 volts. The voltage required to saturate the Sb<sub>2</sub>SO<sub>2</sub> junction current, as well as other amorphous insulators such as As<sub>2</sub>S<sub>3</sub> and As<sub>2</sub>Se<sub>3</sub>, indicate that large fields  $\geq 10^5$  volts are required across the insulators. No such limitation is present in the amorphous p-type SeTeAs alloy junctions because of their good hole transport.<sup>10</sup>

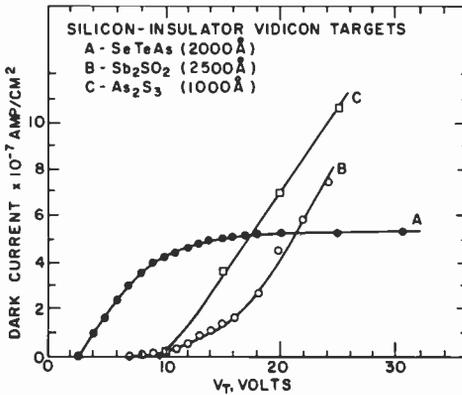


Fig. 4—Electron-beam dark currents of three Si-insulator vidicon targets. The thickness of insulator films are: SeTeAs-2000 Å, Sb<sub>2</sub>SO<sub>2</sub>-2500 Å, As<sub>2</sub>S<sub>3</sub>-1000 Å.

The dark currents obtained with electron-beam scanning measurements were in agreement with electrode current measurements. The vidicon dark currents for a SeTeAs (2000 Å), Sb<sub>2</sub>SO<sub>2</sub> (2500 Å) and As<sub>2</sub>S<sub>3</sub> (1000 Å) targets are shown in Fig. 4. Saturated dark currents are obtained with the Si-SeTeAs targets at ~10 volts. The difference between this saturation voltage and that measured with metal con-

tacts is due to the high currents ( $\geq 100$  nA/cm<sup>2</sup>) and the large impedance of the beam in the demountable system. Although the Sb<sub>2</sub>SC<sub>2</sub> and As<sub>2</sub>S<sub>3</sub> targets exhibited lower dark currents at low target voltages, at high voltages they exceeded the saturated currents of the SeTeAs alloy targets.

The dark currents through the insulator junctions, other than the SeTeAs junction, are complicated by their dependence on the large fields present. Although these junctions included more resistive materials, their quasi-saturated values were greater than those of the SeTeAs junction. Low reverse currents could be obtained with the insulator junctions, but these were in a region of insignificant depletion of the Si because almost all of the applied voltage was across the insulator films. The SeTeAs junctions, however, can unambiguously be related to the carrier generation in the target. The reverse currents obtained in these Si heterojunctions are two to three orders of magnitude higher than those of the p-n junctions grown on similar wafers, and cannot be associated with general currents generated in the bulk Si. The *saturated* reverse currents, particularly those of the SeTeAs junctions, can only be interpreted in terms of interface surface-state generation from Eq. [2]. The saturated currents of the SeTeAs junctions at  $2\text{--}10 \times 10^{-7}$  A/cm<sup>2</sup> and their thermal activation energy indicate a junction barrier,  $E_1 + \Delta E_v$ , of only 1.1 to 1.2 eV and a surface-state density  $\sim 10^{13}$  cm<sup>-2</sup> at the midpoint of the barrier. Such densities of surface states have been reported for Si surfaces<sup>11</sup> in the region 0.6 to 0.8 eV from the conduction band, the region where the steady-state Fermi levels are located.

Because the current generated depends on the junction barrier and on the interface states, Eq. [2], the bandgap of the semiconductor is important in determining the dark currents in heterojunction targets. These currents in Si ( $E_g = 1.1$  eV) junctions can in principle be lowered by having a larger  $\Delta E_v$ . This, however, makes the injection of the photogenerated holes into the insulator valence band more difficult, as discussed later, so it is more advantageous to increase the junction barrier by using a wider-bandgap semiconductor. Significant reduction in reverse-biased currents is obtained with heterojunctions formed on GaAs ( $E_g = 1.4$  eV) and CdSe ( $E_g = 1.8$  eV) but the decrease due to the larger bandgaps is modified by the  $\Delta E_v$  and interface states in the junctions. The currents obtained with SeTeAs junctions formed on GaAs are an order of magnitude lower than those formed on Si, and a further reduction of about two orders of magnitude is obtained in heterojunctions formed on CdSe. These currents exhibit saturation characteristics at voltages  $\leq 3$  volts, but at larger reverse biases they are no longer saturated. This can be attributed to

the electric fields and nonuniformities at the interface. For a dielectric constant of 10, a density of  $10^{12}$   $\text{cm}^{-2}$  charged states results in a field  $\sim 10^5$  V/cm. In heterojunctions formed on CdSe, air baking the CdSe prior to the SeTeAs deposition greatly reduces the currents. The oxidation does not form an electron blocking contact but is sufficient to change the resultant CdSe-SeTeAs interface. Since large densities of surface states have been found on single-crystal CdSe surfaces,<sup>12</sup> control of these states can significantly reduce the electric field due to their ionized surface charge. The dark currents obtained from such a heterojunction are shown in Fig. 5. The data are for a

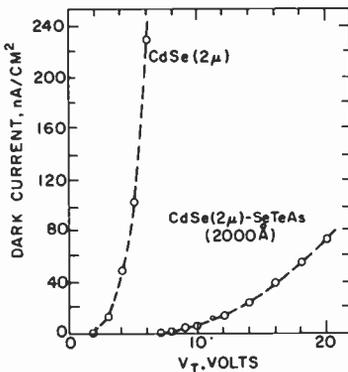


Fig. 5—The dark current of an evaporated CdSe (2- $\mu\text{m}$ -thick)-SeTeAs (2000- $\text{\AA}$ -thick) heterojunction target. Also shown is the dark current of the CdSe layer without the heterojunction.

2000  $\text{\AA}$  SeTeAs film deposited onto a 2- $\mu\text{m}$  CdSe layer. Also shown in Fig. 5 is the dark current obtained from the air-baked CdSe layer without the heterojunction. The heterojunction reduces the quasi-ohmic dark current obtained with the CdSe layer. The current is unsaturated, as are the junction reverse currents at large reverse biases, but is in the nA/cm<sup>2</sup> range over an extended region of voltage. Such low dark currents can be obtained over a wider voltage range with heterojunctions formed using insulators such as As<sub>2</sub>S<sub>3</sub> and Sb<sub>2</sub>SO<sub>2</sub>, but these targets also require large electric fields and operating voltages for efficient photocurrent operation.

Granular-metal films, obtained by the co-deposition of metals with insulators, offer a wide range of controlled resistivities<sup>13</sup> by adjusting

the relative volume fraction of the insulator. The structure of films having resistivities  $\sim 10^9$  ohm-cm is that of metallic particles  $\sim 20$  Å in size, embedded in the insulator.<sup>14</sup> Characteristics of the granular-metal-semiconductor contacts can be explained in terms of a Schottky barrier with a well-defined barrier height, even with as little as 10 vol. % of metal. The granular metal acts as a "high-resistance metal" characterized by an effective Fermi level as shown in Fig. 2(b), and the junction can be represented in terms of a parallel combination of the granular-metal-film resistance and capacitance in series with the barrier.<sup>5</sup> The barrier height does not depend on the resistivity of the granular metal or the grain size. However, there is a dependence on the semiconductor surface and on the changes that take place before or during the deposition.

Granular-metal junctions formed on Si and GaAs n-type crystals have barrier heights that are essentially independent of the metal-insulator combinations used (Au, Ni, SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>). These barrier heights are  $\sim 0.6$  eV, which result in reverse bias currents in excess of  $\sim 10$   $\mu$ A/cm<sup>2</sup>. Such currents are too high for any practical application. The barrier heights obtained on CdSe, both single crystal and layers evaporated onto TIC faceplates, were  $\sim 0.7$  eV. Surface treatment of the semiconductors, similar to that used for the SeTeAs vidicon targets discussed earlier, increased the barrier heights to  $\sim 1.0$  eV. This reduced the reverse-bias currents to  $\sim 1$  nA/cm<sup>2</sup>. These reverse-bias currents, however, were not truly saturated as one would expect for an ideal Schottky barrier. This is probably due to nonuniformities in  $\phi_B$  found between small-area contacts both on single crystals and evaporated layers of CdSe. The dark current of a polycrystalline CdSe-granular-metal target scanned by an electron beam is shown in Fig. 6. The CdSe layer is 1.6  $\mu$ m thick and the AuSiO<sub>2</sub> granular metal is 4000 Å thick with a composition of 10 vol. % Au and 90 vol. % SiO<sub>2</sub> and a resistivity of  $\sim 3 \times 10^9$  ohm-cm. Also shown in the figure is the current density obtained from an area of the target not covered by the granular metal. The ability of the granular metal junction to block the electrons from the beam is demonstrated by the dramatic reduction in the current corresponding to the area coated by the granular metal.

#### 4. Photocurrents

The conditions under which a steady-state photocurrent can be drawn through a fully depleted semiconductor are determined by the hole transit time, the level of illumination, and the hole-trapping centers below midgap. The photocurrent,  $j_i$ , corresponding to a flux  $F$  of incident photons is given by

$$j_i = qF = q\mu\bar{p}_1\bar{\epsilon}_1, \quad [3]$$

where  $\mu$  is the hole mobility,  $\bar{p}_1$  is the mean minimum density of free holes and  $\bar{\epsilon}_1$  is the mean electric field,  $V_1/L$ , across the semiconductor. For  $F = 3 \times 10^{12} \text{ cm}^{-2} \text{ sec}^{-1}$  ( $j_i = 5 \times 10^{-7} \text{ A/cm}^2$ ),  $\mu = 100$ , and  $\bar{E}_1 = 5 \times 10^4 \text{ V/cm}$ , the minimum  $\bar{p}_1$  that is required is  $\sim 10^6 \text{ cm}^{-3}$ . This corresponds to a quasi-Fermi level  $\sim 0.7 \text{ eV}$  from the valence band. For  $E_1 \gtrsim 1.4 \text{ eV}$  the Fermi level is displaced from its position in the dark. This displacement also increases the occupation of the hole-trapping centers giving an increase in the positive space charge. If the density of these centers,  $N_t$ , is comparable to or greater than the donor

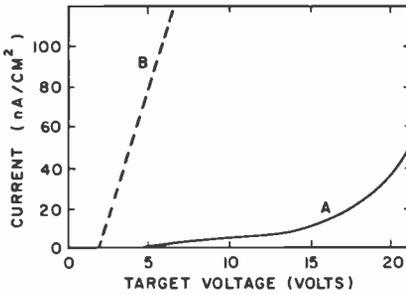


Fig. 6—(A) The dark current of a granular metal (Au:SiO<sub>2</sub>)-semiconductor CdSe vidicon target. The Au:SiO<sub>2</sub> film is 90 vol. % SiO<sub>2</sub>, has a thickness of 4000 Å and a resistivity  $\sim 3 \times 10^9 \text{ ohm-cm}$ . The CdSe is a 1.6- $\mu\text{m}$ -thick polycrystalline layer. (B) is the dark current of the CdSe layer with no granular-metal junction present.

and ionized electron trap densities,  $N_+$ , the voltage required to deplete  $L$  is greater than that in the dark. Consequently, the potential drop,  $V_1$ , that must exist across the semiconductor must be  $\gtrsim (N_+ + N_t) L^2 \times 10^{-7}$  for a dielectric constant of 10.

In the semiconductor-granular-metal targets, all the holes reaching the junction interface recombine with electrons there, and no additional voltage is required across the blocking contact for vidicon operation. In the insulator heterojunctions, however, the hole photocurrent must be injected into the insulator and transported across it. The electric fields and voltages across the insulator that are required to accomplish this depend on the alignment of valence bands,  $\Delta E_v$ , the thickness of the film, and its hole transport characteristics. These voltages can be significantly higher than the potential required across the semiconductor and, thus, dominate the characteristics, performance, and operating conditions of the heterojunction target.

The hole current that is thermally injected from the semiconductor into the insulator valence band for the junction shown in Fig. 2(a) is given by

$$\begin{aligned}
 j_s &= \frac{1}{4} qv_{th}p_1 \exp\left\{\frac{q\Delta E_v}{kT}\right\} \\
 &= \frac{1}{4} qv_{th}N_{v_1} \exp\left\{-\frac{q}{kT}(E_{v_1} + \Delta E_v - E_{FS})\right\}, \quad [4]
 \end{aligned}$$

where  $E_{v_1} - E_{FS}$  is the distance of the quasi-Fermi level at the interface from the semiconductor valence band,  $E_{v_1}$ , which has a density of states  $N_{v_1}$ . For  $v_{th} = 10^7$  cm/sec,  $N_{v_1} = 10^{19}$  cm $^{-3}$ , a photocurrent  $j_s = 5 \times 10^{-7}$  A/cm $^2$  will be injected into the insulator when  $(E_{v_1} + \Delta E_v - E_{FS}) \lesssim 0.75$  eV. For a junction barrier,  $E_1 + \Delta E_v \gtrsim 1.5$  eV a displacement of the quasi-Fermi level at the interface occurs under illumination and there is a corresponding filling of surface states by the trapped holes. This trapping increases the positive charge and the electric field; an increase in the electric field of  $\Delta\epsilon_s$  at the interface results in an additional potential drop,  $\Delta\epsilon_s \times d$ , across the insulator. Since the displacement of the quasi-Fermi level depends on  $\Delta E_v$ , it is desirable to minimize this discontinuity. This not only ensures that the hole photocurrent can be thermally injected without Schottky-barrier lowering and tunneling, but also minimizes the additional voltage drop across the insulator film. The transport of the photocurrent across the insulator film requires a minimum voltage across the film that is determined by its hole transport characteristics, its thickness, and the level of illumination. For most insulators these photocurrents are in the form of space-charge-limited hole currents,<sup>15</sup> which have a high power dependence on thickness and are determined by the hole mobility and trap distribution of the insulator. The more p-type the insulator and better its hole transport, the lower the fields required to carry these currents. Even though the voltage required across the insulator can be lowered by decreasing its thickness, a lower limit on its thickness is imposed by the required electron-beam blocking capability of the film.

The effect that the insulator can have on the photocurrent characteristics of a heterojunction structure is shown in Fig. 7 for an Si-SeTeAs (3500 Å) and an Si-As $_2$ S $_3$  (1000 Å) junction with evaporated gold contacts. Both junctions were illuminated by a microscope lamp, but in order to obtain comparable photocurrents, a 5% neutral density filter had to be used with the SeTeAs junction. The SeTeAs junction shows a saturated photocurrent at  $\sim 0.5$  V reverse bias, the same voltage at which the dark current of this junction (shown in Fig. 3)

reaches its saturation value of  $5 \times 10^{-7}$  A/cm<sup>2</sup>; 50% of the saturated photocurrent was obtained under short-circuit conditions ( $V = 0$ , Fig. 7). In this case there are no serious limitations imposed on the injection of photogenerated holes into the alloy and their transport through it. The saturated photocurrents had a linear dependence on intensity and their quantum efficiencies (Q.E.) were similar to those of the silicon vidicon after corrections are made for hole lifetime and

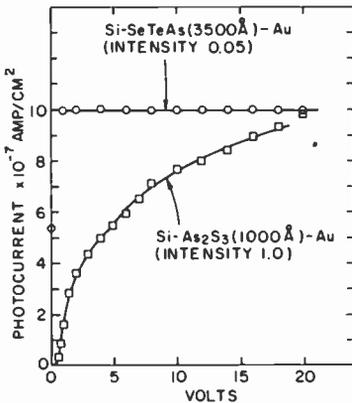


Fig. 7.—Photocurrent characteristics of an Si-SeTeAs junction (SeTeAs film thickness 3500 Å) and an Si-As<sub>2</sub>S<sub>3</sub> junction (As<sub>2</sub>S<sub>3</sub> film thickness 1000 Å). Intensity of illumination used on the As<sub>2</sub>S<sub>3</sub> junction is 20 times that used on SeTeAs junction.

thickness of the Si wafer (0.65 mm). In the As<sub>2</sub>S<sub>3</sub> junction the photocurrent was not saturated even at 20 volts, where the dark current was  $4 \times 10^{-6}$  A/cm<sup>2</sup>, and the Q.E. is much less than for the Si-SeTeAs junctions. The difficulty of collecting the photogenerated holes from the Si and their transport through the As<sub>2</sub>S<sub>3</sub> limit the sensitivity of the structure. In this case, the voltage drop across the insulator is greater than that across the semiconductor. Although in some cases this may not limit the sensitivity, it does affect the operating voltage of the target. An example of this is shown in Fig. 8 for CdSe-SeTeAs (3000 Å) and CdSe-As<sub>2</sub>S<sub>3</sub> (1000 Å) heterojunction targets scanned by an electron beam. The junctions were formed on separate areas of the same CdSe layer,  $\sim 2 \mu\text{m}$  thick, which had a space charge density  $\sim 8 \times 10^{15}$  cm<sup>-3</sup>. The photocurrent in the SeTeAs junction saturates at  $\sim 60\%$  Q.E. at  $\sim 30$  volts, a voltage close to that required to fully deplete the CdSe layer. The photocurrent in the As<sub>2</sub>S<sub>3</sub> junction, how-

ever, does not saturate until a target voltage of  $\sim 50$  volts is reached, although the  $\text{As}_2\text{S}_3$  film is three times thinner than the alloy. Also, the currents in this junction, prior to saturation, exhibited large transient decays to a steady state, which indicates space-charge effects in the  $\text{As}_2\text{S}_3$  and the interface. The additional 20 volts required for the  $\text{As}_2\text{S}_3$  junction saturation, over that for the SeTeAs junction, corresponds to an additional field  $\sim 2 \times 10^6$  V/cm across the  $\text{As}_2\text{S}_3$  film that is required to transport the photocurrent.

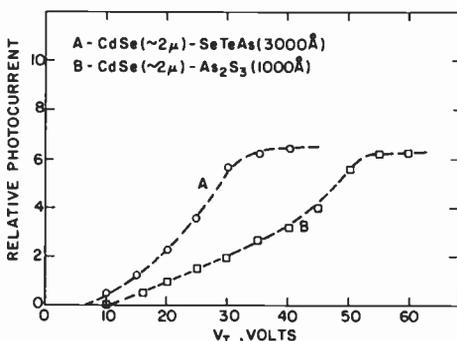


Fig. 8—The photocurrents of two heterojunction targets formed on the same CdSe evaporated layer. (A) is a junction formed with 3000 Å of SeTeAs. (B) is a junction formed with 1000 Å of  $\text{As}_2\text{S}_3$ .

In the granular-metal-semiconductor junctions there is no injection or transport of the photogenerated holes through an insulator. The photocurrent is obtained by their recombination with electrons that migrate under low fields to the granular-metal-semiconductor interface. This metal-like behavior of the granular metal for electron transport eliminates any difficulties associated with hole trapping at the interface and any electric field requirements across an insulator. This results in a negligible potential drop across the blocking contact, so that the target operating voltage is then only the voltage required to efficiently generate and draw the photocarriers across the semiconductor. The vidicon photocurrents of a CdSe-granular-metal target, formed on a 2- $\mu\text{m}$  polycrystalline CdSe layer deposited on TIC, are shown in Fig. 9, where the currents show saturation at a voltage less than 10 volts. These currents have a flat response across the visible spectrum,  $\lambda = 400$  nm to 700 nm, and a quantum efficiency close to 100%. This, together with the measured space charge density of  $\sim 10^{15}$

$\text{cm}^{-3}$ , is in agreement with the condition of complete depletion of the CdSe layer and short hole lifetimes ( $<10^{-9}$  sec).

### 5. Resolution and Lag

For adequate resolution, the time constant for lateral leakage must be greater than the scanning period. In the case of a uniform photoconductor target, having a capacitance of  $1500 \text{ pF/cm}^{-2}$ , a sheet resist-

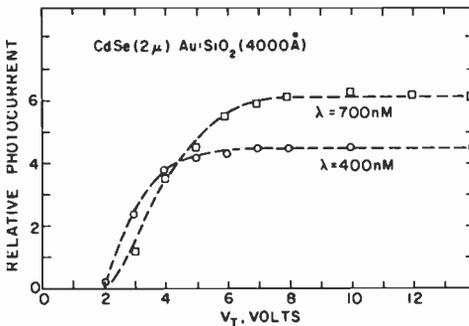


Fig. 9.—The saturated photocurrents of a CdSe-Au:SiO<sub>2</sub> granular-metal target for wavelengths  $\lambda = 400$  and  $700 \text{ nm}$ .

tance,  $R \sim 2 \times 10^{13}$  ohms per square is required.<sup>4</sup> For an insulating film of thickness  $\sim 1000 \text{ \AA}$ , a resistivity  $\geq 2 \times 10^8 \text{ ohm-cm}$  is sufficient. When the semiconductor is fully depleted, it is intrinsic and there is negligible contribution to the lateral conductance from its bulk. However, because there is a barrier at the interface of the heterojunction, the lateral surface conductance of the semiconductor must be considered. For a surface field  $\epsilon_s$  of  $5 \times 10^4 \text{ V/cm}$  and  $\mu = 100$ , to a good approximation, the sheet resistance is given by<sup>16</sup>

$$R = \frac{1}{qp_1\mu \exp\left\{\frac{(kT)}{q\epsilon_s}\right\}} = \frac{10^{23}}{P_{01}}, \tag{5}$$

where  $P_{01}$  is the density of holes in the valence band at the semiconductor surface. For  $R > 2 \times 10^{13}$  ohms per square,  $P_{01}$  must be less than  $\sim 5 \times 10^9 \text{ cm}^{-3}$ , which corresponds to a dark Fermi level more than  $\sim 0.55 \text{ eV}$  from the valence band of the semiconductor. For the insulator heterojunction, this means that  $(E_1 - \Delta E_v)/2$  must be  $\geq 0.55 \text{ eV}$ , which imposes a limit on the mismatch of the valence

bands at the interface depending on the semiconductor bandgap  $E_1$ . In the semiconductor-granular-metal junctions, the Fermi level is  $E_1 - \phi_B$  from the valence band of the semiconductor. Since  $\phi_B \geq 0.9$  eV is necessary for low dark currents, the bandgap of the semiconductor must be greater than  $\sim 1.4$  eV in order to simultaneously maintain the required resolution.

The resolution of the silicon heterojunctions was poor, the best being 300 TV lines for the SeTeAs junction targets. Other silicon-insulator targets had poorer and sometimes non-existent resolution, even though the resistance of the insulator films was greater than that of the SeTeAs. This lack of good resolution can be associated with the narrow bandgap of Si (1.1 eV) and, in the case of the SeTeAs junctions, the close alignment of the valence bands at the interface. The position of the quasi-Fermi level, even at midpoint of the junction, is marginal in fulfilling the resolution requirements. In the case of the other insulators, the filling of the interface states occurring during the photocurrent flow brings the quasi-Fermi level too close to the silicon valence band to satisfy Eq. [5]. The observed resolution on all of the CdSe heterojunction targets was 500 TV lines, which was limited by the demountable test equipment. This is consistent with the resistivities of the insulator and granular metal films used ( $\geq 10^9$  ohm-cm) and the larger junction barrier obtained with the wider-bandgap semiconductor.

The rise and decay times of the photocurrent signal depend on the charging characteristics of the target by the beam and on the trapping of photogenerated holes. To ensure that the charging of the target occurs in one frame time, a capacitance of  $\sim 1500$  pF/cm<sup>2</sup> is desirable. This corresponds to a depleted thickness of  $\geq 5$   $\mu$ m in a semiconductor with a dielectric constant of 10. Such a capacitance and depletion width is obtained at  $\leq 10$  volts across the semiconductor when its space-charge density is  $\leq 2 \times 10^{14}$  cm<sup>-3</sup>. Fig. 10 shows the (capacitance)<sup>-2</sup> versus voltage plot of an Si-SeTeAs (2000 Å) heterojunction that exhibits the uniform doping of the Si,  $2 \times 10^{14}$  cm<sup>-3</sup>, and a capacitance of 1500 pF/cm<sup>2</sup> at a reverse bias of  $\sim 8$  volts. If there is no significant trapping of the photogenerated carriers in the target, the capacitance under illumination remains essentially the same as in the dark. Extensive hole trapping and the subsequent change in electric fields at the operating voltage can decrease the depletion width in the semiconductor and hence increase the capacitance of the target. In addition to this, the kinetics of the filling and emptying of the hole traps and their location in the structure affect the rise and decay of the photocurrent signal.

Trapping of photogenerated holes can occur in the semiconductor;

for the insulator heterojunction trapping can also occur in the insulator and the interface. The extent of the trapping is determined by the position of the quasi-Fermi levels under illumination, and by trapping-center densities, cross sections, and relative energies. The thermal release of the trapped holes after the illumination determines their contribution to the decay of the photocurrent signal. This de-

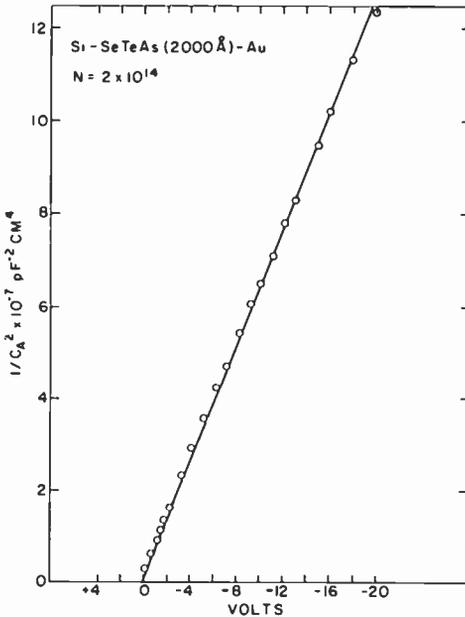


Fig. 10—The  $(\text{capacitance})^{-2}$  versus voltage plot of an Si-SeTeAs (2000 Å) junction.

pend on a large number of parameters, as in the case of insulator vidicons, and will not be discussed here, but it can be seen that the greater the trapping, the greater the probability of photoconductive lag. Minimizing the trapping is therefore important, and the use of granular metal limits this trapping to only that occurring in the semiconductor.

Although no quantitative lag measurements were carried out, the observed lag on CdSe was comparable to that reported for other CdSe vidicons of comparable thickness,  $\sim 2 \mu\text{m}$ .<sup>17</sup> In addition to the limit imposed by the capacitance of the fully depleted targets,  $\sim 4000 \text{ pF/cm}^2$ , hole trapping in the CdSe layers was found in measurements on the various heterojunctions used. However, meaningful lag measure-

ments on Si heterojunction targets, in which there is essentially no hole trapping in the semiconductor, could not be carried out because of the extremely high dark currents.

## 6. Conclusion

The semiconductor-insulator and semiconductor-granular-metal junctions discussed here illustrate the basic concepts of the semiconductor-heterojunction vidicon and the dependence of some of its characteristics on the junction properties. Properly designed large-area semiconductor-heterojunction vidicon targets have the advantages of low saturated dark currents; quantum efficiency of 100% across the visible spectrum; low lag and fast response; operation over a large voltage range; and absence of mosaic structure. In such structures, however, the advantages offered by the semiconductor must be fully utilized. This can only be achieved in junctions where the blocking contact to the electron beam satisfies the stringent camera-tube requirements discussed.

Some of the limitations that insulator blocking contacts can impose on heterojunction properties and vidicon characteristics have been demonstrated. In particular, the serious effect of high electric fields ( $10^5 - 10^6$  volts/cm across the blocking contact) required for efficient photocurrent generation limit the operating voltage range of the target; in addition, accompanying effects such as Schottky-barrier lowering, tunneling and breakdown make uniform target properties difficult to obtain. In the case of the semiconductor-insulator junctions, the surface states at the interface and the alignment of the valance bands play an important role in determining the required electric fields and the resulting dark currents. In the granular-metal-semiconductor junctions the Schottky barrier height determines the dark currents, and the voltages across the granular metal are negligible. The requirements of the semiconductor are that it have a low density of hole traps so that there is little or no space charge and the target can be operated in a completely depleted mode. The semiconductor bandgap is important in determining the spectral response of the target; in addition, it plays an important role in the interface state current generation and the resolution.

The use of granular-metal-semiconductor junctions offers distinct advantages over insulator film blocking contacts. The fundamental difference between the granular metal and the insulator targets is that in the case of the granular metals, the electrons deposited by the beam migrate under low fields to the granular-metal-semiconductor interface. In the insulator-semiconductor vidicons, the beam elec-

trons are located close to the scanned surface of the insulator and this often leads to objectionably high fields and voltage drops in the insulator. These difficulties are eliminated in the semiconductor-granular-metal junction since the granular metal acts like a metal for electron and hole transport, while at the same time it has an electron mobility low enough to maintain lateral insulation for the charge pattern.

### Acknowledgments

It is a pleasure to acknowledge the numerous helpful discussions with A. Rose in the course of this work. The author would also like to acknowledge B. Abeles and A. D. Cope for help and advice, R. Daniel and Y. Arie, for technical assistance, and W. M. Yim and L. Ekstrom for supplying the CdSe films.

### References:

- <sup>1</sup> M. H. Crowell and E. F. Labuda, in *Photoelectronic Imaging Devices*, ed. by L. M. Biberman and S. Nudelman, Plenum Press, New York (1971), Vol. II, p. 301.
- <sup>2</sup> R. W. Redington, see Ref. (1), p. 263.
- <sup>3</sup> E. H. Stupp and R. W. Levitt, see Ref. (1), p. 275.
- <sup>4</sup> P. K. Weimer, in *Advances in Electronics and Electron Physics*, Vol. XIII, Academic Press, New York (1960), p. 387.
- <sup>5</sup> C. R. Wronski, B. Abeles, R. E. Daniel and Y. Arie, "Granular Metal-Semiconductor Schottky Barriers," *J. Appl. Phys.*, **45**, p. 295 (1974).
- <sup>6</sup> Hiroo Hari and Tadashi Tsutsumi, *Japan. J. Appl. Phys.*, **10**, p. 1276 (1971); T. Yamato, S. Yoshikawa, K. Kobayashi, J. Matsuzaki, and I. Tagoshima, "A New Silicon Vidicon with a CdTe-Si Target," *IEEE Trans. Electron Devices*, **ED19**, p. 385 (1972); J. A. Hall, in *Advances in Electronics and Electron Physics*, Vol. 33A, Academic Press, New York (1972), p. 229.
- <sup>7</sup> E. Spenke, *Electronic Semiconductors*, McGraw Hill Book Co., New York (1958), p. 84.
- <sup>8</sup> A. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, Inc., New York (1967), p. 174.
- <sup>9</sup> A. G. Milnes and D. L. Fencht, *Heterojunctions and Metal Semiconductor Junctions*, Academic Press, New York (1972), p. 45.
- <sup>10</sup> M. N. Roilos, "Conductivity and Hall Effect in Vitreous  $As_2(Se,Te)_3$ ," *J. Non-Crystal. Solids*, **6**, p. 5 (1971); J. Schottmiller, M. Tabak, G. Locovsky, and A. Ward, "The Effects of Valency on Transport Properties in Vitreous Binary Alloys of Selenium," *J. Non-Crystal. Solids*, **4**, p. 80 (1970).
- <sup>11</sup> A. Many, Y. Goldstein, and N. Grover, *Semiconductor Surfaces*, Amsterdam, North-Holland Pub. Co.; New York, Interscience Pub. (1965).
- <sup>12</sup> R. L. Cosigny, III, and J. R. Madigan, "Surface Properties of Cadmium Sulfide," *Solid State Elec.*, **13**, p. 113 (1970).
- <sup>13</sup> C. A. Neugenbauer, "Resistivity of Cermet Films Containing Oxides of Silicon," *Thin Solid Films*, **6**, p. 443 (1970); B. Abeles, P. Sheng, and Y. Arie, *Advances in Physics* (in press).
- <sup>14</sup> R. W. Cohen, G. D. Cody, M. D. Coutts, and B. Abeles, "Optical Properties of Granular Silver and Gold Films," *Phys. Rev. B*, **8**, p. 3689 (1973).
- <sup>15</sup> A. Rose, *Concepts in Photoconductivity and Allied Problems*, Interscience Pub., New York (1963).
- <sup>16</sup> A. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Son, Inc., New York (1967), p. 267.
- <sup>17</sup> K. Shimizu, O. Yoshida, S. Aihara, and Y. Kiuchi, in *Advances in Electronics and Electron Physics*, Vol. 33A, Academic Press, New York (1972), p. 293.

# Simplicity in Theory: An Anecdotal Account of Current Injection in Solids—In Tribute to Albert Rose

Murray A. Lampert

Professor, \* Electrical Engineering, Princeton University, Princeton, N.J. 08540

**Abstract**—The "facts" of a scientist's contribution to a particular field can be counted, weighed, or otherwise reasonably quantified. The spirit infusing the work and the influence that the work has on others are not to be measured so easily. This article is concerned with Albert Rose's approach to physics, with the uncommon philosophy that underlay all of his diverse researches, and the profound influence which that philosophy has had on another person, namely the author. The framework for the narrative is the area of current injection in solids; for that is the area in which we collaborated for almost two decades, up to the present. The author attempts to show, as he was shown by Rose, that not all of solid-state physics is impenetrable by simple reasoning. On the other hand, the magically simple can be devilishly elusive; thereby hangs the tale, and quite possibly the author.

The dedication in "Current Injection in Solids"<sup>1</sup> (henceforth referred to as "CIS"), a book many years in the making, reads:

*To ALBERT ROSE  
who chose simplicity over precision  
and thereby gave to all of us insight.*

That dedication is also appropriate to set the theme of this article, which might well have been subtitled, "Simplicity—A Most Praise-

---

\* inactive status.

worthy Goal, Not Easily Achieved," or, "There's Many a Slip 'Twixt the Intent and the Manuscript." For, indeed, the article is mostly an account of a few simplicities achieved, and assorted difficulties encountered en route, in an area of solid-state physics in which Al Rose and I collaborated over the better part of two decades, namely the area of volume-controlled current injection in solids. Because I wished to emphasize the evolutionary aspects of some basic results in this field, rather than the results themselves, which are already in the literature, I obtained a license for an anecdotal narrative (a prose analog of the more renowned poetic license).

Generally, in writing a technical paper, one certainly does not have the luxury of detailing the steps in the all-too-often random walk by which one arrived at one's conclusions. There are, of course, many powerful reasons why this has to be so, not the least of which is that it boggles the imagination to envisage the technical audience, already mercilessly exposed to an unending deluge of less-than-cogent technical expositions, being further harassed by rambling descriptions of the blind alleys, exit-less mazes, shifting sands, and sundry other impedimenta the authors experienced on the path to their less-than-cogent results. Down such a road lies certain chaos, and more than likely, mass insanity.

Well then—why this anecdotal narrative? Most obviously, because the opportunity presented itself, and such occasions are rare. Beyond that, I've waited a long time to write this article. For, I'm afraid the article comes with a message, dear to myself and even more dear to Al Rose. That message is: Simplicity is Good, it is Useful, it is Important, and it is Neglected. Be there a reader skeptical of this last proposition, let him (her) pick up any recent inch-thick volume of *Physical Review B, Solid State* and make his (her) own evaluation. (For an unusually good statistical sample, this same skeptical reader might go directly to the 15 February, 1974 volume which came in at 2 $\frac{1}{8}$  inches.) The author finds it a curious aberration that there is so little regard for simplicity when it could be so useful. My ego insists that most, if not all, of the authors generating the vast sea of nearly unreadable papers must have the same difficulties as myself and the reader coping with the exponential torrent. Is some basic deficiency in empathy responsible? Can it be a desire for revenge lurking just beneath the surface of consciousness (an eye for an eye, a tooth for a tooth, and a Green's function for a Green's function)? Or is it a terror of nakedness of the mind, stripped of the warm, protective blanket of impenetrable equations?

Before proceeding, I had best make two points clear. First, although the content and style are indeed somewhat anecdotal, the

“anecdotes”, such as they are, are not of the personal variety. There are no revelations here of quaint, lovable behavior. And the wilder side of Al Rose’s professional career, such as it might have been, will have to be revealed by Rose himself, I suppose in his memoirs. Second, and more important, the ruts in the royal road to simplicity that I am able to write about are those that, for the most part, I, not Al, ran into. For better or worse, I have never asked Rose, “Al, how many nights and days, or weeks, or months, did you wander in the wilderness before you espied the Holy Grail?”<sup>2</sup>

With the reader now adequately warned, I can proceed with the narrative. Early in the game, it was appreciated by Mott and Gurney<sup>3</sup> that the energy-band picture of a crystalline insulator provided a conceptual framework for understanding current injection into the insulator. This is clearly discussed in Chapter V of their 1940 classic,<sup>3</sup> which also contains the derivation of the perfect-insulator, square-law analog to Child’s 3/2-power law for the vacuum diode. However, the real interest in current injection in insulators derives from the drastic effects, on the current, of localized states in the forbidden gap. Of course, Mott and Gurney knew such states had to be there—a good part of their book was about just such states. It is interesting to read their comments on this problem, excerpted from p. 173 of Ref. (3):

“In any real crystal of, for instance, silver bromide, there exist electron traps where, at sufficiently low temperatures, an electron finds a lower energy level than in the conduction band and can be held in a stable position. If electrons flowed out from our metal electrode into the non-metal, these traps would acquire electrons and the resulting space charge would prevent the passage of any further electrons into the crystal. Even if the number of traps is as low as  $10^{12}$  per  $\text{cm}^3$  (cf. Chap. IV, § 5.1), then, if all the traps in a layer  $10^{-5}$  cm thick were charged, this would produce a potential difference of 1 volt between the interior of the crystal and the surface. Space charges of this type would probably prevent the passage of any current at low temperatures.

“Probably, when a crystal is bombarded by cathode rays, the electrons entering the crystal get trapped in a similar way.

“At moderately high temperatures, however, a trapped electron is not stable but is released again after a short interval of time; and under these conditions it seems that a small current should flow.”

At “sufficiently low” temperatures they are mostly, but not entirely, right (see the discussion of the trap-filled-limit problem, below, and then footnote 15b). At higher temperatures they simply walked away from the problem after noting that it existed.

The field of current injection in insulators took off, in earnest, with

the post-war publications of Rose and his collaborators.<sup>4,5</sup> As is characteristic of most of Rose's work, the studies reported in these publications were born out of technological imperatives, in this instance out of the need to understand the electronic behavior of vidicon target materials.

It illuminates the theme of this article to compare Rose's formulation of steady-state, one-carrier space-charge-limited (SCL) current theory for solids with a conventional formulation. This comparison is provided in Table 1. Here, and throughout the entire article, the discussion is confined to steady-state currents flowing in a planar, one-dimensional configuration. Discussion of both transient effects and other geometries can be found in CIS. All equations are written in MKS units.

Rose's form of the theory is a space average of the "exact" theory, with suitable approximations for each of the three equations. It *always* gives the correct form of the  $J$ - $V$  characteristic, except for a numerical constant which is rarely off by more than a factor of two.<sup>8</sup> This is an extraordinarily small price to pay for the simplifications achieved, particularly since sample inhomogeneities and the effects of finite sample geometry don't really allow for higher accuracy, no matter how exact the starting theory—an important fact frequently buried under the mountain of computer print-out. This direct track to the physical content of a theory, without undue concern for some loss of precision, is the very hallmark of Rose's approach to physics. And his formulation of one-carrier SCL current theory is a very successful example of his approach. Another sparkling example is provided at the end of this article.

For those who like to muse on unanswerable questions, an interesting one is: If Rose hadn't published his formulation in the early 1950's, would anybody else have gotten to it in the intervening two decades? I, myself, suspect not. Rose's formulation is useful *only* because of the approximations that go along with it. There is a free-wheeling quality about the approximations which runs against the grain of the contemporary physicist. The approximations seem to come right out of intuition, rather than from closely cropped differential equations or carefully truncated Hamiltonians.<sup>9</sup> His approach yields Child's famous 3/2-power law with no effort whatever. Replace Eq. [1] by  $T_0 = L/v$ , with  $v \simeq (eV/m_0)^{1/2}$  corresponding to a ballistic trajectory in the vacuum. Now, using Eq. [2] with  $Q_{tot} = Q_f$ —there are no electron traps in the vacuum—we obtain  $J \simeq \epsilon_0(e/m_0)^{1/2}V^{3/2}/L^2$ , which is Child's Law, with only the numerical factor  $4\sqrt{2}/9 = 0.63$  missing. Who else, before Rose, thought to do something so utterly simple?

Table 1—Comparison of Rose's Formulation of One-Carrier SCL Current Theory with the Conventional Formulation

Rose	Conventional ('Exact')
$J = \frac{Q_f}{T_0} = e\mu n \frac{V}{L} \left( T_0 = \frac{L^2}{\mu V} \right) \quad [1]$	$J = e\mu n(x) \varepsilon(x) \quad [5]$
$Q_{tot} = CV, C = \frac{\varepsilon}{L}, Q_{tot} = Q_f + Q_t \quad [2]$	$V = \int_0^L \varepsilon(x) dx. \quad [6]$
<p><math>Q_t</math> is determined by the quasi-Fermi level <math>F</math>, which is defined by <math>Q_f</math> (i.e., by <math>n</math>)</p> 	$\frac{\varepsilon d\varepsilon}{e dx} = n(x) - n_0 + \sum_j (n_{tj}(x) - n_{tj,0}) \quad [7]$
$Q_t = eL \int_{F_0}^F \pi(E_t) dE \quad [3]$ <p>The cathode is an injecting contact [4]</p>	$n(x) = N_c \exp \frac{F(x) - E_c(x)}{kT} \quad [8]$
	$n_{tj}(x) = \frac{N_{tj}}{1 + \frac{1}{g_j} \frac{N_j}{n(x)}} \quad [9]$
	$N_j = N_c \exp \frac{E_{tj} - E_c}{kT}$
	$\varepsilon(0) = 0 \text{ (cathode at } x = 0) \quad [10]$

Notation for Table I

$J$ = current density	$n_0$ = free-electron concentration at thermal equilibrium
$Q_f$ = total free charge, per unit area, between cathode and anode	$n_{tj}(x)$ = trapped-electron concentration (position-dependent) in traps of type $j$
$Q_t$ = total trapped charge, per unit area, between cathode and anode	$n_{tj,0}$ = trapped-electron concentration, at thermal equilibrium, in traps of type $j$
$Q_{tot}$ = total (free plus trapped) charge, per unit area, between cathode and anode	$N_c$ = density of states in the conduction band
$C$ = capacitance per unit area of diode structure	$F(x)$ = quasi-Fermi level (position-dependent)
$V$ = applied voltage	$E_c(x)$ = bottom energy level (position-dependent) of the conduction band
$L$ = cathode-anode separation	$E_t$ = trap energy level
$T_0$ = transit time of injected carriers	$E_{tj}$ = trap energy level for trap of type $j$
$\epsilon$ = dielectric constant of insulator	$k$ = Boltzmann's constant
$e$ = charge on the electron	$T$ = temperature (degrees Kelvin)
$\mu$ = electron mobility	$\mathfrak{N}(E_t)$ = concentration of traps, per unit energy interval, at the energy $E_t$
$x$ = position variable (cathode is at $x = 0$ , anode at $x = L$ )	$N_{tj}$ = concentration of traps of type $j$
$n(x)$ = free electron concentration (position-dependent)	$g_j$ = statistical weight (degeneracy factor) of the traps of type $j$
$n$ = free-electron concentration (space-averaged: $n = n(V)$ )	
$\mathcal{E}(x)$ = electric field intensity (position-dependent)	

Although Rose's theory is complete, in that it contains implicitly the entire content of one-carrier SCL current theory, Rose missed bringing to light one of the more dramatic effects observable with one-carrier SCL injection, and that is the trap-filled-limit (TFL) phenomenon, observed when the quasi-Fermi level passes through a discrete set of traps, or a set of traps more or less localized in energy in the forbidden gap. The reason that Rose missed the TFL phenomenon was simply that all of the trap distributions he studied had one property in common, namely, the trap concentrations increased monotonically<sup>11</sup> from the Fermi level all the way to the band edge,<sup>12</sup> as in the figure in Table 1, which is Fig. 3 of Rose's 1955 paper.<sup>5</sup> There was no energy cutoff in the trap distribution, hence the quasi-Fermi level could not "escape" from the distribution.

How I discovered—"tripped over" is a more precise description—the TFL phenomenon provides a model illustration of how circuitous the road to simplicity can be. I had noticed that Rose had not looked into the matter of an energy cutoff in the trap distribution, so, in late 1955 or thereabouts, I decided to take an analytic look at a relatively

simple example of such a situation, namely, a uniform distribution of traps cut off at an energy  $E_U$  below the conduction-band edge  $E_c$  as illustrated in Fig. 1. So long as the quasi-Fermi level  $F(V)$  was below  $E_U$ , I found the expected exponential dependence of  $J$  on  $V$ , as predicted by Rose.<sup>5</sup> However, the same mathematics showed that, as soon as  $F(V)$  emerged from the distribution, the current rose even more steeply with voltage! My first reaction was that this "bizarre" result probably came from some gross error in the mathematics. So, I

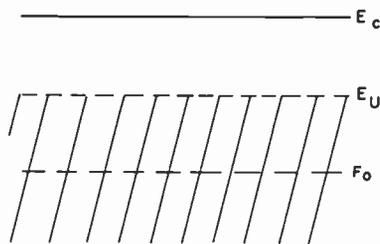


Fig. 1—Schematic energy-band diagram for a uniform trap distribution (area in the energy gap covered by diagonal lines) with an upper energy cutoff  $E_U$  below the conduction-band edge  $E_c$ .

re-did, and re-re-did, the calculation, but the effect wouldn't go away. Then I came to suspect that, perhaps, there was a subtle error associated with one or another of the approximations needed to complete the calculation. I couldn't find such an error and finally decided that I would try to get out of the impasse by finding a simpler trap distribution, which I could solve exactly. This led me to look at the single, discrete trap level. I was delighted to find that this latter problem did indeed have an exact, analytical solution. It was tedious, but it could be written down and studied. The exact solutions for this case exhibited the same effect I had found earlier for the uniform set of traps with a cutoff, namely, the TFL phenomenon. Results of my calculations were presented in Fig. 5 of Ref. (10), reproduced here as Fig. 2. In this same paper are a number of useful, mathematical simplifications which shed a kind of mathematical light on the TFL phenomenon, but no simple, direct physical explanation was put forth.<sup>13</sup>

In 1961 I was asked to write a review paper on Injection Currents in Insulators for the *Proceedings of the Institute of Radio Engineers*. In thinking about the paper before actually writing it, I decided that it was high time that I came up with a simple, convincing, physical explanation of the TFL phenomenon. The argument which I finally settled on is reproduced here from that article.<sup>14</sup> (In order to obtain

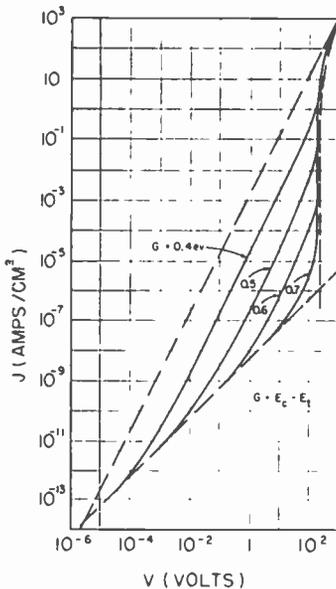


Fig. 2—Exact solutions for the simplified theory of space-charge-limited currents in a CdS crystal with traps at a single discrete energy level. The dashed curves are the three limiting  $J$ - $V$  characteristics. For CdS,  $\epsilon/\epsilon_0 = 11$  and  $\mu = 200$  cm<sup>2</sup>/volt sec at  $T = 300^\circ\text{K}$ . The solid curves are the calculated solutions. For all curves,  $N_t = 10^{14}$  cm<sup>-3</sup>,  $L = 5 \times 10^{-3}$  cm,  $E_c - F_0 = 0.75$  eV, corresponding to  $n_0 \approx 10^6$  cm<sup>-3</sup>, and resistivity  $\rho_0 \approx 3 \times 10^{10}$  ohm cm at  $T = 300^\circ\text{K}$ .

compatibility with the present article, there have been minor changes in notation and in the numbering of equations.)

“The current–voltage characteristic is profoundly affected by the crossing of a trap level by the quasi-Fermi level  $F$ , that is, by the filling of a set of traps. In order to see this, consider, for example, the situation where there is a single set of traps, concentration  $N_t$ , at level  $E_t$ . Further, assume that the thermodynamic Fermi level  $F_0$  lies just below  $E_t$ :  $E_t - F_0 < kT$ . The onset of injection occurs, as always, when the applied voltage is sufficient to double the density of free carriers, that is, when  $n \approx 2n_0$ . From Eq. [8], this corresponds to upward motion of  $F$  by about  $kT$ . Thus,  $F$  has moved through the trap level  $E_t$ , and the traps are now full. The corresponding voltage  $V_{TFI}$ , determined, as usual, from  $Q_{TFI} = eN_tL = CV_{TFI}$ , is given by:  
*trap-filled-limit voltage:*

$$V_{TFI} \approx \frac{eN_tL^2}{\epsilon} \quad [11]$$

In the present case,  $V_{TFL}$  coincides with the voltage at which the transition from Ohm's law to SCL current flow takes place. If, now, the applied voltage is doubled,  $V = 2V_{TFL}$ , the injected charge must likewise be doubled,  $Q(2V_{TFL}) = 2Q_{TFL}$ . Since the traps were already filled at the voltage  $V_{TFL}$ , the additional injected charge,  $Q(2V_{TFL}) - Q(V_{TFL}) = Q(V_{TFL}) \approx eN_tL$ , must all appear in the conduction band. We have, then, for the ratio of the currents at the two voltages:

$$\frac{J(2V_{TFL})}{J(V_{TFL})} \approx \frac{2n(2V_{TFL})}{n(V_{TFL})} \approx \frac{N_t}{n_0}, \quad [12]$$

where  $n(V)$  denotes the total free-electron density, injected plus thermal, at voltage  $V$ . Because  $E_t - F_0 < kT$ ,  $n(V_{TFL}) \approx 2n_0$ . The additional factor of two appears in the numerator of the middle expression in Eq. [12] because the applied field is doubled when the voltage is doubled. The ratio  $N_t/n_0$  may easily be many powers of ten in insulators. We see then that a factor of two change in voltage, following the filling of a set of traps, may produce an enormous change in current."

I don't know whether that review article was useful to anybody else; it was very useful to myself in that it provoked me finally to get an adequate handle on the TFL phenomenon. I would like to believe that the above argument is just the one that Rose himself would have come up with directly, in the early 1950's, had he addressed himself to the particular problem.<sup>15a,b</sup>

This brings me, now, to problems of double injection into solids. In early 1959, while Bob Parmenter was at RCA Laboratories in Zurich, he and Wolfgang Ruppel succeeded in getting an analytical handle on, if not a complete solution to, a very difficult mathematical problem—that of double injection into a perfect insulator, with bimolecular recombination of the injected electrons and holes. As they showed,<sup>16</sup> there is a square-law dependence of current on voltage,

$$J = \frac{3\sqrt{2\pi}}{4} \epsilon \left\{ \frac{\mu_n \mu_p (\mu_n + \mu_p)}{\mu_0} \right\}^{1/2} \frac{V^2}{L^3} \quad [13]$$

with  $\mu_0 = \epsilon\alpha/2e$  and  $\alpha = v_T s$ , with  $v_T$  the free-carrier thermal (relative) velocity and  $s$  the bimolecular recombination cross-section. Further, Eq. [13] is valid only under the conditions  $(\mu_n/\mu_0)^{1/2} \gg 1$  and  $(\mu_p/\mu_0)^{1/2} \gg 1$ , which correspond to the injected-plasma case. In the general case, without any restrictions on  $\mu_n$ ,  $\mu_p$ , and  $\mu_0$ ,  $J \propto V^2/L^3$  still holds, but the coefficient multiplying  $V^2/L^3$  is a more complicated function of the three mobilities.<sup>16</sup> Unfortunately, the Parmenter-

Ruppel analysis, though a very impressive tour de force of mathematical manipulation, did not yield physical insight into the problem. For example, it was unable to exhibit the spatial distribution of the injected carriers and electric field intensity.

Spurred on by the mathematical success of Parmenter and Ruppel, the first published solution of a volume-controlled (drift-dominated) double-injection problem, I worked up the courage, in mid-1959, to look at the problem of double injection into an "almost-perfect" insulator, with a constant lifetime for the injected carriers, independent of injection level (say, due to a small concentration of recombination centers in the insulator). I did not find a mathematical solution to the problem (that came later), but reasoned that, at sufficiently high voltage, namely, high enough that both transit-time-to-lifetime ratios ( $t/\tau$ ) would be small compared to unity:  $t_n/\tau \ll 1$  and  $t_p/\tau \ll 1$ , a plasma would be injected into the insulator, with little attrition of the injected carriers through recombination as they transited the insulator. In effect, I guessed that the injected free carriers would have the spatial dependence exhibited in Fig. 2 of Ref. (17), reproduced here as Fig. 3, where the electron-injecting contact is at  $x = 0$  and the hole-injecting contact at  $x = L$ . Algebraically, the guess is equivalent to taking

$$\begin{aligned} n(x) &\approx n_M \exp \left\{ \frac{(x_M - x)}{L} \left( \frac{t_n}{\tau} \right) \right\} \\ &\approx n_M \left\{ 1 + \frac{t_n}{\tau} \frac{(x_M - x)}{L} \right\}, \end{aligned} \quad [14]$$

$$\begin{aligned} p(x) &\approx p_M \exp \left\{ \frac{(x - x_M)}{L} \left( \frac{t_p}{\tau} \right) \right\} \\ &\approx p_M \left\{ 1 - \frac{t_p}{\tau} \frac{(x_M - x)}{L} \right\}, \end{aligned} \quad [15]$$

and, subtracting Eq. [15] from [14],

$$e(n - p) \approx en_M \left( \frac{t_n}{\tau} + \frac{t_p}{\tau} \right) \frac{(x_M - x)}{L}, \quad [16]$$

where we have used the fact that  $n_M = p_M$  at the mid-plane,  $x_M = L/2$ , to ensure overall neutrality of the material. Between  $x = 0$  and the mid-plane there is net negative charge; between the mid-plane and  $x = L$  there is net positive charge. Defining  $|Q_-|$  by  $\int_0^{x_M} e(n - p)dx$  and  $Q_+$  by  $\int_{x_M}^L e(p - n)dx$ , and using Eq. [16], we obtain:

$$|Q_-| = Q_+ \approx \frac{1}{8} \left( \frac{t_n + t_p}{\tau} \right) e n_M L, \quad [17]$$

where, as usual,

$$t_n = \frac{L^2}{\mu_n V} \text{ and } t_p = \frac{L^2}{\mu_p V}. \quad [18]$$

Since both electrons and holes carry current, Eq. [1] gets replaced by

$$J = J_n + J_p = e(\mu_n + \mu_p) n_M \frac{V}{L}. \quad [19]$$

The space-charge distribution [16] satisfies the condition of footnote

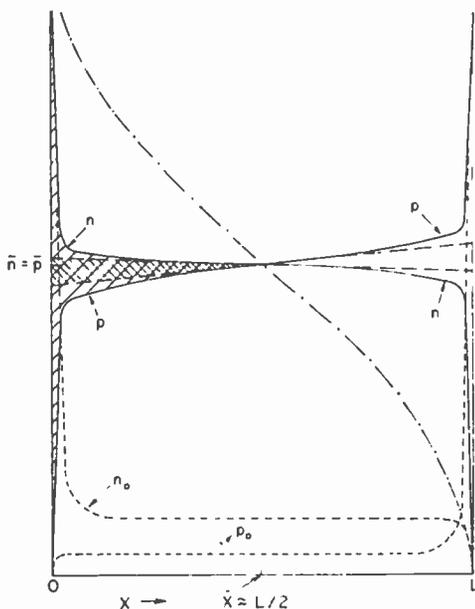


Fig. 3—Schematic representation of free-electron and free-hole density distributions,  $n(x)$  and  $p(x)$  respectively (the solid lines), that were guessed for an insulator with an injected plasma. The short-dashed lines are the schematic free-carrier distributions,  $n_0$  and  $p_0$  respectively, in thermal equilibrium. The electron-injecting contact is at  $x = 0$ , the hole-injecting contact  $x = L$ .

[9], which assures us that the usual capacitance relation, Eq. [2], is valid:

$$|Q_-| = Q_+ \approx \frac{\epsilon}{L} V. \quad [20]$$

Putting together Eqs. [17]–[20] we obtain the final result:

$$J \approx 8\epsilon\tau\mu_n\mu_p \frac{V^3}{L^5}. \quad [21]$$

which appears as Eq. [25] in Reference (17).

Although the final result, Eq. [21], was certainly interesting, it was, after all, based only on a guess (Eqs. [14] and [15]); furthermore, in 1959 there were no experimental data bearing on the problem. I don't think I would have gone ahead and published it if it were not for the already available, analytical result of Parmenter and Ruppel, Eq. [13]. Although my analysis was based on the assumption of a constant lifetime  $\tau$ , independent of the free-carrier densities, whereas the Parmenter–Ruppel analysis was based on a bimolecular-recombination process with electron and hole lifetimes given, respectively, by  $\tau_n = 1/\alpha p$  and  $\tau_p = 1/\alpha n$ , I was able to connect the two analyses by working with the space-averaged plasma density  $n_M(V)$ , along the line of Rose's original treatment of one-carrier SCL currents. The end result was that Eq. [21] yielded Eq. [13], except that the numerical constant  $3\sqrt{2}\pi/4$  in Eq. [13] was replaced by 2, a difference of only 6%. The details are given in Ref. (17). I concluded that Eq. [21] had to be right,<sup>18</sup> that sheer coincidence could not reduce an incorrect Eq. [21] to the correct Eq. [13], and so I went ahead and published.

The extraordinary thing is that the guess on which the entire edifice was built was actually a downright poor one. When a completely analytical solution was available,<sup>18</sup> the correct spatial distributions of  $n(x)$  and  $p(x)$  came out looking like those exhibited in Fig. 4, which is a reproduction of Fig. 5 of Ref. (14). However—marvelous to behold—although  $n(x)$  and  $p(x)$  are *separately* poor guesses, it comes out that their difference  $n(x) - p(x)$  is *very accurately* given by Eq. [16]! The reader can readily check that only their difference enters the derivation of Eq. [21]. The moral of this episode, I suppose, is: If brains don't carry you through, there's always the hope for a spot of timely luck.

Not surprisingly, some years earlier Rose had already discovered the main result Eq. [21] for the special case of equal mobilities:  $\mu_n = \mu_p = \mu$ . His derivation, literally on one line, ran:

$$J \simeq e\mu n \frac{V}{L}, Q \simeq \frac{t}{\tau} enL \simeq \frac{\epsilon}{L}, t = \frac{L^2}{\mu V} \longrightarrow J \simeq \epsilon\tau\mu^2 \frac{V^3}{L^5}. \quad [22]$$

This derivation was given in a letter to me that Rose wrote from Zurich in August, 1955.<sup>19</sup> I hastily scanned the letter in August, went on vacation, read it again, more carefully, in December of 1955 (after

Rose inquired as to why I hadn't answered it), filed its contents in my mind, and promptly forgot them. Not quite, or rather, not entirely. Although a year later I could certainly not remember anything said in the letter, it's entirely reasonable to assume that its contents were drifting around, unnoticed and undisturbed, in that huge memory bank known as the unconscious mind. In 1959, provoked into engagement with the problem by the Parmenter-Ruppel solution, I dredged up Rose's argument from my murky unconscious and re-fashioned it

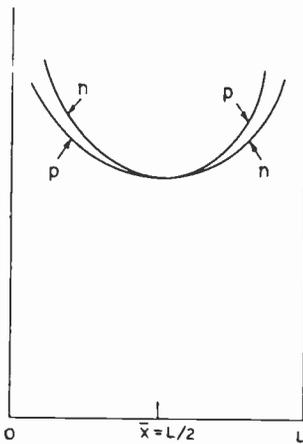


Fig. 4—Schematic representation of the injected free-carrier distributions,  $n(x)$  and  $p(x)$ , for a plasma injected into an insulator, as given by the analytic solution. The contacts are as for Fig. 3.

to my own uses. (I acknowledged Rose's unpublished priority in the matter.) Like it or not, that is how we are constructed. It is a ubiquitous, omnipresent feature of the human condition that, barring singular instances (one is cited in footnote 24), we remember a lot more clearly what we said (or wrote) five or ten years ago than what was said (or written) to us at the same time.

In the summer of 1960, Rose decided to tackle the problem of injection of a plasma into a semiconductor, a problem that had not yet been considered, or at least not solved, for field-driven injection currents. In his usual manner, he was looking for a solution through pure physical insight. However, on this occasion the solution kept eluding him through a long week of struggle. At this point, I joined forces with him and, although our conjoint efforts brought no deeper insights, we quickly found the solution by an alternative, mathematical route.<sup>18</sup> For an n-type semiconductor the  $J$ - $V$  characteristic is<sup>20</sup>

$$J \simeq e(n_0 - p_0)\tau\mu_n\mu_p \frac{V^2}{L^3}. \quad [23]$$

Beyond obtaining Eq. [23], we also managed to come up with a compact, unified theory for the injection of plasmas into solids embracing both the insulator case and the semiconductor case, the latter under the two extremes of field control and diffusion control (classical transistor problem). Within about one week after showing our solution Eq. [13] to R. D. Larrabee he had experimentally verified it.<sup>21</sup> I mention this because it quite possibly belongs in Guinness' book of records.

It is useful, for our later discussion, to sketch briefly the unified theory for injected plasmas. Five equations define the theory: the two current-flow equations:

$$J_p = ep\mu_p\varepsilon - eD_p \frac{dp}{dx}, D_p = V_T\mu_p, V_T = \frac{kT}{e}. \quad [24a]$$

$$J_n = en\mu_n\varepsilon + eD_n \frac{dn}{dx}, D_n = V_T\mu_n. \quad [24b]$$

the Poisson Eq.:

$$\frac{\epsilon}{e} \frac{d\varepsilon}{dx} = (p - p_0) + (n - n_0). \quad [25]$$

and the two particle-conservation equations:

$$\begin{aligned} \frac{1}{e} \frac{dJ_n}{dx} &= \mu_n \frac{d}{dx}(n\varepsilon) + V_T\mu_n \frac{d^2n}{dx^2} = \frac{n - n_0}{\tau} \\ &= \frac{p - p_0}{\tau}, \end{aligned} \quad [26a]$$

$$\begin{aligned} -\frac{1}{e} \frac{dJ_p}{dx} &= -\mu_p \frac{d}{dx}(p\varepsilon) + V_T\mu_p \frac{d^2p}{dx^2} = \frac{n - n_0}{\tau} \\ &= \frac{p - p_0}{\tau}. \end{aligned} \quad [26b]$$

Multiplying Eq. [26b] by  $b = \mu_n/\mu_p$ , and adding to [26a], gives, after using Eq. [25]:

$$-\frac{\epsilon}{e} \frac{d}{dx} \left( \varepsilon \frac{d\varepsilon}{dx} \right) + (n_0 - p_0) \frac{d\varepsilon}{dx} + 2V_T \frac{d^2p}{dx^2} = \frac{(b+1)(p-p_0)}{\mu_n\tau}. \quad [27]$$

(I)                      (S)                      (D)                      (R)

The left-hand side (LHS) of Eq. [27] has three terms, labeled (I), (S),

and (D). If only the first term (I) is retained, we end up describing the insulator problem and obtaining Eq. [21]. If only the second term (S) is retained, we end up describing the "long" semiconductor problem (field-controlled current flow) and obtaining Eq. [23]. Finally, if only the third term is retained we end up describing the "short" semiconductor problem (diffusion-controlled current flow) with its classical diffusion solution.

Although Rose and I were not dissatisfied with the mathematical simplifications yielded by the unified theory, we were not happy over our inability to find a purely physical handle on the problem.<sup>13</sup> Happy, or not, we couldn't find it and had to publish without it.<sup>22</sup> Not long afterwards I was asked to give a seminar at Brooklyn Polytechnic Institute on the injected-plasma problem. Again, I looked for the physical argument, and again I couldn't find it. As a hedge against embarrassing questions, I cooked up a tortured explanation for Eq. [23], something that might conceivably convince an audience with other pressing problems on their minds. It was my luck that day that, sure enough, a gentleman in the audience asked me how one understood Eq. [23] physically. I went through the tortured explanation, whereupon he announced that it wasn't very convincing, whereupon I congratulated him on his perspicacity.

The occasion for finally finding the magical physical argument was once again a review article, namely one I agreed, in 1963, to write for the British annual, "Reports on Progress in Physics."<sup>23</sup> In preparing that article, I thought it judicious to compare the "master" equation for the unified theory, Eq. [27], with the corresponding equation that was the conventional basis for current-flow analysis in transistor problems,<sup>24</sup> namely,

$$-\mu_a \varepsilon \frac{dp}{dx} + D_a \frac{d^2 p}{dx^2} = \frac{p - p_0}{\tau}, \quad [28]$$

with  $\mu_a$  and  $D_a$  the ambipolar mobility and ambipolar diffusion constant respectively:

$$\mu_a = \frac{n - p}{\frac{p}{\mu_n} + \frac{n}{\mu_p}}, \quad D_a = \frac{n + p}{\frac{p}{D_n} + \frac{n}{D_p}}. \quad [29]$$

Eq. [28] is obtained from Eqs. [24]–[26] by straightforward algebraic manipulations found in any standard text on semiconductor electronics.

If diffusion is unimportant, Eq. [28] simplifies to:

$$-\mu_a \mathcal{E} \frac{dp}{dx} = \frac{p - p_0}{\tau} \quad [30]$$

If dimensional analysis is performed on Eq. [30], namely, if  $\mathcal{E}$  is replaced by  $V/L$ , and  $x$  by  $L$ , then Eq. [30] yields  $\mu_a V/L^2 = 1/\tau$ , or:

$$t_a = \tau, \quad t_a = \frac{L^2}{\mu_a V} \quad [31]$$

where  $t_a$  is obviously the ambipolar transit time.

Eq. [31] is the long-sought-after physical argument. It determines the level of the injected plasma.<sup>25</sup> The plasma level simply self-adjusts so as to make Eq. [31] true. The reason it does this is to allow the electric field to build up to the point that it exactly "uses up" the applied voltage, i.e., such that  $\int_0^L \mathcal{E} dx = V$ . A smaller plasma level will give too small a value for  $\int_0^L \mathcal{E} dx$ , and a higher plasma level will give too high a value for this integral. My presumption is that a dyed-in-the-wool plasma physicist would write down Eq. [31] out of hand,<sup>26</sup> the rationale being that, in dealing with plasma problems, the ambipolar mobility is the relevant mobility. In any case, Eq. [23] comes easily out of Eq. [31] if  $(n - p)$  in the numerator of  $\mu_a$  (Eq. [29]) is replaced by  $n_0 - p_0$ . And Eq. [21], except for the factor of 8, comes easily out of Eq. [31] if the same  $(n - p)$  is replaced by  $Q/eL$ . (If one, more fastidiously, identifies  $n - p$  with Eq. [16], one also recovers the factor 8.)

One final chapter in our current-injection saga merits attention here, and that is an approximation scheme for handling otherwise unsolvable problems which has come to be known as the Regional Approximation Method. In this scheme I carried Rose's philosophy ("Get the physics straight—don't worry about a factor of two") to its logical conclusion, or, as others might say, to the bitter end. The idea of the method is conveniently illustrated by a schematic simple example. Suppose a situation we are studying thrusts upon us the unsolvable equation:

$$M_1(x) + M_2(x) = N(x) \quad [32]$$

where  $x$  is the independent variable (typically  $x$  might be a position variable, or an energy variable) and  $M_1(x)$ ,  $M_2(x)$ , and  $N(x)$  are all positive functions, including possibly differential or integral operator functions.<sup>27</sup> The domain of  $x$ ,  $0 \leq x \leq L$ , is now broken up into non-overlapping regions such that in any one region one of  $M_1(x)$ ,  $M_2(x)$  is everywhere greater than the other; in that region we retain only the dominant member of the two. For example,  $M_2(x)$  may be dominant near the contacts, at  $x = 0$  and  $x = L$  respectively. Thus,  $M_2(x) >$

$M_1(x)$  for  $0 \leq x < x_a$  and  $x_b < x \leq L$ , whereas  $M_1(x) > M_2(x)$  for  $x_a < x < x_b$ . At the common boundary of two adjacent regions, continuity requires that the two functions be equal. Instead of trying to cope with the complicated, unsolvable Eq. [32], we work with the simpler, solvable equations:

$$\text{Region I. } 0 \leq x < x_a: \quad M_2(x) = N(x), \quad [33a]$$

$$\text{Region II. } x_a < x < x_b: \quad M_1(x) = N(x), \quad [34]$$

$$\text{Region III. } x_b < x \leq L: \quad M_2(x) = N(x), \quad [33b]$$

The boundary points  $x_a$ ,  $x_b$  are determined by the cross-over conditions:

$$M_1(x_a) = M_2(x_a), \quad M_1(x_b) = M_2(x_b). \quad [35]$$

A single-region, unsolvable problem has, in this case, been replaced by a three-region solvable problem. Note that, although there are three regions, there are only two differential equations, Eqs. [33a] and [33b] being the same equation. Further, as compared to the complicated Eq. [32], the relatively simple Eqs. [33] and [34] yield far greater insight into the physics underlying the solution.

A scheme exactly like the prototype Eq. [33], Eq. [34] applies to the analysis of injected plasmas when the contacts are put into the problem realistically. For the insulator problem ( $n_o$ ,  $p_o$  negligible) Eq. [27] becomes, symbolically

$$(I) + (D) = (R), \quad [36]$$

The contact that promotes the injection of electrons presents a barrier to the exit of holes. Consequently, there is a pile-up of holes near this contact, neutrality being preserved by the freely available electrons. A corresponding (dual) situation obtains near the hole-injecting contact. As a result of the pile-up of plasma at the injecting contacts there are low electric fields and large concentration gradients in the neighborhood of the contacts. This situation favors the diffusion term (D), as compared to the drift term (I), near the contacts. On the other hand, in the interior of the insulator, well away from the contacts,<sup>28</sup> the field is relatively high and the concentration gradients relatively small, a situation favoring (I) over (D). Thus, identifying (I) with  $M_1(x)$ , (D) with  $M_2(x)$ , and (R) with  $N(x)$  we see that the insulator injected-plasma problem can be conveniently studied as a three-region problem. The semiconductor injected-plasma problem is han-

dled in precisely the same manner, only with the neutrality-related drift term (S) replacing the space-charge-related drift term (I) in the above argument. A comprehensive review of applications of the Regional Approximation Method to the solution of diverse problems of injection, including negative-resistance problems, was published in 1970.<sup>29</sup> Recently, the method has been applied to the study of an extracted-plasma problem namely, that of the bleaching of  $\text{WO}_3$  electrochromic films by an electric field.<sup>30</sup> As with any method of approximation, the Regional Approximation Method runs into its own peculiar brand of difficulties.<sup>31</sup>

Although this article has, throughout, dealt with the area of current injection in solids, I should like to add one final example of simplicity, one of Rose's best, as well as most recent. This example deals with a quite different class of problems, namely, amplification and instability phenomena in solids deriving from distributed interactions. Examples are the acoustoelectric effect in piezoelectric materials and plasma instabilities in semiconductors. The following remarks are, in essence, excerpted from a book review written by myself a few years ago.<sup>32</sup> I have made some slight changes to up-date the remarks.

"I have one complaint, which is actually directed to the instability literature in general, and that is the highly mathematical nature of the usual discussion of instabilities. In some vague, intuitive sense one feels that instability phenomena are intensely "physical" in nature and origin; yet, comes the chapter on instabilities, and normally one immediately finds oneself deposited onto the complex plane. Now, although the ensuing procedure is known to be very powerful and very useful, I think that most physicists would acknowledge that, somewhere along the way, a lot of physics gets lost. Yet, the situation is not totally hopeless in regard to this problem of acquiring some insight into what one is doing. Consider the following, extraordinarily simple conceptual scheme developed by Albert Rose.<sup>33</sup>

"A system A, moving at *constant* velocity  $v_A$ , interacts through a mutual frictional force  $F$  with a second system B, which may consequently be at some particular velocity  $v_B$ . By Newton's law, an external agency must apply the force  $F$  to system A to maintain the constant velocity of A. This agency expends the power  $P_A = Fv_A$ . The power  $P_B$  expended on system B, at the given instant of time, is  $P_B = Fv_B$ . Energy is conserved by power dissipation  $P_D$  into frictional (zero-net-momentum) processes:  $P_D = F(v_A - v_B)$ . In an actual problem, it is straightforward to calculate *directly* the dissipation  $P_D$ .  $F$  is then eliminated to yield Rose's relation:  $dE_B/dt = P_B = P_D v_B / (v_A - v_B)$ , where  $dE_B/dt$  is the rate of change of the energy in system B. A crucial property of the above argument is that it is independent of the relative magnitude of  $v_A$  and  $v_B$ . For  $v_A < v_B$  the energy content of system B decays in time; for

$v_A > v_B$  it grows in time. (The apparent singularity at  $v_A = v_B$  is a purely mathematical artifact. In a real problem, at synchronism there is either no interaction at all or it must simply be handled more circumspectly. In the latter case the difficulty relates to a frequency that changes, by Doppler shift, down to zero frequency, in which case averaging over an ac cycle is no longer a possible operation.) Now, let us apply the above scheme, which might appropriately be labeled Rose's "*reductio ad physicum*" scheme, to the simplest conceivable problem of a plasma-wave interaction in a solid. A very unsophisticated, 19th-century type, purely transverse, uniform plane electromagnetic wave propagates in an infinite dielectric medium ( $\epsilon/\epsilon_0 \gg 1$ ) containing free electrons. Everybody knows that the wave will be attenuated because the electrons, in responding to the ac electric field of the wave, make random collisions with phonons or impurities or whatever. In the above scheme, let the electromagnetic wave be the system B and the aggregate of free electrons the system A. In the problem as stated thus far,  $v_A = 0$ ,  $v_B = c(\epsilon_0/\epsilon)^{1/2}$ , and  $dE_B/dt$  is negative, corresponding to attenuation of the wave. Now cause the electrons to drift with a velocity  $v_A > v_B$  (Cerenkov electrons). Rose's relation tells us that  $dE_B/dt$  is now positive—we have a growing wave, the source of energy being, of course, the drifting electrons. For those in doubt, this result is confirmed by a direct, elementary calculation of strictly conventional type.<sup>33</sup> Back around 1966 I used, not-so-innocently, to ask assorted solid-state plasma physicists whether sufficiently fast, drifting electrons could amplify a purely transverse, plane electromagnetic wave traveling in a semiconductor. I never once received the right answer (granted, my acquaintance with solid-state plasma physicists is limited). There was a natural—or, at least, long-standing—prejudice (which I fully subscribed to, prior to Rose's argument) in favor of the need to bunch the electrons, which the purely transverse wave could not do. The Rose relation has far wider application than to the purely academic, prototypical problem just discussed. For example, it leads very quickly, and painlessly, to the well known Hutson-White small-signal formula for the amplification of piezoelectric waves by free carriers drifting faster than the velocity of sound, leaving out the effects of free-carrier diffusion.<sup>33</sup> (It takes just slightly more work to insert the latter effects properly.) I might take the opportunity here to note that the above-cited paper by Rose was the first of a series of five<sup>34</sup> by him in which he also succeeded in bringing very simple physics to bear on the important and diverse class of problems of energy loss by a fast electron in a solid by various mechanisms, a subject invariably treated by more-or-less formal mathematical methods, typically via perturbation theory. This latter work is based on a sweeping extension, to contemporary solid-state problems, of Bohr's famous, early (1913) calculation of the energy loss by fast electrons in their passage through matter."

I should like to conclude this tribute to Al Rose by wishing him well in the days and years ahead, and by saying to him: "Thank you, Al. It

was one of the happiest experiences of my life that I met, and worked with, you."

## References:

- <sup>1</sup> M. A. Lampert and P. Mark, *Current Injection in Solids*, Academic Press, New York, 1970.
  - <sup>2</sup> I must say, we have discussed just about everything else, with a particular fondness for the Larger Questions, such as: "Will the passage from the five-day work-week to the no-day work-week be continuous and without trauma?" and "Is intelligent life consistent with the known laws of physics?" (The answer to the first question is "No;" and to the second, "No, with high probability.")
  - <sup>3</sup> N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals*, Oxford University Press (Clarendon), London, 1940.
  - <sup>4</sup> A. Rose, "An Outline of Some Photoconductive Processes," *RCA Rev.*, **12**, p. 362 (1951); P. K. Weimer and A. D. Cope, "Photoconductivity in Amorphous Selenium," *RCA Rev.*, **12**, p. 314 (1951); S. V. Fergue, R. R. Goodrich, and A. D. Cope, "Properties of Some Photoconductors, Principally Antimony Trisulfide," *RCA Rev.*, **12**, p. 335 (1951); R. W. Smith, "Some Aspects of the Photoconductivity of Cadmium Sulfide," *RCA Rev.*, **12**, p. 350 (1951). These papers are all contained in the same issue of *RCA Review*, namely, an issue devoted to the science and technology of vidicons.
  - <sup>5</sup> A. Rose, "Space-Charge-Limited Currents in Solids," *Phys. Rev.*, **97**, p. 1538 (1955); R. W. Smith, "Properties of Ohmic Contacts to Cadmium Sulfide Single Crystals," *Phys. Rev.*, **97**, p. 1525 (1955); R. W. Smith and A. Rose, "Space-Charge-Limited Currents in Single Crystals of Cadmium Sulfide," *Phys. Rev.*, **97**, p. 1531 (1955).
  - <sup>6</sup> In writing the "exact" theory we have omitted the diffusive contribution to the current, hence the quotation marks around exact. If one includes the missing diffusion current, one has truly an *exact* theory (well, at least "a more exact" theory) and, at the same time, a truly useless theory. The resulting equations are so hopelessly nonlinear that only a single, physically trivial, case is solvable (the perfect insulator problem), and that solution is so opaque that, (1) it has appeared in the published literature on at least eight separate occasions (references on p. 198 of CIS), and (2) it cannot possibly be used to any good purpose. We have here an inspiring analog of the well-known gain-bandwidth trade-off, namely, a "utility-exactitude" trade-off—the greater the exactitude the smaller the utility.
  - <sup>7</sup> Rose has generally used the nomenclature "steady-state Fermi level." I switched to "quasi-Fermi level" when confronted with the problem, in CIS, of using a common notation for the description of both transient and steady-state injection phenomena. A phrase like "the transient behavior of the steady-state Fermi level" sounds like a loser. On the other hand, almost any nomenclature would have to be an improvement over Shockley's "imref," which is not quite Shockley-spelled-backwards, and which certainly earned the instant non-recognition it received.
  - <sup>8</sup> In one case, the uniform distribution of traps, the factor-of-two error appears in an exponent:  $J \propto \exp \gamma V$ , with  $\gamma \propto C$ . Here it is judicious to replace  $C$  by its exact, space-charge-perturbed value  $2\epsilon/L$ , instead of using the purely geometric value,  $\epsilon/L$ . There is a small moral worth noting here: Simplicity is very valuable, and most certainly worth the candle, but it doesn't come completely free. If you give up your equations, you must lean more heavily on your wits. The reader is not promised a rose garden.
  - <sup>9</sup> One of these approximations initially bothered me. Why is the fixed geometric capacitance such a good approximation to a voltage-dependent capacitance? Rose did not involve himself with this question. In an early paper<sup>10</sup> I decided to nail this point down and, indeed, succeeded in proving quite rigorously, and under very general conditions, that  $\epsilon/L \leq C(V) \leq 2\epsilon/L$ . The proof rests on the convexity of  $\delta(x)$ , and is purely geometric in nature. For several years I was quite proud of my proof; I thought it lent a touch of mathematical elegance to the theory. Later it dawned on me that it is a rather trivial point. So long as the free and trapped carrier densities,  $n(x)$  and  $n_t(x)$ , decrease monotonically going from the injecting contact toward the exiting contact (which is a physically obvious situation, in the absence of constraints at the exiting contact) the center of gravity of the injected charge must obviously lie on the cathode side of the mid-plane,  $x_M = L/2$ , which is all that the inequality says. This was obviously obvious to Rose.
- The generalization to double-injection space-charge situations is fairly painless. Let the hole-injecting anode be at  $x = 0$  and the electron-injecting cathode at  $x = L$ . Denote the space-charge density by  $\rho(x)$ . Then  $\rho(x)$  is positive near the anode, i.e., near  $x = 0$ , and negative near the cathode, i.e., near  $x = L$ . Assume only that  $\rho(x)$  decreases monotonically, including the algebraic sign, between the anode and the cathode, with  $\rho(x)$  going through 0 at some neutrality point  $x_N$ :  $\rho(x_N) = 0$ . In the interval  $0 \leq x < x_N$ ,  $\rho(x)$  is positive, and in the interval  $x_N < x \leq L$ ,  $\rho(x)$  is negative. It is now obvious that the center of gravity of the positive charge  $Q_+$  must lie closer to the anode than the "positive mid-plane,"  $x_{M, \text{pos}} = x_N/2$ , and, correspondingly, the center of gravity of the negative charge  $Q_-$  must lie closer to the cathode than the "negative mid-plane,"  $x_{M, \text{neg}} = (L + x_N)/2$ . Thus the separation of  $Q_+$  and  $Q_-$  must exceed  $x_{M, \text{neg}} - x_{M, \text{pos}} = L/2$ , and the desired inequality has been proved.

<sup>10</sup> M. A. Lampert, "Simplified Theory of Space-Charge-Limited Currents in an Insulator with Traps," *Phys. Rev.*, **103**, p. 1648 (1956).

<sup>11</sup> Strictly speaking, the uniform distribution, which Rose did study, is monotonic non-decreasing. No matter; without an energy cutoff, before the levels merge with the band, the TFL phenomenon does not come into play.

<sup>12</sup> The reason Rose confined himself to such trap distributions is that they were entirely appropriate for characterizing the materials he was interested in at that time, namely those suitable for vidicon targets. A good example is amorphous  $Sb_2S_3$ . In this regard, Rose was to become a victim of a well-known syndrome to which unusually creative people are sometimes prone: premature discovery. At the time Rose introduced his trap distributions, i.e., in the early fifties, the attention of almost all solid-state physicists was literally riveted on single-crystal physics, mostly that of Ge. Indeed, in zone-purified, single-crystal Ge, the dominant trap levels tend to be discrete. As a result, Rose's trap distributions were regarded as an anachronism, a throwback to those unsophisticated days before pure, single crystals were available. Strange to behold, after twenty years of single-crystal physics, amorphous materials arrived, meteor-like, on the scene, via a sensational public-relations coup (front page of the *N. Y. Times*) fashioned out of a dream and a sprinkling of prominent names. Following this, the remarkable discovery was made that the electronic properties of amorphous semiconductors can only be understood by invoking traps distributed-in-energy throughout the gap, the concentrations monotonically decreasing with distance (in energy) from the band edge. Nowadays, these are usually called band-tailing states. Rose, unfortunately, had exhibited such poor judgment as to recognize this state of affairs far too early and in a language not later certified for acceptance, two mistakes for which a chap is not readily forgiven.

<sup>13</sup> This was not the last time I was to arrive at mathematical simplification while physical simplicity still eluded me. An equally striking example is furnished by the injected-plasma problem discussed later in the text. Rose has the dubious honor of sharing with me the difficulties in understanding that problem. The relationships between mathematical simplicity and physical simplicity would certainly make for an interesting essay. On the one hand, they are closely related, and on the other, they are most assuredly not always the same thing. I'll be happy to let some interested reader take it on from there!

<sup>14</sup> M. A. Lampert, "Injection Currents in Insulators," *Proc. IRE*, **50**, p. 1781 (Aug. 1962).

<sup>15</sup> (a) An interesting way of looking at the TFL phenomenon is to regard it as a consequence of Fermi-Dirac statistics, namely, Eq. [9]. As everybody knows, the Fermi occupation function makes a rather sharp transition from approximately unity (the "filled" Fermi sea) to approximately zero (the Boltzmann tail) in an energy interval of width  $\approx kT$  about the Fermi level. It might be said that the nearly vertical transition in the I-V characteristic, Fig. 2, is simply mirroring this sharp transition in the Fermi function. This is not incorrect, but I personally do not favor this set of words as an explanation for the TFL phenomenon for two reasons: (1) it is a bit too mathematical and (2) it carries the suggestion that the TFL phenomenon is peculiarly quantum mechanical, which I feel is not true.

The latter point is worth some discussion. I think that most physicists identify Fermi-Dirac statistics with fermions. This identification is not always correct. Fermi-Dirac statistics is an outcome, mathematically, of the constraint that only a single particle can occupy any one state. In the familiar situations (e.g., electrons in atoms, or in bands in solids) this constraint is indeed a consequence of the Pauli exclusion principle, and therefore a property of fermions. However, in the case of localized trapping states in the forbidden gap of insulating crystals, this constraint may be the result simply of electrostatic repulsion, a purely classical phenomenon (although the existence of the state and, indeed, of stable atoms in general, is a distinctive consequence of quantum mechanics). In this case, the Fermi function will describe the occupancy of the trapping state by the charged carrier whether the latter is a fermion or a boson. Since charged bosons resembling electrons are hard to come by (excitons are neutral bosons; Cooper pairs are doubly charged bosons, but, up to the present, their existence is confined to highly conducting materials, which do not exhibit the phenomena discussed in this article), the above assertion may seem strange to the reader. In order to convince a skeptical reader, we need only note that protons can be injected into ice as excess carriers and trapped in interstices in the ice lattice. The resulting I-V characteristic is a quite typical, one-carrier SCL characteristic (see H. Engelhardt and N. Riehl, *Phys. Letters*, **14**, p. 20 (1965) or CIS, pp. 154-155), and so is indeed a consequence of Fermi-Dirac statistics. But, note that, although the experiment appears not yet to have been done, there is no doubt whatever that injected deuterons, which are bosons, would give pretty much the same result, making allowance for the higher mass and concomitantly lower mobility. Deuteron substitution for protons is a standard labeling technique, through mass difference, in chemistry and biochemistry.

(b) We can see immediately, from the physical argument, why Mott and Gurney "are mostly, but not entirely, right" about their prediction for the effect of traps at low temperatures. At voltages below  $V_{TF_L}$  the quasi-Fermi level is pinned near the trap level  $E_t$  and, from Eq. [8], negligible current will flow in the solid. However, between  $V_{TF_L}$  and  $2V_{TF_L}$  there will be the usual near-vertical rise of current. At  $V = 2V_{TF_L}$ ,  $J \approx e\mu N_t V_{TF_L}/L$  independent of temperature (except for the temperature dependence of  $\mu$ ).

<sup>16</sup> R. H. Parmenter and W. Ruppel, "Two-Carrier Space-Charge-Limited Current in a Trap-Free Insulator," *J. Appl. Phys.*, **30**, p. 1548 (1959).

<sup>17</sup> M. A. Lampert, "A Simplified Theory of Two-Carrier, Space-Charge-Limited Current Flow in Solids," *RCA Rev.*, **20**, p. 682 (1959).

<sup>18</sup> M. A. Lampert and A. Rose, "Volume-Controlled, Two-Carrier Currents in Solids: The Injected Plasma Case," *Phys. Rev.*, **121**, p. 26 (1961). In this paper the insulator cube law Eq. [21] is derived analytically, without the need to make any guesses, and only the numerical factor 8 is changed—the current numerical factor is  $125/18 \approx 7$ .

<sup>19</sup> Rose made no attempt to publish, so far as I know. Perhaps he was deterred by the novel logistics of publishing a one-line derivation. Also, it is really only one line because of the equal-mobilities assumption. If  $\mu_n \neq \mu_p$  one must deal with Eq. [17], which requires at least another line or two of explanation!

<sup>20</sup> It appears from Eq. [23] that, for a truly intrinsic sample:  $p_0 = n_0$ ,  $J = 0$ —a result that defies common sense, and that is also incorrect. What happens in this case is that there is simply no square-law regime in the overall  $J$ - $V$  characteristic. At a critical voltage there is a transition from the low-level Ohm's-law regime to the insulator cube-law regime. This is discussed in Ref. (18) and in CIS.

<sup>21</sup> R. D. Larrabee, "Current Voltage Characteristics of Forward Biased Long p-i-n Structures," *Phys. Rev.*, **121**, p. 37 (1961).

<sup>22</sup> Section III of Ref. (18) is titled "Physical Arguments." Probably, nothing in that Section is wrong—it's just that the center-piece is missing.

<sup>23</sup> M. A. Lampert, *Reports on Progress in Physics*, **27**, p. 329 (1964).

<sup>24</sup> One of the referee's criticisms of the 1961 paper by Rose and myself<sup>18</sup> was that we had based the entire paper on the unfamiliar Eq. [27] instead of the familiar Eq. [28], and how come? We replied that we were unfamiliar with any legislative act that mandated that we work with Eq. [28], and that we had found it very useful to work with Eq. [27] instead. I found the referee's way of stating his point so peculiar that I never forgot it. This probably played a role in my decision to compare the two approaches when I was brooding over the content of Ref. (23).

<sup>25</sup> Equations looking like Eq. [31] are quite familiar in injection theory. However, in the usual case, such an equation characterizes a point of transition from one regime in a  $J$ - $V$  characteristic to the next regime. For example, for a plasma injected into an n-type semiconductor, the relation  $t_p = \tau$  ( $t_p$  the hole transit time and  $\tau$  the lifetime) determines the voltage of the transition from Ohm's law to the injected-plasma square law Eq. [23], and the relation  $t_p = t_{\Omega} = \epsilon / en_0 \mu_n$  (the ohmic relaxation time of the majority-carrier electrons) determines the voltage of transition from Eq. [23] to the insulator cube law Eq. [21]. The unusual thing about Eq. [31] is that it determines an entire regime in the  $J$ - $V$  characteristic, and this is so because  $\mu_a$  is not a physical constant characterizing the material;  $\mu_a$  depends explicitly on the free-carrier densities  $n$  and  $p$ .

<sup>26</sup> From this, the reader might properly conclude that both Rose and myself must have been dyed in some other fabric than wool—not necessarily the same one. The reader can go even further, and conclude that physical arguments are all well and good, provided one has a keen *prior* appreciation of the underlying physics.

<sup>27</sup> We require that  $M_1(x)$  and  $M_2(x)$  both be positive in order to avoid having to deal with a small difference between two large quantities. In the latter situation you lose touch with reality very quickly neglecting the smaller quantity.

<sup>28</sup> "Well away from," in practice, means "a few ambipolar lengths removed from."

<sup>29</sup> M. A. Lampert and R. B. Schilling in *Semiconductors and Semimetals*, Vol. 6, edited by R. K. Willardson and A. Beer, Academic Press, New York, 1970.

<sup>30</sup> B. W. Faughnan, R. S. Crandall, and M. A. Lampert, to be published. The bleaching is associated with the extraction of an electron-proton plasma from the film.

<sup>31</sup> Since nothing comes for free, not even the Regional Approximation Method, we should mention two annoying features of the Method, one real and the other somewhat imaginary, i.e., psychological. A real annoyance is the algebraic complexity, which derives, ultimately, from the matching conditions Eq. [35]. These conditions inevitably involve messy algebraic and/or transcendental expressions. Not surprisingly, the algebraic complexity seems to exponentiate with the number of regions. When one has dealt with three or four regions simultaneously one knows, first-hand, the meaning of the word "tedious." On the other hand, this situation can always be alleviated by the observation that, although there may, *in principle*, be three, or four, or even five, regions in the problem, in practice, at any given current (voltage), only two regions are really important—of course, not the same two regions at all currents (voltages). The second annoyance is an intrinsic feature of the method, namely, the presence of discontinuities, at the joining points  $x_a$ ,  $x_b$ , etc. in the values of some physical parameters of secondary importance. It seems to bother people to be confronted with discontinuities in the values of physical parameters, even when the parameters in question are of only minor significance. These discontinuities are, of course, an artifact of the method. It is inherent in the

method that it be somewhat (factor-of-two) inaccurate at the joining points. Physical parameters of primary importance are always continuous at the joining points, because, in effect, the joining conditions Eq. [35] are an expression of this continuity. The author regards both annoyances as a small price to pay for the solution of an otherwise unsolvable problem, and, at that, a solution which exposes all of the underlying physics.

<sup>32</sup> M. A. Lampert, *Physics Today*, p. 49, April 1972.

<sup>33</sup> A. Rose, "The Acoustoelectric Effects and the Energy Losses by Hot Electrons—Part I," *RCA Rev.*, **27**, p. 98 (1966).

<sup>34</sup> Parts II, III, IV, and V of this paper (Ref. [33]) appeared, respectively, in *RCA Review*, **27**, p. 600 (1966); **28**, p. 634 (1967); **30**, p. 435 (1969); and **32**, p. 463 (1971).

# Photoconductivity in Amorphous Semiconductors\*

Richard H. Bube

Department of Materials Science and Engineering, Stanford University, Stanford, Calif. 94305

**Abstract**—The September 1951 issue of *RCA Review* that contained Albert Rose's first major attempt to describe in detail the nature of photoconductivity processes in insulators and semiconductors, also contained articles on photoconductivity in amorphous selenium and amorphous antimony trisulfide in connection with vidicon research at RCA. Earlier believed to be non-photoconducting, amorphous selenium had been utilized in the first developments of Xerography in 1948 because of its photoconductive properties, and research at RCA in 1950 had investigated the photoconductivity of amorphous selenium. Since that time many investigations have been reported on photoconductivity in amorphous selenium, which we shall not review here. Interest in amorphous semiconductors was revived, beginning in the 1960's, by the realization that switching phenomena are observable in a wide variety of amorphous chalcogenide semiconductors, and by basic interest in the amorphous state of elemental and III-V semiconductors as a major perturbation on the crystalline state through the loss of long-range order. In this paper we review some of the more recent understanding of the nature of the photoconductivity process in amorphous semiconductors involving elements from Groups IV, V, and VI of the periodic table.

## Energy States in Amorphous Semiconductors

The existence, nature, and density of energy states in amorphous semiconductors is a problem very much in the center of research activity. The presence of structural disorder in amorphous materials early led Mott<sup>1,2</sup> and others to predict the possible tailing of states

\* This research was supported by the Army Research Office, Durham, N.C.

from the equivalent crystalline band edges into the forbidden gap. Considerable experimental evidence exists that the Fermi level is pinned at a single energy over a wide temperature range (dark conductivity with a single thermal activation energy over a broad temperature range, formation of ohmic electrical contacts with many metals, small magnitude of the field effect) and that low-temperature dc and ac conductivity is characteristic of hopping.<sup>3</sup> The attempt to circumscribe theoretical and empirical inputs into a consistent energy-band diagram has led to a variety of suggestions, some of which are shown in Fig. 1.

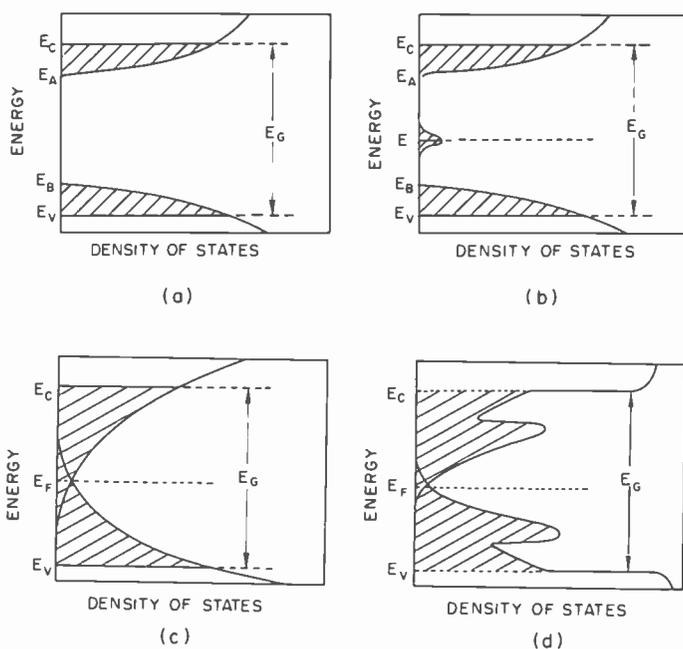


Fig. 1—Four proposed density-of-states distributions for amorphous semiconductors. (a) and (b) are after Mott and Davis (Refs. [3] and [8]); (c) is after Cohen, Fritzsche, and Ovshinsky (Ref. [4]); (d) is after Adler (Ref. [9]).

Cohen, Fritzsche, and Ovshinsky<sup>4</sup> postulated that the band tailing extends so far into the forbidden gap that valence and conduction bands may overlap in energy near the middle of the gap, thus providing a pinning of the Fermi level at that point. They followed the work of Anderson<sup>5</sup> and Mott<sup>6,7</sup> to postulate that the single conductivity thermal activation energy arises because the conduction-band electron mobility decreases sharply for energies below some energy  $E_c$ ,

and the valence-band hole mobility decreases sharply for energies above some energy  $E_v$ . Thus the concept of a "mobility gap" tends to replace that of a "density of states gap" in the crystalline material.

For semiconductors like Ge, Si,  $\text{As}_2\text{Se}_3$ , etc., which have p-like or d-like conduction- and valence-band states, Mott and Davis<sup>3,8</sup> proposed the possibilities shown in Figs. 1a and 1b. In Fig. 1a, the density of states simply goes to zero at energies  $E_A$  and  $E_B$  within the gap. In Fig. 1b an additional set of states at the Fermi level is postulated in order to account for the pinning of the Fermi level and other transport properties described above; these localized states at the Fermi level were attributed to defects such as dangling bonds, interstitials etc.

A modified form of the Cohen, Fritzsche, and Ovshinsky model, shown in Fig. 1c, is given in Fig. 1d after Adler;<sup>9</sup> it takes into account possible electronic correlation effects involving coulombic repulsive splitting of localized states, which makes the density of states a function of electron occupation.

Other examples that have been proposed are a gaussian tail of states due to statistical density fluctuations by Gubanov,<sup>10</sup> and a "chaotic-band model" by Fritzsche and Böer,<sup>11,12</sup> which is based on long- or short-range electrical potential fluctuations of the band edges resulting from charged defects.

All of these possibilities are significant for an understanding of photoconductivity in these materials for two complementary reasons: (1) analysis of photoconductivity must start with some energy-state model, and (2) the observed variations of photoconductivity can be used to give insights into the nature of the model that is suitable.

One possible use of photoconductivity is in the direct determination of energy-state parameters by measuring the spectral response. The problems here, however, are equivalent to those encountered in attempting to achieve the same goal by optical absorption measurements. For  $\text{As}_2\text{SeTe}_2$ , for example,<sup>13</sup> the absorption constant increases monotonically from  $3 \times 10^{-1} \text{ cm}^{-1}$  at 0.6 eV to  $2 \times 10^4 \text{ cm}^{-1}$  at 1.3 eV, and the choice of an absorption edge or the distinction between nonlocalized and localized states is very difficult. Fig. 2 shows the corresponding photoconductivity measurements on both a bulk sample and a sputtered 1- $\mu\text{m}$  film of  $\text{As}_2\text{SeTe}_2$ . Three features of interest are indicated. (1) The spectral dependence of photoconductivity response is essentially the same in the film as in the bulk. (2) The spectral dependence of photoconductivity is the same as that of optical absorption down to photon energies less than or equal to the optical gap (about 0.9 eV at 300°K), indicating that the quantum efficiency is constant over this energy range, quite probably equal to

unity. This behavior has been reported for other chalcogenides.<sup>14,15</sup> (3) The spectral response decreases only slightly for the higher energies corresponding to strong optical absorption, indicating that the surface recombination rate is not significantly different from that in the volume of the material.

Another possible use of photoconductivity is in an attempt to describe the intensity and temperature variations of photoconductivity

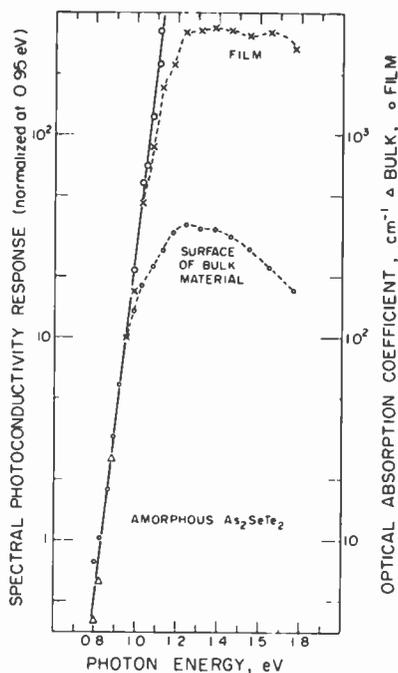


Fig. 2—Spectral photoconductivity response and optical-absorption coefficient as a function of photon energy for amorphous bulk and sputtered 1- $\mu$ m film  $\text{As}_2\text{SeTe}_2$  at 300°K. Values of photoresponse are per incident photon. Data for film and bulk samples are normalized at 0.95 eV. After Arnoldussen et al. (Ref. [13]).

by a consistent recombination model involving nonlocalized and localized states, provided that such variations appear to be reliable indicators of material properties. It is the fact that a consistent pattern of photoconductivity behavior is characteristic of a surprisingly large number of different types of amorphous semiconductors that makes this approach attractive.

### Type-I Photoconductivity

In this section we describe the empirical characteristics of the pattern of photoconductivity behavior that we call Type I photoconductivity. It is found in a variety of IV-V-VI and V-VI amorphous semiconductors,<sup>13,16-19</sup> and in IV-VI amorphous semiconductors of the  $\text{Ge}_x\text{Te}_{1-x}$  type for lower values of  $x$ . Examples are given in Figs. 3 and 4 for  $\text{As}_2\text{SeTe}_2$  and  $\text{As}_2\text{Te}_3$ .

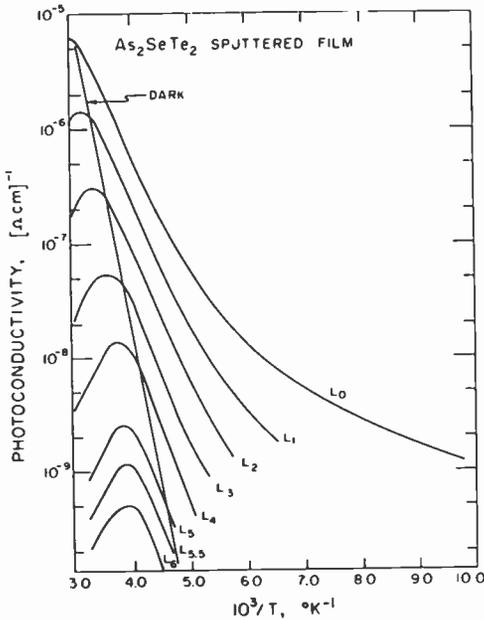


Fig. 3—Temperature dependence of dark conductivity and ac photoconductivity for different photoexcitation intensities in a sputtered 1- $\mu\text{m}$  film of amorphous  $\text{As}_2\text{SeTe}_2$ , as an example of Type-I photoconductivity. Subscripts on the  $L$ 's represent the optical density of the attenuation filters used. The total incident photon flux is  $10^{18} \text{ cm}^{-2} \text{ sec}^{-1}$  for a white-light source evaluated at its maximum at 1.1  $\mu\text{m}$ . After Arnoldussen et al. (Ref. [13]).

Type I photoconductors have a dark conductivity with a single activation energy over many orders of magnitude; in general they have a positive thermoelectric power that varies with temperature in a manner characteristic of one-carrier conductivity. They are relatively insensitive to annealing below the crystallization temperature, and show a large increase in both dark conductivity and photoconductivity upon crystallization.<sup>19</sup>

The photoconductivity in a Type I material exhibits a maximum as a function of temperature, which for high light intensities is about the same magnitude as the dark conductivity at that same temperature. For temperatures above the maximum, the photoconductivity increases with increasing  $1/T$  with a variation that can be characterized by a single activation energy  $E^+$ . This activation energy is independent of intensity over a wide range of intensities; for  $\text{As}_2\text{SeTe}_2$ ,

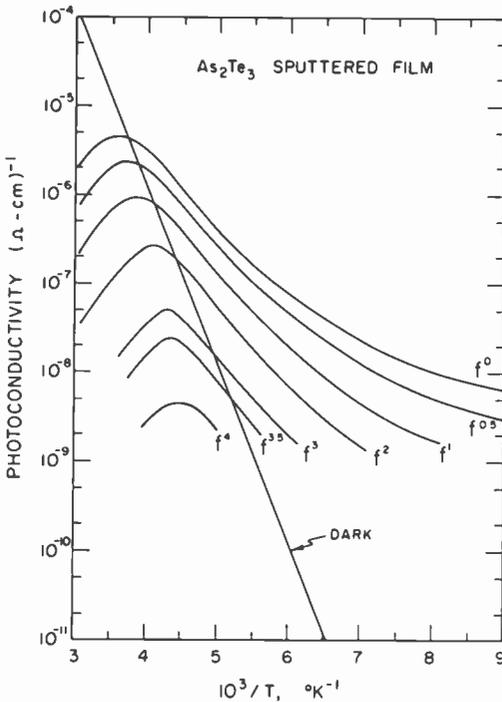


Fig. 4—Dark conductivity versus temperature, and photoconductivity at different light intensities versus temperature for sputtered film of  $\text{As}_2\text{Te}_3$ , as an example of Type-I photoconductivity. Superscripts on the  $f$ 's represent the optical density of the attenuation filters used. The total incident flux is  $2.5 \times 10^{17} \text{ cm}^{-2} \text{ sec}^{-1}$  for a white-light source evaluated at its maximum at  $1.1 \mu\text{m}$ . After Bube et al. (Ref. [18]).

for example,  $E^+ = 0.204 \pm 0.006 \text{ eV}$  for a range of intensities of  $10^3$ . In this high-temperature region, the photoconductivity varies linearly with light intensity.

In the intermediate temperature range just below the temperature maximum, the photoconductivity varies as the square root of light intensity for high intensities (and in this regime the maximum temper-

ature shifts with intensity) and linearly with light intensity for low intensities (in this regime the maximum temperature is independent of intensity). For all intensities, the photoconductivity decreases with increasing  $1/T$  with a variation that can be characterized by a single activation energy  $E^-$ ; for  $\text{As}_2\text{SeTe}_2$ ,  $E^- = 0.252 \pm 0.006$  eV for a range of intensities of  $10^6$ .

Finally at low temperatures, the photoconductivity approaches an asymptotic value independent of temperature, and varies linearly with the light intensity.

### Photoconductivity Model

A variety of *ad hoc* and more basic models exist that describe portions of the above Type-I behavior adequately,<sup>15,20-24</sup> but no model does justice to all these phenomena and is at the same time consistent with dark conductivity data on a variety of materials as well as that proposed by Arnoldussen et al.<sup>16,17</sup> This model has been applied to analyze behavior in three IV-V-VI compounds,<sup>17</sup> and four V-VI compounds,<sup>13,18</sup> providing the basis for the calculation of a variety of energy, density, and recombination parameters. Recently the general framework of the model has been applied to a total of 20 amorphous covalent semiconductors<sup>19</sup> with results that are summarized later in this paper.

Since the development of the model was evolutionary, with constant feedback between various forms of a basic physical model and empirically observed behavior, a somewhat chronological description may be worthwhile to distinguish it from other models that are either totally theoretical or totally empirical.

The model started with the simple assumption that there are conduction nonlocalized states and valence nonlocalized states, and that there are localized states near the conduction and valence mobility edges. Occupancies of localized states are assumed to control recombination and to be describable in terms of Fermi statistics following the pattern of the standard Shockley-Read model.<sup>25</sup> The simplest type of model of this kind is one in which only nonlocalized (NL) to localized (L) state transitions occur, and in which some continuous distribution of L states is assumed, e.g., a uniform or exponential distribution in density with energy away from the mobility edges. Simmons and Taylor<sup>22</sup> have also examined this form of the model in some detail. It is readily shown that the uniform distribution of "tail" states is inadequate to describe Type-I photoconductivity. An exponential distribution of "tail" states is also inadequate if a "slow" tailing of density with depth is chosen; if a "rapid" tailing is chosen, the

first signs of agreement with experiment are found: (a) the carrier density varies linearly with light intensity in the small-signal case and as the square root of intensity in the large-signal case, and (b) at temperatures above the photoconductivity maximum, the carrier density increases exponentially with  $1/T$ . Other features of Type I behavior are not reproduced by simple tailing states at the mobility edges, however.

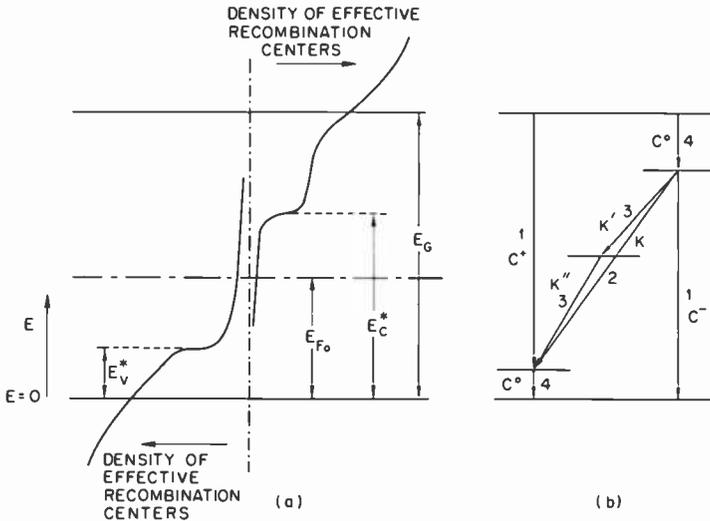


Fig. 5—(a) Schematic energy-level diagram for the photoconductivity model and (b) typical recombination transitions considered in the photoconductivity model.

This experience leads to the conclusion that the way to obtain an exponential variation of photoconductivity with  $1/T$  above and below the photoconductivity maximum is to limit the distribution of effective L states controlling recombination to regions of the mobility gap near the mobility edges. Fig. 5a illustrates this picture. Nonlocalized states for negative carriers lie above  $E_G$ ; C states (neutral when electron unoccupied) have a high density or high probability of controlling recombination above  $E_c^*$  and a much smaller one below  $E_c^*$ ; V states (neutral when electron occupied) have a high density or high probability of controlling recombination below  $E_v^*$  and a much smaller one above  $E_v^*$ ; nonlocalized states for positive carriers lie below  $E = 0$ . The actual shape of the localized state distributions is neither specified nor critical, as long as there is a rapid change at  $E_c^*$  and  $E_v^*$  compared to changes at higher or lower energies; in the extreme case  $E_c^*$  and  $E_v^*$  could respond to discrete energy levels.

Fig. 5b shows all of the electronic transitions finally involved in the model. At this stage of the evolution, only transitions 1 and 4 have been discussed. It is generally found, however, that consistency between the predictions of the model based on the photoconductivity data and the necessary correlation with the activation energy of the dark conductivity are not satisfied if only NL-to-L transitions are considered. Taylor and Simmons<sup>24</sup> do claim satisfactory correspondence for  $\text{Si}_{40}\text{Te}_{40}\text{As}_{20}$ , but examination of data on amorphous covalent semiconductors in general shows that the model-predicted dark conductivity activation energy and the measured value of that energy differ by an average of 20 percent.<sup>19</sup> Consistency is achievable, however, if L-to-L transitions are included, transitions 2 in Fig. 5b with recombination coefficient  $K$ . Other evidence suggests the importance of transitions 2 also; Weiser et al.,<sup>21</sup> for example, have observed luminescence emission that they interpret as arising from an L-to-L transition.

Even including the L-to-L transitions, however, does not account for the linear variation of photoconductivity at low light intensities in the intermediate temperature range just below the photoconductivity maximum. One apparent option is to associate these transitions with L states at the equilibrium Fermi level in such a way that transitions 3 with recombination coefficients  $K'$  and  $K''$  occur between these states and L states near the mobility edges.

An evaluation of the total spectrum of possibilities afforded by this model leads to the conclusion that the following relationships are the only ones consistent with the model and its application to a variety of V-VI materials. They may be written as follows for a material in which  $\Delta p \gg \Delta n$ , the alternative situation being readily achieved by simply inverting the energy diagram and its labeling.

In the high-temperature range,  $\Delta p \ll p_0$ , and it is formally possible for either transitions 1 or transitions 2 to dominate. In either case, charge neutrality is controlled by states at the Fermi level.

$$\Delta\sigma_1 = \frac{fe\mu_0}{2kTC_h^-G_c} \exp[(E_c^* - E_{f_0} - E_\mu)/kT], \quad [1]$$

$$\Delta\sigma_2 = \frac{fe\mu_0N_v}{4(kT)^2KG_cG_v} \exp[(E_c^* - E_v^* - E_{f_0} - E_\mu)/kT], \quad [2]$$

where  $f$  is the photoexcitation intensity in  $\text{cm}^{-3} \text{sec}^{-1}$ ,  $G_c$  is the density of localized states at and above  $E_c^*$ ,  $G_v$  is the density of localized states at and below  $E_v^*$ ,  $E_{f_0}$  is the equilibrium Fermi energy, and  $N_v$  is an "effective density of states" at the valence edge. The mobility is

assumed to be given by  $\mu = \mu_0 \exp(-E_\mu/kT)$ ; measured differences between  $E_{f_0}^0$  from thermoelectric power data and  $E_\sigma$  are interpreted to mean that  $E_\mu = E_\sigma - E_{f_0}^0$ . General trends of behavior suggest that transitions 2 dominate at high temperatures as well as at high intensities at intermediate temperatures, although profound differences in the analysis for densities and recombination coefficients do not result from using either Eq. [1] or [2] in place of the other with  $\text{As}_2\text{SeTe}_2$ .<sup>13</sup>

For the intermediate temperature range at high light intensities,  $\Delta p \gg p_0$ , transition 2 dominates, and neutrality is determined by localized states near the mobility edges.

$$\Delta\sigma_{ht} = \left(\frac{f}{2K}\right)^{1/2} \frac{N_c}{G_c kT} e\mu_0 \exp[-(E_c^* + E_\mu)/kT]. \quad [3]$$

At low light intensities in the same temperature range,  $\Delta p \ll p_0$ , transitions 3 dominate, and neutrality is determined by states at the Fermi level.

$$\Delta\sigma_{10} = \frac{f e \mu_0 N_c}{N_{f_0} K'' G_c kT} \exp[-(E_c^* + E_\mu)/kT], \quad [4]$$

where  $N_{f_0}$  is the density of states at the Fermi level.

Finally for low temperatures,  $\Delta p \gg p_0$ , transitions 4 dominate, and neutrality is determined by states near the mobility edges.

$$\Delta\sigma_{LT} = \frac{f e \mu_{LT}}{C_h^0 G_c E_c^{*0}}, \quad [5]$$

where  $\mu_{LT}$  is the asymptotic mobility reached at low temperatures.

From Eqs. [2] through [4] it follow that

$$E_c^{*0} = E^- + E_{f_0}^0 - E_\sigma = E^- - E_\mu, \quad [6]$$

$$E_c^{*0} = E^+ + E^- + E_{f_0}^0. \quad [7]$$

The use of the superscript zero emphasizes that it is the value of the energy at absolute zero that is determined from a measurement of activation energy, if the energy varies linearly with temperature.

It is evident that the above equations are consistent with the characteristics of Type-I photoconductivity. If they are applied to a series of photoconductivity curves as a function of light intensity and temperature for a given amorphous material, values of  $N_v$ ,  $G_v$ ,  $G_c$ ,  $K$ ,  $N_{f_0} K''$ ,  $E_c^{*0}$  and  $E_c^{*0}$  can all be derived provided that values are assumed for  $\mu_0$  and  $C_h^0$ . For Type-I photoconductors, using  $\mu_0 = 10$

$\text{cm}^2/\text{V}\cdot\text{sec}$  and  $C_h^0 = 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$ ,  $N_v \sim 5 \times 10^{19} \text{ cm}^{-3}$ ,  $G_v \sim 5 \times 10^{18} \text{ cm}^{-3} \text{ eV}^{-1}$ ,  $G_c \sim 10^{20}\text{--}10^{22} \text{ cm}^{-3} \text{ eV}^{-1}$ ,  $K \sim 10^{-6}\text{--}10^{-4} \text{ cm}^3 \text{ sec}^{-1}$ , and  $N_{j_0}K'' \sim 3 \times 10^7 \text{ sec}^{-1}$ . The variation in these quantities between different members of V-VI and similar Type-I materials is relatively small.

The energy parameters of these materials are found to scale closely with the absorption gap  $E_G^0$  regardless of the specific composition of the material. It has been known for some time that the conductivity activation energy for amorphous covalent semiconductors tends to be about one-half of  $E_G^0$ .<sup>26</sup> A survey of some 20 amorphous semiconductors of IV-V-VI, V-VI, IV-V and IV-VI types<sup>19</sup> (showing both Type-I and Type-II photoconductivity) shows that all of them fall with some small scatter about a line representing  $E_\sigma = 0.46 E_G^0$ . Furthermore, data obtained on a smaller group of seven of these materials, on which complete photoconductivity and thermoelectric power data were available, show that  $E_c^{*0} = 0.79 E_G^0$ ,  $E_{j_0}^0 = 0.36 E_G^0$ ,  $E_v^{*0} = 0.11 E_G^0$ ,  $E_\sigma = 0.49 E_G^0$ , and consequently  $E_\mu = 0.13 E_G^0$ .

It is of interest to note that research on certain forms of amorphous silicon indicate that a quite similar model is also adequate for that situation.<sup>27-32</sup>

### Type-II Photoconductivity

In some amorphous covalent semiconductors, the photoconductivity does not follow the characteristics attributed to Type I above. Instead the photoconductivity maximum is absent, the photoconductivity simply increasing slowly and monotonically with increasing temperature, and in general the photoconductivity is much less than the dark conductivity. In essence the behavior is like that when transition 3 dominates the photoconductivity and as if this behavior continued over the whole measurable range without transforming to a transition-2-dominant regime. This approach, with some approximation, can be used to analyze Type-II behavior at least partially.

Definite Type-II behavior is found in IV-V semiconductors such as  $\text{Sn}_x\text{As}_{1-x}$  and  $\text{Ge}_x\text{Sb}_{1-x}$ , and in  $\text{Ge}_x\text{Te}_{1-x}$  IV-VI materials<sup>19</sup> for  $x \geq 0.67$ . Of these materials,  $\text{Sn}_x\text{As}_{1-x}$  ( $x = 0.55$ ) is the only one to show a positive thermoelectric power over the whole temperature range.  $\text{Ge}_x\text{Sb}_{1-x}$  ( $x = 0.41, 0.57$ ) exhibits a negative thermoelectric power at higher temperatures, and a positive thermoelectric power at lower temperatures, the transition temperature depending on  $x$ .  $\text{Ge}_{0.67}\text{Te}_{0.33}$  has a negative thermoelectric power over the whole measured range.<sup>33</sup> In general the dark conductivity of these materials

does not vary exponentially with  $1/T$  except at high temperatures, but shows concavity upwards.

Analysis of the photoconductivity of these materials suggests that a major difference occurs in the magnitude of the localized state density at the mobility edge, as compared to Type-I materials. As indicated above, in p-type Type-I materials,  $G_c \gg G_v$ . In p-type Type-II materials  $G_v > G_c$ , and in n-type Type-II materials  $G_c \gg G_v$ . Furthermore the value of the localized state density at the mobility edge of Type-II materials is so small that the  $G_c G_v$  product is much smaller than in Type-I materials; this is a consistent description of why transitions 2 do not dominate over transitions 3 in Type-II materials.

A number of materials appear to be intermediate between Type I and Type II; these include the  $\text{Ge}_x\text{As}_x$  ( $x = 0.33, 0.50$ ) materials<sup>34</sup> and the  $\text{Ge}_x\text{Te}_{1-x}$  materials<sup>19</sup> for  $0.3 \leq x \leq 0.5$ . The  $\text{Ge}_x\text{As}_x$  materials exhibit a photoconductivity maximum and a well defined  $E^-$ , but the magnitude of the photoconductivity is much less than the dark conductivity. The  $\text{Ge}_x\text{Te}_{1-x}$  materials in the composition range indicated show a generally Type-I behavior as prepared, but gentle annealing below the crystallization temperature makes their behavior shift toward Type II.  $\text{Ge}_{0.50}\text{Te}_{0.50}$  actually transforms from Type-I to Type-II behavior upon annealing. Materials that are definitely either Type I or Type II are relatively unaffected by annealing.

A full photoconductivity analysis on  $\text{Ge}_x\text{Te}_{1-x}$  materials in the composition range where they exhibit an intermediate behavior between Type I and Type II cannot be carried out, since the thermoelectric power on these materials does not exhibit a temperature dependence characteristic of one-carrier behavior. This is not surprising since the conductivity type of  $\text{Ge}_x\text{Te}_{1-x}$  materials shifts from p-type to n-type with increasing  $x$ . Since these materials do show a photoconductivity maximum, however, it is possible to analyze them in terms of the location of that maximum in the same way applied to Type-I materials; such an analysis is capable of yielding a value of  $(E_c^* - E_{f0})/E_G^0$ . For Type-I materials this quantity has the value 0.43, as indicated above. The analysis was applied to two IV-V-VI materials<sup>17</sup> and ten IV-VI  $\text{Ge}_x\text{Te}_{1-x}$  materials;<sup>19</sup> the average value obtained was  $0.43 \pm 0.04$ . At the least, this result indicates that the temperature for maximum photoconductivity in these materials scales with  $E_G^0$  in the same way as found for Type-I materials.

### States at the Fermi Energy

An additional photoconductivity technique is available to obtain an estimate of the density of states at the Fermi energy. This is the tech-

nique of photoconductivity decay. If the density of states within  $kT$  of the steady-state Fermi level under photoexcitation is  $N_{fo}$ , then  $N_{fo} = f\tau_0$ , where  $\tau_0$  is the measured photoconductivity decay time in a range where  $\tau_0$  is much greater than the free-carrier lifetime.

Such measurements were made on three V-VI materials and six IV-VI  $\text{Ge}_x\text{Te}_{1-x}$  materials.<sup>19</sup> The range in bandgaps represented was 0.77 eV ( $\text{Sb}_2\text{Te}_3$ ) to 1.36 eV ( $\text{As}_2\text{Se}_2\text{Te}$ ), and the corresponding range in  $N_{fo}$  was  $6 \times 10^{19} \text{ cm}^{-3} \text{ eV}^{-1}$  to  $4 \times 10^{18} \text{ cm}^{-3} \text{ eV}^{-1}$ . A general trend toward higher values of  $N_{fo}$  for smaller  $E_G^0$  was exhibited by most of the materials. If we take  $kT$  for the width controlling  $N_{fo}$  by recombination, comparison with the previously cited value of  $N_{fo}K'' \sim 3 \times 10^7 \text{ sec}^{-1}$  indicates a value of  $K'' \sim 2 \times 10^{-11} \text{ cm}^3 \text{ sec}^{-1}$ .

Field-effect measurements on these materials indicate localized state densities of about the same order of magnitude.<sup>35</sup> A definite correlation has been obtained between two-carrier conductivity and a control of the field effect by localized states at the Fermi level in  $\text{Ge}_x\text{As}_{1-x}$  materials.<sup>34</sup> The results are a localized state density of  $1.2 \times 10^{20} \text{ cm}^{-3} \text{ eV}^{-1}$  with  $\sigma_n/\sigma_p = 1.10$  in  $\text{GeAs}$ , and  $9.0 \times 10^{19} \text{ cm}^{-3} \text{ eV}^{-1}$  with  $\sigma_n/\sigma_p = 1.02$  in  $\text{GeAs}_2$ .

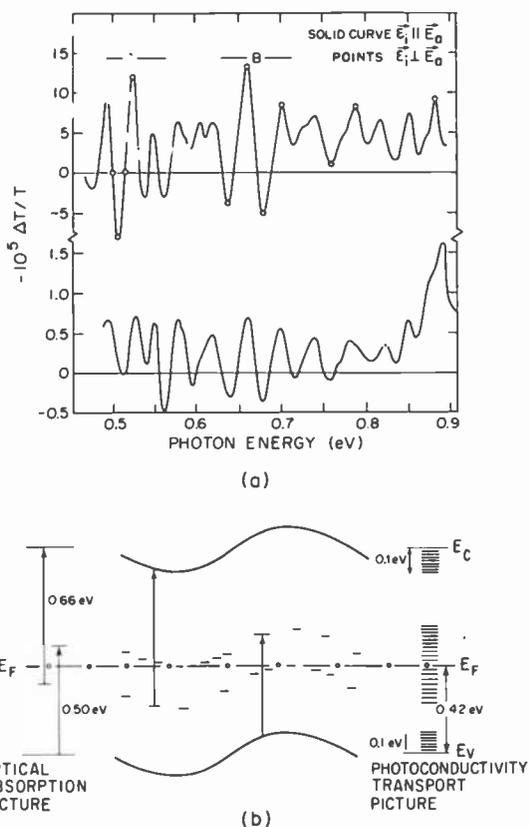
### Physical Interpretation of the Model

The physical significance of the various energy parameters of the model cannot be said to be understood. Several other interpretive inputs may be helpful.

One of the key questions concerns the interpretation of the mobility activation energy  $E_\mu$ . If the model assumes transport by free carriers in nonlocalized states, then what is the significance of a thermally activated mobility? Is it only coincidence that on the average  $E_\mu^0 \sim E_\mu$ ?

One answer that has been proposed is that conductivity is not by free carriers but by small polarons and that the mobility activation energy is the activation energy for hopping of these polarons.<sup>36-38</sup> Dark-conductivity, Hall-effect, thermoelectric-power, and field-effect measurements may all be interpreted in terms of a small polaron mechanism only, without reference to localized states in the gap. It seems quite possible that the actual charge-transporting species is a small polaron. However, there is considerable evidence that suggests that such polarons must also interact with localized states in the material: (a) the details of the photoconductivity dependence on intensity and temperature, particularly the role of transitions 3, seems to require interaction with localized states; (b) photoconductivity decay seems to require emptying of localized states; (c) general correlation

between field-effect density of states and photoconductivity density of states suggests that the field effect may also measure localized state densities; and (d) evidence has been found for two quite different field-effect phenomena: one in which the field effect is controlled by mobile carriers (perhaps the polaron range), and a second in which the field effect is controlled by localized states.<sup>35</sup>



**Fig. 6**—(a) Electroabsorption spectra for amorphous  $As_2SeTe_2$  at  $80^\circ K$  for first harmonic ( $V$  applied of 500 V) in upper curve, and second harmonic ( $V$  applied of 800 V) in lower curve. Curve drawn for first-harmonic spectrum is for electric vector of light parallel to applied ac electric field; dots superposed on curve are for first-harmonic spectrum with electric vector of light perpendicular to applied electric field, i.e., there are no polarization effects. (b) "Chaotic potential" model showing localized defect energy levels corresponding to peaks A and B of electroabsorption spectrum. The left-hand side shows the energy-level spectrum as seen by electroabsorption; the right-hand side as seen by electrical transport measurements. After Voronkov et al. (Ref. [39]).

A second answer to the question builds on the chaotic-band model of Fritzsche and Böer<sup>11,12</sup> and seeks a correlation with electroabsorption measurements<sup>39</sup> on  $\text{As}_2\text{SeTe}_2$ . The data and the interpretation are given in Fig. 6. The two main peaks, A and B, in the electroabsorption spectrum are interpreted as being the consequence of a transition from the valence edge to an empty donor 0.50 eV above the edge, and of a transition from an occupied acceptor 0.66 eV below the conduction edge, respectively. If these defect densities are large and roughly equal, as well as uniformly distributed spatially, the Fermi level would lie about 0.42 eV from the valence edge, which compares favorably with the measured  $E_{fo}^0 = 0.41$  eV. The defect states would be charged, and if their density were of the order of  $10^{19} \text{ cm}^{-3} \text{ eV}^{-1}$  near  $E_{fo}$ , they would cause a potential fluctuation of  $\pm 0.1$  eV that would bring the levels closer to and further away from the Fermi level at different positions in the material, effectively pinning it. The edges also fluctuate in energy by  $\pm 0.1$  eV, giving rise to spatially isolated pools of carriers which in transport measurements behave like localized states extending down to about 0.1 eV; this effect would then account for  $E_v^{*0} = 0.11$  eV  $\sim E_\mu = 0.14$  eV. Electroabsorption measurements see only the discreteness of the states, as shown on the left side of Fig. 6b, whereas photoconductivity and transport measurements integrate over the material and see a distribution of states near the edges and near the Fermi level, as shown on the right side of Fig. 6b.

### Acknowledgments

I would like to thank the Army Research Office, Durham, and Dr. Charles Boghosian for the six years of support they have given to us to permit much of the research described in this paper. I would also like to thank the many individuals who contributed to its success: Dr. Thomas C. Arnoldussen, Dr. Cornelius A. Menezes, Dr. Ralph T.-S. Shiah, Dr. Edward A. Fagen, Yoichiro Nagakawa, John E. Mahan, Hubert A. Vander Plas, Chan Jo Park, and Scott Holmberg.

### Appendix—Science as a Human Endeavor: A Tribute to Albert Rose

Science is derived from the facts of experience by the process of human interpretation. Although the equations, laws, and theories of science are based on experimental data (i.e., upon facts), the interpretations of scientists (i.e., science) are based upon the ideas of human beings. No fact is self-interpreting; facts never speak for themselves. Each fact that a scientist establishes must be interpreted by him in

accord with his presuppositions and contextual perspective at that moment. Every experiment is already designed and planned on the basis of certain theoretical presuppositions which, to some extent, predetermine both the outcome and the interpretation of the experiment. Depending on the field of science, the scientist is more or less part of the experimental context. Good science seeks to minimize these presuppositional effects; real science cannot exist without them.

It is perhaps not as fashionable today as it once was to think of science as an infallible and impersonal process—somewhat like a vast meatgrinder—that mechanically churns out significant understanding and technological results if only objective data are fed in the other end. But I think there is still often an underestimation of the personal and human ingredients implicit in a successful career in science. Few students in a scientific field realize the large role to be played in their future by their ability to communicate effectively to other human beings through speaking and writing. It is only relatively recently that the ethical role that the concerned and responsible scientist must play has been generally evaluated. The part played by faith in the pursuit of science is often ignored, as though human beings could successfully undertake any commitment without a basic faith commitment.

J. T. Davies suggests that one of the necessities for creative thought in science is a curiosity about things rather than people.<sup>40</sup> Curiosity, perhaps; but if creative thought in science includes the wider goals of a fully human life, it cannot exclude sensitivity to people. In his life and work, Albert Rose has combined a profound curiosity about things and a deep sensitivity to persons. As much as I have been constantly impressed by Rose's technical insight—and much of my work has been based on the contributions he has made from the time we first met in 1948—I never think of him as a creative thinker without also thinking of him as a gentle and perceptive human being. His determination to understand the essence of a problem and to cut through elaborations that obscure basic physical reality never prevent him from being charitable and patient. He looks for merit in a weak presentation rather than an opportunity for personal belittling; he thinks that truth is strong enough to win its own way without his participation as ultimate judge. To know Albert Rose is to know not just a scientist, but a human being.

#### References:

<sup>1</sup> N. F. Mott and W. D. Twose, "The Theory of Impurity Conduction," *Adv. Phys.*, 10, p. 107 (1961).

- <sup>2</sup> N. F. Mott, "Electrons in Disordered Structures," *Adv. Phys.*, **16**, p. 49 (1967).
- <sup>3</sup> N. F. Mott and E. A. Davis, *Electronic Properties in Non-Crystalline Materials*, Oxford University Press, London, England (1971).
- <sup>4</sup> M. H. Cohen, H. Fritzsche, and S. R. Ovshinsky, "Simple Band Model for Amorphous Semiconducting Alloys," *Phys. Rev. Lett.*, **22**, p. 1065 (1969).
- <sup>5</sup> P. W. Anderson, "Absence of Diffusion in Certain Random Lattices," *Phys. Rev.*, **109**, p. 1492 (1958).
- <sup>6</sup> N. F. Mott, "The Basis of the Electron Theory of Metals, with Special Reference to the Transition Metals," *Proc. Phys. Soc. Lond.*, **A62**, p. 416 (1949).
- <sup>7</sup> N. F. Mott, "The Transition to the Metallic State," *Philos. Mag.*, **6**, p. 287 (1961).
- <sup>8</sup> E. A. Davis and N. F. Mott, "Conduction in Non-crystalline Systems," *Philos. Mag.*, **22**, p. 903 (1970).
- <sup>9</sup> D. Adler, *Amorphous Semiconductors*, CRC Press, Cleveland, Ohio (1971).
- <sup>10</sup> A. I. Gubanov, "Local Fluctuation Levels in Amorphous Semiconductors," *Sov. Phys.-Solid State*, **4**, p. 2104 (1963).
- <sup>11</sup> H. Fritzsche, "Optical and Electrical Energy Gaps in Amorphous Semiconductors," *J. Non-Cryst. Solids*, **6**, p. 49 (1971).
- <sup>12</sup> K. W. Boer, "The Electrical Conduction Mechanisms in Highly Disordered Semiconductors," *J. Non-Cryst. Solids*, **2**, p. 444 (1970).
- <sup>13</sup> T. C. Arnoldussen, C. A. Menezes, Y. Nakagawa, and R. H. Bube, "Localized States and Carrier Transport in Amorphous Chalcogenide Semiconductors," *Phys. Rev. B*, **9**, p. 3377 (1974).
- <sup>14</sup> H. Rockstad, "Hopping Conduction and Optical Properties of Amorphous Chalcogenide Films," *J. Non-Cryst. Solids*, **2**, p. 192 (1970).
- <sup>15</sup> E. A. Fagen and H. Fritzsche, "Photoconductivity of Amorphous Chalcogenide Alloy Films," *J. Non-Cryst. Solids*, **4**, p. 480 (1970).
- <sup>16</sup> T. C. Arnoldussen, R. H. Bube, E. A. Fagen, and S. Holmberg, *J. Non-Cryst. Solids*, **8-10**, p. 933 (1972).
- <sup>17</sup> T. C. Arnoldussen, R. H. Bube, E. A. Fagen, and S. Holmberg, "Analysis of Photoconductivity in Amorphous Chalcogenides," *J. Appl. Phys.*, **43**, p. 1798 (1972).
- <sup>18</sup> R. H. Bube, J. E. Mahan, R. T.-S. Shiah, and H. A. Vander Plas, "Transport and Localized Levels in Amorphous Binary Chalcogenides," *Appl. Phys. Lett.*, **25**, p. 419 (1974).
- <sup>19</sup> R. T.-S. Shiah and R. H. Bube, *J. Appl. Phys.*, to be published.
- <sup>20</sup> B. T. Kolomiets and V. M. Lyubin, "Electric and Photoelectric Properties of Arsenic Selenide Layers," *Sov. Phys.-Solid State*, **4**, p. 291 (1962).
- <sup>21</sup> K. Weiser, R. Fischer, and M. H. Brodsky, *Proc. 10th Internat. Semicon. Conf.*, Cambridge, Mass. (1970), p. 667.
- <sup>22</sup> J. G. Simmons and G. W. Taylor, *J. Non-Cryst. Solids*, **8-10**, p. 947 (1972).
- <sup>23</sup> J. G. Simmons and G. W. Taylor, "Theory of Photoconductivity in Amorphous Semiconductors Containing Slowly-Varying Trap Distributions," *J. Phys. C: Solid State Phys.*, **6**, p. 3706 (1973).
- <sup>24</sup> G. W. Taylor and J. G. Simmons, "Photoconductivity and the Determination of Trapping Parameters in Amorphous Semiconductors," *J. Phys. C: Solid State Phys.*, **7**, p. 3067 (1974).
- <sup>25</sup> W. Shockley and W. T. Read, "Statistics of the Recombination of Holes and Electrons," *Phys. Rev.*, **87**, p. 835 (1952).
- <sup>26</sup> H. Fritzsche, *Ann. Rev. Mat. Sci.*, **2**, p. 697 (1972).
- <sup>27</sup> W. E. Spear and P. G. LeComber, *J. Non-Cryst. Solids*, **8-10**, p. 727 (1972).
- <sup>28</sup> P. G. LeComber, A. Madan, and W. E. Spear, "Electron Transport and State Distribution in Amorphous Si Films," *J. Non-Cryst. Solids*, **11**, p. 219 (1972).
- <sup>29</sup> W. E. Spear, *Proc. 5th Internat. Conf. on Amorphous and Liquid Semicon.*, Garmisch-Partenkirchen (1973).
- <sup>30</sup> W. E. Spear, R. J. Loveland, and A. Al-Sharbaty (unpublished).
- <sup>31</sup> A. K. Malhotra and G. W. Neudeck, "Field Effect Conductance Change in Amorphous Silicon," *Appl. Phys. Lett.*, **24**, p. 557 (1974).
- <sup>32</sup> G. W. Neudeck and A. K. Malhotra, *J. Appl. Phys.*, **46**, p. 2662 (1975).
- <sup>33</sup> H. A. Vander Plas and R. H. Bube (unpublished data on thermoelectric power).
- <sup>34</sup> C. J. Park, J. E. Mahan, R. T.-S. Shiah, H. A. Vander Plas, and R. H. Bube, *J. Appl. Phys.* (to be published).
- <sup>35</sup> J. E. Mahan and R. H. Bube (unpublished data on field effect).

- <sup>36</sup> M. Roilos, "Conductivity and Hall Effect in Vitreous  $As_2(Se,Te)_3$ ," *J. Non-Cryst. Solids*, **6**, p. 5 (1971).
- <sup>37</sup> D. Emin, C. H. Seager, and R. K. Quinn, "Small Polaron Hopping Motion in Some Chalcogenide Glasses," *Phys. Rev. Lett.*, **28**, p. 813 (1972).
- <sup>38</sup> C. H. Seager, D. Emin, and R. K. Quinn, "Electrical Transport and Structural Properties of Bulk As-Te-I, As-Te-Ge, and As-Te Chalcogenide Glasses," *Phys. Rev.*, **B8**, p. 4746 (1973).
- <sup>39</sup> E. Voronkov, A. D. Jonath, T. C. Arnoldussen, and R. H. Bube, "Extrinsic Electroabsorption of Amorphous  $As_2SeTe_2$ ," *J. Non-Cryst. Solids*, **15**, p. 275 (1974).
- <sup>40</sup> J. T. Davies, *The Scientific Approach*, Academic Press, N.Y. (1973).

# Speed of Response of Photocurrents in CdSe

H. Kiess and B. Binggeli

Laboratories RCA Ltd., Zurich, Switzerland

**Abstract**—The experimental investigations involve measurements of the photocurrent as a function of light intensity, I-V characteristics with light intensity as parameter, and the rise and decay of photocurrents at different light levels. The photocurrents were measured with a logarithmic amplifier so that the rise of the current could be observed from the dark to the steady-state value, the latter exceeding the dark value by several orders of magnitude. The photocurrent-light-intensity characteristic shows a supralinear and, at high light intensities, a saturation regime. The analysis of this characteristic gives trap distributions for the electrons and the holes. Since the traps determine the time constants for the rise and decay of the photocurrents, measurements of the time constants should confirm the derived trap distributions. Indeed, consistency can be obtained if certain assumptions of the de-trapping of electrons and holes are made.

## 1. Introduction

In the outline of photoconductive processes published by A. Rose<sup>1</sup> in 1951, reasons are given for the observation that the response time of the photocurrents is different from the steady-state lifetime of the photogenerated charge carriers. The response time, defined as the time for the photocurrent to rise or to decay to half of the steady-state value, was found to increase with respect to the steady-state lifetime by the ratio of the trapped to the free majority carrier densities. If the electrons are taken to be the majority carriers then the assumption for this result is that during the decay process the conduction electrons are in thermal equilibrium with the trapped electrons and that the holes are exclusively bound to recombination centers

through which recombination takes place. In order for the current to decay to half value, the steady-state Fermi level must drop by approximately  $kT$  and the electrons trapped in this  $kT$  slice of energy must be emptied via the conduction band to the recombination centers, thus determining the time constant.

Normally, however, the quasi-Fermi level of the electrons may shift in energy during the complete rise or decay of the current by many  $kT$ . Furthermore, the holes are not only located in recombination centers, but also in hole traps. The question then arises as to what modifications of the model of A. Rose are necessary in order to describe the response of the photoconductor appropriately.

This paper first shows results of rise and decay measurements over several decades in currents, as well as with time, in order to illustrate the observation of uncommon features in the response behavior of photoconductors. Then changes to the model are suggested to account for the observed behavior.

## 2. Sample Characterization and Experimental Technique

In order to pursue our intention of investigating the large-signal behavior it was necessary to use a photoconductor in which the quasi-Fermi level shifts by several  $kT$  under illumination. Also, the trap distribution and recombination should be known. The latter information can be obtained in part by evaluating the photocurrent versus light intensity characteristic. A material very well suited for these measurements seemed to be CdSe.

The vapor-phase-grown crystals, obtained from the Institut für angewandte Physik, Karlsruhe, were provided with indium contacts. The electrode separation was 80  $\mu\text{m}$ , the length of the electrodes about 500  $\mu\text{m}$ . Between 0.1 and 10 volts, the I-V characteristics indicated ohmic behavior for both the dark current and the photocurrent. However, scanning the samples with a light spot 100  $\mu\text{m}$  in diameter revealed that the CdSe was quite inhomogeneously photoconducting. This is shown in Fig. 1. We decided, therefore, to restrict the data given in this article on trap density, etc., to this particular crystal and to the spot B of this crystal indicated in Fig. 1, because our intention was to elucidate the response behavior of photoconductors rather than to give numbers on the densities of traps and recombination centers in CdSe.

The light source used in the experiments was a GaAs/P light-emitting diode (Fairchild FLV 100) the light from which was focused onto the crystal giving a spot 100  $\mu\text{m}$  in diameter. The wavelength of the light was 6500 Å, and the light intensity could be varied with neutral

density filters between  $10^{-10}$  and  $10^{-5}$  watt. The rise and decay time of the emitted light on applying an electrical pulse to the diode was below 5 nanoseconds, hence several orders of magnitude smaller than the response time of the photoconductor.

The decay of the photocurrents, involving times greater than  $10^{-3}$  sec could be measured with a logarithmic amplifier having a current range between  $10^{-4}$  and  $10^{-10}$  A. The rise of the photocurrents was measured with broadband linear amplifiers, since the speed of response of the logarithmic amplifier was too slow.

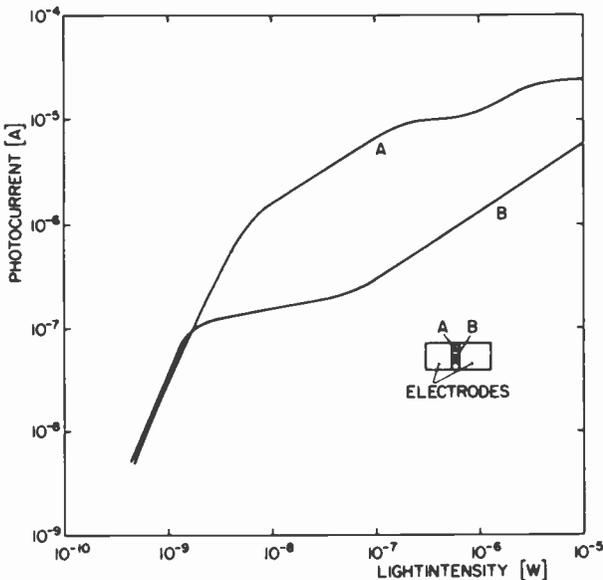


Fig. 1—Photocurrent as a function of light intensity at two different spots on the CdSe single crystal. The light intensity varied between  $10^{-10}$  and  $10^{-5}$  W, which corresponds to a variation in generation rate between  $6 \times 10^{16} \text{ cm}^{-3} \text{ sec}^{-1}$  and  $6 \times 10^{21} \text{ cm}^{-3} \text{ sec}^{-1}$ , figures which were calculated assuming a penetration depth of the light of  $10^{-4}$  cm.

### 3. Photocurrent Versus Light Intensity Characteristic

The photocurrent as a function of light intensity, as shown in Fig. 1 curve B, is characterized by a supralinear increase of the photocurrent at low light intensities and a pronounced saturation of the photocurrent at medium light levels. At high light intensities, the photocurrent is proportional to the square root of the light intensity. Similar results have been published by Bube<sup>2</sup>; the interpretation, how-

ever, follows the proposals of Stöckmann<sup>3</sup> and of Heinz and Stöckmann.<sup>4</sup>

Supralinearity of photocurrents is due to the shift of the quasi-Fermi level of the majority carriers (in the case of CdSe, of the electrons) across a trap with increasing light intensity. As in the case of the trap-filled limit in space-charge-limited currents, the free-electron concentration increases strongly, since the photogenerated electrons can no longer be stored in a trap but are shifted to the conduction band when the Fermi level moves across the trap with increasing light intensity. The argument is only valid if the photogenerated holes are predominantly stored in traps and not in recombination centers, and if their total concentration increases linearly with light intensity.

The recombination rate is given by

$$R_{stat} = v\sigma_p R_n p = g. \quad [1]$$

where  $R_{stat}$  is the recombination rate in the stationary state,  $v$  is the thermal velocity of the holes,  $\sigma_p$  is the capture cross section of the recombination centers  $R$  for the holes when occupied with an electron,  $p$  is the density of the free holes, and  $g$  is the generation rate.  $R_n$  is approximately equal to the total concentration  $R$ .

Therefore the density of trapped holes  $Y_p^I$  is proportional to the generation rate;

$$Y_p^I = \frac{\theta_p^I}{v\sigma_p R} g. \quad [2]$$

where  $\theta_p^I$  is the ratio of trapped to free holes. We find, using the neutrality condition,

$$n + X_n^I = p + Y_p^I = \frac{(\theta_p^I + 1)}{v\sigma_p R} g. \quad [3]$$

If the quasi-Fermi level for the electrons is below the electron trap level,  $X_n^I$ , then the density of occupied traps  $X_n^I$  is proportional to the electron density  $n$ , and therefore with  $\theta_n^I = X_n^I/n$ ,

$$n = \frac{(\theta_p^I + 1)}{(\theta_n^I + 1)} \frac{1}{Rv\sigma_p} g \propto g. \quad [4]$$

If the Fermi level is above the electron trap, then  $X_n^I$  is constant and the electron concentration is again a linear function of the generation rate;

$$n = \frac{\theta_p^I + 1}{v\sigma_p R} g \propto g. \quad [5]$$

The proportionality factor is now much greater, since the trapping factor  $\theta_n^I$  is missing. Therefore the transition range between the two linear regimes, Eq. [4] and [5], must be nonlinear.

The conclusions to be drawn from the existence of the supralinear regime are the following. From the electron density and from the fact that the quasi-Fermi level shifts across an electron trap level  $X$ , we calculate the energy level of  $X^I$  to be 0.3 eV below the conduction band. Recombination takes place via recombination centers that are practically filled with electrons, so that Eq. [1] holds. In order to satisfy the neutrality condition, a hole trap  $Y^I$  must exist that is located between the quasi-Fermi level of the holes and the valence band.

Saturation of the photocurrent occurs if the recombination rate increases strongly with increasing light intensity. Assuming that Eq. [1] still holds, the enhanced recombination rate is only possible if the concentration of free holes increases supralinearly with light intensity. This requires, using the arguments for the explanation of supralinearity of the photocurrent, that the quasi-Fermi level for holes shifts across the hole trap level  $Y^I$ . For neutrality reasons, we have to assume the existence of a further electron trap  $X^{II}$  whose energy separation from the conduction band is calculated from the electron density at the highest generation rate to be smaller than 0.15 eV.

The saturation ends whenever the occupation of hole traps  $Y^{II}$  located between  $Y^I$  and the valence band becomes comparable to  $Y^I$ . Then  $n$  will increase again with light intensity with a linear law, if Eq. [1] still holds, and with a square root law, if recombination takes place predominantly through the centers  $Y^{II}$ . In the latter case we find from the neutrality condition,

$$n = \frac{Y_p^{II}}{\theta^{II} + 1} + \frac{Y^I - X^I}{\theta^{II} + 1}, \quad [6]$$

where  $\theta^{II} = X_n^{II}/n$ . Neglecting the second term,

$$n \propto Y_p^{II}, \quad [7]$$

In the steady state

$$g = rnY_p^{II}$$

and therefore

$$n \propto (g)^{1/2}$$

[8]

In summary, in order to explain the  $J_{ph}$ -light-intensity characteristic we have to assume that at least two electron traps, two hole traps, and a recombination center exist. The energy-band scheme with these centers is shown in Fig. 2.

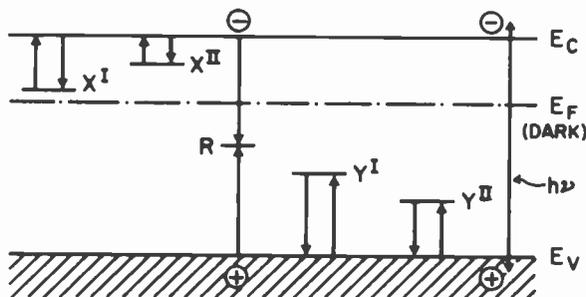


Fig. 2—Schematic representation of the trapping levels and recombination centers, which are needed to explain the  $J_{ph}$ -light-intensity characteristic and which will be used for the explanation of the rise and decay curves.

#### 4. The Rise and Decay of the Photocurrents

The experimentally observed curves of the rise in photocurrent on step illumination are shown in Fig. 3. Since the photocurrents as well as the time cover several orders in magnitude, the curves were plotted logarithmically. The photocurrent became comparable to the dark current only after the density of photo-excited electrons in the conduction band reached about  $10^4$  times the density in the dark. This is due to the different geometrical cross sections of the current carrying path in the dark and under illumination. The totally generated density of electron-hole pairs is then  $10^{15} \text{ cm}^{-3}$  ( $= g \cdot t$ ). The free-electron density increases with a power law in time whose exponent is greater than unity;

$$n_{ph} \propto t^a, \text{ with } a \geq 1.$$

At low light intensities (below  $10^{-8} \text{ W}$ )  $n_{ph}$  levels off to its stationary concentration at a value of  $gt$  of approximately  $10^{16} \text{ cm}^{-3}$ ; at higher light intensities the curve flattens out at the same electron-hole pair density of  $gt \sim 10^{16} \text{ cm}^{-3}$  and continues to increase with an approximate power law in time having an exponent smaller than unity. The stationary value of the free-electron density is reached only after a

significantly larger density of electron-hole pairs ( $gt = 10^{18} \text{ cm}^{-3}$ ) has been generated.

The decay curves on switching off the light depend also on the light intensity. Examples of decay curves are shown in Fig. 4 in a semi-logarithmic plot. At low light intensity, the current decays nearly exponentially. At medium and at high light intensities the decay curves show three time regimes. The first regime, immediately after switching off the light, is characterized by a low decay rate; then follows a regime with a much higher decay rate; and finally, in the third regime, the photocurrent decays with a rate approximately equal to that in the first regime. This behavior is quite unusual—the expectation being that the lower the photocurrent the larger the time constant for the decay.

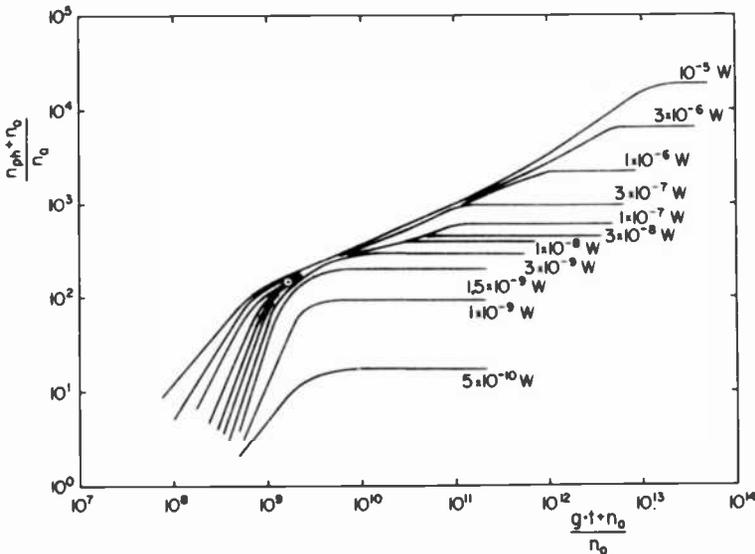


Fig. 3—Rise curves of the photocurrent switching on step illumination; light intensity is a parameter. The photogenerated carrier density is approximately  $10^4$  times the carrier density in the dark in order for the photocurrent to become measurable. This can be traced back to the fact that only a small percentage of the volume carrying the dark current is utilized for the photocurrent. The carrier density in the dark is  $10^7 \text{ cm}^{-3}$ . The curves were measured at spot B (Fig. 1).

## 5. Discussion of the Rise and Decay Curves

### 5.1 The Rise Curves

The complete set of equations describing the response of a photoconductor to light is practically intractable. Böer and Vogel<sup>5</sup> have dis-

cussed the differential equations of models with various recombination mechanisms and they derived qualitatively different possible shapes of rise curves. In the case of the CdSe crystals, however, our knowledge of the traps and of the recombination mechanism obtained from the  $I_{ph}$ -light-intensity characteristic justifies a simplification that reduces the complexity of the rate equations and gives solutions for the rise curves. Despite the simplification some general in-

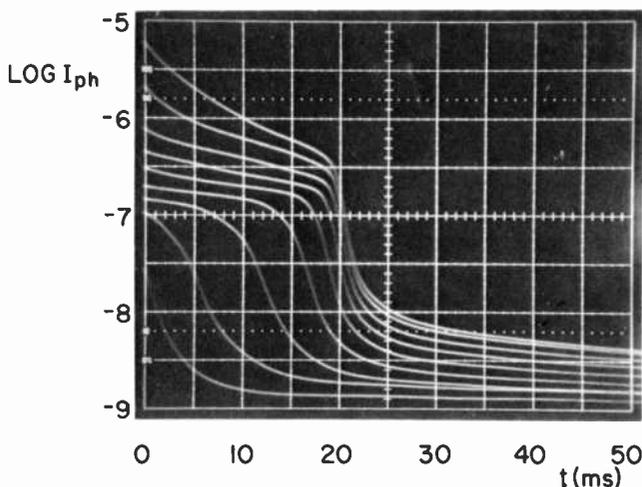


Fig. 4—Decay curves of the photocurrent switching off the illumination. Parameter is the light intensity, the curves were measured at spot B (Fig. 1).

sight into the dynamic processes is obtained. Up to a light intensity of  $\sim 10^{-7}$  watt, the recombination rate in the stationary state is given by Eq. [1] and is independent of the free-electron density. If Eq. [1] also holds during the rise of the photocurrent, then the rise curve for the free and for the trapped holes can be calculated exactly (Appendix 1), and a time limit can be given below which recombination can be neglected and the density of trapped holes increases linearly with time. Since the neutrality condition holds, the density of free and trapped electrons also increases linearly in time. Under these conditions it is again possible to calculate the rise curves of the free-electron density assuming that the traps are shallow (Appendix 2). The result of this calculation is shown in the lower half of Fig. 5. In fact, the curve shown is valid for all generation rates since it represents the dependence of the free-electron density on the total number of generated carriers, i.e., on the product of  $gt$ . Since the  $I_{ph}$ -light-intensity characteristic shows that there are two dominating electron traps, we

included the traps  $X^I$  and  $X^{II}$  in our calculation, causing the two shoulders in the rise curve (these shoulders may be broadened if the capture cross section, trap density, and the mutual energetic separation of the traps are not appropriate). The linear increase of  $n_{ph}$  with time is determined by the parameters of the deepest of the two traps. If only  $X^{II}$  existed, then the dashed curve would have been obtained. Obviously, whenever the trap  $X^I$  fills up, the lower curve must join

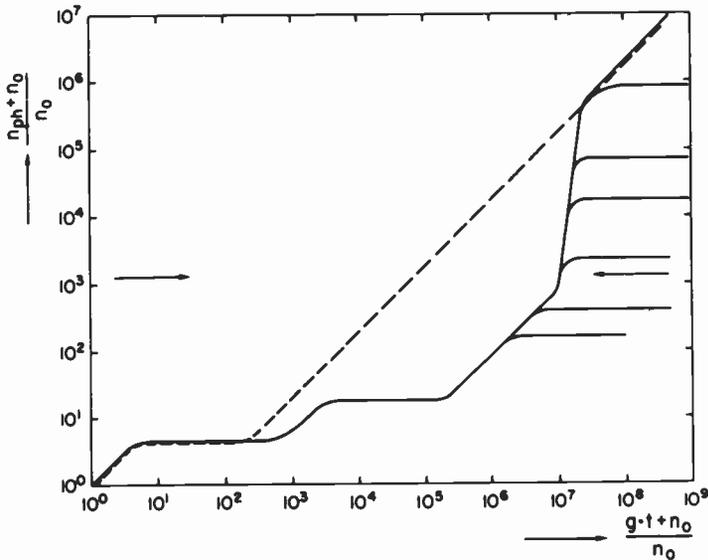


Fig. 5—Rise curves of the free electron density (or photocurrent) at various light intensities calculated under the assumptions specified in the text. Since the  $I_{ph}$ -light-intensity characteristic starts with a supralinear regime at low light intensities, we expect that only the upper half of the figure as indicated by the two arrows can be compared with the experimental rise curves.

the dashed curve so that  $n_{ph}$  increases more than linearly with time (upper half of Fig. 3), i.e., the supralinearity in the  $I_p$ -light-intensity characteristic shows up also in a supralinear part of the rise curve. The rise curves of stationary photocurrents greater than those of the supralinear regime should consequently show first the linear, then the supralinear, and finally again a linear time law before the curves level off to the stationary values.

If a saturation regime occurs in the  $I_{ph}$ -light-intensity curve, then the rise curves of the photocurrent measured in that range of light intensities should level off at the same value of  $gt$ , since only the density of free holes (being, however, negligibly small compared to the

trapped carrier densities) increases significantly with increasing generation rate. An increase of the stationary current beyond the saturation value at still higher light intensities has been interpreted in Section 2 in terms of the neutrality condition: if the concentration of holes in trap  $Y''$  becomes comparable to that in  $Y'$ , then  $n + X_n'' = n(1 + \theta'')$  starts to increase again. Since now recombination can no longer be neglected, the assumption for the calculation of the rise curves breaks down and an exact solution of the problem cannot be found. Recombination, however, will cause the rise curve to become sublinear.

Qualitatively the theoretical predictions are reflected in the experimental rise curves. We observe the supralinear time dependence of the photocurrents in correspondence with the supralinearity of the  $I_{ph}$ -light-intensity characteristic. In the range of current saturation, the value of  $gt$  is in fact practically constant, and at still higher generation rates the rise curves are sublinear. However, the curves obtained for different light intensities do not coincide in the plot of  $n_{ph}$  versus  $gt$ , especially at low values of  $gt$ , where a considerable spread of the curves is observed. To date, no explanation for this deviation can be given.

## 5.2 The Decay Curves

The decay curves can be explained in terms of the shift of the quasi-Fermi level of the electrons. For the sake of clarity it is best to consider first the decay of photocurrents equal to or smaller than the saturation current and second the decay of higher photocurrents.

If the stationary photocurrent is equal to or smaller than the saturation current, we expect the recombination rate to drop significantly upon switching off the light, since the density of free holes reduces by recombination and the only source of supply are the holes trapped in  $Y'$ . These traps are deep compared to the electron traps, and therefore the emission rate of holes from  $Y'$  to the valence band will become the time-limiting process for the recombination; hence it will be identical to the decay rate of the free and trapped electrons, assuming that direct recombination of electrons with trapped holes in  $Y'$  is still negligible. As the decay rate of electrons is small compared to the emission and recapture rate of the electron traps, we can consider the free electrons to be in quasi-equilibrium with their traps. Therefore the electrons in the conduction band and in the shallow traps can also be described by a quasi-Fermi level during the decay process.

If during the decay process the quasi-Fermi level is in the void between  $X'$  and  $X''$ , then the electron density (in  $X''$  and in the con-

duction band) will decay with the same rate as the holes in  $Y^I$ , the time law for the decay of the photocurrent being therefore exponential. The decay of the photocurrent will accelerate whenever the quasi-Fermi level approaches  $X^I$  since during the shift of the Fermi level across  $X^I$  the free electron density drops strongly. Once the Fermi level is sufficiently below  $X^I$  the free electron density is proportional to the electron density in  $X^I$  and therefore the decay rate will be again equal to the emission rate of holes out of  $Y^I$ .

The decay of photocurrents greater than the saturation current will be given initially by the biomolecular recombination process through the centers  $Y^{II}$ . It will be therefore hyperbolic until the hole concentration in  $Y^{II}$  drops to a sufficiently small value so that the emission rate of the holes from  $Y^I$  into the valence band becomes again the time-limiting step. Therefore, except for an initially faster decay, we expect in this case the same shape of the decay curves.

The experimental results show, indeed, the time dependence predicted by these considerations. Within the accuracy of these measurements the decay laws are exponential for the three different time regimes. It cannot be determined whether or not the time law for the decay of the highest photocurrents is first hyperbolic and then exponential. In fact more accurate measurements will be needed to establish these details in the decay curves.

## 6. Conclusions

These results show that the shape of the rise and decay curves reflect the trapping and the recombination of the photoexcited carriers. If worked out in more detail, these measurements, in combination with the  $I_{ph}$  versus light intensity characteristic, allow a determination of the trap densities and energies as well as of the recombination mechanism.

The interpretation of the rise curves was essentially based on the neutrality condition and, up to a certain elapsed time, also on neglecting recombination. The time limit for these assumptions to hold was calculated from the trapping and the lifetime of the holes to be  $T = \theta_p \cdot \tau_p$ , a time equal to the rise time of the photocurrent in terms of the model of A. Rose. The additional feature obtained here is the structure appearing in the rise curves.

For the explanation of the decay curves, the following points were considered to be important: the detrapping of the holes, a recombination rate different from that in the stationary state, and the assumption that the electrons are in quasi-equilibrium with their traps. If these assumptions are lumped together, the overall decay time is

again identical to the response time as given by A. Rose; however, the shape of the decay curve can only be explained if the assumptions are taken into account in detail.

## Appendix 1

The assumption for this calculation, namely that the recombination rate can be expressed by  $R_{stat} = r_p R_n p$  and that the hole traps are shallow, have been confirmed in the case of our CdSe crystals by measurements of the  $I_{ph}$ -light-intensity characteristic for currents below saturation. The concentration of recombination centers occupied by electrons was approximately equal to the total concentration of recombination centers. Therefore the recombination rate depends only on the density of free holes. Assuming two shallow hole traps  $Y^I$  and  $Y^{II}$ , the rate equations for the holes are

$$\frac{dp}{dt} = g - r_1 Y^I p + \gamma^I Y_p^I - r_2 Y^{II} p + \gamma^{II} Y_p^{II} - r_p R_p p. \quad [9]$$

$$\frac{dY_p^I}{dt} = r_1 Y^I p - \gamma^I Y_p^I \quad [10]$$

$$\frac{dY_p^{II}}{dt} = r_2 Y^{II} p - \gamma^{II} Y_p^{II} \quad [11]$$

Here  $r_1$ ,  $r_2$ , and  $r_p$  are the recombination coefficients of the centers  $Y^I$ ,  $Y^{II}$ , and  $R$ , respectively,  $\gamma^I$  and  $\gamma^{II}$  are the thermal emission coefficients of the traps, and  $Y_p^I$  and  $Y_p^{II}$  are the densities of  $Y^I$  and  $Y^{II}$  when occupied with a hole. The other symbols have their usual meaning.

This system of coupled differential equations can be solved by elementary methods. After some lengthy calculation one finds from Eq. [9] to [11] the following equation for  $p$ :

$$\frac{d^3 p}{dt^3} + A \frac{d^2 p}{dt^2} + B \frac{dp}{dt} + C p - D = 0, \quad [12]$$

with

$$A = (r_1 Y^I + r_2 Y^{II} + r_p R + \gamma^I + \gamma^{II})$$

$$B = \gamma^{II} r_1 Y^I + \gamma^I r_2 Y^{II} + (\gamma^I + \gamma^{II}) r_p R$$

$$C = \gamma^I \gamma^{II} r_p R$$

$$D = \gamma^I \gamma^{II} g.$$

The boundary conditions are: for  $t < 0$ ,  $g = 0$ ; and for  $t = 0$ ,  $g = g_0$ . Therefore at  $t = 0$ ,  $p = 0$  and

$$\frac{dp}{dt} = g_0$$

$$\frac{d^2p}{dt^2} = -g_0(r_1Y^I + r_2Y^{II}).$$

The solution of Eq. [12] is

$$p = \frac{g}{r_p R} [1 - \exp\{-\rho_1 t\}] + C_3 [\exp\{-\rho_3 t\} - \exp\{-\rho_1 t\}]$$

$$+ C_2 [\exp\{-\rho_2 t\} - \exp\{-\rho_1 t\}]. \quad [13]$$

Since  $\rho_2, \rho_3 \gg \rho_1$  and  $g/r_p R \gg C_3, C_2$ , only the first term in Eq. [13] is important for our consideration. We see immediately that recombination can be neglected whenever  $\rho_1 T \ll 1$ . With  $\rho_1 = (\gamma_1/r_1 Y^I) r_p R$ , we find  $T = (r_1 Y^I / \gamma_1^I) (1/r_p R) = \theta_p \cdot \tau_p$  since  $r_1 Y^I / \gamma_1^I$  is equal to the trapping factor  $\theta_p = Y_p/p$  and since the lifetime of the free holes is  $\tau_p = 1/r_p R$ . Below  $T$  the density of trapped holes increases also linearly with time and obviously  $(Y_p^I + Y_p^{II}) \approx gt$ .

## Appendix 2

Since the electrons are the majority carriers, the rise curve of the density of free electrons is identical to the rise curve of the photocurrent if changes in the mobility can be neglected. In Appendix 1 a time limit was derived below which recombination can be neglected. Then the rate equations for the electrons are, if we include two electron traps in agreement with our interpretation of the  $I_{ph}$ -light-intensity characteristic,

$$\frac{dn}{dt} = g - \gamma_1 X^I n + \gamma^I X_n^I - r_2 X^{II} n + \gamma^{II} X_n^{II}, \quad [14]$$

$$\frac{dX_n^I}{dt} = r_1 X^I n - \gamma^I X_n^{II}, \quad [15]$$

$$\frac{dX_n^{II}}{dt} = r_2 X^{II} n - \gamma^{II} X_n^{II}. \quad [16]$$

The symbols have their usual meaning. This set of differential equations is the same as before except for the missing recombination term. Using the analogous boundary conditions as before in the case of the holes, the solution is found to be

$$n = |C_2| [1 - \exp\{-\rho_2 t\}] + |C_1| [1 - \exp\{-\rho_3 t\}] + \frac{k}{B} t. \quad [17]$$

with

$$\begin{aligned} \rho_3 &= r_1 X^I + r_2 X^{II} + \gamma^I + \gamma^{II}, \\ C_3 &= \frac{1}{\rho_3(\rho_3 - \rho_2)} [g(\rho_2 - r_1 X^I - r_2 X^{II}) - \frac{k}{B} \rho_2] \\ \rho_2 &= \frac{\gamma^{II} r_1 X^I + \gamma^I r_2 X^{II} + \gamma^I \gamma^{II}}{\rho_3}, \\ C_2 &= \frac{1}{\rho_2(\rho_3 - \rho_2)} [g(r_1 X^I + r_2 X^{II} - \rho_3) + \frac{k}{B} \rho_3] \\ k &= g \gamma^I \gamma^{II} \\ B &= \gamma^{II} r_1 X^I + \gamma^I r_2 X^{II} + \gamma^I \gamma^{II}. \end{aligned}$$

If  $\rho_3 > \rho_2$ , then  $C_2 > C_3$ , so that the first two terms show up as shoulders in the rise curves at  $t = 1/\rho_3$  and  $t = 1/\rho_2$ , respectively (Fig. 5). The linear part of the curve is given by the last term, which is essentially determined by the reciprocal of the trapping factor  $\gamma_1^I/\gamma_1 X^I$  of the deepest of the two traps, i.e., of  $X^I$ . If  $X^I$  fills up and if recombination can still be neglected, then  $X^{II}$  becomes important. Since the trapping factor of  $X^{II}$  is smaller than that of  $X^I$ , the two linear curves have to join by a nonlinear transition region, which corresponds to the supralinear time law in the rise curve of the photocurrent.

### Acknowledgment

We are indebted to W. J. Merz and J. Sandercock for critical readings of the manuscript and to F. Stöckmann for helpful discussions.

### References:

- <sup>1</sup> A. Rose, "An Outline of Some Photoconductive Processes," *RCA Review*, **12**, p. 362 (1951).
- <sup>2</sup> See, e.g., R. H. Bube, *Photoconductivity of Solids*, John Wiley and Sons, Inc., New York (1960), p. 343.
- <sup>3</sup> F. Stöckmann, "On the Dependence of Photocurrents on the Excitation Strength," *Phys. Stat. Sol.*, **34**, p. 741 (1969).
- <sup>4</sup> K. Heinz and F. Stöckmann, "A Direct Method to Measure Trap Parameters in Photoconductors," *Phys. Stat. Sol.*, **A20**, p. 469 (1973).
- <sup>5</sup> K. W. Boer and H. Vogel, "New Method for the Examination of Systems of Differential Equations in the Reaction Kinetics and Related Fields," *Z. Physik. Chem. (Leipzig)*, **206** (1956): "Part I. Mathematical Part," p. 1, and "Part II. Physical Applications," p. 17.

# On The Concept of "Lifetimes" in Photoconductors

F. Stockmann

Institut für angewandte Physik der Universität Karlsruhe, West Germany

**Abstract**—In this paper the different types of "lifetime," which are necessary to understand the performance of a photoconductor, are defined and their relations to each other are discussed. It proves necessary to distinguish between the steady-state lifetimes of free and trapped carriers, a recombination lifetime, the response time, and several relaxation times. Many other concepts that are being used in the literature are not really necessary.

## 1. Introduction

From the standard books on photoconductivity<sup>1-3</sup> newcomers in this field are usually worried about the many different meanings of "lifetime." The electron and the hole lifetimes  $\tau_n$  and  $\tau_p$ , seem to be simple concepts although  $\tau_n$  and  $\tau_p$  may be different in photoconductors in contrast with common semiconductors. In addition, free lifetime, excited lifetime, pair lifetime, microscopic lifetime, effective lifetime, response time, relaxation time, and others are all used. This paper discusses how much of this terminology is really necessary and how the remaining concepts are related to each other.

For simplicity, in the following discussion the free-carrier concentrations in the illuminated photoconductor are assumed to be large as compared with the equilibrium concentrations in the dark. This is a simplification that constitutes no real restriction for the arguments that follow.

In a photoconductor the difficulties with the various lifetimes arise

from the fact that photogenerated free carriers do not remain free until they recombine. Instead they become trapped and must be released from the traps thermally before they can recombine. Since in most photoconductors there are many more trapped carriers than free ones, the trapped carriers must not be forgotten in any consideration of lifetimes.

## 2. Steady-State Lifetimes

If, in a photoconductor, both free electrons and holes are photogenerated at a generation rate  $G$  ( $\text{cm}^{-3} \text{s}^{-1}$ ), in the steady state they recombine at the same rate:  $\mathcal{R} = G$ . In general, the recombination rate  $\mathcal{R}$  depends on the free-carrier concentrations  $n$  and  $p$  as well as on other parameters,  $\mathcal{R} = \mathcal{R}(n, p, \dots)$ , but details are not important at present.

Then the steady-state lifetimes  $\tau_n$  and  $\tau_p$  of the free electrons and holes, respectively, are defined by

$$\begin{aligned} n &= G\tau_n = \mathcal{R}(n, p, \dots)\tau_n, \\ p &= G\tau_p = \mathcal{R}(n, p, \dots)\tau_p. \end{aligned} \quad [1]$$

Obviously  $\tau_n$  and  $\tau_p$  can be determined experimentally by measuring the generation rate  $G$  and the steady-state concentrations  $n$  and  $p$ . Actually, however, only the majority carrier concentration and lifetime can be evaluated from the photoconductivity alone.

One may also include the trapped carriers,  $n_t$  and  $p_t$ , in such definitions. However, since the total number of excited electrons equals the total number of excited holes ( $n + n_t = p + p_t$ ), only one extra lifetime,  $\tau^*$ , is necessary; this may be termed the steady lifetime of the excited state:

$$(n + n_t) = (p + p_t) = G\tau^* = \mathcal{R}(n, \tau, \dots)\tau^*. \quad [2]$$

From Eqs. [1] and [2] one deduces immediately the very important ratio:

$$(n + n_t = p + p_t) : n : p = 1 : \theta_n : \theta_p = \tau^* : \tau_n : \tau_p. \quad [3]$$

It relates the three steady-state lifetimes  $\tau^*$ ,  $\tau_n$ , and  $\tau_p$  to the trapping factors  $\theta_n$  and  $\theta_p$  defined by

$$\theta_n = \frac{n}{n + n_t}, \quad \theta_p = \frac{p}{p + p_t}. \quad [4]$$

Note that the recombination rate  $\mathcal{R}(n, p, \dots)$  in Eqs. [1] and [2] cancels in the ratio [3]. Thus Eq. [3] applies to any case regardless of the recombination mechanism.

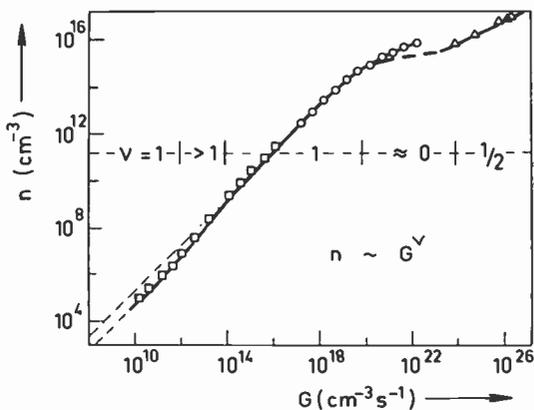
If there are no trapped carriers, it is obviously not necessary to distinguish between the different steady-state lifetimes because then  $\tau^* = \tau_n = \tau_p$ . That is the usual case in semiconductors such as Si and Ge at room temperature. In photoconductors, on the other hand, trapped carriers usually play a dominant role. Furthermore, in general the ratio of free-to-trapped carriers is different for electrons and holes, i.e.,  $\theta_n \neq \theta_p \neq 1$  and hence  $\tau_n \neq \tau_p \neq \tau^*$ . And finally,  $\tau_n$ ,  $\tau_p$ , and  $\tau^*$  are usually not constant; they depend on  $n$  and/or  $p$  because of  $\mathcal{R} = \mathcal{R}(n, p, \dots)$  in Eqs. [1] and [2].

### 3. Recombination Lifetimes

The steady-state lifetimes  $\tau_n$  and  $\tau_p$  of the free carriers are sometimes referred to as their "recombination lifetimes." Originally the two terms were synonymous. The term "recombination lifetime," of course, is intended to indicate that the steady state is a quasi-equilibrium between generation and recombination. However, the relationship between the recombination lifetimes and the recombination mechanism is not trivial. This fact has been overlooked quite frequently, and many doubtful conclusions in the literature arise from it. Therefore one should be careful to apply the term "recombination lifetime" only to those situations where it is physically meaningful—usually either for  $\tau_n$  or for  $\tau_p$ . Sometimes, however, neither  $\tau_n$  nor  $\tau_p$  are closely related to the recombination mechanism.

As an example, assume an n-type photoconductor in which the recombination is due to recombination centers according to the Shockley-Read mechanism. If most of the recombination centers are occupied by electrons, the recombination rate is  $\mathcal{R}(n, p, \dots) = rRp$ , where  $r$  is the recombination coefficient for hole capture and  $R$  the concentration of the centers. Then from Eq. [1] the steady-state lifetime of holes is  $\tau_p = 1/rR$ , which is independent of the generation rate. Hence  $\tau_p$  can be interpreted reasonably as the recombination lifetime of the holes.

On the other hand, from Eq. [3] the steady-state lifetime of the electrons is  $\tau_n = \tau_p \theta_n / \theta_p$ . Since  $\theta_n$  and  $\theta_p$  do depend on the generation rate, whenever the quasi-Fermi-energies  $E_{Fn}$  and  $E_{Fp}$  move across localized energy levels,  $\tau_n$  is not constant. An experimental example has been discussed recently in some detail.<sup>4</sup> For convenience, a typical result is shown in Fig. 1. Because of the nonlinearities in the  $n(G)$  curve, the steady lifetime  $\tau_n = n/G$  of the electrons depends on their concentration  $n$  in a quite curious manner. But that is not related to any peculiarities in the recombination mechanism if the hole lifetime  $\tau_p$  is constant throughout the full regime of  $G$ . In the original



**Fig. 1**—Free-electron concentration  $n$  versus generation rate  $G$  at room temperature for a copper-doped CdS crystal (from Ref. [4]). Within limited regimes,  $n(G)$  may be approximated by a power law:  $n \sim G^v$ . Five different regimes must be distinguished in the present case.

meaning  $\tau_n$  denoted the recombination lifetime of the electrons. It is misleading, however, to use this notation here, because it has no physical significance with respect to the recombination mechanism. Actually the  $\tau_n(G)$  dependence yields information only about the trapping centers.

Frequently the recombination rate  $\mathcal{R}(n, p, \dots)$  depends on powers of  $n$  and  $p$ :  $\mathcal{R} = Cn^i p^j$ . In the preceding example ( $\mathcal{R} = rRp$ ) the special case  $i = 0, j = 1$  has been discussed. If on the other hand most of the recombination centers are occupied by holes,  $\mathcal{R} \sim n$ , i.e.,  $i = 1, j = 0$ . The Shockley-Read mechanism also provides the possibility  $\mathcal{R} \sim n \cdot p$ , i.e.,  $i = 1, j = 1$ , which is more typical for a direct recombination. For an Auger recombination,  $\mathcal{R} \sim n^2 \cdot p$  or  $\mathcal{R} \sim n \cdot p^2$ , etc. The corresponding values of  $\tau^*$  from Eq. [2] are given in Table 1 as examples for the algebraic structure of such expressions.

#### 4. Response Time

In a photoconductor that is not in the steady state, all concentrations

**Table 1**—Lifetimes  $\tau^*$  for recombination rates  $\mathcal{R} = Cn^i p^j$ .

$(i, j)$	(0,1)	(1,0)	(1,1)	(2,0)	(2,1)	(1,2)
$1/\tau^*$	$C\theta_p$	$C\theta_n$	$Cn\theta_p = Cp\theta_n$	$Cn\theta_n$	$Cnp\theta_n$	$Cnp\theta_p$

are functions of the time. Instead of  $G = \mathcal{R}(n, p, \dots)$  one now has

$$\frac{d(n + n_t)}{dt} = \frac{d(p + p_t)}{dt} = G - \mathcal{R}(n, p, \dots). \quad [5]$$

For definiteness assume that the photoconductor has been illuminated long enough to get into the steady state and that the illumination is then switched off suddenly at  $t = t_0$ . The initial response time  $\tau_{resp}$  of an n-type photoconductor is defined by

$$\left. \frac{dn}{dt} \right|_{t_0} = - \frac{n(t_0)}{\tau_{resp}} \quad [6]$$

where  $n(t_0)$  is the steady-state concentration  $n_{st} = G\tau_n$ . By comparison with an exponential decay it is seen that this definition of  $\tau_{resp}$  indicates how fast the photoconductor responds to time-dependent generation rates. When operating a photoconductor with modulated light, the highest frequency  $\omega_{max}$  that can be handled by the photoconductor is  $\omega_{max} \approx 1/\tau_{resp}$ . This is the physics underlying the initial response time given by Eq. [6].

Usually the photocurrent does not decay exponentially. Then  $\tau_{resp}$  does not remain constant during the decay; in general it increases. Nevertheless, in such cases the momentary value  $\tau_{resp}(n)$  of the response time which then is  $n$  dependent, yields useful information.

Eq. [5] is the differential equation of the decay curve with  $G = 0$ . Its initial slope is

$$\left. \frac{d(n + n_t)}{dn} \right|_{t_0} = \left. \frac{d(n + n_t)}{dn} \right|_{st} \cdot \left. \frac{dn}{dt} \right|_{t_0} = - \mathcal{R}_{st} = - \frac{n_{st}}{\tau_n}.$$

Here the subscript  $st$  indicates steady-state conditions. Introducing  $\tau_{resp}$  from Eq. [6], one obtains

$$\left( \frac{d(n + n_t)}{dn} \right)_{st} = \frac{\tau_{resp}}{\tau_n}. \quad [7]$$

The special case in which the steady-state concentration of trapped electrons  $n_t$  is proportional to the free-electron concentration  $n$ , yields

$$\left( \frac{d(n + n_t)}{dn} \right)_{st} = \left( \frac{n + n_t}{n} \right)_{st} = \frac{1}{\theta_n}$$

and hence with Eqs. [1] and [2],

$$\tau_{resp} = \frac{\tau_n}{\theta_n} = \frac{\tau_p}{\theta_p} = \tau^*. \quad [8]$$

This is another important result. The initial response time of a photoconductor equals the steady-state lifetime  $\tau^*$  of the excited state. Hence, because in general  $\theta_n \neq \theta_p \neq 1$ , the response time of a photo-

conductor is not the steady lifetime of the majority carriers.  $\tau_{\text{resp}}$  is longer by the factor  $1/\theta$ , which is the ratio (free + trapped)/free majority carriers.

Eq. [8] is commonly used in the literature. Remember, however, that the more general expression Eq. [7] applies if  $n_t$  is not proportional to  $n$ .

According to Eq. [7], measurements of  $\tau_n$  and  $\tau_{\text{resp}}$  seem to yield straightforward information on the trapped electrons. Actually, however, one must be careful. There are cases in which  $\tau_{\text{resp}}$  cannot be determined reliably from the initial slope of the decay curve. Fast electronic processes within the photoconductor may alter the recombination rate significantly before it can be distinguished from the photocurrent. However, in a slightly more sophisticated approach using a modulated excitation which is sketched in Sec. 5, this difficulty can be eliminated. One can measure  $(dn_t/dn)_{st}$  from Eq. [7] with a minimum of simplifying assumptions.<sup>5</sup>

## 5. Relaxation Times

Thus far, only the response time has been introduced in describing transient photocurrents. In fact, other characteristic times are also involved. The most suitable name for them is "relaxation times." It is the exception rather than the rule when any of the steady-state lifetimes  $\tau_n$  and/or  $\tau_p$  of the free carriers equals a relaxation time. It occurs more frequently for the steady-state lifetime  $\tau^*$  of the excited state, i.e., the response time  $\tau_{\text{resp}}$ .

As an example, consider a photoconductor that contains one class each of electron traps, recombination centers, and hole traps. Then, according to Fig. 2, there are eight electronic transitions (besides the photogeneration  $G$ ) which must be taken into account for the kinetics of this system. As in the preceding section, one characteristic time can be ascribed to each of the eight transition rates. Fortunately, however, these eight times are not independent. Because of detailed balance they are pairwise equal in the equilibrium state. Nevertheless, four relaxation times remain which are different in general.

This can also be seen formally. In Fig. 2 there are five different energy states for electrons. Therefore five rate equations are necessary to describe the kinetics of the system. One integral is the neutrality condition. Thus the complete solution consists of four other independent functions. From this argument it is apparent that each extra class of imperfections introduces one other relaxation time.

Actually not much is known about these relaxation times. There are several reasons. First, usually the system of the five rate equa-

tions is strongly nonlinear. Therefore the solutions are not sums of four exponentials with four significant decay times. Instead, along a decay curve they are functions of the time and hence cannot be evaluated from experiments directly.

One can avoid this difficulty by investigating transient photocurrents that deviate only slightly from the steady state. Then the steady-state values of the relaxation times also apply to the transient processes to a first approximation. A convenient experimental meth-

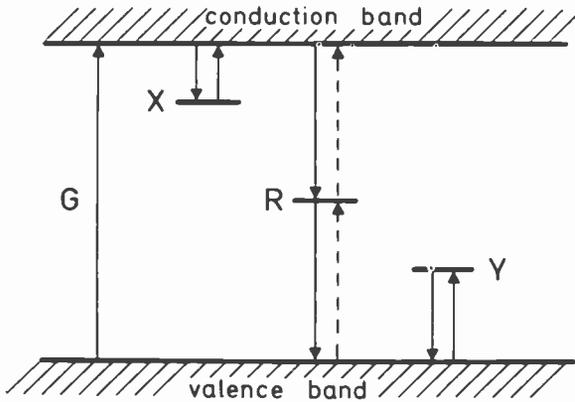


Fig. 2—Electron transitions in a photoconductor that contains one class each of electron traps  $X$ , recombination centers  $R$ , and hole traps  $Y$ .  $G$  denominates the photo-generation.

od is to modulate the generation rate sinusoidally and to measure the ac component of the photocurrent as a function of the modulation frequency. Any relaxation process depends on the frequency as  $a/(1 + i\omega\tau)$ . Thus a log-log plot of the ac photocurrent versus the modulation frequency  $\omega$  has a step for each relaxation time  $\tau_n$ , where  $\omega\tau_n = 1$ , provided that all coefficients  $a_n$  are large enough for each step to be detected. Unfortunately, that is in general not the case. Preliminary results of such measurements are shown in Fig. 3. Three steps can clearly be seen. There are good reasons why a fourth step, which should also be there, cannot be observed. In the present case its coefficient is  $a_4 \approx 0$ . A detailed discussion is beyond the scope of this paper.

For small deviations from a steady state, the mathematics are greatly simplified—the rate equations may be linearized. Hence the solutions are sums of exponentials. But unfortunately the relaxation times, although they are now physically significant, usually are quite

complicated expressions in which the parameters of different electron transitions are mixed. That is another reason why so little is known about the relaxation times.

The physics of the relaxation times, on the other hand, can be explained quite simply, at least qualitatively. In a steady state there are quasi-equilibria between all free and trapped carriers. When the generation rate varies, these quasi-equilibria must be readjusted. This

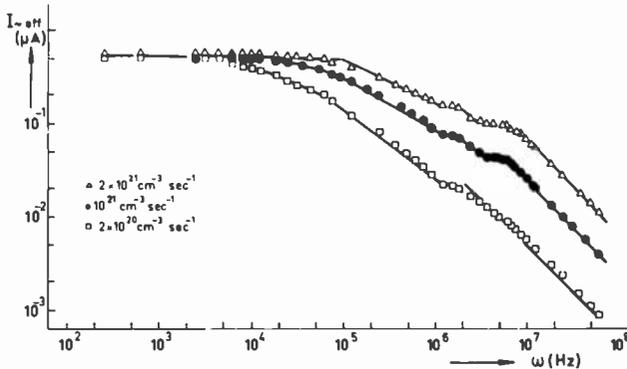


Fig. 3—AC photocurrent versus frequency  $\omega$  in a CdS crystal when the generation rate is modulated sinusoidally with the frequency  $\omega$ :  $G = G_0 + G_1 \exp\{i\omega t\}$  where  $G_1 \ll G_0$  (from Ref. [6]). The three curves belong to three different generation rates  $G_0$ , which are shown in the figure.

requires a time that is usually different for each class of centers. Hence for three classes of centers as in the model of Fig. 2 three readjustment times are involved. For the electron-to-hole ratio, a fourth time is necessary that is essentially the response time if the recombination channel is the bottleneck.

For the performance of a photoconductor, usually only the response time is important which is the largest of the relaxation times. This is a third reason why the relaxation times have not been investigated more thoroughly. On the authors team, some work on this problem is being done. The motivation is to check experimentally, at least once, that the logical arguments are correct (nature is malicious and wise!).

## 6. Summary

For photoconductors, it is necessary in practice to distinguish between several characteristic times: steady-state lifetimes, response

time, and relaxation times. These, however, are straightforward results of the physics of the electronic transitions involved.

### References:

- <sup>1</sup> R. H. Bube, *Photoconductivity of Solids*, John Wiley, New York (1960).
- <sup>2</sup> A. Rose, *Concepts in Photoconductivity and Allied Problems*, Interscience Publ., New York (1963).
- <sup>3</sup> S. M. Rywkin, *Photoelectric Effects in Semiconductors*, Consultants Bureau, New York (1964).
- <sup>4</sup> K. Heinz and F. Stockmann, "A Direct Method to Measure Trap Parameters in Photoconductors," *Phys. Stat. Sol. (A)*, **20**, p. 469 (1973).
- <sup>5</sup> F. Stockmann, to be published.
- <sup>6</sup> D. Stremme, Ph.D. Thesis, Karlsruhe (1972).

# Oxygen Chemisorption and Photodesorption Processes on ZnO Surfaces\*

D. Eger, Y. Goldstein and A. Many

The Racah Institute of Physics, The Hebrew University of Jerusalem,  
91000 Israel.

**Abstract**—Very strong accumulation layers with excess surface-electron densities of  $\Delta N = 5 \times 10^{13} \text{ cm}^{-2}$  and higher have been obtained on single-crystal ZnO surfaces by photodesorption of oxygen. The kinetics of oxygen chemisorption under such extreme accumulation conditions has been studied as a function of oxygen pressure, excess surface-electron density  $\Delta N$ , and temperature. The chemisorption rate is extremely slow ( $\lesssim 10^{13} \text{ cm}^{-2} \text{ sec}^{-1}$  at atmospheric pressure and room temperature) and is directly proportional to the oxygen pressure. It is thermally activated with an activation energy of about 0.25 eV. Both chemisorption and photodesorption rates vary *exponentially* with  $\Delta N$  (an Elovich-type relation). This is most surprising in view of the fact that an accumulation layer is present at the surface throughout the adsorption range studied. Possible models to account for the observed adsorption characteristics are discussed.

## 1. Introduction

In the course of studies of ZnO surfaces we succeeded in producing accumulation layers with excess surface-electron densities of  $5 \times 10^{13} \text{ cm}^{-2}$  or higher. To our knowledge such phenomenally strong accumulation (or inversion) layers have not yet been reported for any semiconductor. Quantization effects<sup>1</sup> in these layers are very pronounced and are presently under investigation in our laboratory. In

\* This work was supported in part by The Israel National Commission for Basic Research of The Israel National Academy of Sciences and Humanities.

this paper we present one aspect of this work, namely, the kinetics of oxygen chemisorption and photodesorption under conditions of extreme accumulation. The results obtained shed light on the mechanism involved in the charge-transfer process between adsorbate and adsorbent, and provide values for the characteristic cross sections. The main objective of the investigation was, however, to establish optimal procedures for producing and maintaining strong accumulation layers on the ZnO surface. More important still, our aim has been to obtain a surface that is easily accessible to the variety of experiments<sup>1</sup> one would like to perform on the two-dimensional electron gas system associated with the surface layer. For this reason, all the measurements reported herein were carried out on "real," etched surfaces maintained in ordinary vacuum ( $10^{-5}$  Torr), rather than on clean and well-controlled surfaces as in ultra-high vacuum systems. Results on the latter type of surfaces, which are necessarily more limited in scope, will be reported elsewhere.

ZnO belongs to the hexagonal wurtzite structure.<sup>2</sup> The Zn-O bond is partly ionic and partly covalent, the effective charge of each ion being estimated as about one electronic charge.<sup>3</sup> Each cation ( $\text{Zn}^+$ ) is surrounded by four anions ( $\text{O}^-$ ) situated at the corners of a nearly regular tetrahedron, and each anion is similarly surrounded by four cations. The structure can be described as a series of atomic double layers stacked normal to the *c*-axis. The series terminates on both sides with two *different* {0001} surfaces, one composed of zinc ions, the other of oxygen ions. The two different surfaces can easily be distinguished by suitable etching procedures.<sup>4,5</sup> The prism surfaces (parallel to the *c*-axis) contain an equal number of cations and anions.

By suitable growth and doping procedures, ZnO crystals can be made either insulating ( $\geq 10^{10}$  ohm-cm) or n-type semiconducting (typically 1–100 ohm-cm). So far, no p-type samples could be produced. The forbidden energy gap is 3.2 eV. In relatively pure samples the room-temperature electron mobility ranges between 100 and 300  $\text{cm}^2/\text{V sec}$ .

Oxygen adsorption on ZnO surfaces has been studied by several workers.<sup>6–10</sup> It is now well established that oxygen is chemisorbed by capture of conduction electrons from the underlying semiconductor. Such trapping very likely effects a transition from physical to chemical adsorption. This process shifts the energy bands at the surface upwards and can be detected by a drop in surface conductance and an increase in work function. Desorption of chemisorbed oxygen by thermal processes is extremely slow and cannot be detected even at temperatures of 200°–300°C. However, desorption takes place readily

under illumination in vacuum with bandgap light (near-UV). It appears that photodesorption proceeds by the recombination of photo-generated holes with electrons trapped at the chemisorbed oxygen.<sup>11</sup> In this process the oxygen is again converted into a physically adsorbed species, which can then be easily pumped off. Desorption can also be obtained by exposure of the ZnO surface to atomic hydrogen,<sup>3</sup> which probably removes the chemisorbed oxygen by reduction.

The chemisorption kinetics on single-crystal ZnO samples has been studied by Van Hove and Luyckx<sup>9</sup> and by Arijs and co-workers.<sup>10</sup> The former group observed an Elovich-type relation<sup>12</sup> for the chemisorption process, i.e., surface conductance decreasing logarithmically with exposure time to oxygen. A behavior of this sort is expected when a depletion layer is formed at the surface during the chemisorption process, and indeed this is how Medved<sup>8</sup> explained similar results he obtained on ZnO-powder samples. Arijs et al.,<sup>10</sup> on the other hand, find that the conductance decay is inversely proportional to exposure time to oxygen. In this case the surface was characterized by relatively weak accumulation layers (less than  $10^{11}$  cm<sup>-2</sup> excess electrons at the surface) and the results could be accounted for by the standard rate equations applicable to this regime. The experimental results of both groups show that the process of oxygen adsorption on ZnO surfaces is very slow. This was found to be the case by Heiland<sup>3</sup> even for clean ZnO surfaces produced by cleavage in ultra-high vacuum.

Our preliminary measurements indicate that the adsorption characteristics of the three types of surfaces (Zn, O, and prism) are different, but the differences do not appear to be drastic. The results presented here pertain to the oxygen surface only. They cover the variation of the chemisorption rate with oxygen pressure, excess surface-electron density ( $\Delta N$ ), and temperature. The chemisorption rate, which is proportional to pressure, is found to be thermally activated. A rather surprising feature of the data is the exponential dependence of both the chemisorption and photodesorption rates on  $\Delta N$ . The cross section for chemisorption is found to be extremely small, smaller than that estimated on the basis of the data of Arijs et al.,<sup>10</sup> and of Heiland.<sup>3</sup> Possible mechanisms for explaining these observed characteristics of the adsorption kinetics are discussed.

## 2. Experimental

The measurements were carried out on single crystals of ZnO grown by Airtron Co. The bulk resistivity was about 20 ohm-cm. The samples were rectangular parallelepipeds with typical dimensions of  $8 \times$

$1.5 \times 0.5 \text{ mm}^3$ , having their *c*-axis perpendicular to the two larger surfaces (the polar surfaces). After cutting, the samples were polished to a flatness of about  $1 \mu\text{m}$  and etched in a 30% solution of HCl. (Other etches, such as  $\text{HNO}_3$  and  $\text{H}_3\text{PO}_4$ , gave similar results as far as the adsorption characteristics are concerned.) After etching the oxygen surface became matt, while the zinc surface became quite shiny. All measurements were performed on the oxygen surface. However, the other faces were not masked, and it is possible that they too contributed somewhat to the results.

The samples were mounted in a small chamber provided with a quartz window. The pressure in the chamber could be varied between  $10^{-5}$  Torr and atmospheric pressure. The pumping system was a conventional one, consisting of an oil-diffusion pump with a liquid-nitrogen trap to prevent back-streaming of oil. Oxygen desorption was produced by focusing light from a 200-watt mercury arc lamp on the surface under study. The pressure in the chamber during this illumination was  $\sim 10^{-5}$  Torr. For adsorption measurements above room temperature, the chamber was immersed in a constant-temperature bath, where the temperatures could be maintained constant to within  $1^\circ\text{C}$ . For measurements below room temperature a suitable cryostat was used. In either case photodesorption was produced by illumination (in vacuum) at room temperature.

The experimental quantity measured was the sample resistance as a function of time under different surface conditions. The electrical contacts were prepared by first wetting the sample with indium amalgam and then using indium solder. A constant-current source was connected to the two end contacts of the sample and a pair of potential probes, soldered in the mid-region of the sample, were employed to determine the sample's resistivity. This procedure eliminates contact effects, which are quite severe in the case of ZnO. A recorder was used to monitor the time variation of the voltage drop  $V_p$  across the pair of probes,  $V_p$  being directly proportional to the sample resistivity. Typical plots of  $V_p$  as a function of exposure time to oxygen (at  $300^\circ\text{K}$ ) are shown in Fig. 1 for different oxygen pressures. The measurements were carried out as follows. The sample was first illuminated for about two hours in a vacuum of  $10^{-5}$  Torr. After illumination, the sample was kept in the dark (in vacuum) for about 15 minutes to allow for the decay of any photoconductance and/or thermal gradients present. After this, at time  $t = 0$ , oxygen was introduced at a certain pressure, and  $V_p$  was recorded as a function of exposure time. The various curves in the figure correspond to different oxygen pressures, as marked.

On the assumption that in the process of chemisorption, each oxy-

gen molecule (or atom) captures one conduction electron from the space charge-region, one has

$$\frac{dN_a}{dt} = - \frac{d(\Delta N)}{dt}, \quad [1]$$

where  $N_a$  is the surface density of chemisorbed oxygen species and  $\Delta N$  is the number of excess electrons (with respect to flat bands conditions) per unit surface area in the space-charge region.<sup>13</sup>  $\Delta N$  can be obtained from the surface conductance  $\Delta\sigma$ ;

$$\Delta N = \frac{\Delta\sigma}{q\mu_s}, \quad [2]$$

which in turn can be calculated from the measured sample resistance;

$$\Delta\sigma = \frac{l}{w} \left( \frac{1}{R} - \frac{1}{R_b} \right). \quad [3]$$

Here  $q$  is the absolute value of the electronic charge and  $\mu_s$  the electron surface mobility (which is usually a function of  $\Delta N$ );  $w$  is the width of the sample and  $l$  the separation between the potential probes;  $R$  is the sample resistance between the probes, and  $R_b$  the corresponding bulk resistance. As a good approximation for  $R_b$  we took the maximum value of  $R$  obtained after long exposure (in the dark) to an atmospheric pressure of oxygen.

In order to derive  $\Delta N$  from  $\Delta\sigma$  (Eq. [2]) one has to know the surface mobility  $\mu_s$  as a function of  $\Delta N$ . The surface mobility is expected to be smaller than the bulk mobility  $\mu_b$  because the carriers in the accumulation layer are scattered by the surface in addition to the normal bulk scattering. Preliminary results obtained by field-effect and galvanomagnetic measurements show that in the  $\Delta N$  range studied here,  $\mu_s$  usually *increases* with increasing  $\Delta N$ , approaching the value of  $\mu_b$  in strong accumulation conditions. While this rather unusual behavior is typical of most of the samples studied, the magnitudes of  $\mu_s$  vary from sample to sample, being very sensitive to surface treatment. Unfortunately, no surface-mobility measurements were carried out on the samples used in the present adsorption measurements, so that  $\Delta N$  cannot be derived accurately. Accordingly, all adsorption data must necessarily involve the directly measurable surface conductance  $\Delta\sigma$ . However, to obtain a feel for the surface-electron densities involved, it would be useful to define the quantity  $\Delta N'$  as

$$\Delta N' \equiv \frac{\Delta \sigma}{q \mu_b} = \frac{\mu_s}{\mu_b} \Delta N. \quad [4]$$

The bulk mobility in our samples was  $150 \text{ cm}^2/\text{Vsec}$ . All the adsorption measurements presented below are expressed in terms of  $\Delta N'$ .

While  $\Delta N'$  may differ from  $\Delta N$  (but by not more than a small factor), it serves as the lower bound for  $\Delta N$ . The time derivative of  $\Delta N$ , which is a direct measure of the chemisorption rate (see Eq. [1]), can readily be expressed in terms of the time derivative of  $\Delta N'$ ;

$$\frac{d(\Delta N)}{dt} = \left[ \frac{\mu_b}{\mu_s} + \Delta N' \frac{d\left(\frac{\mu_b}{\mu_s}\right)}{d(\Delta N')} \right] \frac{d(\Delta N')}{dt}. \quad [5]$$

It should be noted that as long as  $\Delta N$  (or  $\Delta N'$ ) is maintained con-

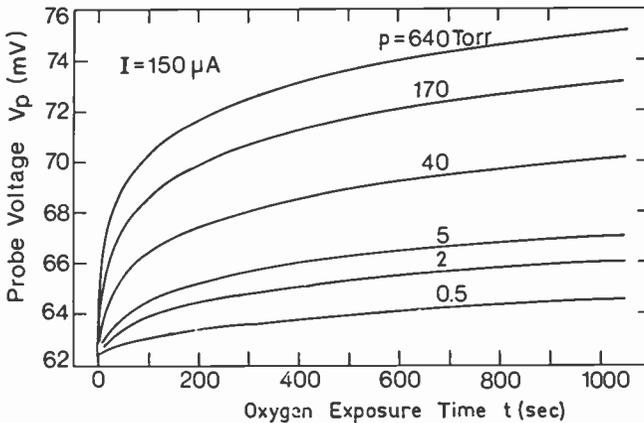


Fig. 1—Voltage drop  $V_p$  across pair of potential probes on ZnO sample as function of exposure time to oxygen for various oxygen pressures. Measurements performed at room temperature. Current through the sample  $150 \mu\text{A}$ . Bulk resistivity  $20 \text{ ohm-cm}$ , sample dimensions: width,  $1.5 \text{ mm}$ ; thickness,  $0.5 \text{ mm}$ ; distance between probes,  $3 \text{ mm}$ .

stant,  $d(\Delta N)/dt$  is directly proportional to  $d(\Delta N')/dt$ . This is made use of in the measurements of the pressure and temperature dependences of the chemisorption rate.

### 3. Results

The data in Fig. 1 are represented in a different form in Fig. 2. Here,  $\Delta N'$  (as obtained by the use of Eqs. [3] and [4]) is plotted as a func-

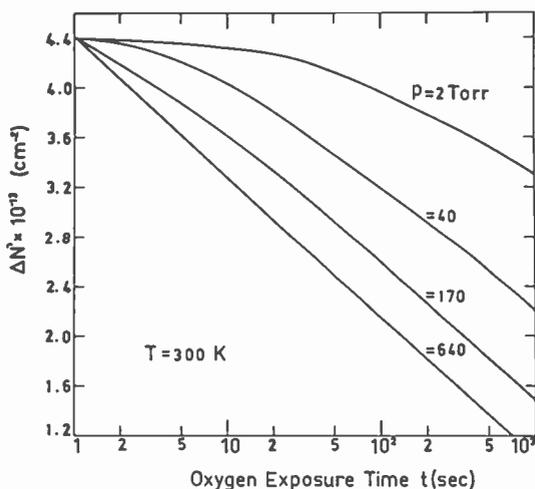


Fig. 2— $\Delta N'$  as defined by Eq. [4] and derived from the data of Fig. 1, plotted on semi-log scale as function of time.

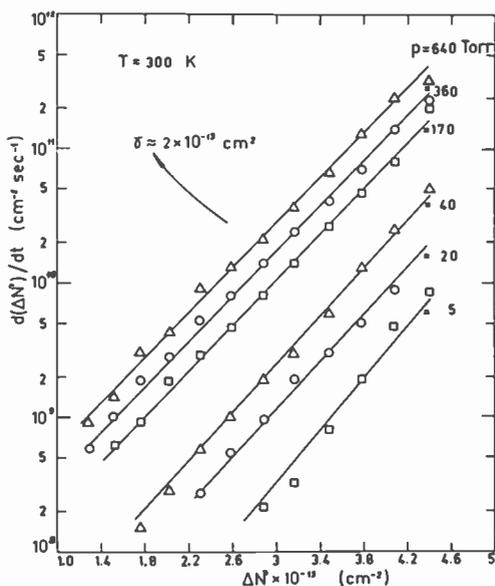


Fig. 3—Semi-log plot of rate of change of  $\Delta N'$  during chemisorption as function of  $\Delta N'$  for different oxygen pressures. The lines for the different pressures are seen to be approximately parallel with a common slope of  $\gamma = 2 \times 10^{-13} \text{ cm}^2$ .

tion of the logarithm of the exposure time to oxygen for different oxygen pressures. It is seen that at all pressures,  $\Delta N'$  decreases more or less logarithmically with exposure time, corresponding to an Elovich-type relation.<sup>12</sup> The nature of the chemisorption kinetics is, however, seen much more clearly if one plots the rate of change of  $\Delta N'$  against  $\Delta N'$ . Such plots are shown (on a semi-log scale) in Fig. 3, again for different ambient pressures. All lines are seen to be straight and parallel, indicating that at each pressure,  $d(\Delta N')/dt$  decreases exponentially with decreasing  $\Delta N'$ , and that for any fixed value of  $\Delta N'$ ,  $d(\Delta N')/dt$  is proportional to the oxygen pressure. The latter characteristic is shown in more detail by Fig. 4 where  $d(\Delta N')/dt$  is plotted

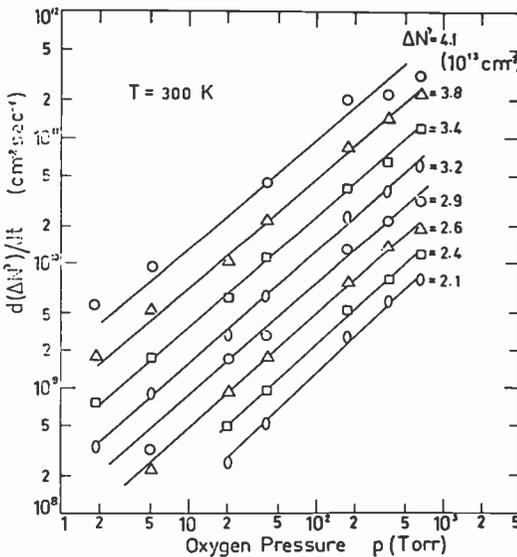


Fig. 4—Log-log plots of rate of change of  $\Delta N'$  during chemisorption as function of oxygen pressure, for different fixed values of  $\Delta N'$ . The lines for the different values of  $\Delta N'$  are essentially parallel, with slope of unity, showing that the rate of chemisorption is proportional to pressure.

on a log-log scale against oxygen pressure for different fixed values of  $\Delta N'$ . The common slope of about unity characterizing the various straight lines indicates direct proportionality between  $d(\Delta N')/dt$  and pressure in the range of  $\Delta N'$  shown ( $2-4 \times 10^{13}$  cm<sup>-2</sup>). Such proportionality is maintained from about 1 Torr to atmospheric pressure. Below 1 Torr the chemisorption rate becomes too slow to permit meaningful measurements. As was pointed out above, for  $\Delta N'$  constant,  $d(\Delta N')/dt$  is proportional to  $d(\Delta N)/dt$ . Hence one concludes

from Fig. 4 that the chemisorption rate ( $-d(\Delta N)/dt$ ) is also directly proportional to pressure. No such clear-cut conclusion can be arrived at concerning the dependence of the chemisorption rate on  $\Delta N$  because of the lack of information on the variation of the surface mobility with  $\Delta N$  (see Eq. [5]). However, since  $\mu_s$  very likely increases with  $\Delta N$ , the chemisorption rate is expected to increase even more rapidly with  $\Delta N$  than indicated by the exponential relation between  $d(\Delta N')/dt$  and  $\Delta N'$  (Fig. 3).

The room-temperature data of Figs. 3 and 4 can be summarized by the following empirical relation

$$\frac{d(\Delta N')}{dt} = -K \left( \frac{p}{p_0} \right) \exp \{ \gamma \Delta N' \}, \quad [6]$$

where  $p_0$  is atmospheric pressure. The value of  $K$  was found to vary considerably from sample to sample and, as shown below, for a given sample it is temperature dependent. The value of  $\gamma$ , on the other hand, is essentially the same for all samples studied (lying in the range  $2-3 \times 10^{-13} \text{ cm}^2$ ) and is temperature independent.

It should be noted that the chemisorption process is an extremely inefficient one. Even the maximum chemisorption rate (at atmospheric pressure, during the initial stages of chemisorption) is of the order of  $10^{12} \text{ cm}^{-2} \text{ sec}^{-1}$  or less, a very slow rate indeed. At atmospheric pressure it takes on the order of an hour before an appreciable part of the accumulation layer is removed. It is this highly inefficient chemisorption process that makes the attainment of strong accumulation layers possible at moderate vacuum.

In order to compare the chemisorption rates at different temperatures, some preliminary studies were necessary. These consisted of measurements of the temperature dependence of the sample conductivity  $\sigma$  for different fixed values of  $\Delta N'$ . All such measurements were carried out at a pressure of  $10^{-5}$  Torr, at which chemisorption is negligible, so that each chosen value of  $\Delta N'$  remains constant as the temperature is varied. (The constancy of  $\Delta N'$  was checked by ascertaining that upon cooling to room temperature, the sample conductance regained its original value.) The temperature dependence of  $\sigma$  is illustrated in Fig. 5 for two values of  $\Delta N'$ ,  $\Delta N' = 0$  (bulk conductivity) and  $\Delta N' = 3.3 \times 10^{13} \text{ cm}^{-2}$ . Inspection of the figure shows that the *ratio* of the conductivities is independent of temperature. This feature has been verified by measurements at several other values of  $\Delta N'$  (not shown), and one concludes that the conductivity ratio is a function of  $\Delta N'$  only. Hall-effect measurements show that the electron bulk concentration in the temperature range studied is constant

(the donor impurities are completely ionized). It therefore follows that for any given value of  $\Delta N'$ , the ratio of surface to bulk mobilities is temperature independent. This rather surprising result simplifies the analysis of chemisorption data obtained at different temperatures, as will be discussed below.

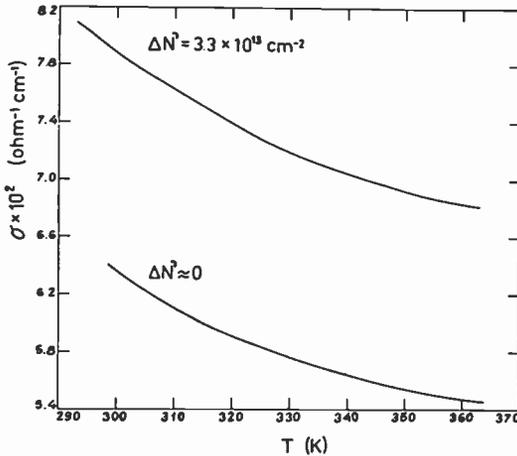


Fig. 5—Temperature dependence of conductivity of ZnO sample for  $\Delta N' = 0$  (bulk conductivity) and  $\Delta N' = 3.3 \times 10^{13} \text{ cm}^{-2}$ . The ratio of the conductivities at the different values of  $\Delta N'$  is seen to be temperature independent.

In Fig. 6, measured values of  $d(\Delta N')/dt$  are plotted against  $\Delta N'$  for four different temperatures. Since the chemisorption rate increases quite rapidly with temperature, the measurements were carried out at a relatively low pressure (4 Torr). It is seen that the curves for the different temperatures are linear and parallel. Thus in the temperature range indicated, the exponential dependence of  $d(\Delta N')/dt$  on  $\Delta N'$  is characterized by the *same* slope. In other words, the value of  $\gamma$  (see Eq. [6]) is temperature independent in this range. It is only the proportionality factor  $K$  that is temperature dependent. This is seen more clearly in Fig. 7, which shows semi-log plots of  $d(\Delta N')/dt$  versus  $1000/T$  for three values of  $\Delta N'$ . The three curves are linear and parallel, being characterized by the *same* activation energy of  $E = 0.25 \text{ eV}$ . It should be noted that for each plot in Fig. 7, the chemisorption rate is *directly proportional* to  $d(\Delta N')/dt$ . This follows from the observation discussed above that for any fixed value of  $\Delta N'$ ,  $\mu_s/\mu_b$  is temperature independent. Thus, the value of  $E = 0.25 \text{ eV}$  actually represents the activation energy for chemisorption.

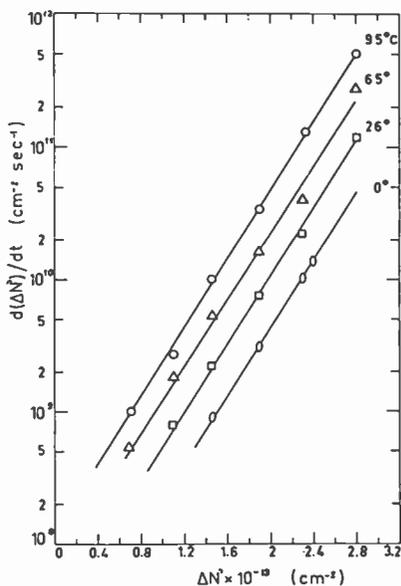


Fig. 6—Semi-log plot of rate of change of  $\Delta N'$  during chemisorption as function of  $\Delta N'$  for four different temperatures. Measurements performed at oxygen pressure of 4 Torr. The parallelism of the lines shows that the exponential coefficient  $\gamma$  is temperature independent.

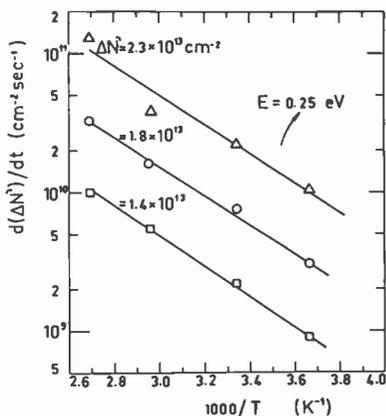
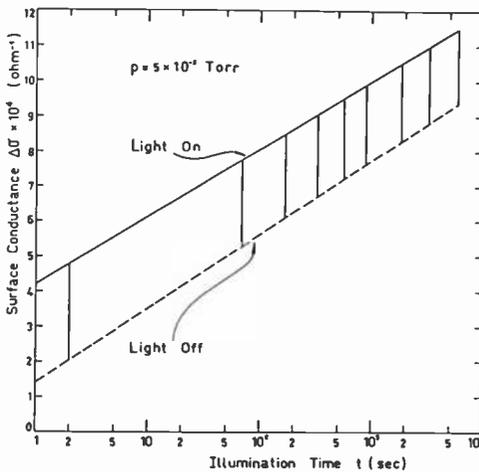


Fig. 7—Semi-log plot of rate of change of  $\Delta N'$  during chemisorption as function of  $1000/T$  for different fixed values of  $\Delta N'$ . Measurements performed at oxygen pressure of 4 Torr. The linearity and parallelism of the lines show that the rate of chemisorption is thermally activated with a  $\Delta N$ -independent activation energy  $E$ .

Including the temperature dependence of the chemisorption rate in Eq. [6] we obtain the empirical relation

$$\frac{d(\Delta N')}{dt} = -K_0 \left( \frac{p}{p_0} \right) \exp \left\{ -\frac{E}{kT} \right\} \exp \{ \gamma \Delta N' \}, \quad [7]$$

where  $K_0 = 10^{11} - 10^{14} \text{ cm}^{-2} \text{ sec}^{-1}$ ,  $E \approx 0.25 \text{ eV}$ ,  $\gamma = 2-3 \times 10^{-13} \text{ cm}^2$  and  $k$  is Boltzmann's constant.  $K_0$  is a constant for a given sample, but may vary appreciably from sample to sample. This is illustrated by the results of Figs. 3 and 6, obtained for two different samples. Comparing the line for 5 Torr in Fig. 3 with the line for  $26^\circ\text{C}$  in Fig. 6 (obtained at 4 Torr), we see that the value of  $K_0$  for the latter sample is higher by about 3 orders of magnitude. The value of  $\gamma$ , on the other hand, is relatively constant.



**Fig. 8**—Semi-log plot of surface conductance during photodesorption as function of illumination time. Upper (solid) line represents measurements during illumination and includes changes of  $\Delta\sigma$  due to both photoconductivity and photodesorption. Vertical lines show drop in  $\Delta\sigma$  upon interrupting the illumination. Lower (dashed) line represents changes in  $\Delta\sigma$  due to photodesorption only.

We turn now to measurements of oxygen photodesorption. These measurements were carried out at room temperature only. After the sample had been exposed to atmospheric pressure for some time, the chamber was evacuated and the rise in surface conductance with illumination time was monitored. Upon illumination, the conductance shows a fast initial rise followed by a much slower increase. Turning off the light (in vacuum) results in a corresponding fast drop in con-

ductance, after which the conductance stays nearly constant. The fast initial conductance rise is associated with photoconductivity processes; only the slower increase represents the oxygen-desorption process. In order to separate the two components, the illumination was interrupted at regular intervals for a short time to allow for the dark conductance to be recorded. Fig. 8 shows the variation of  $\Delta\sigma$  with illumination time obtained in this manner. The upper (solid) curve represents the conductance during illumination. The vertical lines

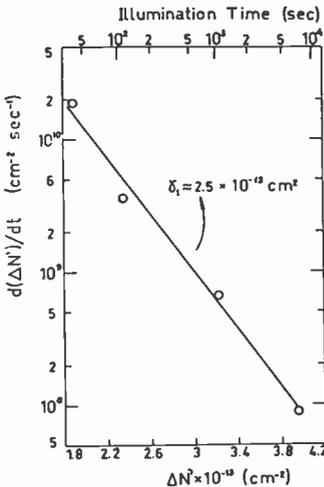


Fig. 9—Semi-log plot of rate of change of  $\Delta N'$  during photodesorption as function of  $\Delta N'$ . Values of the illumination time are also marked at the top of the figure. The exponential coefficient  $\gamma_1$  obtained from slope of straight lines is  $\gamma_1 = 2.5 \times 10^{-13} \text{ cm}^2$ .

show the drop in  $\Delta\sigma$  upon interrupting the illumination. The lower (dashed) curve thus represents the photodesorption kinetics without the complications introduced by the photoconductivity. Actually, however, the difference between the two curves remains fairly constant throughout the illumination so that either curve can be used in the analysis of the photodesorption process.

The variation of  $d(\Delta N')/dt$  with  $\Delta N'$ , as derived from the logarithmic curves of Fig. 8, is shown by the semi-log plot of Fig. 9. The straight line means that  $d(\Delta N')/dt$  decreases exponentially with increasing  $\Delta N'$ :

$$\frac{d(\Delta N')}{dt} = K_1 \exp\{-\gamma_1 \Delta N'\}. \tag{8}$$

where  $K_1 \approx 10^{13} \text{ cm}^{-2} \text{ sec}^{-1}$  and  $\gamma_1 \approx 2.5 \times 10^{-13} \text{ cm}^2$ . The value of  $\gamma_1$  is seen to be very close to that of  $\gamma$  characterizing the chemisorption process (see Eq. [7]). The magnitude of  $K_1$  depends of course on the light intensity, and the value quoted is typical for fairly strong illumination.

#### 4. Discussion

The chemisorption kinetics associated with the relaxation in surface conductance will be analyzed on the premise that oxygen adsorption introduces acceptor-like surface states. Physically adsorbed molecules are identified with unoccupied states, chemisorbed molecules with occupied states. Actually, the term physically adsorbed molecules used here has a broader sense. It represents any neutral species which, on the one hand, can easily be pumped off and, on the other hand, is able to capture a conduction electron and thereby become strongly bonded to the surface (chemisorbed). Such species may be associated with neutral oxygen molecules or atoms adsorbed in either ordinary or special lattice sites; they may well be more tightly bound to the surface than ordinary physisorbed molecules. The only requirement is that the physisorbed species are in good communication with the ambient gas, so that at any given pressure  $p$  their density  $N_p$  at the surface is maintained constant. The rate equation for chemisorption can then be written as

$$\frac{dN_a}{dt} = - \frac{d(\Delta N)}{dt} = A_n \Phi N_p, \quad [9]$$

where  $N_a$  is the density of chemisorbed species,  $A_n$  is the electron capture cross section for chemisorption, and  $\Phi$  is the electron flux, i.e., the number of electrons striking a unit area of the surface per unit time. It is reasonable to expect that the density  $N_p$  of physisorbed species capable of electron capture, and hence the chemisorption rate  $dN_a/dt$ , will be proportional to the ambient pressure. This behavior is exhibited by our data as well as by those of other workers.<sup>8-10</sup> Eq. [9] should include also a term representing the rate of thermal emission back into the conduction band of electrons trapped at the chemisorbed oxygen. As pointed out above, however, there is no noticeable thermal desorption of chemisorbed oxygen at all temperatures studied, so that this term can be neglected. The electron flux is usually expressed as

$$\Phi = v_s n_s, \quad [10]$$

where  $n_s$  is the volume density of conduction electrons at the surface and  $v_e$  their mean velocity.

Oxygen adsorption measurements carried out by Morrison,<sup>6</sup> Melnick,<sup>7</sup> and by Medved<sup>8</sup> on powdered ZnO samples yielded an Elovich-type relation for the chemisorption process, i.e., surface conductance decreasing logarithmically with exposure time to oxygen. A similar relation has been observed by Van Hove and Luyckx<sup>9</sup> on single-crystal samples. A behavior of this sort is expected<sup>8</sup> when a depletion layer is formed during the chemisorption process. The barrier height  $V_s$  then becomes increasingly more negative, so that  $n_s$  and hence the chemisorption rate  $dN_a/dt$  (Eqs. [9] and [10]) decrease very rapidly with exposure time. It can be shown<sup>13</sup> that in this case  $n_s$  is given approximately by

$$n_s \propto \exp[-B(\Delta N)^2], \quad [11]$$

where  $B$  is a constant. An approximate solution of Eq. [9], using Eqs. [10] and [11], yields

$$\sigma \propto \Delta N \propto [\ln(t + t_0)]^{1/2}. \quad [12]$$

This is close to the behavior observed experimentally by the above workers.<sup>6-9</sup> Unfortunately, however, the unknown surface-to-volume ratio in the case of the powdered samples<sup>6-8</sup> and insufficient details in the data on single crystals<sup>9</sup> make it very difficult to determine whether a depletion layer was indeed present.

Arijs et al.,<sup>10</sup> on the other hand, find that the conductance decay is inversely proportional to exposure time to oxygen. In this case the surface was characterized by an accumulation layer (the variation in  $\Delta N$  during the chemisorption process being typically  $10^{11}$ – $10^7$  cm<sup>-2</sup>). The surface electron concentration  $n_s$  is then<sup>13</sup> approximately proportional to  $(\Delta N)^2$ , and the rate equation (Eq. [9]) reduces to

$$\frac{dN_a}{dt} = -\frac{d(\Delta N)}{dt} \propto (\Delta N)^2. \quad [13]$$

The solution of this equation is  $\Delta\sigma \propto \Delta N \propto 1/t$ , which is just the relation observed by Arijs et al.<sup>10</sup> These workers did not derive the electron capture cross section  $A_n$  characterizing the chemisorption process. A rough estimate based on their published data yields the order of magnitude of  $10^{-26}$  cm<sup>2</sup> for  $A_n$  if one assumes a monolayer of physisorbed oxygen at atmospheric pressure ( $N_p \approx 10^{15}$  cm<sup>-2</sup>). This value of  $A_n$  is much smaller than an atomic cross section,  $10^{-16}$ – $10^{-15}$  cm<sup>2</sup>, defined as the area of a physisorbed oxygen molecule or atom. It

should be pointed out that the results presented by Arijs et al.<sup>10</sup> should be viewed with some caution. The conductance measurements were carried out without the use of potential probes. Contact effects, which are very pronounced in the case of ZnO, especially when in the insulating form, could very well have distorted their data.

A completely different behavior is exhibited by our data, which also involve accumulation layers, even though of much greater strength. The chemisorption process in this case does not follow the simple rate equation expected for accumulation layers (Eq. [13]), the chemisorption rate increasing *exponentially* with  $\Delta N$  rather than as  $(\Delta N)^2$ . Under these conditions, a straightforward assignment of a cross section to the chemisorption process is not very meaningful, and a different approach to the problem is discussed below. Nevertheless, in order to get an idea of the orders of magnitude involved, we used Eqs. [9] and [10] to derive from the data the cross sections at the two limits of the variation in  $\Delta N$ . The cross sections obtained in this manner vary from  $\sim 10^{-28}$  cm<sup>2</sup> for  $\Delta N' = 5 \times 10^{13}$  cm<sup>-2</sup> to  $\sim 10^{-31}$  cm<sup>2</sup> for  $\Delta N' = 1 \times 10^{13}$  cm<sup>-2</sup>. These values are even smaller than those estimated from the data of Arijs et al.<sup>10</sup>

A surprising feature of the ZnO surface is that even in the clean state the chemisorption rate appears to be extremely slow. Heiland<sup>3</sup> has studied the polar surfaces of ZnO produced by cleavage in ultra-high vacuum at 90°K. After cleavage, no surface conductivity was detected on either the Zn or O surface. Transient heating (in vacuum) to room temperature produced no change on the O surface but gave rise to a high surface conductance on the Zn surface. Such conductance corresponds to a very strong accumulation layer ( $\Delta N \approx 10^{13}$  cm<sup>-2</sup>). Admission of oxygen was found to remove the high surface conductance of the Zn surface but the chemisorption rate was very slow. No detailed results of the chemisorption kinetics are given by Heiland<sup>3</sup> but a rough estimate indicates that the electron capture cross section is of the order of  $10^{-26}$ – $10^{-25}$  cm<sup>2</sup>, comparable to that derived from the data of Arijs et al.<sup>10</sup> Both the O and Zn surfaces develop strong accumulation layers by exposure to *atomic* hydrogen.<sup>3</sup> It is not clear whether the hydrogen atoms are adsorbed as donors or whether they remove chemisorbed oxygen, leaving excess Zn donors and thus a positive layer at the surface.

The adsorption-desorption characteristics of the ZnO/O system pose two fundamental questions: (a) what is the origin of the huge accumulation layer on the oxygen-free surface, and (b) what is the mechanism of charge transfer responsible for the extremely slow capture rate of electrons by the physisorbed oxygen, and for the exponential dependence of the chemisorption and photodesorption rates

on  $\Delta N$  observed under strong accumulation conditions. As to the first question, it has been suggested by Arijs et al.<sup>10</sup> and Heiland<sup>3</sup> that the accumulation layer is produced by donor impurities that are either introduced by etching or have diffused to the surface from the underlying bulk. The attainment of huge accumulation layers in the present studies, with close to  $10^{14}$  free electrons per  $\text{cm}^2$ , would then mean that there is an equally large density of ionized donors at the surface. This corresponds roughly to a tenth of a monolayer, which is not very likely. It is more probable that the positive surface charge is associated with positive Zn ions. We recall that ZnO is partly ionic, so that the double layer consists of positively and negatively charged Zn and O ions. If the double layers at the Zn and O surface are completely stoichiometric, one might expect a net positive charge on the double layer terminated by the Zn surface, while the double layer terminated by the O surface could be negatively charged. Thus, on surfaces obtained by cleavage in ultra-high vacuum, a compensating accumulation layer would be likely to form on the Zn surface and not on the oxygen surface. This appears to be the situation in the measurements of Heiland,<sup>3</sup> at least when the cleaved surfaces are warmed from 90°K (at which cleavage was carried out) to room temperature. As to the oxygen surface, an accumulation layer was produced only after exposure to atomic hydrogen. Very likely, lattice oxygen is removed by the reducing action of hydrogen, leaving a positive excess of Zn ions. The effect of illumination with bandgap light is equivalent: photogenerated holes recombine with electrons in oxygen ions at the surface.<sup>11</sup> The neutral oxygen so produced is easily pumped off. The photogenerated electrons constitute the accumulation layer produced in this manner. In principle, these electrons could be trapped in surface or bulk traps and thus become immobile. Perhaps one of the distinctive features of ZnO is that electron trapping is negligible, as evidenced by the huge accumulation layers that can actually be attained. Upon admission of oxygen the accumulation layers on both the Zn and O surface are removed, their electrons being used up in the chemisorption process.

The second question posed by the available data is the reason for the extremely inefficient process of charge transfer characterizing both clean surfaces produced by cleavage in ultra-high vacuum<sup>3</sup> and real surfaces prepared by etching. If we proceed on the assumption that the value of the electron capture cross section should be of the order of an atomic cross section, we must look for processes that inhibit electron capture by physisorbed oxygen, thereby reducing the chemisorption rate. Such processes can be divided into two categories—those that reduce the effective density of the physisorbed oxy-

gen species available for electron capture and those that reduce the electron flux  $\Phi$  (see Eq. [9]). The strong (exponential) dependence of the chemisorption rate on  $\Delta N$  observed for strong accumulation layers (see Eq. [7]) can be explained only by a process that reduces  $\Phi$ . An obvious reason for a reduced electron flux would be the presence of a retarding potential barrier at the surface. Contrary to the case of depletion layers, such a retarding barrier cannot be associated with the space-charge region under accumulation conditions. On the other hand, on real surfaces it could very well be associated with a buffer layer ( $\sim 10 \text{ \AA}$  thick) existing between the ZnO surface proper and the adsorbed oxygen. Such a layer may have been produced during the etching process or by the subsequent exposure to room air. Preliminary Auger-spectroscopy measurements indeed indicate the existence of a surface layer, composed mostly of carbon atoms. If the buffer layer is assumed to be impregnable to oxygen, then one must also assume that the Zn excess is present beneath the layer, at least in the case of the O surface. Otherwise, no accumulation layer could form, since neutral oxygen atoms produced by hole capture would not be able to desorb to form the Zn excess.

Electron transfer across a potential barrier can take place by thermal excitation, tunneling, and thermally assisted tunneling. In principle, an  $\exp(\gamma\Delta N)$  dependence could represent charge transfer by thermal excitation *over* a potential barrier if the barrier height decreases linearly with increasing  $\Delta N$ . This mechanism, however, is ruled out by the observation that  $\gamma$  in Eq. [7] is temperature independent. The temperature dependence of the chemisorption rate is characterized entirely by a *fixed* ( $\Delta N$ -independent) activation energy.

The second possibility, tunneling through a potential barrier, is illustrated by Fig. 10 for the case of strong accumulation layers. This is a simplified energy-level diagram at two stages of the chemisorption process. The accumulation layer is shown together with the energy barrier assumed to be introduced by a buffer layer of thickness  $b$ . For the sake of simplicity it is assumed that the initial condition (solid lines) corresponds to no chemisorbed oxygen at the outer surface of the buffer layer. The negative (free) space charge in the accumulation layer is thus entirely produced and balanced by the positive charge of the excess Zn ions residing on the ZnO surface proper. One need consider only tunneling of electrons at the Fermi level into states of the same energy on the other side of the buffer layer assumed to be introduced by the physisorbed oxygen. The tunneling probability for these electrons is proportional<sup>14</sup> to  $\exp(-D\eta^{1/2})$  where  $\eta$  is the effective barrier height for tunneling as shown in Fig. 10 and  $D = 2(2m^*/\hbar^2)b$  ( $m^*$  is the effective electron mass in the buffer layer and  $\hbar$  is

Planck's constant). Thus, for the case of tunneling, the electron flux  $\Phi$  in Eq. [9] is given by  $v_e n_s \exp(-D\eta^{1/2})$  rather than by Eq. [10]. The dashed curves in the figure represent the situation after a certain amount of oxygen has been chemisorbed on the outer surface of the buffer layer. In this process electrons (of density  $\delta(\Delta N)$ ) have been transferred from the space-charge layer into the chemisorbed species. This has two consequences: (a) the potential barrier  $V_s$  in the space-charge layer is reduced (by  $\delta V_s$ ), and (b) a double-charge layer is

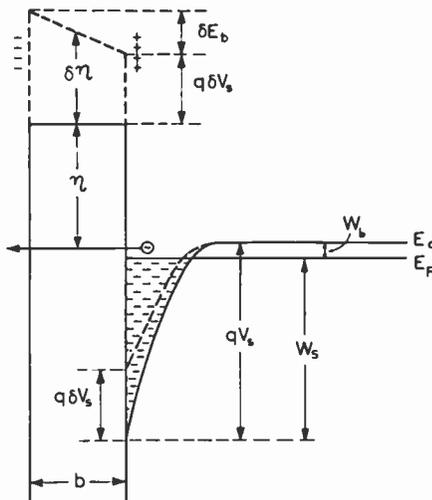


Fig. 10—Simplified energy-level diagram illustrating tunneling model for oxygen chemisorption on ZnO. Solid and dashed curves correspond, respectively, to initial and some later stage of the chemisorption process.  $E_c$  is the conduction-band edge and  $E_F$  the Fermi level.  $V_s$ ,  $W_s$  and  $W_b$  are defined as positive for the accumulation layer shown. The effective barrier height for tunneling is denoted by  $\eta$ .

formed on both sides of the buffer layer, with a surface charge density of  $q\delta(\Delta N)$  and a barrier enhancement of  $\delta E_b = -q^2 b \delta(\Delta N) / \kappa_b \epsilon_0$  ( $\kappa_b$  being the relative dielectric constant of the buffer layer and  $\epsilon_0$  the permittivity of free space). Thus, there is an overall enhancement in the effective barrier height for tunneling of  $\delta\eta = -q\delta V_s + 1/2\delta E_b$ . It can be shown<sup>15</sup> that for strong accumulation layers,  $W_s = qV_s - W_b$  is approximately proportional to  $\Delta N$ ;

$$W_s \approx \theta \Delta N, \quad [14]$$

where  $\theta$  is a constant and  $W_s$  and  $W_b$  are the distances between the Fermi level and the conduction-band edge at the surface and in the bulk, respectively (see Fig. 10). Since, moreover,  $\delta E_b$  is also proportional to  $\delta(\Delta N)$ , we can write for strong accumulation layers

$$\eta \approx \eta_0 - W_b - (\theta + \alpha)\Delta N, \quad [15]$$

where

$$\alpha = \frac{q^2 b}{2\kappa_b \epsilon_0} \quad [16]$$

and  $\eta_0$  is the value of  $\eta$  at flat bands ( $V_s = 0$ ). For  $(\theta + \alpha)\Delta N \ll \eta_0 - W_b$ , one can expand  $\eta^{1/2}$  and obtain for the tunneling exponent

$$-D\eta^{1/2} \approx -D(\eta_0 - W_b)^{1/2} + \beta\Delta N, \quad [17]$$

where

$$\beta = \frac{D(\theta + \alpha)}{2(\eta_0 - W_b)^{1/2}}. \quad [18]$$

Thus, for electron transfer by tunneling, Eq. [10] is replaced by

$$\Phi \approx v_s n_s \exp[-D(\eta_0 - W_b)^{1/2}] \exp(\beta\Delta N) \quad [19]$$

Substitution of Eq. [19] into Eq. [9] yields an  $\exp(\beta\Delta N)$  dependence for the chemisorption rate, with a temperature-independent coefficient  $\beta$ , which agrees in this respect with the experimental results (Eq. [7]). The tunneling model considered does not explain, however, the observed thermal activation for chemisorption. Attempts to account for the activation energy in terms of a thermally assisted tunneling mechanism<sup>14</sup> were unsuccessful; one must assume an unreasonably thick buffer layer (over 100 Å). On the other hand, the value of 0.25 eV obtained for the activation energy seems too low to be attributed to the dissociation of a physically adsorbed oxygen molecule into oxygen atoms. Yet, such dissociation may be a necessary step prior to chemisorption (as O<sup>-</sup> species). This is supported by preliminary measurements showing that the chemisorption of ozone is several orders of magnitude faster than that of (molecular) oxygen, indicating that atomic oxygen does indeed play a very important role. At the same time, while ESR measurements<sup>16</sup> seem to point to the existence of O<sub>2</sub><sup>-</sup> on the ZnO surface after exposure to oxygen, they do not rule out the presence of O<sup>-</sup> as well.

Alternatively, the observed activation energy may be due to a sit-

uation in which only the physisorbed molecules excited into higher energy states or occupying some special lattice sites are active in the capture of electrons. Some kind of an equilibrium distribution between such active centers and the ordinary physisorbed oxygen molecules (or the gas phase) is maintained at the surface, with the result that the number of active centers is a small, pressure-independent fraction of the total density of physisorbed molecules. The activation energy may then be associated with the promotion of the physisorbed oxygen molecules into the active state.

We shall now show that by introducing reasonable values for the characteristic parameters of the buffer layer, one obtains the right orders of magnitude for the experimental values of both  $\gamma$  and  $K_0$  in Eq. [7]. We shall arbitrarily take for the effective mass the free-electron value, while the dielectric constant  $\kappa_b$  in the buffer layer will be assumed to be around 3. Taking a buffer layer of, say, 12 Å thickness, one obtains  $D \approx 13$  (eV)<sup>1/2</sup>. The value of  $\theta$ , as estimated from the known properties of the space-charge layer under strong accumulation conditions,<sup>15</sup> is about  $3.5 \times 10^{-14}$  eV cm<sup>2</sup>. The value of  $\alpha$ , as given by Eq. [16], is about the same. Finally, taking 5 eV for  $\eta_0 - W_b$ , the order of magnitude of the ZnO work function, one obtains from Eq. [18]  $\beta \approx 2 \times 10^{-13}$  cm<sup>2</sup>. This is quite close to the experimental value of  $\gamma$  (see Eq. [7]).

Substituting for  $\Phi$  in Eq. [9] its value from the tunneling equation, Eq. [19], we can compare the resulting expression with the empirical relation, Eq. [7]. In this comparison we take  $\Delta N' = \Delta N$  and  $\gamma = \beta$ . Furthermore, we associate the experimentally observed activation energy with promotion of the physisorbed oxygen molecules into active species, as discussed above. In other words we assume that

$$N_p = N_p^0 \left( \frac{p}{p_0} \right) \exp \left( - \frac{E}{kT} \right), \quad [20]$$

where  $N_p^0$  is the total density of physisorbed oxygen molecules at atmospheric pressure ( $\approx 10^{15}$  cm<sup>-2</sup>). The comparison then yields

$$K_0 = v_e n_s N_p^0 A_n \exp[-D(\eta_0 - W_b)^{1/2}]. \quad [21]$$

For typical accumulation layers,<sup>15</sup>  $\Delta N \sim 10^{13}$  cm<sup>-2</sup>,  $n_s \sim 10^{20}$  cm<sup>-2</sup>. The average electron velocity  $v_e$  in a degenerate accumulation layer is about 10<sup>8</sup> cm/sec, while the attenuation due to the barrier (the exponential term) is about 10<sup>-13</sup>. Hence a typical experimental value of  $K_0$  ( $\approx 10^{14}$  cm<sup>-2</sup> sec<sup>-1</sup>) yields for the capture cross section  $A_n \sim 10^{-16}$  cm<sup>2</sup>, a reasonable value.

The tunneling model accounts also for the observed photodesorp-

tion kinetics (Eq. [8]). Here one must consider the tunneling of photogenerated holes from the valence band into chemisorbed oxygen species. It is not unreasonable to expect the effective barrier height for such tunneling to be comparable to that for conduction electrons in the case of chemisorption. However, the barrier height for hole tunneling now *increases* with increasing  $\Delta N$ , so that the tunneling probability becomes proportional to  $\exp(-\beta\Delta N)$ , as observed experimentally.

So far we have considered strong degenerate accumulation layers, the relevant case for our data. The situation is different for weak accumulation layers such as those studied by Arijs et al.<sup>10</sup> ( $\Delta N < 10^{12}$  cm<sup>-2</sup>). In this case the conduction-band edge at the surface is above the Fermi level, and the electrons involved in the tunneling process lie close to the band edge. Also, the changes  $\delta E_b$  across the buffer layer are negligible for the small values of  $\Delta N$  involved (see Eq. [16]). Hence, the barrier height for tunneling is very nearly independent of  $V_s$  (or  $W_s$ ), being approximately equal to  $\eta_0 - W_b$  (see Fig. 10). In other words, the tunneling model introduces a constant,  $\Delta N$ -independent attenuation factor and the rate equation reduces to

$$\frac{dN_a}{dt} = -\frac{d(\Delta N)}{dt} = v_n n_s N_p A_n \exp[-D(\eta_0 - W_b)^{1/2}]. \quad [22]$$

This is equivalent to Eqs. [9] and [10], but with a much reduced effective cross section. The reducing factor in our case has been estimated as  $\sim 10^{-13}$ , whereas for the data of Arijs et al.<sup>10</sup> one obtains an atomic cross section for a reducing factor of  $10^{-11}$ – $10^{-10}$ . Obviously the reducing factor is very sensitive to small changes in the barrier thickness and height. In fact, as pointed out above, our own measurements on different samples exhibit variations in the chemisorption rate amounting to 2–3 orders of magnitude.

In conclusion, the tunneling model considered above seems to explain the extremely slow chemisorption rate observed for real ZnO surfaces, as well as the functional dependence of both the chemisorption and photodesorption rates on  $\Delta N$ . The observed activation energy characterizing the chemisorption process has been explained by an ad hoc assumption involving the physisorbed species. The model cannot account, however, for the very slow chemisorption process reported by Heiland<sup>3</sup> for clean surfaces, since a buffer layer cannot be envisaged in this case. More work on both clean and real surfaces is required before a better understanding of the chemisorption process in ZnO is obtained.

## References:

- <sup>1</sup> G. Dorda, *Advances in Solid State Physics*, Edited by H. J. Queisser, Vol. 13, p. 215, Pergamon Press, Oxford, (1973). (See also references therein.)
- <sup>2</sup> W. L. Roth, *Physics and Chemistry of II-VI Compounds*, Edited by M. Aven and J. S. Prener, North Holland, Amsterdam, (1967).
- <sup>3</sup> G. Heiland and P. Kunstman, "Polar Surfaces of Zinc Oxide Crystals," *Surface Sci.*, **13**, p. 72, (1969).
- <sup>4</sup> G. Heiland, P. Kunstman, and H. Pfister, "Polar Eigenschaften von Zinkoxyd-Kristallen," *Z. Physik*, **176**, p. 485, (1963).
- <sup>5</sup> A. N. Marino and R. E. Hanneman, "Crystallographic Polarity of ZnO Crystals," *J. Appl. Phys.*, **34**, p. 384 (1963).
- <sup>6</sup> S. R. Morrison, *Advances in Catalysis*, Edited by W. G. Frankenburg, V. I. Kamarewsky and E. K. Rideal, Vol. 7, Academic Press, N. Y., (1955).
- <sup>7</sup> D. A. Melnick, "Zinc Oxide Photoconduction, an Oxygen Absorption Process," *J. Chem. Phys.*, **26**, p. 1136, (1957).
- <sup>8</sup> D. B. Medved, "Photodesorption in Zinc Oxide Semiconductor," *J. Chem. Phys.*, **28**, p. 870, (1958).
- <sup>9</sup> H. J. van Hove and A. Luyckx, "Photoconductivity Decay of ZnO Crystals in Oxygen," *Solid State Commun.*, **4**, p. 603, (1966).
- <sup>10</sup> E. Arijs, F. Cardon, and W. Maenhout-Van Der Vorst, *Surface Sci.*, **17**, p. 387, (1969); *J. Solid State Chem.*, **6**, (1973), "The Influence of Surface Donor States on the Chemisorption Kinetics of Oxygen at the Surface of ZnO Single Crystals." "Part II. Experimental Results," E. Arijs, F. Cardon, and W. Maenhout-Van Der Vorst, p. 319; "Part I. Theoretical Model," E. Arijs and F. Cardon, p. 310.
- <sup>11</sup> R. J. Collins and D. G. Thomas, "Photoconduction and Surface Effects with Zinc Oxide Crystals," *Phys. Rev.*, **112**, p. 388, (1958).
- <sup>12</sup> F. Charachovin and S. Elovich, *Acta Physicochim, USSR*, **20**, p. 227, (1945).
- <sup>13</sup> A. Many, Y. Goldstein, and N. B. Grover, *Semiconductor Surfaces*, Ch. 4, North-Holland, Amsterdam, (1965).
- <sup>14</sup> See, for example, E. Burstein, and S. Lunquist, *Tunneling Phenomena in Solids*, Plenum Press, N. Y., (1969).
- <sup>15</sup> A. Many, International Summer Institute in Surface Science, University of Wisconsin-Milwaukee, Aug. 1973, *Critical Reviews in Solid State Sciences*, Vol. 4, p. 515 (1974).
- <sup>16</sup> R. D. Ivjengar, V. V. Subba Rao, and A. C. Zettlemoyer, "ESR Studies of the Interactions of O<sub>2</sub>, NO<sub>2</sub>, N<sub>2</sub>O, NO, and Cl<sub>2</sub> with Zinc Oxide," *Surface Sci.*, **13**, p. 251, (1969). (See also references therein.)

# A Criterion for the Onset of Hot Electron Transport

Richard S. Crandall

RCA Laboratories, Princeton, N. J. 08540

**Abstract**—By solving the Boltzmann transport equation in the diffusion approximation, a general criterion for the onset of non-ohmic transport can be derived. The departure of the electron distribution function from its thermal equilibrium value is completely specified by the ratio  $R$  of the energy input rate to the energy loss rate. These rates are defined as point functions in energy space and are thus not the same as the more familiar averaged gain and loss rates that are equal in the steady state. As  $R$  approaches unity, the electron system becomes hotter than its surroundings, and non-ohmic transport is observable. Thus  $R$  is the physical determinant for the observation of hot-electron transport. For the special case where the energy and momentum loss are determined by acoustic-phonon scattering,  $R \approx (V_D/c)^2$ , where  $V_D$  is the low-field drift velocity and  $c$  the sound velocity. This expression gives the physical basis for the Shockley criterion for acoustic phonons that transport becomes non-ohmic when the sound and drift velocities are equal.

## Introduction

As far as the solid-state experimentalist is concerned, hot-electron phenomena began with Shockley's<sup>1</sup> discussion of the non-ohmic mobility measurements of Ryder.<sup>2</sup> This paper set the stage and gave some of the physical ideas behind hot electrons. One of the concepts was that hot-electron effects begin when the electron drift velocity equals the sound velocity in a solid. This criterion has since been taken as general, even though it only holds in the special situation where the momentum and energy relaxation are due to acoustic pho-

nons whose interaction can be described by the same potential. It would be better to have a general criterion that is valid for a combination of scattering mechanisms and has an understandable physical basis.

It turns out that there is a general solution of the Boltzmann transport equation, in the diffusion approximation, that can be written as the integral of the ratio of the energy input and loss rates. In this context an electron gains energy only from the electric field and loses energy only to its surroundings. If  $P_i(E)$  is the rate of energy input for an electron of energy  $E$  and  $P_L(E)$  is its rate of energy loss, then  $R(E) = P_i(E)/P_L(E)$ . It is this ratio  $R(E)$  that determines the electron distribution function and therefore the hot-electron effects. These gain and loss rates should not be confused with the familiar *average* gain and loss rates, which are always equal in steady state. When  $R(E) \ll 1$  the electrons are nearly in thermal equilibrium with their surroundings and Ohm's law is obeyed. However, when  $R(E)$  approaches unity, the electron system is out of thermal balance, and non-ohmic transport effects are observable. The ratio  $R(E)$  is the fundamental quantity that determines whether the electrons are hot, and thus should be used as the criterion. Actually, in some cases this ratio can be written as  $V_D^2/V_R^2$ , where  $V_D$  is the electron drift velocity at low electric fields where Ohm's law applies, and  $V_R$  is the random velocity of the particles or collective modes that take energy from the electron gas. In a solid they are the phonons; in a liquid, phonons or other collective modes; and in a gas, atoms or molecules. Thus, the physical basis of the Shockley criterion ( $V_D = V_R$ ) is that the rate of energy input is the order of the rate of energy loss.

In what follows I give a formal on solution of the Boltzmann transport equation in the regime where the diffusion approximation is valid, and show that  $R(E)$  is the parameter that determines the deviation of the electron distribution from its low-field value.

### Boltzmann-Equation Solution

Here I use the standard approach<sup>3</sup> to solve the Boltzmann transport equation in what is termed the diffusion approximation. With mild restrictions, the electron distribution function  $f(\mathbf{K})$  can be written in terms of a single integration over energy. Here  $\mathbf{K}$  is the electron wave vector. What I show is that the argument of this integral is just the ratio  $R(E)$ .

To obtain a solution of the Boltzmann equation in closed form it is convenient to expand the electron distribution function into a spherical harmonics and retain only the first two terms. Therefore,

$$f(\mathbf{K}) = f(E) + \mathbf{g} \cdot \mathbf{K}, \quad [1]$$

where  $f(E)$  is the part of the distribution function that is isotropic in wave vector space and  $\mathbf{g}$  is a vector in the direction of the electric field  $\mathbf{F}$ . This two-term expansion is sufficient if all collisions are momentum randomizing and energy exchange in a collision is a small fraction of the electron energy.<sup>4</sup> Therefore, optical phonon and electron-electron scattering are not included in the theory. For clarity I assume spherical symmetry and classical statistics. Fermi statistics are treated in the appendix.

The time-independent Boltzmann equation can be written, in the absence of a magnetic field, as

$$\frac{e}{\hbar} \mathbf{F} \cdot \frac{\partial}{\partial \mathbf{K}} f(\mathbf{K}) - \hat{C} f(\mathbf{K}) = 0, \quad [2]$$

where  $\hat{C}$  is the collision integral operator, which is defined by

$$\hat{C} f(\mathbf{K}) = \sum_{\mathbf{K}'} [W(\mathbf{K}' \rightarrow \mathbf{K}) f(\mathbf{K}') - W(\mathbf{K} \rightarrow \mathbf{K}') f(\mathbf{K})]. \quad [3]$$

Here  $W(\mathbf{K}' \rightarrow \mathbf{K})$  is the probability per unit time that an electron in the state  $\mathbf{K}'$  makes a transition to the state  $\mathbf{K}$ .

If Eq. [1] for  $f(\mathbf{K})$  is substituted into Eq. [3], then the resulting collision integral contains a term that is isotropic in wave vector space and one that is anisotropic in wave vector space. A similar substitution of  $f(\mathbf{K})$  into the Eq. [2] produces terms that are isotropic and anisotropic in wave vector space. Eq. [2] then becomes two coupled equations; one is isotropic and the other anisotropic in wave vector space.

Now the introduction of two characteristic times simplifies the solution of Eq. [2]. The first time is the familiar momentum relaxation time defined by

$$\tau_M^{-1} \equiv -\hat{C}(\mathbf{g} \cdot \mathbf{K})/\mathbf{g} \cdot \mathbf{K}. \quad [4]$$

Making the usual assumption of quasi-elastic scattering, Eq. [4] can be expressed by

$$\tau_M^{-1} = \frac{-2\pi}{\hbar K \cos(\gamma)} \int \cos(\gamma) |V|^2 \rho(E) |\Delta \mathbf{K}| d\Omega, \quad [5]$$

where  $V$  is the perturbation causing the scattering,  $\Omega$  the solid angle,  $\gamma$  the angle between  $\mathbf{F}$  and  $\mathbf{K}$ , and  $\rho(E)$  the density of electron states. The momentum transfer in the scattering is  $\Delta \mathbf{K} = \mathbf{K}' - \mathbf{K}$ . Of course,

I have assumed that first-order perturbation theory is sufficient to describe the scattering.

The second characteristic time is the energy loss time, which has a parallel definition. It is defined by

$$\tau_E^{-1} = \frac{2\pi}{E\hbar} \int |V|^2 \rho(E) \Delta E \, d\Omega, \quad [6]$$

where  $\Delta E$  is the energy transfer in a collision.

This time can be given physical meaning in terms of the energy-loss rate  $P_L(E)$ . Using first-order time-dependent perturbation theory,

$$P_L(E) = \sum_{K'} \Delta E W(K' \rightarrow K) \quad [7a]$$

$$= \frac{2\pi}{\hbar} \int \rho(E') |V|^2 \Delta E \delta(E' - E) \, dE' \, d\Omega \quad [7b]$$

$$= \frac{E}{\tau_E} \quad [7c]$$

Thus  $\tau_E$  is a measure of the ability of an electron to lose energy.

Using these two times, Eq. [2] can be rewritten and separated into two terms. The one that is isotropic in wave vector space is

$$\frac{e\hbar}{m} \mathbf{F} \cdot \mathbf{K} \frac{\partial f(E)}{\partial E} + \frac{\mathbf{g} \cdot \mathbf{K}}{\tau_M} = 0, \quad [8]$$

and the one that is anisotropic in wave vector space is

$$\frac{\partial}{\partial E} \left[ \frac{E}{\tau_E} \left( f(E) + k_B T \frac{\partial f(E)}{\partial E} \right) \right] + \frac{\partial}{\partial E} \left( \frac{e^2 F^2 \tau_M E}{m} \frac{\partial f(E)}{\partial E} \right) = 0, \quad [9]$$

where  $T$  is the lattice temperature.

In the Appendix, I show the steps leading to Eqs. [8] and [9] in more detail. These two coupled equations can now be solved for  $f(E)$  and  $\mathbf{g}$ . Eq. [9], with the boundary conditions that  $f(\infty) = 0$  and  $f(0) = 1$ , has the solution

$$f(E) + k_B T \frac{\partial f(E)}{\partial E} + \frac{e^2}{M} F^2 \tau_M \tau_E \frac{\partial f(E)}{\partial E} = 0, \quad [10]$$

which can be integrated to give

$$f(E) = \exp \left\{ -\frac{1}{k_B T} \int_0^E \frac{dE'}{1 + e^2 F^2 \tau_M \tau_E / m k_B T} \right\} = \exp(-J). \quad [11]$$

Eq. [11] takes on physical meaning when we realize that it can be written in terms of the energy input and loss rates. The rate of energy input is

$$P_i(E) = \frac{e}{4\pi} \int d\Omega \mathbf{g} \cdot \mathbf{K} \frac{F\partial E}{\hbar\partial K} / f(E). \quad [12]$$

or the power input per electron at energy  $E$ . The total power input would be  $P_i(E)f(E)$  summed over all electron energies. Using Eq. [8] for  $\mathbf{g} \cdot \mathbf{K}$  and Eq. [11] for  $f(E)$ , Eq. [12] becomes,

$$P_i(E) = \frac{e^2 F^2 \tau_M E}{m k_B T (1 + e^2 F^2 \tau_M \tau_E / m k_B T)}. \quad [13]$$

Using Eq. [7c] for  $P_L(E)$  and Eq. [13] for  $P_i(E)$ , the ratio  $R(E) = P_i(E)/P_L(E)$  is

$$R(E) = 1/(1 + m k_B T / e^2 F^2 \tau_M \tau_E). \quad [14]$$

Using this expression for  $R(E)$  we can write  $f(E)$  in the compact form.

$$f(E) = \exp\left(-\frac{1}{k_B T} \int_0^E [1 - R(E')] dE'\right) \quad [15a]$$

$$= f_0(E) \exp\left(\frac{1}{k_B T} \int_0^E \frac{P_i(E')}{P_L(E')} dE'\right). \quad [15b]$$

where  $f_0(E) = \exp(-E/k_B T)$ . This form of  $f(E)$  shows that the deviations from Maxwellian behavior are determined by the ratio of the rates of energy input to loss. If an electron can readily lose energy (large  $P_L(E)$ ), then it will be difficult for the electron system to be disturbed from thermal equilibrium.

The magnitude and energy dependence of  $R(E)$  can not be determined until  $\tau_M$  and  $\tau_E$  have been evaluated. However, at least two general statements can be made. If  $R(E) = R_0$ , which is a constant, then the distribution function is Maxwellian, with an electron temperature different from  $T$ . The electron temperature  $T_e$  is expressed by  $T_e = T(1-R(E))^{-1}$ . If  $R(E)$  is an increasing function of energy, then  $f(E)$  cannot be normalized, and the electron system is unstable against runaway; i.e., all the electrons are at high energy. Thus we see the central role played by the ratio  $P_i(E)/P_L(E)$  in determining the behavior of the distribution function. This ratio is a measure of the ability of an electron to lose the energy it gains from the electric field. At this point I want to stress again that energy balance is not obtained by equating  $P_i(E)$  to  $P_L(E)$ . Rather, the energy balance, which

obtains in the steady state, is defined by

$$\sum_{\mathbf{K}} E \left( \frac{e}{\hbar} \mathbf{F} \cdot \frac{\partial(\mathbf{g} \cdot \mathbf{K})}{\partial \mathbf{K}} - \hat{C}f(E) \right) = 0, \quad [16]$$

which, using Eqs. [37], [12], [13], and [7c] and integrating over angles and once by parts over energy becomes

$$\int dE \rho(E) \left[ P_i(E) f(E) - P_L(E) \left( f(E) + k_B T \frac{\partial f(E)}{\partial E} \right) \right] = 0. \quad [17]$$

Thus we see that the energy balance is an average over energy of the gain and loss rates weighted by the distribution function. The result that  $P_L(E)$  is weighted by  $(f(E) + k_B T \partial f(E)/\partial E)$ , not just  $f(E)$ , arises from the fact that  $\hat{C}f(E)$  includes scattering in and out of the state  $\mathbf{K}$ , whereas  $P_L(E)$  is defined in terms of scattering out of the state  $\mathbf{K}$ . This becomes clear when an electron temperature  $T_e$  can be defined. Then  $(f(E) + k_B T \partial f(E)/\partial E) \approx (1 - (T/T_e))f(E)$ , which shows that there is energy loss only if the electron temperature is higher than the lattice temperature.

Returning now to the question of a criterion for the observation of non-ohmic effects, we see that it is contained in  $R(E)$ , since  $R(E)$  is a measure of the deviation of the distribution function from its thermal equilibrium value. For simplicity assume that at low field  $R(E)$  can be expressed as a power of  $E$ ; i.e.,

$$R(E) = R_0(E/k_B T)^\alpha \quad [18]$$

where  $R_0$  is a constant proportional to  $F^2$ . With this definition,  $f(E)$  becomes

$$f(E) = f_0(E) \exp \left[ \frac{R_0}{1 + \alpha} \left( \frac{E}{k_B T} \right)^{1 + \alpha} \right]. \quad [19]$$

At low fields the electron transport is described by the warm electron coefficient<sup>5</sup>  $\beta$ . It is defined in terms of mobility  $\mu$ ;

$$\mu = \mu_0 [1 + \beta F^2], \quad [20]$$

where  $\mu_0$  is the electron mobility at zero field. The mobility is calculated from<sup>3</sup>

$$\mu = - \frac{e}{m} \int_0^\infty \tau_M E \rho(E) \frac{\partial f(E)}{\partial E} / \int_0^\infty f(E) \rho(E) dE. \quad [21]$$

If  $\tau_M$  is proportional to  $E^p$ , then Eq. [20] becomes

$$\mu = \mu_0 \left( 1 + R_0 \left[ \frac{(1+p+\alpha)!}{(1+\alpha)p!} - \alpha! \right] \right). \quad [22]$$

Thus  $R_0$  determines the magnitude of the mobility deviation, whereas the sign of the deviation is determined by the sign of  $p$ . In terms of  $R_0$ , then the warm electron coefficient is

$$\beta = \frac{R_0}{F^2} \left[ \frac{(1+p+\alpha)!}{(1+\alpha)p!} - \alpha! \right]. \quad [23]$$

To show the connection between the Shockley criterion and the parameter  $R(E)$ , we write  $R(E)$  in its low-field limit as

$$R(E) \simeq \frac{e^2 F^2 \tau_M \tau_E}{m k_B T}. \quad [24]$$

For acoustic phonon scattering, the energy loss time can be obtained from the argument that<sup>6</sup>

$$P_L = \frac{E}{\tau_E} = \frac{\hbar \omega}{\tau_M (2n_p + 1)}, \quad [25]$$

where  $n_p$  is the phonon occupation number and  $\omega$  the phonon frequency. Using momentum conservation to equate the phonon momentum to the electron momentum and assuming equipartition, one finds that

$$\tau_E = \tau_M \frac{kT}{mc^2}. \quad [26]$$

where  $c$  is the sound velocity and  $m$  the electron mass. Thus  $R(E)$  becomes

$$R(E) \approx \frac{e^2 F^2 \tau_M^2}{m^2 c^2} \approx \frac{V_D^2}{c^2} \quad [27]$$

apart from a factor the order of unity that relates the drift velocity  $V_D$  to  $e\tau_M/m$ . Thus as Shockley pointed out, non-ohmic transport should be observable when the drift velocity is on the order of the sound velocity.

However, if we consider energy loss to acoustic phonons and momentum loss to ionized impurities, then  $V_D \approx V_R$  is no longer the criterion. In this case

$$R(E) \approx \frac{V_D^2}{c^2} \frac{\tau_{MP}}{\tau_{MI}} \quad [28]$$

where  $\tau_{MP}$  and  $\tau_{MI}$  are the momentum relaxation times for phonon

and impurity scattering, respectively. Since  $\tau_{MP}/\tau_{MI} \gg 1$ , hot-electron effects will be observable when the drift velocity is much less than the sound velocity  $V_R$ . Nevertheless,  $R(E)$  approaching unity is still the criterion for hot electrons.

The dominant energy dependence of the energy input and loss rates for some of the common scattering mechanisms are listed in Table 1. In all cases, equipartition holds and the system is considered to be isotropic. The letters DP and PP refer to acoustic phonon interactions via the deformation and piezoelectric potentials, respectively.

Table 1—Dominant Energy Dependence of the Energy Input and Loss Rates and the Integral  $I(E)$

Scattering Mechanism	$P_i(E)$	$P_L(E)$	$I(E)$ DP	$I(E)$ PP
ionized impurities	$E^{5/2}$ (7)		$E^2$	$E^3$
neutral impurities	$E$ (8)		$E^{1/2}$	$E^{3/2}$
dipoles	$E^{1/2}$ (9)		$\ln(E)$	$E$
acoustic phonons (DP)	$E^{1/2}$ (10)	$E^{3/2}$ (11)	$\ln(E)$	$E$
acoustic phonons (PP)	$E^{3/2}$ (12)	$E^{1/2}$ (13)	$E$	$E^2$

Notations for Table 1: The symbols DP and PP refer to the interaction with acoustic phonons via the deformation and piezoelectric potentials, respectively. The numbers in parenthesis are References for the various interactions.

The energy dependence of the integral

$$I(E) = \frac{1}{k_B T} \int_0^E dE' R(E') \quad [29]$$

for the low electric field limit is also listed in the table. The subscript on  $I(E)$  refers to the mechanism of energy loss. If  $I(E)$  increases with energy faster than linearly, then a runaway situation results.

## Summary

I have shown how the general solution of the Boltzman transport equation, in the diffusion approximation, can be expressed in terms of a single physical quantity, namely, the ratio of the rate of energy input from the electric field to the rate of energy loss to the lattice. This ratio, which measures the deviation of the electron distribution from its thermal equilibrium value, is thus the determinant for hot-electron effects. If it is small the system remains near equilibrium; if it approaches unity the electrons become hot.

## Acknowledgments

I am indebted to A. Rose who gave me my start in non-ohmic transport and is always ready to offer the much needed physical insight. I also wish to thank Ping Sheng for an illuminating discussion.

## Appendix

To include Fermi statistics in the Boltzmann equation, Eq. [3] must be modified to reflect the fact that transitions can occur only to unoccupied states. This is done by replacing  $f(\mathbf{K}')$  by  $f(\mathbf{K}') (1 - f(\mathbf{K}))$ . A similar substitution is made for  $f(\mathbf{K})$ . As an illustrative example of how the Boltzmann equation is solved in the diffusion approximation, I consider electron-acoustic-phonon scattering. Using Eq. [1] for  $f(\mathbf{K})$  the collision integral operator is

$$\widehat{C}f(\mathbf{K}) = \widehat{C}f(E) + \widehat{C}(\mathbf{g} \cdot \mathbf{K}) \quad [30]$$

$$\begin{aligned} \widehat{C}f(E) = & \frac{2\pi}{\hbar} \int \rho(E') dE' d\Omega |V_0|^2 \times \\ & [[f(E')(1 - f(E))(n_p + 1) - f(E)(1 - f(E'))n_p] \\ & \times \delta(E' - E - \hbar\omega) + [f(E')(1 - f(E))n_p - f(E) \\ & \times (1 - f(E'))(n_p + 1)] \delta(E' - E + \hbar\omega)] \quad [31] \end{aligned}$$

$$\begin{aligned} \widehat{C}(\mathbf{g} \cdot \mathbf{K}) = & \frac{2\pi}{\hbar} \int \rho(E') dE d\Omega |V_0|^2 \times \\ & [[\mathbf{g} \cdot \mathbf{K}'(1 - f(E))(n_p + 1) - \mathbf{g} \cdot \mathbf{K}(1 - f(E'))n_p] \\ & \times \delta(E' - E - \hbar\omega) + [\mathbf{g} \cdot \mathbf{K}'(1 - f(E))n_p - \mathbf{g} \cdot \mathbf{K} \\ & \times (1 - f(E'))(n_p + 1)] \delta(E' - E + \hbar\omega)] \quad [32] \end{aligned}$$

and

$$|V|^2 = |V_0|^2(n_p + 1/2 \pm 1/2). \quad [33]$$

The + and - signs refer to phonon emission and absorption respectively. Eqs. [31] and [32] are correct to  $\mathcal{O}(g)$ . It is convenient in reducing the above integrals to expand the  $\delta$  function to second order, i.e.,

$$\begin{aligned} \delta(E' - E - \hbar\omega) \simeq & \delta(E' - E) - \hbar\omega \frac{\delta}{\delta E'} \delta(E' - E) \\ & + \frac{1}{2}(\hbar\omega)^2 \frac{\partial^2}{\partial E'^2} \delta(E' - E) \quad [34] \end{aligned}$$

Since only terms  $\mathcal{O}(g)$  need be retained in  $\widehat{C}\mathbf{g} \cdot \mathbf{K}$ , it reduces to

$$\begin{aligned} \widehat{\mathbf{c}} \cdot \mathbf{K} &= \frac{2\pi}{\hbar} \int \rho(E') dE d\Omega |V_0|^2 \\ &\times \mathbf{g} \cdot (\mathbf{K}' - \mathbf{K}) f(E) (2n_p + 1) \delta(E' - E) \end{aligned} \quad [35]$$

which is the same as Eq. [4].

The symmetry of the integrand in Eq. [31] makes the first term containing  $\delta(E' - E)$  vanish on integration. The remainder can be integrated by parts to remove the derivatives of the  $\delta$  function. This results in

$$\begin{aligned} \widehat{C}f(E) &= \frac{2\pi}{\hbar} \int \delta(E' - E) dE d\Omega \times \\ &\left[ \hbar\omega \frac{\partial}{\partial E'} [f(E')(1 - f(E)) - f(E)(1 - f(E'))] \right. \\ &+ \frac{(\hbar\omega)^2}{2} \frac{\partial^2}{\partial E'^2} [f(E')(1 - f(E)) \\ &\left. - f(E)(1 - f(E'))] (2n_p + 1) \rho(E) |V_0|^2 \right] \end{aligned} \quad [36]$$

which, upon integration over energy and solid angle becomes

$$\widehat{C}f(E) = \frac{\partial}{\partial E} \left[ \frac{E}{\tau_E} \left\{ f(E)[1 - f(E)] + (2n_p + 1) \frac{\hbar\omega}{2} \frac{\partial f(E)}{\partial E} \right\} \right]. \quad [37]$$

I have used the fact that  $\hbar\omega = \Delta E$  to simplify this equation using Eq. [6] for  $\tau_E$ .

Now the part of Eq. [2] that is anisotropic in wave vector space becomes, using Eqs. [36] and [4],

$$\begin{aligned} \frac{\partial}{\partial E} \left( \frac{E}{\tau_E} f(E)(1 - f(E)) \right) + (2n_p + 1) \frac{\hbar\omega}{2} \frac{\partial f(E)}{\partial E} \\ + \frac{e^2}{m} F^2 \frac{\partial}{\partial E} \left( E \tau_M \frac{\partial f(E)}{\partial E} \right) = 0 \end{aligned} \quad [38]$$

In the limit of Boltzmann statistics where  $f(E) \ll 1$ , in the high temperature limit ( $n_p \sim kT/\hbar\omega$ ), Eq. [38] is just Eq. [9]. In its general form, Eq. [38] has the solution

$$f(E) = \frac{1}{1 + N_0 e^{+J}} \quad [39]$$

where  $J$  is defined by Eq. [11]. Now the constant  $N_0$  is determined by the condition that  $f(E)$  must be the Fermi function in zero electric field. Since  $J = E/k_B T$  in this limit,

$$N_0 = \exp \{-E_f/k_B T\}, \quad [40]$$

where  $\bar{E}_f$  is the Fermi energy.

#### References:

- <sup>1</sup> W. Shockley, "Hot Electrons in Germanium and Ohms Law," *Bell System Tech. J.*, **30**, p. 990 (1951).
- <sup>2</sup> E. J. Ryder, "Mobility of Holes and Electrons in High Electric Fields," *Phys. Rev.*, **90**, p. 766 (1953).
- <sup>3</sup> E. G. S. Paige, "Electrical Conductivity of Germanium," in *Progress in Semiconductors*, ed. by A. F. Gibson and R. E. Burgess, John Wiley and Son, New York (1964), Vol. 8, Chap. 2.
- <sup>4</sup> See Ref. [3], p. 14.
- <sup>5</sup> See Ref. [3], p. 179.
- <sup>6</sup> A. Rose, "The Acoustoelectric Effects and the Energy Losses by Hot Electrons," *RCA Review*, **30**, p. 435 (1969).
- <sup>7</sup> See Ref. [3], p. 75.
- <sup>8</sup> See Ref. [3], p. 72.
- <sup>9</sup> A. D. Boardman, "The Theory of Dipole Scattering in Semiconductors," *Proc. Phys. Soc.*, p. 141 (1965).
- <sup>10</sup> See Ref. [3], p. 55.
- <sup>11</sup> E. M. Conwell, "High Field Transport in Semiconductors," in *Solid State Physics*, edited by F. Seitz, D. Turnbull and H. Ehrenreich, Academic Press Inc., New York (1967), Suppl. 9, p. 126.
- <sup>12</sup> See Ref. [11], p. 112.
- <sup>13</sup> See Ref. [6], p. 440.

# Structure and Energy of the Interface Between an Insulator and an Electrolyte

Richard Williams

RCA Laboratories, Princeton, N.J. 08540

**Abstract**—The surface tension of water is increased by the addition of a soluble ionic salt. In this work it is shown that there is a similar increase in the interfacial free energy between the aqueous solution and a polymer. The origin of these effects can be traced to electrostatic forces that deplete the ions from a thin layer of water near the interface. In an idealized model of interface, we represent it as a thin layer of pure water sandwiched between the bulk solution and an insulator. Generally speaking, the free energy of pure water is higher than that of water containing dissolved ions. By analyzing this idea quantitatively it is possible to calculate the energy of the interface as a function of the concentration of dissolved salt.

## Introduction

Electronic technology makes frequent use of the interface between insulators and electrolytes. Examples include the widespread use of photoresists, electroless plating,<sup>1</sup> charging of insulators by liquid electrodes for the purpose of writing<sup>2</sup> or subsequent recording,<sup>3</sup> and the processing of thin films of silicon dioxide.<sup>4</sup> The electrical properties of the interface will be affected by the distribution of ions in the electrolyte near the interface. Electrical properties, with emphasis on charge injection and transport within the insulator, have been reviewed by Mehl and Hale.<sup>5</sup> The properties of spreading and adhesion are determined by the interfacial free energy. When ions are dissolved in water both the surface tension and the interfacial free ener-

gy are strongly affected. Both of these effects are due to the same root cause—depletion of ions from the layer of solution near the interface. In this article, experiments and analysis are presented to illustrate these effects.

### Electrostatic Image Force

Ions dissolved in water are repelled from an interface with an insulator, such as air or a polymer, that has a dielectric constant smaller than that of water. In general, at the interface between a dielectric medium, 1, with a large dielectric constant,  $\epsilon_1$ , and another dielectric medium, 2, having a smaller dielectric constant,  $\epsilon_2$ , a charge,  $q$ , in medium 1 is repelled from the interface. If the interface lies in the  $x$ - $y$  plane at  $z = 0$ , the image potential,  $U(z)$  is

$$U(z) = \left( \frac{q^2}{4\epsilon_1 z} \right) \frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + \epsilon_2} = \left( \frac{q^2}{4\epsilon_1 z} \right) \alpha \quad [1]$$

where  $\alpha = (\epsilon_1 - \epsilon_2)/(\epsilon_1 + \epsilon_2)$ . For water at 25°C,  $\epsilon_1 = 72$ . For an interface with air,  $\epsilon_2 = 1$  and  $\alpha = 0.98$ . For the interface with polymers,  $\epsilon_2$  is typically 2.0–2.3 and  $\alpha = 0.94$ . Due to the positive potential energy,  $U(z)$ , ions of both signs will be depleted from a thin zone near the interface. We assume, for simplicity, that the ion concentration,  $n(z)$ , can be related to the bulk concentration,  $n_0$ , by a simple Boltzmann factor:

$$n(z) = n_0 \exp\{-U(z)/kT\}. \quad [2]$$

A more careful analysis would have to take account of the screening of the image field by the ions in the solution. Fig. 1 shows, qualitatively, the behavior of  $U(z)$  and  $n(z)$  near an interface with an insulator of low dielectric constant. Also shown is the distance,  $\delta$ , at which the ion concentration is depleted to half its value in the bulk. Since  $\alpha$  is nearly the same for water–air and water–polymer interfaces,  $\delta$  is nearly the same for both cases. For singly charged ions  $\delta$  is 2.5 Å, while for doubly charged ions  $\delta$  is 10 Å.

### Effects of Ion Depletion

The primary effect of ion depletion near the interface is to increase the surface tension or interfacial free energy. These are equivalent terms expressing simply the additional free energy needed to form a square centimeter of new surface or interface. A review of the literature on effects of dissolved ions on the surface tensions of aqueous so-

lutions has been published by Randles.<sup>6</sup> More recently, similar results have been found for the interfacial free energy at polymer-solution interfaces.<sup>7</sup> In both cases the interfacial free energy is increased when dissolved ions are present. Significantly, the increase in surface tension produced by a given concentration of ions is the same as the increase in interfacial tension for the interface with Teflon or polyethylene.

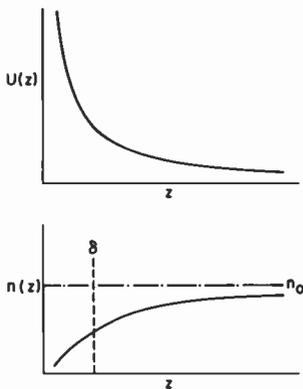


Fig. 1—The image potential energy  $U(z)$  and the equilibrium ion concentration  $n(z)$  as a function of the distance  $z$  from interface of an aqueous solution with an insulator such as air or a polymer.

The interfacial free energy is inferred from measurements of the contact angle of a drop of aqueous solution in contact with a polymer surface. In the simplest case,<sup>8</sup> the contact angle,  $\theta$ , is determined by three interfacial free energies.  $\gamma_{LA}$  is the surface tension of the liquid.  $\gamma_{SA}$  and  $\gamma_{SL}$  are the interfacial free energies for the solid-air interface and the solid-liquid interface, respectively. For a liquid that does not wet the surface, but stands up in discrete drops, the relation between  $\theta$  and the interfacial energies is given by the equation

$$\gamma_{LA} \cos\theta = \gamma_{SA} - \gamma_{SL}. \quad [3]$$

Consider now what happens when the contact angle on a given polymer surface is measured for a series of aqueous solutions of a given ionic salt over a wide range of concentrations. For solution of potassium carbonate, for example, the surface tension,  $\gamma_{LA}$ , varies from 72 ergs/cm<sup>2</sup> for pure water to 110.8 ergs/cm<sup>2</sup> for a concentrated solution at 8.3 mol/L. The contact angle,  $\theta$ , and the interfacial free energy,  $\gamma_{SL}$ , are also strongly dependent on the salt concentration. All the terms in Eq. [3] are then functions of the salt concentration,  $c$ , with

the exception of  $\gamma_{SA}$ . We incorporate this feature by using this notation;

$$\gamma_{LA}(c) \cos\theta(c) = \gamma_{SA} - \gamma_{SL}(c). \quad [4]$$

We now assume that the contribution of the ions to the surface tension is the same as their contribution to the interfacial tension,  $\gamma_{LA}$ . This is reasonable if the effect is dominated by the image force since

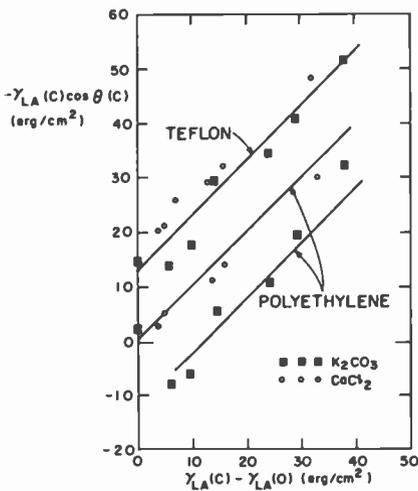


Fig. 2— $\gamma_{LA}(c) \cos\theta$  vs.  $\gamma_{LA}(c) - \gamma_{LA}(0)$  for aqueous solutions on Teflon and polyethylene. The data for  $K_2CO_3$  all refer to  $20^\circ C$ . The surface tension data reported in the literature was available only for these temperatures and the contact angle measurements were made at the corresponding temperatures. Note that, on a given substrate, there is only value for the contact angle at zero concentration. (Experimental details in Ref. [7].)

this effect is almost identical in the two cases. We further assume that the effect is simply additive, i.e.,

$$\gamma_{SL}(c) - \gamma_{SL}(0) = \gamma_{LA}(c) - \gamma_{LA}(0). \quad [5]$$

Eq. [4] then becomes

$$\begin{aligned} \gamma_{LA}(c) \cos\theta(c) &= \{\gamma_{SA} - \gamma_{SL}(0)\} - \{\gamma_{LA}(c) - \gamma_{LA}(0)\} \\ &= \gamma_{LA}(0) \cos(0) - \{\gamma_{LA}(c) - \gamma_{LA}(0)\}. \end{aligned} \quad [6]$$

Fig. 2 shows a test of this equation, using our own data<sup>7</sup> for the contact angles of solutions of  $K_2CO_3$  and  $CaCl_2$  on surfaces of Teflon and

polyethylene. The data are plotted in the form suggested by Eq. [6]. Since the points fall approximately on lines of unit slope, the assumption of Eq. [5] appears to be confirmed. The data indicate that the image force may be the dominant cause for the increase in surface and interfacial tension when ions are dissolved in water. A more detailed discussion will be given below.

### Work of Adhesion

The work of adhesion,  $W_{ad}$ , is defined as the work needed to separate two materials that are in intimate contact. It can be expressed by the equation<sup>8</sup>

$$W_{ad} = \gamma_{SA} + \gamma_{LA} - \gamma_{SL} \quad [7]$$

Similarly, the work of cohesion,  $W_{co}$ , to separate a volume of liquid into two parts is

$$W_{co} = 2\gamma_{LA} \quad [8]$$

per cm<sup>2</sup>. Combining Eq. [7] with Eq. [3] we get

$$W_{ad} = \gamma_{LA}(1 + \cos\theta) \quad [9]$$

For the interfaces studied here, the work of adhesion is independent of salt concentration, even though the individual surface tensions and interfacial energies are changing by substantial amounts. The reason for this is seen in Eq. [7]. The part of this equation that depends on salt concentration is the combination,  $\gamma_{LA}(c) - \gamma_{SL}(c)$ . However, from Eq. [5] we see that this combination is independent of concentration. To show the behavior of  $W_{ad}$  with concentration, the quantity  $\gamma_{LA}(c)(1 + \cos\theta(c))$  is plotted as a function of salt concentration  $c$  in Fig. (3). The work of adhesion, then, is unaffected by dissolved salts because the effect of the salt is identical in two cancelling terms that enter into the definition of  $W_{ad}$ . On the molecular level, the explanation for this observation is probably that the adhesion is determined by the last layer or two of water molecules near the interface. Due to the image potential, ions are excluded from this region, even though they may be present in high concentration in the bulk of the solution.

### How the Ion Distribution Affects the Interfacial Tension

A molecular model for the effect of ions on the surface tension of aqueous solutions was formulated by Onsager and Samaras.<sup>9</sup> Accord-

ing to this picture, the image potential depletes the ion concentration near the surface and those that remain have an extra potential energy, as illustrated in Fig. 1. By taking a Boltzmann distribution and averaging the energy over all space one gets the additional contribution to the energy by those ions near the surface. To obtain a convergent solution to the problem, the Debye screening of the image potential by the ions in solution must be considered. The result is straightforward but can be handled only by a numerical method that

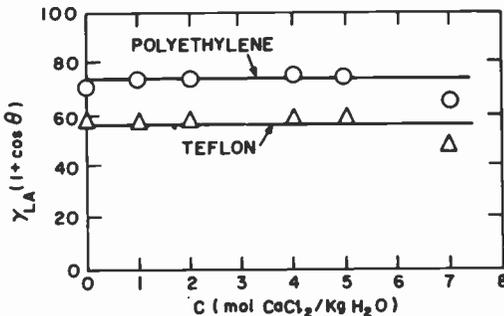


Fig. 3—The work of adhesion  $W_{ad} = \gamma_{LA}(1 + \cos \theta)$  as a function of concentration for aqueous solutions of  $\text{CaCl}_2$  on Teflon and polyethylene.

obscures some of the physics. We adopt here an approximate thermodynamic approach that is in reasonable quantitative agreement with experiment. A similar approach has been discussed by Schmutzer.<sup>10</sup> The basic idea is that the free energy or chemical potential  $\mu$  of a pure liquid is higher than that of the same liquid when something is dissolved in it. The solute dilutes the solvent and lowers its chemical potential. This is the basis of so-called colligative properties such as boiling-point elevation, freezing-point depression, and osmotic pressure. The mol fraction of the solvent in the solution is denoted by  $x$  and that of the solute by  $x_s$ ; the chemical potential of the pure solvent is  $\mu(0)$  and that at mol-fraction,  $x$ , is  $\mu(x)$ . According to the theory<sup>11</sup> of ideal solutions:

$$\mu(x) = \mu(0) + kT \ln x = \mu(0) + kT \ln(1 - x_s); \quad [10]$$

for  $x_s \ll 1$  this becomes

$$\mu(0) - \mu(x) \approx kT x_s. \quad [11]$$

Consider now the idealized picture of the surface or interfacial layer shown in Fig. (4). As a result of the image potential, a layer near the interface, of thickness  $\delta$  is completely depleted of ions of both signs.

The underlying layer is assumed to have the bulk concentration of ions up to the point where it joins the surface layer. This is clearly an idealization of the more realistic picture of ion concentration sketched in Fig. (1). As stated above,  $\delta$  is 2.5 Å for a singly charged ion and 10 Å for a doubly charged ion. In the surface layer the free energy of each water molecule is increased by the quantity  $kTx_s$ .  $x_s$

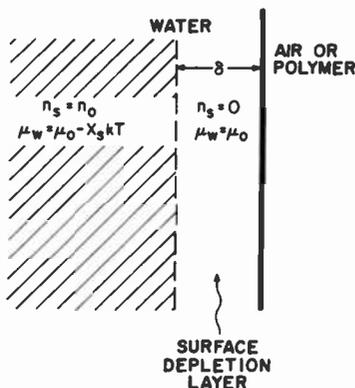


Fig. 4—Idealized model of an aqueous solution of an ionic salt near an interface with air or a polymer. The chemical potential  $\mu_w$  of the water is shown for the bulk and for the ion-depleted layer of thickness  $\delta$  near the interface.

can be expressed in terms of the particle concentrations  $n_s$  and  $n_w$  for salt and water molecules, respectively;

$$x_s = \frac{n_s}{n_s + n_w} \approx \frac{n_s}{n_w}. \quad [12]$$

The increase in surface tension or interfacial tension  $\Delta\gamma$  caused by a given concentration of ions is simply the total increase in free energy for all the water molecules in the surface layer. This, in turn, is the increase in free energy per molecule times the number of molecules in a square centimeter of the surface layer;

$$\Delta\gamma = x_s kT n_w \delta \approx n_s \delta kT. \quad [13]$$

To get the magnitude of  $\Delta\gamma$ , the contributions of both positive and negative ions are added, using the appropriate values of  $n_s$  and  $\delta$ .

Fig. 5 shows the results of Eq. [12] compared with surface tension data from the literature. There is reasonable agreement between the model and the data over a wide range of concentrations. The range of

concentrations covered is far wider than that over which ideal solution equations are normally valid. This is perhaps due to the fact that we are dealing here with a depleted region of the solution, within which concentrations of the ions are well below their bulk values.

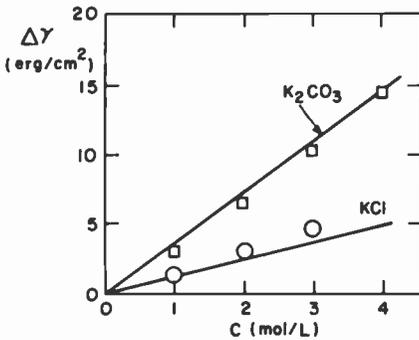


Fig. 5—The increase in surface tension  $\Delta\gamma_{LA}$  as a function of concentration of dissolved salt. The solid lines are theoretical, according to Eq. [12]. Points are experimental data. Data for KCl are taken from Ref. [12]; data for  $K_2CO_3$  are taken from Ref. [13].

### Summary and Conclusions

When an aqueous solution is in contact with an insulator of low dielectric constant, the layer of water next to the insulator is strongly depleted of ions. Based on this model, the effect of dissolved salts on the interfacial energy can be understood as due simply to the additional free energy of the ion-depleted water near the interface.

As an example of the usefulness of this model for specific problems, we take the experiments of Engelbrecht<sup>2</sup>—electrostatic recording on insulators by non-wetting conductive liquids. A thin sheet of polymer was placed on a grounded backing and a charge pattern recorded on it using a movable recorder pen moistened with water. About 100 volts was applied to the pen. Ions in the water are repelled from the polymer surface. How then do they get onto the surface to charge it? We note that an ion in air is strongly attracted to the polymer surface by the same image force that repels it when it is in water. These conditions suggest that the transfer of charge to the surface takes place only at the trailing edge of the drop. Using this concept, a more precise analysis can be made of the optimum shape for the pen and the expected noise properties.<sup>14</sup>

## References:

- <sup>1</sup> I. A. Abu-Isa, "Metal Plating of Polymeric Surfaces," *Polymer-Plast. Technol. Eng.*, **2**, p. 29, (1973).
- <sup>2</sup> R. S. Engelbrecht, "Electrostatic Recording on Insulators by Nonwetting Conducting Liquids," *J. Appl. Phys.*, **45**, p. 3421, (1974).
- <sup>3</sup> R. Williams, U.S. Patent #3,579,332, (1971).
- <sup>4</sup> R. Williams and A. M. Goodman, "Wetting of Thin Layers of SiO<sub>2</sub> by Water," *Appl. Phys. Letters*, **25**, p. 531, (1974).
- <sup>5</sup> W. Mehl and J. M. Hale in *Advances in Electrochemistry and Electrochemical Engineering*, P. Delahay, Editor, Vol. 6, pp. 399-457, Interscience, New York, (1961).
- <sup>6</sup> J. E. B. Randles in *Advances in Electrochemistry and Electrochemical Engineering*, P. Delahay, Editor, Vol. 3, pp. 1-30, Interscience, New York, (1961).
- <sup>7</sup> R. Williams, "Interfacial Free Energies Between Polymer and Aqueous Electrolyte Solutions," *J. Phys. Chem.* **79**, p. 1274 (1975).
- <sup>8</sup> W. A. Zisman, "Contact Angle, Wettability and Adhesion," *Advances in Chemistry Series 43, American Chemical Society, Washington*, pp. 1-52, (1964).
- <sup>9</sup> L. Onsager and N. T. Samaras, "The Surface Tension of Debye-Hückel Electrolytes" *J. Chem. Phys.*, **2**, p. 528, (1934).
- <sup>10</sup> E. Schmutzer, *Z. Phys. Chem. (Leipzig)*, **204**, p. 131, (1955).
- <sup>11</sup> J. H. Hildebrand and R. L. Scott, *The Solubility of Non-Electrolytes*, 3rd Ed., Reinhold, New York, p. 20, (1950).
- <sup>12</sup> G. Jones and W. A. Ray, "The Surface Tension of Electrolytes as a Function of the Concentration II," *J. Amer. Chem. Soc.*, **63**, pp. 288, (1941).
- <sup>13</sup> *Lange's Handbook of Chemistry*, 11th Ed., **10**, McGraw-Hill Co., New York, pp. 267-8, (1973).
- <sup>14</sup> A. M. Goodman, Private Communication.

# Solar-Cell Capacitance

A. R. Moore

RCA Laboratories, Princeton, N. J. 08540

**Abstract**—The silicon photovoltaic solar cell when in the dark exhibits the ordinary Schottky barrier capacitance for abrupt junctions. When illuminated with no external biasing (the ordinary condition for use as a solar power generator), the capacitance depends on whether the cell is open or short circuited. The short-circuit condition is most interesting, for it leads to an extra capacitance proportional to the light intensity and dependent on its spectral distribution. A model is proposed that attributes the effect to a voltage modulation of the space-charge width and the subsequent redistribution of the minority carriers in the base. Comparison with experimental data on several types of cells shows that the correct magnitude, spectral dependence, and effect of circuit test frequency are predicted.

## 1. Introduction

The photovoltaic solar cell is likely to assume great importance in the future harnessing to technical purposes of our one universal source of inexhaustible energy—the sun. While by no means the only way to convert the sun's energy to useful purposes, the photovoltaic cell converts solar to electrical energy directly. It is, therefore, a prime contender for power generation. The characteristics sought and frequently cited are open-circuit voltage, short-circuit current, fill factor, quantum efficiency, and spectral response, since all these go into determination of the ultimate power-conversion efficiency.

Yet sometimes it is forgotten, or at least imperfectly remembered, that the solar cell is a diode like any other. Its characteristics as a power converter derive from extensions of its characteristics as a

diode under illumination. In most ordinary diodes, measurements of diode capacitance, under both forward and reversed bias, are common. The use of capacitance measurements in solar cells has mostly been limited to reverse-bias conditions, the results being used to determine the base-doping profile. Probably the reason for the rarity of capacitance techniques in ordinary solar-cell testing is that solar cells are used with essentially steady currents. The capacitance is not an important operating characteristic of the cell. But the determination of capacitance can yield useful information and insight into the operation of these specialized p-n junctions.

## 2. Capacitance in the Dark

Most solar cells are made by a shallow diffusion of impurity into a moderately doped base of the opposite conductivity type, resulting in an abrupt junction.<sup>1</sup> n+ on p cells have been favored due to improved radiation resistance, although this property is not important for cells designed for terrestrial use. As has been mentioned, measurement of capacitance under reverse bias in the dark yields the base doping level through the application of the Schottky formula for barrier capacitance per unit area:

$$C_d = \epsilon_s/x_j = \{q\epsilon_s N_a/2(\phi + V)\}^{1/2}, \quad [1]$$

The junction space-charge width  $x_j$  is given by

$$x_j = \{2\epsilon_s(\phi + V)/qN_a\}^{1/2}. \quad [2]$$

In these equations conventional notation is used:  $\epsilon_s$  is the dielectric permittivity,  $\phi$  the built-in voltage,  $V$  the applied voltage,  $N_a$  the doping level (here assumed acceptor), and  $q$  the electronic charge. Fig. 1 shows a typical plot made on a commercial cell. The measured capacitance values are normally high because of the large area of solar cells. The conventional plot of  $1/C^2$  versus  $V$  gives  $N_a$  from the slope and  $\phi$  from the intercept. Both these quantities are important in solar-cell design, the first because of its relationship to minority carrier lifetime and base series resistance, the second because it approximates the maximum possible open circuit voltage at high light intensities.

With the cell still in the dark, we may now bias in the forward direction. This brings in additional capacitance incorporated into the forward biased admittance.<sup>2</sup> This additional capacitance is commonly called diffusion capacitance, because it arises from the diffusive flow

of injected minority carriers in the base. It is directly related to the lifetime of minority carriers and is, therefore, of great importance in solar cells.

Diffusion capacitance is conventionally calculated in either of two ways. In one, the time-dependent diffusion equation in the base is solved with an ac (small-signal) voltage on the emitter (surface doped layer in a solar cell) as the driving function, the ac current at the junction ( $x = 0$ ) is evaluated and divided by the ac voltage to give the forward biased admittance, whose imaginary part is the diffusion capacitance. In the other method, the stored charge in the base is calcu-

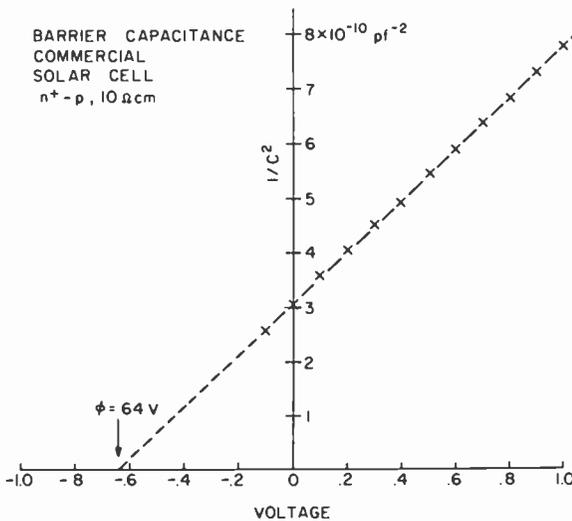


Fig. 1—Schottky barrier capacitance of a solar cell under reverse bias.  $n^+ - p$  commercial cell, 10 ohm-cm nominal. Slope determines  $N_D$ , intercept gives  $\phi$ .

lated by integrating the solution to the time-independent diffusion equation at a fixed current and evaluating the capacitance directly as  $C = dQ/dV$ . We use the first method in connection with forward-biased capacitance in the dark, and the second for the photocapacitance under illumination.

The problem solved in most articles on the subject of diode admittance is the case in which base thickness  $W \gg L$ , the minority-carrier diffusion length. Some solar cells fit this condition, especially those made with ohmic contacts on thick base layers. The differential equation to be solved is

$$\frac{\partial P_n}{\partial t} = -\frac{(P_n - P_0)}{\tau} + D \frac{\partial^2 (P_n - P_0)}{\partial x^2}, \quad [3]$$

subject to the boundary condition  $P = P_n - P_0 = 0$  as  $x \rightarrow \infty$ . If we assume an ac variation  $\tilde{P}_n = P_n \exp\{i\omega\tau\}$ , the solution is

$$\tilde{P}_n = \tilde{P}_0 \exp\left\{-\frac{x}{\sqrt{D\tau^*}}\right\}, \quad [4]$$

where  $\tau^* = \tau/(1 + i\omega\tau)$  and  $\tilde{P}_0$  is the ac value of  $P$  at  $x = 0$ . Evaluation of the ac current and the ac voltage at  $x = 0$  then leads to the admittance

$$Y = \frac{\tilde{i}}{\tilde{v}} = \frac{q}{AkT} I \sqrt{1 + i\omega\tau}. \quad [5]$$

The factor  $A$  has been incorporated to account for the fact that the silicon diffused diodes (and others) do not obey the ideal Shockley diffusion theory, and  $I$  is the dc current. For low frequencies ( $\omega\tau \ll 1$ ) expanding the square root is permissible:

$$Y_L = \frac{q}{AkT} I \left(1 + \frac{i\omega\tau}{2}\right) = G + i\omega C_{diff}, \quad [6]$$

where

$$G = \frac{q}{AkT} I \text{ and } C_{diff} = G\tau/2 \quad [7]$$

The real term in  $Y_L$  is the conductance and the imaginary susceptive term defines the diffusion capacitance. Therefore, a single measurement on an admittance bridge gives the essential parameters  $A$  and  $\tau$ . It is practical to obtain them as functions of forward bias current, as shown in Fig. 2. A major advantage of the admittance method for lifetime determination is that exceedingly small lifetimes can be measured with conventional bridges, as in thin-film silicon-on-sapphire cells. We have had no difficulty in measuring lifetimes down to 1 nanosecond, a difficult task with commonly used pulse methods for lifetime determination.

However, some modern solar cells do not meet the boundary conditions assumed for the solution given in Eq. [6]. In some cells made by epitaxy on heavily doped silicon substrates, the base layer is thin,  $W \approx L$ , and the cell efficiency is maintained by providing a low recombination surface, a so-called minority carrier  $n^+ - n$  reflector, at  $X = W$ . In order to make the admittance method useful for lifetime deter-

mination in this thin base and variable surface recombination situation, Eq. [3] must be solved for the new boundary conditions at  $X = W$ :  $(-1/P)(dP/dX) = S/D$ . No conditions are imposed on  $W$ . Proceeding in the same way, we find for the admittance

$$Y = G\sqrt{1 + i\omega\tau} \left[ \frac{\tanh W/L^* + SL^*/D}{1 + (SL^*/D) \tanh W/L^*} \right] / \left[ \frac{\tanh W/L + SL/D}{1 + (SL/D) \tanh W/L} \right]. \quad [8]$$

Here  $L^* = L/\sqrt{1 + i\omega\tau}$ , i.e.,  $L^{*2} = D\tau^*$  and  $S$  is the surface recombination velocity of the boundary at  $X = W$ .  $SL/D$  is the dimensionless measure of surface recombination.

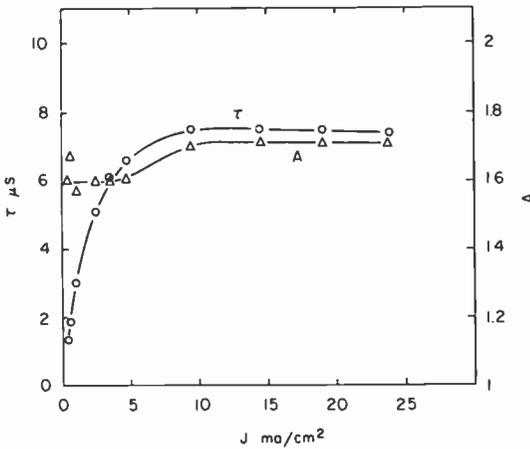


Fig. 2—Dependence of minority carrier lifetime and  $A$  factor on forward bias current as determined from admittance measurements. Experimental solar cell, base resistance 30 ohm-cm.

If  $W/L \gg 1$ , Eq. [8] reverts to Eq. [5], that is to say, the surface recombination is unimportant if the base thickness is sufficiently great. Since it is tedious to separate Eq. [8] into real and imaginary parts without approximations, the equation was separated numerically by computer and the imaginary part representing the diffusion capacitance plotted in Fig. 3. If the diffusion capacitance is written as  $C_{diff} = G\tau M/2$ , the correction factor  $M$  is found from the graph for any value of  $W/L$  and  $SL/D$ . We observe that  $M \rightarrow 1$  as  $W/L \rightarrow \infty$  independent of  $SL/D$ , as already mentioned above.

The effectiveness of the doped substrate in reducing  $S$  can be independently assayed if an additional independent lifetime measurement is combined with the capacitance data. For example, the pulse reverse recovery diode method measures lifetime by reverse biasing the junction after a forward pulse current input and follows the decay of the stored charge. Because of the high minority-carrier gradient near the junction, the measured lifetime in that method is far less influenced by conditions at the back boundary  $X = W$  than the admitt-

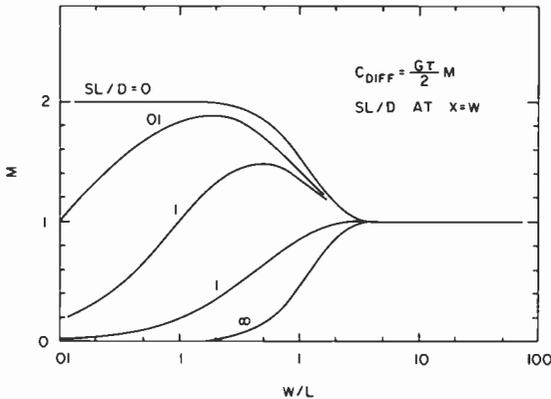


Fig. 3—Diffusion capacitance calculated for variable base width ( $W/L$ ), parameter  $SL/D$  at  $X = W$ .  $\tau$  is the bulk minority carrier lifetime and  $G$  is the conductance at any given forward current.

tance method. If the reverse recovery method is then used to estimate  $W/L$ , the observed value of  $M$  by admittance can be used to obtain  $SL/D$ . Using this approach, it has been demonstrated that the  $n^+ - n$  barrier at  $X = W$  is indeed very effective in reducing surface recombination. Typical values range around  $S \approx 10$  cm/sec.

### 3. Capacitance While Illuminated

Now we consider what happens to capacitance if the cell is illuminated. The results depend on the dc circuit of the solar cell. An illuminated photovoltaic cell in the open-circuit condition by definition passes no current in the external circuit. The entire photocurrent is opposed internally by the forward bias current corresponding to the open-circuit voltage. Thus the cell acts capacitatively equivalent to a junction that is forward biased in the dark to an equivalent current. The illumination, per se, adds nothing to the capacitance effects. Eq.

[7] or [8] (depending on boundary conditions) applies. If we wish, we may substitute the open-circuit voltage for the dc current through the V-I relationship of the diode, but nothing is actually gained.

On the other hand, if the cell is short circuited for dc during illumination, the dc voltage is zero, no forward injection occurs, and it would seem from the foregoing that the capacitance would correspond to only the Schottky-barrier capacitance of the unbiased junction in the dark. W. B. Berry<sup>3</sup> pointed out that this is not quite the case. He measured short-circuit capacitance under illumination and found that it could be represented as

$$C_{sc} = C_d(1 + \delta), \quad [9]$$

where  $C_d$  is the Schottky capacitance in the dark and  $\delta$  is proportional to light intensity (or short-circuit current). For intensities up to 1 sun,  $\delta < 1$ . A model for this effect has been published,<sup>4</sup> and it is some aspects of this model and application to specific cases that will be discussed here.

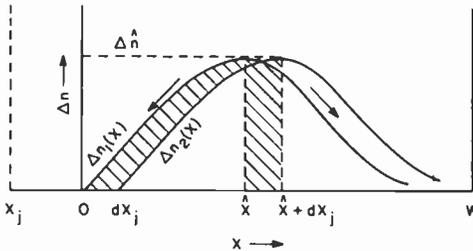


Fig. 4—Distribution of minority carriers in the base of an illuminated solar cell under short-circuit conditions showing the effect of a shift in the space-charge edge from  $X = 0$  to  $X = dx_j$ .

Even though the illuminated cell is short circuited for dc, the circuit must be arranged so that an ac test voltage can be applied. The ac signal modulates the depletion layer width. According to the model, the observed photocapacitance arises from the charge exchange across the junction needed to effect this modulation in the presence of the dc minority-carrier distribution set up by the illumination. In order to calculate this special form of diffusion capacitance we used the second, or charge-storage, method.

Suppose light of a given intensity and spectral distribution incident on a cell has set up a certain excess minority carrier distribution  $\Delta n_1(x)$  in the base as shown in Fig. 4. The usual boundary condition

that the excess minority carrier concentration is zero at the space-charge edge,  $\Delta n_1(0) = 0$ , is assumed. The boundary condition at  $X = W$  is left unspecified for the moment. From Eqs. [1] and [2], a small signal change of applied voltage  $dV$  results in a space-charge widening  $dX_j$  such that

$$dX_j/dV = \epsilon_s/qN_aX_j = C_d/qN_a. \quad [10]$$

The space-charge edge must shift to  $X = dX_j$ . To accommodate this change and maintain the boundary condition a new distribution  $\Delta n_2(X)$  results with  $\Delta n_2(dX_j) = 0$ , and in the process a charge  $dQ$  is swept out by the junction. Since  $W \gg dX_j$ , the new distribution is simply the old distribution shifted by  $dX_j$ , that is,  $\Delta n_2(X) = \Delta n_1(X - dX_j)$ . Both distributions have the same maximum value of  $\Delta n$  which we call  $\Delta \hat{n}$ , the first occurring at  $\hat{X}$ , the second at  $\hat{X} + dX_j$ .

To calculate the charge that crosses the junction, we can write generally

$$dQ = q \int_0^{\hat{X}} \Delta n_1(X) dX - q \int_{dX_j}^{\hat{X} + dX_j} \Delta n_2(X) dX. \quad [11]$$

This is the shaded area on the left of Fig. 4. As long as the distribution does not change rapidly in a distance  $dX_j$ ,  $\Delta n_2(X)$  can be represented to a good approximation by the expansion

$$\Delta n_2(X) = \Delta n_1(X - dX_j) \approx \Delta n_1 - \frac{d\Delta n_1(X)}{dX} dX_j.$$

Since  $\hat{X} \gg dX_j$  and  $\Delta n_2(X) = 0$  for  $0 < X < dX_j$ , we may to the same degree of approximation take the limits of integration of the second integral in Eq. [11] as 0 and  $\hat{X}$ . Then Eq. [11] becomes

$$\begin{aligned} dQ &= q \int_0^{\hat{X}} \Delta n_1(X) dX - q \int_0^{\hat{X}} \left[ \Delta n_1(X) - \frac{d\Delta n_1(X)}{dX} dX_j \right] dX \\ dQ &= q\Delta \hat{n} dX_j, \end{aligned} \quad [12]$$

It will be recognized that Eq. [12] can be simply represented geometrically by the shaded area on the right of Fig. 4, which is equivalent to the shaded area on the left.

The photocapacitance resulting from this charge exchange is then

$$\Delta C_{sc} = dQ/dV = q\Delta \hat{n} dX_j/dV. \quad [13]$$

and using Eq. [10],

$$\Delta C_{sc} = C_d \Delta \hat{n} / N_a. \quad [14]$$

Defining

$$\delta = \Delta \hat{n} / N_a = \Delta C_{sc} / C_d, \quad [15]$$

$$C_{sc} = C_d + \Delta C_{sc} = C_d(1 + \delta), \quad [16]$$

which is the same as Eq. [9] with the interpretation of  $\delta$  given by Eq. [15]. Since  $\Delta \hat{n}$  is proportional to light intensity, Eq. [16] or [9] says that the increase in capacitance of an illuminated cell under short-circuit conditions is proportional to the incident light. Fig. 5 shows how well this law is obeyed for an ordinary silicon cell.

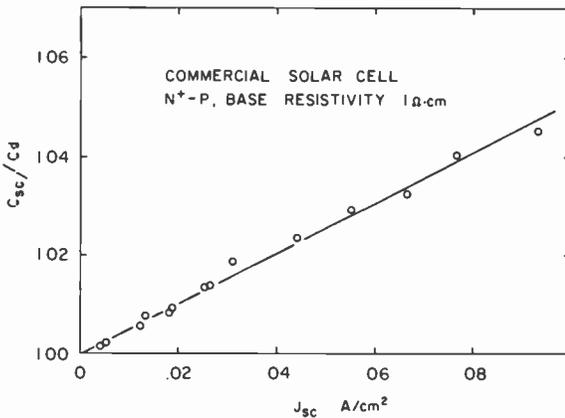


Fig. 5—Capacitance of short circuited illuminated solar cell relative to dark capacitance as a function of light intensity measured as short-circuit current.

Eqs. [14], [15], and [16] are general. To proceed further with a quantitative calculation of the photocapacitance according to the depletion-layer modulation model requires a knowledge of  $\Delta \hat{n}$ . This depends on the specific conditions of the cell and the light source and the boundary condition at the rear contact. We assume in all cases that the boundary condition at the junction is  $\Delta n(0) = 0$ . The equation to be solved is the steady-state diffusion equation for minority carriers with a pair generation function corresponding to the absorbed light. The well-known differential equation is

$$D \frac{d^2 \Delta n}{dX^2} - \frac{\Delta n}{\tau} + N \alpha e^{-\alpha X} = 0. \quad [17]$$

$N$  is the incident light intensity in photons/cm<sup>2</sup>/sec and  $\alpha$  is the absorption coefficient, both functions of wavelength  $\lambda$ .

If the base thickness  $W$  is thick compared to the minority diffusion length  $L$  ( $W \gg L$ ) and the illuminant is weakly absorbed light ( $\alpha W \ll 1$  hence  $\alpha X \ll 1$ ) the solution is simple:  $\Delta \hat{n} = N\alpha\tau$ . The solution in the form best applicable to quantitative comparison with photo-capacitance data has already been given:<sup>4</sup>

$$\frac{\delta}{J_{sc}} = \frac{L}{qDN_a} = (q/kT)(\mu_p/\mu_n)\rho_p L \quad [18]$$

Here  $J_{sc}$  is the short-circuit current density obtained from  $J_{sc} = qD(d\Delta n/dX)|_{X=0}$  and  $\rho_p$  is the base resistivity (assumed here p-type). The other symbols have their conventional meanings. Since both  $\delta$  and  $J_{sc}$  are proportional to light intensity, Eq. [18] is independent of light level and contains only material constants.

In many practical cells  $W$  is not large compared to  $L$ . The back contact, therefore, strongly influences the result. The two most important cases can be idealized to correspond to (1) no recombination at  $X = W$  as for the  $n - n^+$  boundary and (2) infinite surface recombination at  $X = W$  as for an ohmic contact.

Case 1:  $\Delta n = 0$  at  $X = 0$ ,  $d\Delta n/dX = 0$  at  $X = W$ .

The solution to Eq. [17] is

$$\Delta n = \frac{N\alpha\tau}{\alpha^2 L^2 - 1} \left\{ \frac{\cosh((W - X)/L) - \alpha L e^{-\alpha W} \sinh(X/L)}{\cosh(W/L)} - e^{-\alpha X} \right\}. \quad [19]$$

To obtain  $\hat{X}$  we set  $d\Delta n/dX = 0$ , obtaining

$$\frac{\sinh((W - \hat{X})/L) + \alpha L e^{-\alpha W} \cosh(\hat{X}/L)}{\alpha L \cosh(W/L)} - e^{-\alpha \hat{X}} = 0. \quad [20]$$

Since it is impractical to solve explicitly for  $\hat{X}$ , Eq. [20] was solved numerically by computer. The value of  $\hat{X}$  so obtained substituted back into Eq. [19] gives  $\Delta \hat{n}$ . The short circuit current  $J_{sc}$  is also obtained from Eq. [19].

$$J_{sc} = qD \left. \frac{d\Delta n}{dX} \right|_{X=0} = \frac{qD(N\alpha\tau)}{L(\alpha^2 L^2 - 1)} \left\{ \frac{\sinh(W/L) + \alpha L e^{-\alpha W}}{\cosh(W/L)} - \alpha L \right\}. \quad [21]$$

Then

$$\frac{\delta}{J_{sc}} = \left[ \frac{L}{qDN_a} \right] \frac{\left\{ \frac{\cosh((W - \hat{X})/L) - \alpha Le^{-\alpha W} \sinh(\hat{X}/L)}{\cosh(W/L)} - e^{-\alpha \hat{X}} \right\}}{\left\{ \frac{\sinh(W/L) + \alpha Le^{-\alpha W}}{\cosh(W/L)} + \alpha L \right\}} \quad [22]$$

in which it is understood that  $\hat{X}$  is the root of Eq. [20]. We observe that the first factor in square brackets is identical to the solution of the simple case represented by Eq. [18]. The remaining fraction represents the effects of finite  $W$ , the boundary condition at  $X = W$ , and the optical absorption coefficient.

Case 2:  $\Delta n = 0$  at  $X = 0$  and at  $X = W$ .

The solution to Eq. [17] is

$$\Delta n = \frac{N\alpha\tau}{\alpha^2 L^2 - 1} \left\{ \frac{\sinh((W - X)/L) - e^{-\alpha W} \sinh(W/L)}{\sinh(W/L)} - e^{-\alpha X} \right\}. \quad [23]$$

$\hat{X}$  is the root of

$$\frac{\cosh((W - \hat{X})/L) - e^{-\alpha W} \cosh(\hat{X}/L)}{\sinh(W/L)} - \alpha Le^{-\alpha \hat{X}} = 0. \quad [24]$$

obtained numerically as before. The short-circuit current

$$J_{sc} = \frac{qD(N_a\tau)}{L(\alpha^2 L^2 - 1)} \left\{ \alpha L - 1 + \frac{e^{\alpha W} - e^{-W/L}}{\sinh(W/L)} \right\}. \quad [25]$$

Then

$$\frac{\delta}{J_{sc}} = \left[ \frac{L}{qDN_a} \right] \frac{\left\{ \frac{\sinh((W - \hat{X})/L) + e^{-\alpha W} \sinh(W/L)}{\sinh(W/L)} - e^{-\alpha \hat{X}} \right\}}{\left\{ \alpha L - 1 + \frac{e^{\alpha W} - e^{-W/L}}{\sinh(W/L)} \right\}} \quad [26]$$

The computer generated solutions for case 1 (Eq. [22]) and case 2 (Eq. [26]) are presented in Figs. 6 and 7, respectively. One further case, which is really a modification of case 1, has also proved of interest. Long-wavelength light partially transmitted through the base may be reflected backward to traverse the base again if the rear elec-

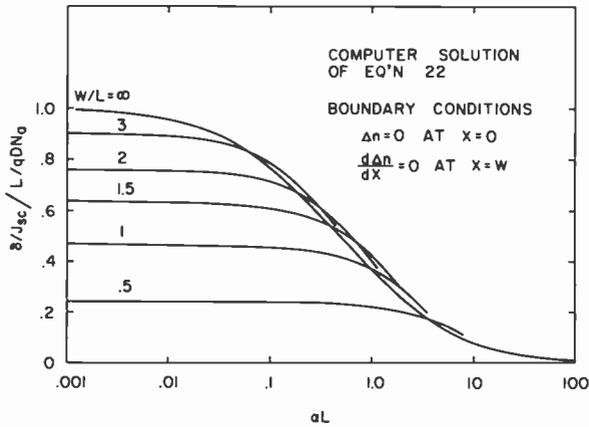


Fig. 6—Computation of  $\delta/J_{sc}$  according to Eq. [22]. Illuminated solar cell, boundary conditions:  $\Delta n = 0$  at  $X = 0$ ,  $d\Delta n/dX = 0$  at  $X = W$ .

trode is smooth and reflecting. This changes the minority carrier distribution, and thus  $\delta/J_{sc}$ , if the boundary condition at  $X = W$  allows appreciable build-up of minority charge there, as in case 1. Since the base may be formed epitaxially on a doped substrate, the reflection need not necessarily occur at  $X = W$ , but at  $X = T$ , where  $T$  is the total thickness of base plus substrate. Without going into details of solution, we present one such situation in Fig. 8, computed similarly from a differential equation modified to take reflection into account. What is plotted in the first two cases is  $\delta/J_{sc}$  divided by  $[L/qDN_a]$

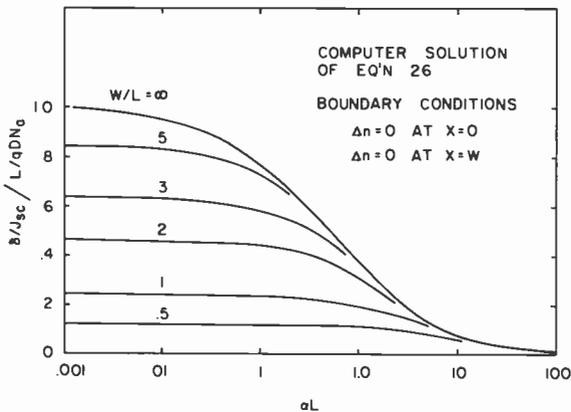


Fig. 7—Computation of  $\delta/J_{sc}$  according to Eq. [26]. Illuminated solar cell, boundary conditions:  $\Delta n = 0$  at  $X = 0$ ,  $\Delta n = 0$  at  $X = W$ .

versus the dimensionless variable  $\alpha L$ , with parameter  $W/L$ . For the reflection case,  $\delta/J_{sc}$  is presented relative to  $\delta/J_{sc}$  at  $\alpha L = 0$  (for each value of  $W/L$ ) versus  $\alpha L$ .

All three figures have the common characteristic that  $\delta/J_{sc}$  decreases as  $\alpha L$  increases. The physical reason for this is that as the absorption coefficient rises, the maximum of  $\Delta n$  moves toward the junction. For any given magnitude of  $\Delta \hat{n}$ , the photocapacitance is

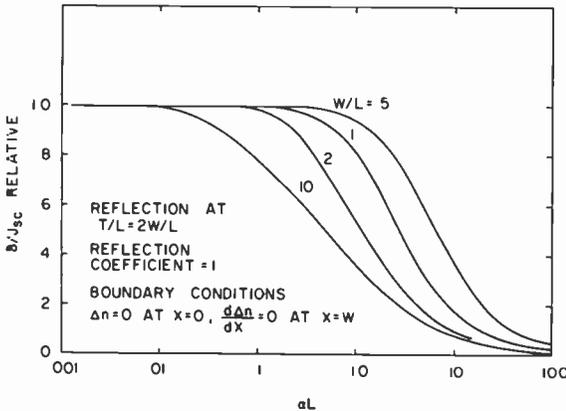


Fig. 8—Computation of  $\delta/J_{sc}$  for illuminated solar cell including the effects of light reflection at the back surface. Cell thickness, including substrate is  $T$ . Base thickness is  $W$ . Reflection occurs at  $T/L = 2W/L$  with reflection coefficient  $R = 1$ . Boundary conditions:  $\Delta n = 0$  at  $X = 0$ ,  $d\Delta n/dX = 0$  at  $X = W$ .  $\delta/J_{sc}$  is given relative to  $\delta/J_{sc}$  at  $\alpha L = 0$  for each value of the parameter  $W/L$ .

constant (Eq. [14]), while the short-circuit current increases with the gradient of  $\Delta n$  at  $X = 0$ . Thus the ratio  $\delta/J_{sc}$  falls as  $\alpha$  increases.

The degree to which the experimental data fit the theory is illustrated by Fig. 9. Since the theory assumes perfectly shorted junctions while experimentally there is always a finite resistance in the junction circuit, the data have been corrected for this. The correction formula is

$$\frac{\delta_n}{J_{sc}} = \frac{\delta_{meas}}{J_{sc}} - \frac{AR}{2\phi} \quad [27]$$

where  $\delta_0$  is the corrected  $\delta$ ,  $\delta_{meas}$  is the measured  $\delta$ ,  $A$  is the diode area,  $R$  is the series resistance composed of resistance internal and external to the diode, and  $\phi$  is the built-in voltage. The assumption has been made that the IR drop is not more than a few  $kT/q$ . Since Fig. 9 represents the comparison of theory with experiment with no adjustable constants, the agreement appears satisfactory.

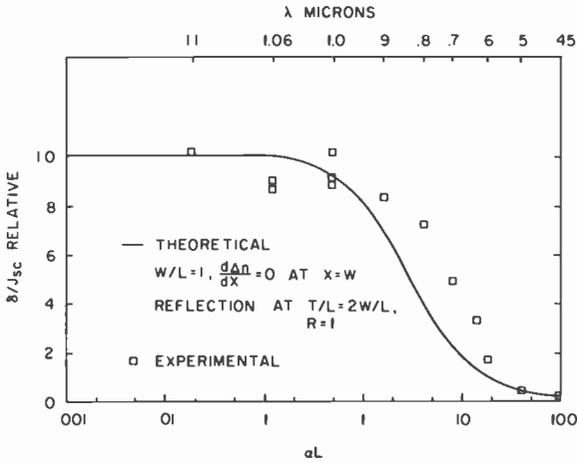


Fig. 9—Comparison of experimental determination of photocapacitance with computation. Calculated curve is from Fig. 8 with  $W/L = 1$ ,  $T/L = 2W/L$ . Data points are from an experimental solar cell made by epitaxy and ion implantation. Base 45 ohm-cm n type,  $L = 50 \mu\text{m}$ ,  $W = 50 \mu\text{m}$ , n+ substrate  $50 \mu\text{m}$  thick.

So far in all our discussion of diffusion capacity, both under forward bias in the dark and short circuited under illumination, we have used the low-frequency approximation ( $\omega\tau \ll 1$ ). Eq. [5] permits the

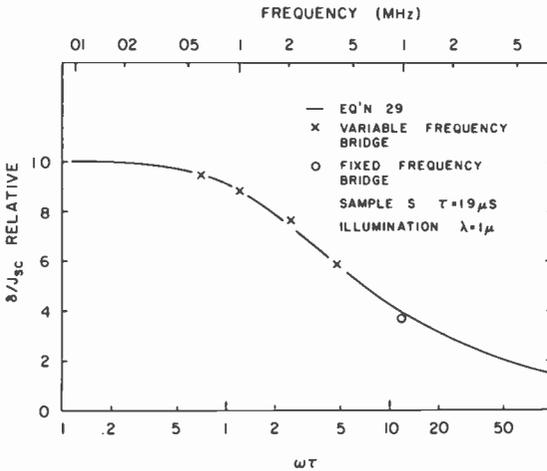


Fig. 10—Frequency dependence of diffusion capacitance. Solid line from Eq. [29]. Data are measurements of short circuit capacitance of an illuminated experimental solar cell expressed as  $\delta/J_{sc}$ .

calculation of the dependence of the diffusion capacity on the test frequency. Let the admittance be represented by a conductance plus a susceptance

$$Y = G(1 + i\omega\tau)^{1/2} \equiv g + i\omega C_d.$$

Square both sides and equate real and imaginary parts

real part:

$$G^2 = g^2 - \omega^2 C_d^2,$$

imaginary:

$$G^2\tau = 2gC_d.$$

Regarding  $G$  as known and  $g$  and  $C_d$  as unknown, solve simultaneously. This leads to a quartic equation in  $C_d$  which is easily solved to give

$$C_d = \frac{(g/\omega)}{\sqrt{2}}[(1 + \omega^2\tau^2)^{1/2} - 1]. \quad [28]$$

For comparison purposes, it is convenient to represent this as a ratio of  $C_d$  at frequency  $\omega$  to  $C_d$  at low frequency ( $\lim \omega\tau \rightarrow 0$ );

$$\frac{C_d(\omega)}{C_d(0)} = \sqrt{2} \left[ \frac{(1 + \omega^2\tau^2)^{1/2} - 1}{\omega^2\tau^2} \right]^{1/2} \quad [29]$$

This function is plotted in Fig. 10. The reason that diffusion capacitance falls off when  $\omega\tau > 1$  is that the redistribution of minority carriers in the base in response to the ac test voltage across the diode does not occur instantaneously, but takes of the order of one lifetime. This is true for biased diodes in the dark or for photocapacitance. A comparison with photocapacitance data ( $\delta/J_{sc}$ ) taken under illumination by light of  $\lambda \approx 1\mu$  at various test frequencies shows that the predicted frequency dependence is well obeyed. This could be used as another method for lifetime determination if  $\tau$  were regarded as an unknown and the best  $\omega\tau$  needed to fit the curve were extracted. Obviously, the method would also work with ordinary diffusion capacitance in biased diodes in the dark.

#### References:

- <sup>1</sup> P. Rappaport, "The Photovoltaic Effect and Its Utilization," *RCA Review*, 20, p. 373 (1959).
- <sup>2</sup> W. Shockley, *Electrons and Holes in Semiconductors*, Van Nostrand, Princeton, N.J. (1950).
- <sup>3</sup> W. B. Berry, *Appl. Phys. Lett.*, 25, p. 195 (1974).
- <sup>4</sup> A. R. Moore, *Appl. Phys. Lett.*, 27, p. 26 (1975).

# Basic Concepts of Charge-Coupled Devices

W. F. Kosonocky and J. E. Carnes

RCA Laboratories, Princeton, N.J. 08540

**Abstract**—Key concepts involved in the operation and design of charge-coupled devices (CCD's) are discussed and described in simple physical terms. These include the relationship between speed and device dimensions, the key concept of complete charge transfer versus the bucket-brigade mode of transfer, a simplified discussion of buried-channel operation, and a qualitative explanation of the available techniques for the introduction of the charge signal into CCD's and output detection methods.

## 1. Introduction

The concept of charge-coupled devices (CCD's) was described by W. S. Boyle and G. E. Smith of Bell Telephone Laboratories in 1970. While the CCD is based upon the familiar MOS capacitor, its operation involves several new concepts. A great deal of literature has been generated over the past few years describing the operation and performance of this device, but much of it involves somewhat complicated mathematical analyses.

Here, we review some of the more important physical concepts involved in the operation of CCD's. The paper is intended primarily for the reader who desires to gain a qualitative insight into the operation, performance limitations, and design of these devices.

## 2. Charge Transfer Concepts

Charge-coupled devices (CCD's) operate by transferring packets of minority charge, which represent analog signals, from one potential well to another. These potential wells are formed by a linear array of deep-depleted MOS capacitors, either on a uniformly doped substrate (surface channel, SCCD) or on a substrate with a thin, depleted layer of opposite conductivity at the surface (buried channel, BCCD). In either case, the minimum potential energy of these wells is determined by the voltage applied to the gate electrode, so that the appropriate manipulation of the phase voltages causes charge packets to transfer along the line, since charge always moves to the local potential minimum, as shown in Fig. 1.

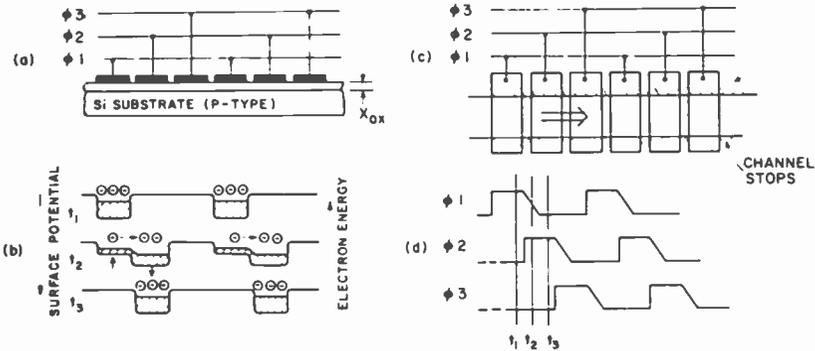


Fig. 1—Basic charge-transfer action with a three-phase, n-surface-channel CCD. (a) Cross-sectional view, (b) surface potential profile at three different times, (c) top view, and (d) clock waveforms.

For a surface channel device, the potential at the surface, i.e., the potential well depth, is related to the gate voltage and other device parameters by the equation

$$V_s = V_G' - B \left[ \left( 1 + \frac{2V_G'}{B} \right)^{1/2} - 1 \right], \quad [1]$$

where

$$V_G' = V_G + \frac{Q_{ss}}{C_{ox}}$$

$$B = \frac{qN_D \epsilon_{ss} x_{ox}^2}{\epsilon_{ox}^2}$$

- $V_G$  = gate voltage  
 $Q_{ss}$  = oxide charge per unit area  
 $C_{ox}$  = oxide capacitance per unit area  
 $q$  = electronic charge in coulombs  
 $N_D$  = substrate doping level in  $\text{cm}^{-3}$   
 $\epsilon_s$  = silicon dielectric constant in F/cm  
 $\epsilon_{ox}$  = oxide dielectric constant in F/cm  
 $x_{ox}$  = oxide thickness in cm.

When charge gathers in this well it resides at the Si-SiO<sub>2</sub> interface and reduces the well depth according to

$$\Delta V_s = \frac{Q_{s,K}}{C_{ox} + C_d} \quad [2]$$

where  $C_{ox}$  is the oxide capacitance and  $C_d$  is the depletion-layer capacitance.  $C_d$  depends upon the signal level but is usually small compared to  $C_{ox}$ , which is independent of signal. Thus the signal level can be estimated using Eq. [1] to calculate the well depth and assuming a constant well capacitance equal to  $C_{ox}$ .

The most important figure of merit for a charge-transfer device is charge-transfer efficiency denoted by  $\eta$ , the fraction of the original charge packet that is transferred from one storage site to the next. The fraction that is not transferred is denoted by  $\epsilon$ , the charge-transfer inefficiency, or charge-transfer loss. Since the potential well used to store and transfer minority carriers also serves to repel majority carriers, recombination is negligible. Thus, minority carriers are lost from the original charge packet only by being left behind; i.e., charge is either transferred or it is not. Thus,

$$\eta + \epsilon = 1. \quad [3]$$

The effect of imperfect transfer efficiency is to erode the amplitude of the signal packet, so that after  $n$  transfers, the ratio of the signal level  $A_n$  to the original level  $A_0$  is given by

$$A_n/A_0 = \eta^n = (1 - \epsilon)^n \approx \exp(-\epsilon n) \quad [4]$$

for small  $\epsilon$ .

Since many CCD applications require in excess of 1000 transfers,  $\epsilon$  must be very small, on the order of  $10^{-4}$  or  $10^{-5}$ , to prevent excessive signal degradation.

Two different effects prevent the attainment of perfect efficiency—the time required to transfer the free charge in the packet from one

site to the next, and trapping effects. First let us consider the limitations on clock frequency and efficiency imposed by the intrinsic process of free-charge transfer. Fig. 2 illustrates the situation that exists at some time during the charge-transfer process. Charge originally stored under the middle gate flows into the deeper potential well on the right. It is important that no potential barrier exist between the two potential wells, so that all of the charge can be transferred. Several driving mechanisms cause the free charge to transfer, including charge repulsion, thermal diffusion, and drift in the externally applied fields (fringing fields). Charge-repulsion effects are unimpor-

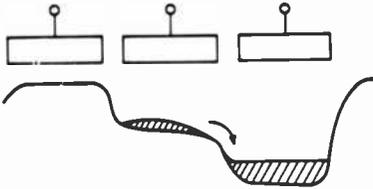


Fig. 2—Complete charge transfer. There are no barriers to hinder the flow of charge into the deeper potential well.

tant after most ( $\sim 99\%$ ) of the charge has transferred, so that the limitations on device speed are essentially determined by either thermal diffusion or fringing field drift. Both of these mechanisms cause an exponential decay of charge under the transferring electrode. The magnitude of the time constants for this decay determines the efficiency that can be achieved at any particular clock frequency, since the clock frequency determines the time available for transfer.

The time constant for thermal diffusion can be estimated by simple physical arguments. The driving force is the thermal voltage  $kT/q$  applied across the gate length  $L$ . The transit time is therefore

$$\tau_{th} \approx \frac{L}{\mu E} = \frac{L}{\mu kT/qL} = \frac{L^2}{D} \quad [5]$$

More rigorous treatments<sup>7</sup> indicate that this time is  $L^2/2.5D$ . To achieve efficiencies of 99.99% ( $\epsilon = 10^{-4}$ ) requires about ten time constants ( $10^{-4} \approx e^{-10}$ ). Thus for typical values ( $L = 10 \mu\text{m}$  and  $D = 10 \text{ cm}^2/\text{sec}$ ), efficiencies of 99.99% can be achieved at approximately 1 MHz clock rates, assuming that thermal diffusion is the only mechanism responsible for charge transfer. This speed is insufficient for many applications.

Charge transfer is also enhanced by carrier drift in the field in-

duced by the externally applied gate voltages. The externally applied field is principally directed perpendicular to the motion of charge, but there is a component of field along the direction of motion. The calculation of these fringing fields involves a rather complex two-dimensional electrostatics problem generally requiring a numerical solution.<sup>2,3</sup> In the simplified case where the silicon substrate is treated as a semi-infinite dielectric with no depletion layer, a closed-form solution for the minimum fringing field is possible:<sup>2</sup>

$$E_{min} = 3.2 \frac{V}{L} \frac{x_{ox}}{L}, \quad [6]$$

where  $V$  is the applied gate voltage,  $x_{ox}$  is the oxide thickness, and  $L$  is the gate length. This represents the highest value that is achievable for surface-channel devices and applies in the limit of low substrate doping ( $\sim 10^{14}$  cm<sup>-3</sup>). The transit time (or decay time constant) for fringing-field drift is then given by

$$\tau_f = \frac{L}{\mu E_{min}} = \frac{L^3}{3.2 \mu V x_{ox}}.$$

For 10- $\mu$ m gates, the thermal time constant is  $10^{-7}$  sec. Assuming  $\mu = 400$ ,  $x_{ox} = 1000$  Å, and  $V = 10$  V, the fringing-field time constant is  $7.8 \times 10^{-9}$  sec, or a factor of 13 higher. Thus, for 10- $\mu$ m gate lengths and low substrate doping, fringing-field drift results in about 10–15 MHz operation. In addition because the fringing-field time constant depends upon  $L^3$ , the increase in speed over that possible with thermal diffusion increases as gate length is decreased. Also, fringing-field drift is relatively more important in buried-channel devices, since the channel is a greater distance from the gates, effectively increasing the  $x_{ox}/L$  factor in Eq. [6].

These concepts only apply for the case where no potential barriers exist between the two potential wells, so that all charge can be transferred; i.e., for the complete charge-transfer mode. When a barrier exists so that a reservoir of charge is held behind in the transferring well, the amount of charge transferred at low clock frequencies is determined by the barrier height, as shown in Fig. 3. The barrier height is partly determined by the amount of charge in the receiving well. If a large signal is present there, charge repulsion will result in a higher barrier. Therefore, less charge will be transferred for a large charge packet than for a small charge packet. This barrier-height modulation effect results in a frequency-independent transfer loss of about  $10^{-3}$  at gate lengths of 10  $\mu$ m. The transfer loss increases as gate length decreases. Thus any charge transfer device in which the

amount of charge that is transferred depends on the height of a potential barrier will have serious limitations on transfer efficiency. This operation is similar to that which occurs in a bucket-brigade device,<sup>9</sup> and therefore is referred to as the bucket-brigade mode. High transfer efficiency ( $\approx 99.99\%$ ) requires a channel potential variation with no barriers so that complete charge transfer occurs.

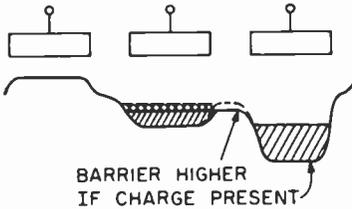


Fig. 3—Bucket-brigade mode of charge transfer. The presence of charge in the right potential well causes the barrier to be higher, which causes more charge to be left behind.

However, even in the complete transfer mode of operation there are other limitations on efficiency caused by trapping effects. In surface-channel devices, trapping and subsequent loss of carriers from the charge packets occurs in fast interface states. These states are distributed rather evenly across the bandgap with a density of about  $10^{10}$  states/cm<sup>2</sup>/eV. They all fill at the same rate depending only upon the number of carriers in the conduction band at the surface (within about 100 Å of the surface). However, the emptying rate depends upon the energy difference between the conduction band and the trap level—the rate decreasing an order of magnitude for every 0.06 eV increase in energy at room temperature. The emptying times vary from about  $10^{-11}$  sec near the band edge to about  $10^{-4}$  sec at mid-gap. Fill times are about  $10^{-10}$  sec for typical signal levels. Thus many states can fill faster than they can empty, resulting in a net loss of signal into trap states. This loss mechanism can be minimized by continuously circulating a small amount of charge ( $\sim 10\%$  of full signal level). This background charge is generally called “fat zero”, and tends to keep the fast states continuously filled so that no states are empty to trap charge when a full well signal arrives.

However, the fat zero is only effective in the deepest parts of the potential well. Since these wells have sloping sides there are regions that never “see” the fat zero. Along each side or edge of the potential well, trapping losses are not affected by fat zero, as shown schematically in Fig. 4. An analysis of the magnitude of this edge-effect trapping shows that the lowest (best) transfer inefficiency achievable for

surface channel devices is given by

$$\epsilon_{\text{edge effect}} = 3.9 \times 10^{-4} \left( \frac{1}{W_{\text{mil}}} \right) \left( \frac{N_{\text{ss}}}{10^{16}} \right) \left( \frac{10^{15}}{N_D} \right)^{1/2} \quad [8]$$

Trapping also occurs in buried-channel devices in mono-energetic bulk states.<sup>12</sup> However, the effect is much smaller, since the number of states that participate in trapping is lower. The number of trapping states per unit area is equal to the volume density times the distance over which signal charge resides in the channel. This distance is on the order of 1  $\mu\text{m}$ , so that  $10^{12}$  states/cm<sup>3</sup> would result in only

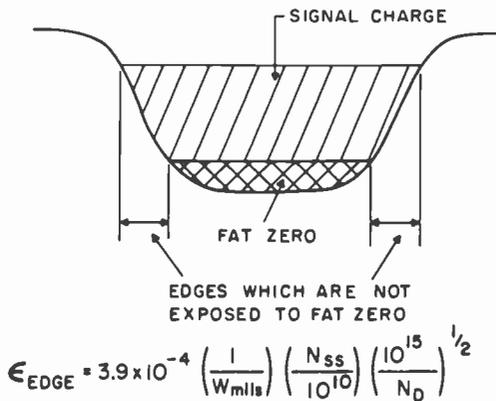


Fig. 4—Cross-sectional view of potential well (charge motion into the paper) indicating the sloping edges, which do not see fat zero. (See Eq. [8].)

about  $10^8$  states/cm<sup>2</sup> of effective trap sites, approximately two orders of magnitude lower than the  $10^{10}$  cm<sup>-2</sup> states effective in surface channel devices. Generally, buried-channel devices can achieve very high efficiencies without fat zero charge, but efficiency is improved somewhat with a small amount of fat zero, indicating that some trapping losses are present.

### 3. Buried Channel

The buried-channel CCD (BCCD)<sup>13</sup> was conceived to avoid the problems of efficiency limitations and noise associated with fast interface states. The objective is to move the potential well minimum away from the Si-SiO<sub>2</sub> interface. This is done by creating a thin oppositely doped layer at the silicon surface. A typical structure might be an n-type layer, 1  $\mu\text{m}$  thick of  $10^{16}$  cm<sup>-3</sup> doping, formed by ion implanta-

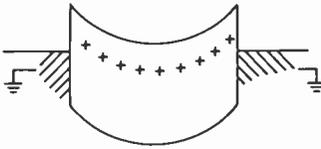


Fig. 5—Band diagram showing parabolic potential profile for a metal-insulator-metal structure with a uniform positive space charge in the insulator region. Both electrodes are grounded.

tion on a p-type substrate. When appropriate positive bias is applied to the source and drain  $n^+$  diffusions, all of the free conduction electrons are removed from the implanted layer region, leaving behind the uniform positive space charge of the donor sites. This creates a parabolic potential profile with a minimum near the middle of the layer region. This structure is similar to the potential profile in a uniform positively charged insulator with grounded electrodes on either side as shown in Fig. 5. The BCCD profile differs somewhat because of the oxide layer and the substrate depletion region, and is shown in Fig. 6. The potential difference between the minimum and the bulk

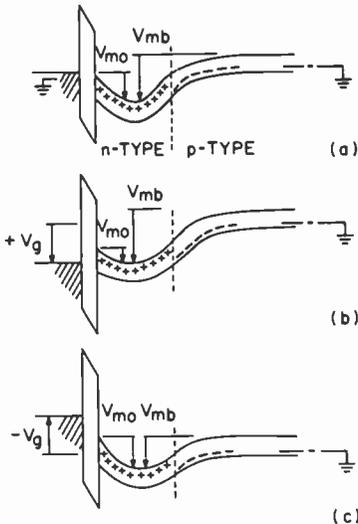


Fig. 6—Band diagram for buried-channel CCD. (a) Gate grounded ( $V_{mb}$  is potential difference between conduction band at the minimum and conduction band in the bulk and  $V_{mo}$  is the difference between the minimum and the potential at the Si-SiO<sub>2</sub> interface). (b) With positive (ON) gate voltage ( $V_{mb}$  increases,  $V_{mo}$  decreases). (c) With negative (OFF) gate voltage ( $V_{mb}$  decreases,  $V_{mo}$  increases until they are equal where pinning occurs).

(ground) will be referred to as  $V_{mb}$ . This value can be made larger by applying positive gate voltages or smaller by applying negative gate voltages (for an n-channel device). The potential difference between the conduction band at the Si-SiO<sub>2</sub> interface and the channel minimum is referred to as  $V_{mo}$ , (see Fig. 6). This value is important because charge must be kept away from the Si-SiO<sub>2</sub> interface to avoid the inefficiency and noise associated with interface states.

When positive gate voltages are applied,  $V_{mb}$  tends to increase and  $V_{mo}$  to decrease as shown in Fig. 6b. When negative gate voltage is applied  $V_{mb}$  tends to decrease and  $V_{mo}$  tends to increase, as shown in Fig. 6c. Devices are built with p<sup>+</sup> channel stops, which are physically adjacent to the channel and are at ground potential; these channel stops are a source of holes for the channel. Thus the valence band in the channel can never move above ground without attracting an unlimited supply of holes from the channel stop. At a certain negative gate voltage the valence band at the Si-SiO<sub>2</sub> reaches the same potential as the valence band in the bulk. This is called the pinning gate voltage. As the gate voltage is made more negative, the source of holes from the channel stop pins the valence band at the Si-SiO<sub>2</sub> interface at ground, and no further changes in  $V_{mo}$  or  $V_{mb}$  occur. Thus the channel minimum can never be raised above ground, and a buried-channel device with a grounded source can not be cut off by applying negative gate voltage alone. To cut off the charge flow in a BCCD, one must apply a positive potential to the source.

One of the most important factors in the design and operation of a buried-channel device is the  $V_{mb}$  and  $V_{mo}$  versus  $V_{gate}$  characteristics. Fig. 7 shows calculations of  $V_{mb}$  and  $V_{mo}$  for a typical structure having a  $1.2 \times 10^{12}$  cm<sup>-2</sup> implant of arsenic in a substrate with  $4 \times 10^{14}$  doping and an oxide thickness of 1000 Å. Several different curves are shown for different signal charge levels. For no charge, the pinning voltage is -5.0 V. No change in  $V_{mb}$  or  $V_{mo}$  occurs to the left of  $V_{pin}$ .  $V_{pin}$  is the most negative useful voltage. The most positive voltage that can be applied is determined by the  $V_{mo}$  branch of the curve. When  $V_{mo}$  becomes too small, the carriers can interact with the fast interface states with all the deleterious effects associated with surface-channel devices. Since the fill rate of fast states depends on surface density of free carriers (included within 100 Å of the surface), the surface density must be kept low enough so that the fill time constant is long compared to the clock period. If the surface density is 10<sup>6</sup> cm<sup>-2</sup> or lower, this condition is fulfilled. Thus if the channel layer doping is 10<sup>16</sup> cm<sup>-3</sup>, then a  $V_{mo}$  of about  $10 kT/q$  (0.26 V) will ensure that the surface density (within 100 Å of the surface) is less than 10<sup>6</sup> cm<sup>-2</sup>. As seen from the figure, if one-half of the maximum charge

(the maximum charge is defined as the implant dose) is present in the channel, something in excess of +10 V can be applied to the gate before  $V_{m0}$  reaches 0.25 V. A 10-volt clock varying between -5 (OFF) to +5 (ON) would result in satisfactory operation and could handle a charge level of one-half of the implant level without going into a surface-channel mode.

In surface-channel devices, two-phase devices can be built utilizing two thicknesses of oxide to provide signal directionality, since the potential wells are deeper under thinner oxides. However, in buried-

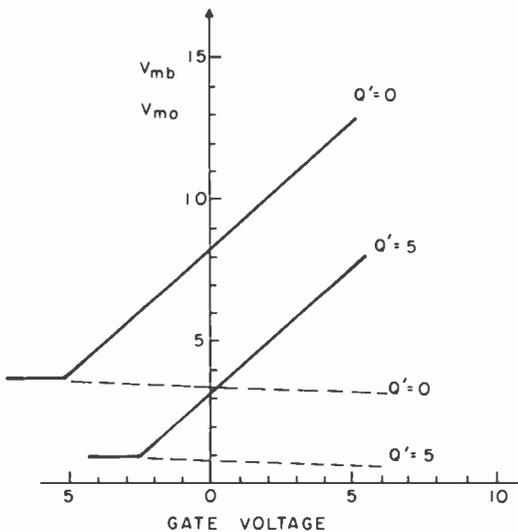


Fig. 7— $V_{mb}$  (solid lines) and  $V_{m0}$  (broken lines) plotted as a function of gate voltage for a typical BCCD structure. The horizontal region on the left corresponds to pinning.

channel devices the depth of the potential well as measured by  $V_{mb}$  decreases as the oxide thickness decreases. This occurs because as oxide thickness increases, less of the positive buried layer charge is neutralized in the gate metal and more is neutralized by the substrate depletion region. A widening depletion region implies a larger  $V_{mb}$ . This effect means that thick-oxide techniques cannot be used for channel confinement in buried-channel devices, and in a two-phase device where oxide thickness variations are used for directionality, charge will move toward the thicker oxide.

Estimating the signal level of a buried-channel device is more difficult than in a surface-channel device. In the latter all of the signal

charge is stored at the same physical location, i.e., at the Si-SiO<sub>2</sub> interface, and the well capacitance is reasonably constant. However, in the buried-channel device, the signal charge spreads out, and the well capacitance constantly increases as signal level increases. Certainly the maximum signal level per unit area is given by the implant dose (the product of the layer doping per cm<sup>3</sup> and the layer thickness). If the signal exceeded this value, the channel would have a net negative charge and no potential well would exist. As shown in Fig. 8, as signal level increases, the depletion layer widths (and capacitance) change and the signal level is given by the product of the layer doping and the distance over which the signal is present (the flat part of the potential well).

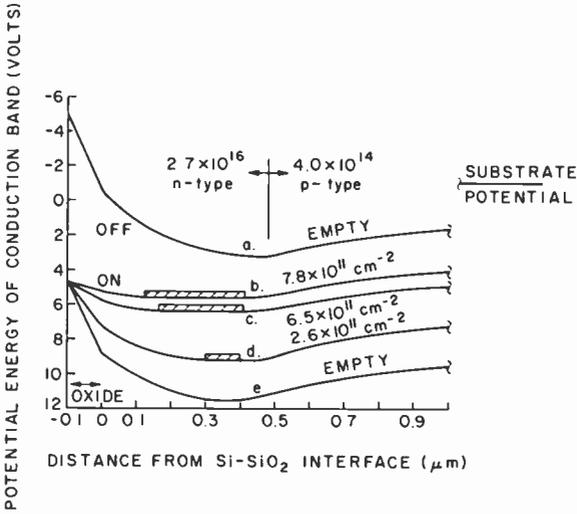


Fig. 8—Potential profiles for a typical buried-channel structure. The hash-marked areas show the region where signal charge resides.

As previously mentioned, buried-channel devices have higher transfer efficiency, require no fat zero, and have higher speed capabilities and lower transfer noise characteristics. An example of the higher efficiency is shown in Fig. 9. On the other hand, buried-channel devices have lower signal-handling capabilities, more involved fabrication procedures, and generally have higher dark-current levels. Thus, a careful evaluation of the needs of a given application is required to determine the most advantageous CCD approach.

## 4. Input Techniques

### 4.1 Current and Voltage Inputs

One possible way of introducing the electrical signal into a CCD is by means of a *current-integration method*. An example of such an input is shown in Fig. 10. The current source is formed by the source diffusion, S, and input gate,  $G_1$ , while the first potential well of the CCD acts as the drain of a MOSFET. With the source at a dc level and

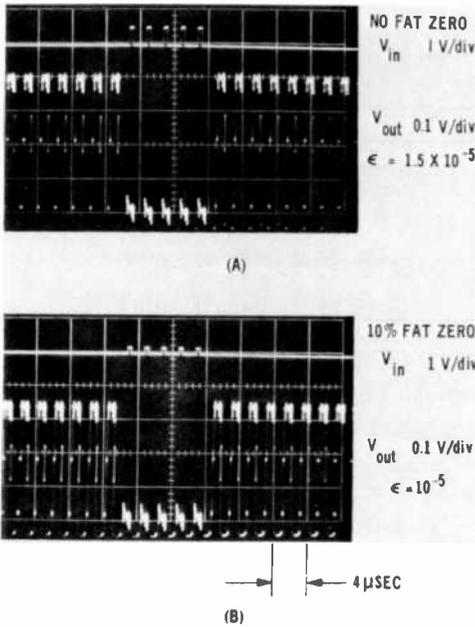


Fig. 9—Output waveforms for a two-phase buried-channel CCD. With no fat zero the charge transfer inefficiency is  $1.5 \times 10^{-5}$ . With 10% fat zero inefficiency drops to  $1.0 \times 10^{-5}$ .

the signal applied to  $G_1$ , the signal charge accumulated in the first potential well will be the time integral over the ON time of  $\phi-1$  of the current. This current is very nonlinear, and varies exponentially with gate voltage for subthreshold currents and parabolically for above-threshold levels.

An alternative approach involves placing a timing pulse on the source so that current flows only during a fixed time rather than the total time  $\phi-1$  is ON.

In either case, this type of input circuit tends to be very nonlinear and critical to control unless additional structure is provided for a more stable current source. An optical input can be used as an alternative current source and, therefore, can be considered as a special form of the current-integration input.

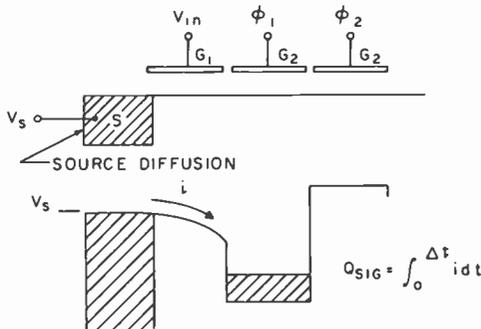


Fig. 10—Current-integration input circuit-construction and operation.

A much less critical way of introducing signal charge is the *voltage-input method* illustrated in Fig. 11. Here the signal is applied to the source diffusion and a strobe or sampling pulse is applied to  $G_1$ . This essentially forms a transmission gate or switch such that the first potential well is filled with charge to the same potential level as the source (see Fig. 11b). Then, as the input gate is turned off (see Fig. 11c) the charge under the input gate is pushed back into the source. If the clock pulse applied to the input gate has relatively slow fall time, the first potential well will tend to come to a thermal equilibrium with the source potential. Otherwise, as the channel collapses under the input gate, some of the charge in the channel will be added to the first potential well. Unlike the current-integration input this method tends to be independent of clock frequency. For a surface-channel CCD the input charge can be approximated as

$$Q_{SIG} = \Delta V_s (C_{ox} + C_d), \quad [9]$$

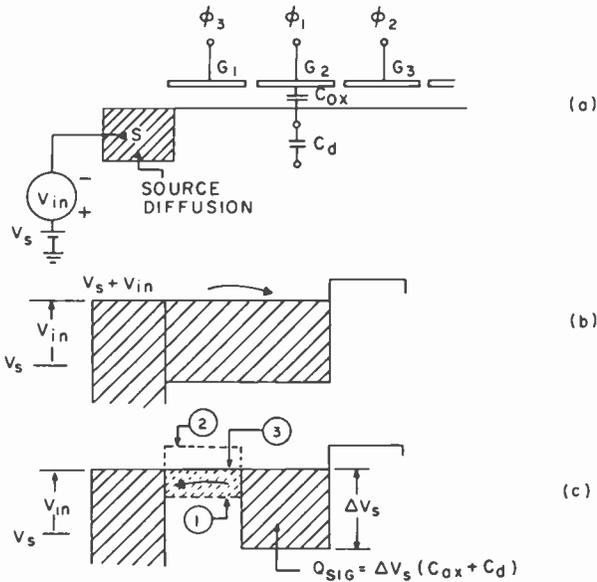
where  $\Delta V_s$  is the change of the surface potential due to the input signal on the source,

$C_{ox}$  is the channel oxide capacitance under gate  $G_2$ , and  $C_d$  is the dc depletion capacitance.

Since the depletion capacitance  $C_d$  is a function of the surface po-

tential, the signal charge  $Q_{SIG}$  is not completely linear with the input voltage  $V_{in}$  and it also depends on the MOS threshold of the gate  $G_2$ . The linearity of the voltage input is better for devices with lower substrate doping.

For the current-integration mode (Fig. 10) the input noise is basically shot noise, where the rms fluctuation in the number of carriers



**Fig. 11—Voltage Input:** (a) Input structure. (b) Potential diagram illustrates the filling of the input well to the source potential,  $V_s + V_{in}$ . (c) 1 During the fall time of  $\phi_3$  clock the charge in the conducting channel under gate  $G_1$  is returned to the source diffusion  $S$ . 2 As the conducting channel under gate  $G_1$  is being cut off, the input well under  $G_2$  is filled to the source potential,  $V_s$ . 3 The clock  $\phi_3$  is off and the charge signal,  $Q_{sig}$ , in the input well is ready for transfer to the CCD channel.

( $\bar{N}_n$ ) introduced into charge packet is

$$\bar{N}_n = \sqrt{N_{SIG}}, \quad [10]$$

where  $N_{SIG}$  is the average number of carriers in the signal packet. This applies anytime the input is introduced optically or by means of a subthreshold current injected over the potential barrier under the input gate  $G_1$ . However, the voltage input (Fig. 11) should have a noise input fluctuation approaching<sup>4,6</sup>

$$\bar{N}_n = \frac{1}{q} \sqrt{kTC} \approx 400 \sqrt{C_{pf}}, \quad [11]$$

where  $C_{pf}$  is in pF. Experiments with this type of input have shown that the input noise is considerably higher than  $\sqrt{kTC}/q$  apparently due<sup>6</sup> to a charge-partitioning noise associated with the collapse of the channel under the input gate  $G_1$ . This somewhat unpredictable channel-partitioning noise can, however, be avoided by several low-noise input techniques.<sup>4,14-17</sup>

#### 4.2 Low-Noise, Charge-Preset Inputs

The basic concept of the *charge-preset input* method illustrated in Fig. 12 is as follows. The input well is overfilled with charge and then the excess charge is allowed to spill back into the source diffusion. During the spill-back process, the charge remaining in the input well under the gate  $G_2$  approaches a thermal equilibrium with the barrier under the gate  $G_1$ . The charge-preset method, therefore, is also referred to as the charge-equilibration,<sup>14</sup> stabilized-input,<sup>18</sup> niche-mode,<sup>6</sup> or fill-and-spill method.<sup>19</sup>

The variance in the number of carriers in the input signal is equal to the average number of carriers corresponding to  $kT/q$  voltage on the capacitance<sup>4</sup> under gate  $G_2$ .

$$\bar{N}_n = \sqrt{\frac{(kT/q)C}{q}} = \frac{1}{q} \sqrt{kTC} = 400 \sqrt{C_{pf}}, \quad [12]$$

where  $C_{pf}$  is the well capacitance under  $G_2$  expressed in picofarads. For SCCD's, this input method shows very good linearity in the relationship between input charge level  $Q_{SIG}$  and input voltage level  $V_{in}$ . Although, this relationship between  $V_{in}$  and  $Q_{SIG}$  can be derived using Eq. [9], a simpler description given below is useful.

Let's focus our attention on the instant of time at which the charge in the potential well under the gate  $G_2$  is in thermal equilibrium with the barrier under  $G_1$ . This means that the surface potentials under  $G_1$  and  $G_2$  are equal. Also, in SCCD's we can assume that the inversion layer containing the charge signal  $Q_{SIG}$  under  $G_2$  is negligibly thin. Thus, equal depletion charge  $Q_d$  must be exposed in the depletion layers under  $G_1$  and  $G_2$  in order to make the surface potentials equal. Therefore, the input voltage  $V_{in}$  must be equal to the difference between the voltages across the channel oxides under the  $G_1$  and  $G_2$  gates, or

$$V_{in} = V_{t2} + \frac{Q_{SIG} + Q_d}{C_{ox-2}} - \left( V_{t1} + \frac{Q_d}{C_{ox-1}} \right), \quad [13]$$

where  $V_{t1}$  and  $V_{t2}$  are the MOS threshold voltages and  $C_{ox-1}$  and  $C_{ox-2}$  are the channel oxide capacitances corresponding to the gates  $G_1$  and  $G_2$ , respectively. From Eq. [13] it is clear that for  $V_{t1} = V_{t2}$  and  $C_{ox-1} = C_{ox-2} = C_{ox}$ ,

$$Q_{SIG} = V_{in} C_{ox}. \quad [14]$$

While the low-noise characteristic of the charge-preset circuit is also applicable to BCCD's, the linearity of this input for BCCD's is not as good as for SCCD's. This is because the spatial distribution of the signal charge and larger edge effects associated with fringing fields play a much larger role in BCCD's.

The charge-preset input can be designed to operate in various modes; however, the following is basic to all modes. The input well is first overfilled by the charge originating from a source diffusion, the input well is then allowed to approach thermal equilibrium and the excess charge spills back to the source, the input well is defined on one side by a precisely controlled barrier (under gate  $G_1$ ) and a significantly larger barrier on the other side (under gate  $G_3$ ), the input signal is applied as a voltage at  $G_2$  relative to  $G_1$  (either  $G_1$  or  $G_2$  can be dc biased), and a gating circuit (involving gates  $G_1$ ,  $G_2$ ,  $G_3$ , and possibly other gates) is provided to transfer the input charge from the input well to the CCD channel. For example, if the source diffusion is maintained at a dc bias, then the voltage applied to the gate  $G_1$  ( $V_1$  in Fig. 12) will be a clock pulse and the input well will be filled by scooping the charge from the source diffusion. However, if a d-c reference voltage or the input voltage is applied to the gate  $G_1$ , the filling of the input well requires pulsing of the source diffusion.

The charge-preset circuit shown in Fig. 12 has the d-c reference voltage,  $V_1$ , applied to gate  $G_1$  and the input voltage,  $V_{in}$ , to  $G_2$ . The advantage of such an arrangement is that since a constant surface potential is maintained for the barrier under gate  $G_1$ , the linearity of the input will not be sensitive to any differences of the MOS threshold characteristics between the gates  $G_1$  and  $G_2$  (for example, if  $G_1$  has a thicker channel oxide or an implanted barrier). Another consideration in designing the charge-preset input is to provide the means to transfer full well charge signal from the input well to the regular CCD channel. In Fig. 12 this is accomplished by increasing the lengths of the gates  $G_1$  and  $G_2$  so that these gates can be operated with the same clocks as applied to the regular CCD channel.

In the operation of the charge-preset inputs described thus far, the input is applied into the CCD channel by completely emptying the input well (under gate  $G_2$ ). Another approach for a low-noise input, shown in Fig. 13, never completely empties the input well.<sup>20</sup> Here, the gate  $G_2$  is dc biased. The input charge packet is formed by first setting the charge level in the input well to the barrier voltage ( $V_{in} - V_{t1}$ ) under the gate  $G_1$  (as in the charge-preset circuit of Fig. 12) and

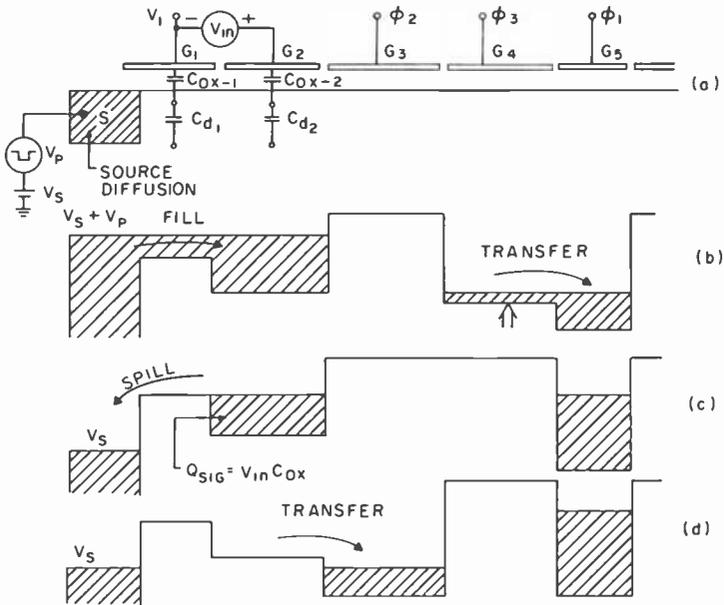


Fig. 12—Charge-preset input: (a) Input gate structure; (b) and (c) potential profiles illustrating the fill-and-spill operations; (d) potential profile showing the transfer of input charge signal,  $Q_{sig}$ , from the input well to the CCD register.

then partially emptying the input well to a new, lower barrier level ( $V_{ref} - V_{t2}$ ) under the gate  $G_3$ . The charge introduced into the CCD channel will be

$$Q_{SIG} = [V_{in} - V_{ref} - (V_{t1} - V_{t3})]C_{ox-2}. \quad [15]$$

Note that the threshold voltage  $V_{t2}$  of the gate  $G_2$  is not involved, since the well under  $G_2$  is not emptied in this operation.

Another version of this low-noise circuit, where the potential well under  $G_2$  is replaced by a floating diffusion,<sup>15</sup> is shown in Fig. 14. The operation of this input circuit, as in the case of the circuit in Fig. 13,

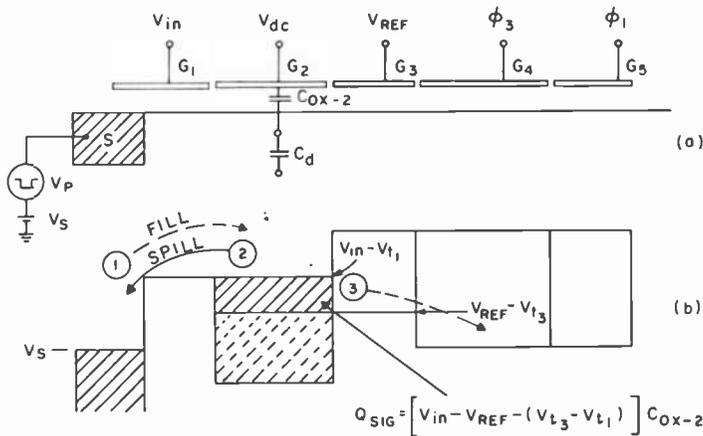


Fig. 13—Charge-preset input in which the input well (under gate  $G_2$ ) is never completely emptied, i.e., a voltage-preset input.

involves setting the floating input diffusion to the potential of the barrier under gate  $G_1$  and then resetting it to a lower barrier potential under the gate  $G_2$ . Since the input charge signal is formed here effectively by setting and resetting a capacitive node between two voltage levels, this method for low-noise input is also referred to as the voltage-preset input.<sup>17</sup>

One interesting property of the input scheme shown in Fig. 14 is that the voltage changes of the floating diffusion due to the input charge signal can be detected by an on-chip MOS amplifier.

In applications that require the introduction of a very uniform signal into a number of parallel channels, such as the fat zero for the

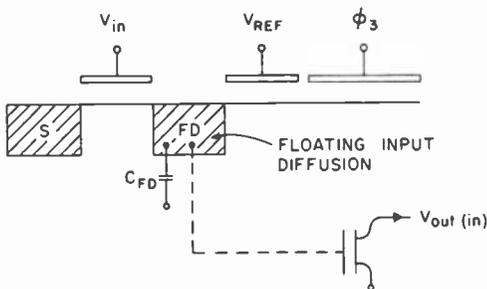
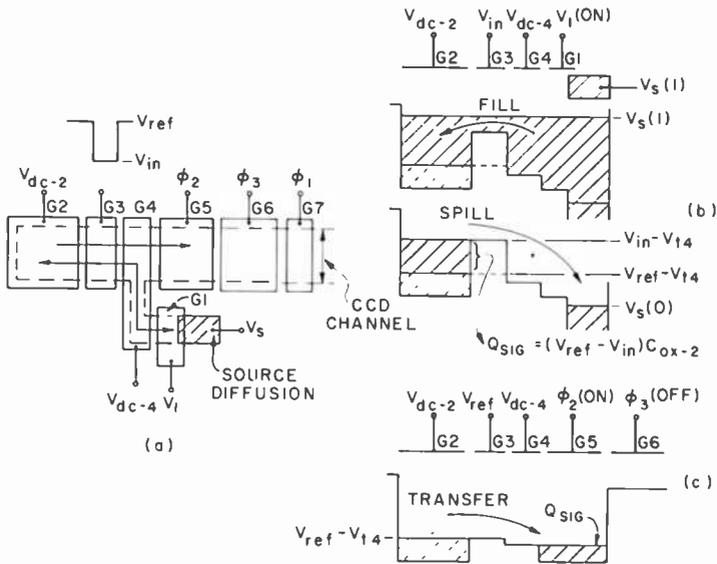


Fig. 14—Voltage-preset floating-diffusion input. This circuit operates virtually the same way as the circuit in Fig. 13. The dc biased gate  $G_2$  of Fig. 13 is replaced here by the floating diffusion (FD).

vertical channels of a charge-coupled imager, it is very desirable to have a low-noise input that is also independent of local voltage threshold variations.<sup>17</sup> Such an input can be obtained by modifying the voltage-preset circuit shown in Fig. 13 or Fig. 14 so that the same gate is involved in setting and resetting the surface potential of the input well under gate  $G_2$  for Fig. 13 or the floating diffusion potential in the case of Fig. 14. The top view of such a threshold-insensitive low-noise input is illustrated in Fig. 15(a). Note in Fig. 15b that when



**Fig. 15**—Voltage-preset input insensitive to local MOS threshold variations: (a) Top view of the gating structure; (b) input well is charge preset to the barrier potential (under gate  $G_3$ )  $V_{in} - V_{t4}$  while a channel-blocking barrier is formed under the gate  $G_5$  due to  $\phi_2$  (off); and (c) signal charge,  $Q_{sig}$ , is transferred under gate  $G_5$  as the barrier under gate  $G_3$  is lowered to the potential of  $V_{ref} - V_{t4}$ .

the surface potential under the d-c biased gate  $G_2$  is set by the input voltage to the potential  $V_{in} - V_{t3}$ , the gate  $G_5$  forms a barrier and isolates the CCD channel from the input stage. Then in the second part of the input cycle, shown in Fig. 15(c), the source diffusion is isolated from the CCD channel by the gate  $G_1$ , the barrier voltage under the gate  $G_3$  is set to a lower level  $V_{ref} - V_{t3}$ , and the input signal charge packet flows into the CCD channel.

Since gate  $G_2$  is maintained at a constant dc bias, the signal charge is

$$Q_{SIG} = (V_{in} - V_{ref})C_{ox-2} \quad [16]$$

An interesting variation of the threshold-insensitive input scheme is shown in Fig. 16. Here, the roles of the gates  $G_1$  and  $G_2$  are reversed. The gate  $G_3$  is dc biased to a fixed barrier voltage while the input and reference voltage levels are applied to the gate  $G_2$ . Now the input voltage, when applied to the gate  $G_3$ , will be the complement of the input voltage previously applied to gate  $G_3$ . Also the barrier voltage under the gate  $G_3$  will be fixed, which may be an advantage in the design of the gating structure connecting the input stage with the CCD channel.

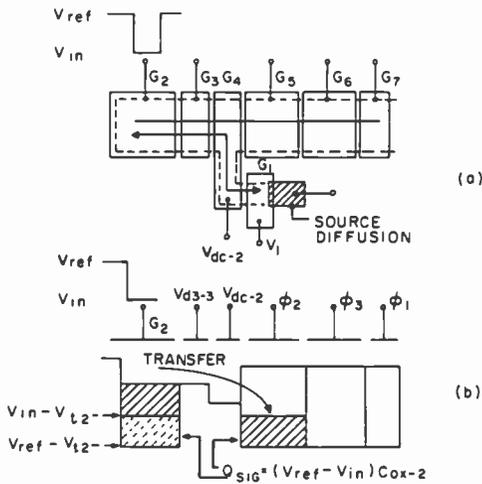


Fig. 16—Voltage-preset input insensitive to local MOS threshold variations in which the input well (under gate  $G_2$ ) is pulsed by the input voltage  $V_{in}$  and the potential barrier under the transfer gate  $G_3$  is fixed. (a) Top view of the input gating structure; (b) potential profile illustrating that as the voltage applied to the gate  $G_2$  is changed from  $V_{ref}$  to  $V_{in}$ , the charge signal  $Q_{sig}$  is pushed out from the input well into the CCD channel.

The main limitation of the voltage-presetting inputs shown in Figs. 13 to 16 is that they involve two equilibrations; therefore they are expected to give rms input signal noise fluctuations.

$$\bar{N}_n = \frac{1}{q} \sqrt{2kTC} \quad [17]$$

Another price that must be paid for obtaining the threshold-insensitive low-noise input is that, unlike the other input schemes, this

type of input structure cannot directly sample input signal. Rather it requires an input signal in the form of a voltage pulse with a precise reference voltage.

### 4.3 Input Considerations

In addition to the matter of linearity and noise, there are differences in how the different input methods we have discussed sample the externally applied input voltage. The current-integration input will, of course, sample the input by integrating the input over a given time. The voltage input on the other hand samples the input at the time when the channel between the input well and the source diffusion is being cut off. The charge-preset input gives still a more complicated sampling function that will tend to sample the input at the beginning of the charge-presetting operation or near the end, depending on how the input signal is changing.

Another design consideration is the method of introduction of a well-defined fat-zero level. We can accomplish this by controlling the voltage levels of the input. An alternative way is to have two or more parallel inputs, all eventually transferring separately measured charge packets into a single collecting well. Here we can also scale the input signals by controlling the effective capacitance (or the dimensions) of the input wells. This approach can also be used to add two separate input signals.

Finally, another interesting concept is the technique of the charge-preset input, i.e., the fill-and-spill operation, to remove a fixed amount of charge from the charge signal already in the CCD channel. In this operation, the charge for overflowing the input well will be supplied not from a source diffusion but from a charge packet of a CCD stage. A fixed amount of this charge will be retained and the excess returned to the CCD stage.

## 5. Output Techniques

### 5.1 Current Output

A circuit for sensing the output of a CCD in the form of high-impedance current pulses is shown in Fig. 17. The drain diffusion,  $D$ , is biased here to act as a sink for the charge signal pushed out of the potential well under the gate  $G_{N-1}$  over the potential barrier formed under the gate  $G_N$ . The dc biased gate  $G_N$  decouples the clock pulses from the drain diffusion. The linearity and the noise of the current output depends only on the capacitance associated with the drain diffusion mode and the external amplifier.

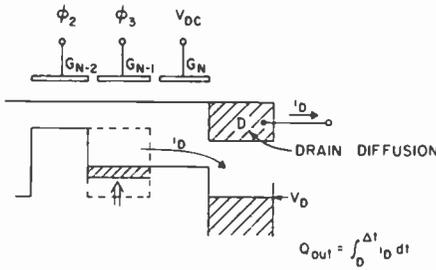


Fig. 17—Current output circuit, construction and operation.

### 5.2 Floating-Diffusion Amplifier

The floating-diffusion amplifier, FDA, is the most popular approach for detecting the output signal with an on-chip amplifier.<sup>21</sup> Construction and operation of an FDA having a negligible capacitance coupling between the gates  $G_N$  and  $G_{N+1}$ , and the floating diffusion, FD, are illustrated in Fig. 18. The floating diffusion is periodically reset to the drain voltage,  $V_D$ . Then, before each charge signal packet is introduced into the floating diffusion, the floating diffusion is isolated from the drain by the reset gate  $G_{N+1}$ . To detect the output signal, the floating diffusion is connected to the gate of an on-chip MOS amplifier. The detected signal at the gate of the MOS device  $\Delta V_{out}$  is

$$\Delta V_{out} = \frac{Q_{SIG}}{C_{FD}}, \quad [18]$$

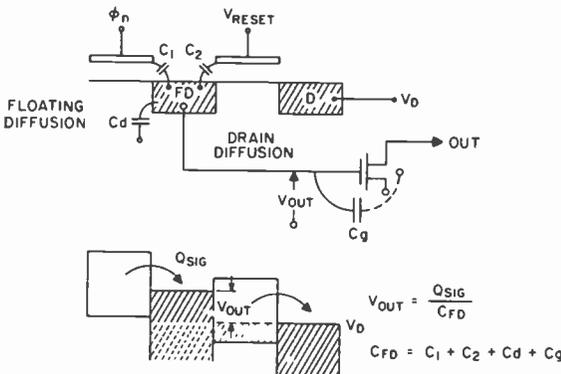


Fig. 18—Floating-diffusion amplifier (FDA) construction and operation.

where  $C_{FD} = C_d + C_1 + C_2 + C_g$  is the total capacitance associated with the floating diffusion.

If the capacitance  $C_1$  represents a large part of the total capacitance  $C_{FD}$ , then the FDA can operate with a dc voltage applied to the reset gate. In this case the floating diffusion is periodically clamped to the barrier potential under the reset gate provided that some charge is transferred at each cycle. Otherwise the potential of the floating diffusion will tend to settle to a value dictated by the thermal emission process of charge over this barrier.

Another type of FDA is shown in Fig. 19. Here, the dc-biased gates  $G_N$  and  $G_{N+1}$  provide a capacitive decoupling of the floating diffusion

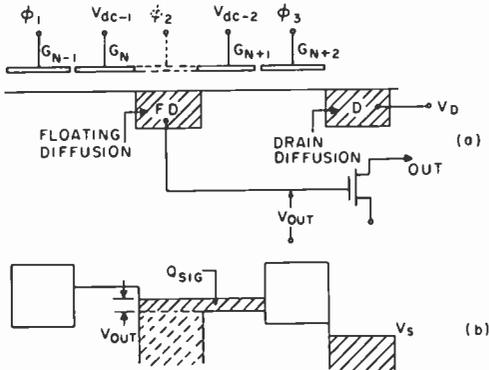


Fig. 19—Floating-diffusion amplifier capacitively decoupled from the clocks  $\phi_1$  and  $\phi_3$  by the dc biased gates  $G_N$  and  $G_{N+1}$ . (The broken lines illustrate an FDA in which the floating diffusion is under a single large gate powered by clock  $\phi_2$ .)

from the clock pulses. Note that, as shown in Fig. 19(b), the floating diffusion for this circuit effectively extends under the dc biased gate  $G_{N+1}$ . The floating diffusion also is periodically reset to the surface potential under gate  $G_{N+1}$  as the signal charge is transferred to the drain diffusion under the control of the gate  $G_{N+2}$ . A special version of this type of circuit is indicated by the dotted lines in Fig. (19a). This version illustrates the case when the gates  $G_N$  and  $G_{N+1}$  are formed into a single gate powered by the clock  $\phi_2$ . Now, the floating diffusion will be overlapped on both sides by this large gate and strongly capacitively coupled with it. In this case, the operation of the floating diffusion could be thought of as nondestructive voltage sensing of the potential well under this large gate.

In general a nondestructive sensing of the charge signal by the floating diffusion requires that the floating diffusion be reset to the

barrier potential under an adjacent gate (on the right side) as part of the charge-transfer process with no external charge introduced into the floating diffusion during this process.

This means that every FDA construction can provide nondestructive voltage sensing provided the means are available for continuing the charge flow to the successive stages. However, the charge transfer out of the floating diffusion must always involve a transfer with a potential barrier, which has the basic characteristics of bucket-brigade transfer. A nondestructive output sensing with the characteristics of the complete charge-transfer process (not involving a transfer barrier) can be obtained only with a properly operated floating gate amplifier.

### 5.3 Floating-Gate Amplifier

Two types of floating-gate amplifiers, FGA's, are illustrated in Figs. 20 and 21. In the FGA shown in Fig. 20, the floating gate is capacitively-coupled to the clock  $\phi_2$ . The MOSFET  $T_1$  is used to clamp the gate N to  $V_{ref}$ . The MOSFET  $T_0$  represents the on-chip output amplifier, whose input gate voltage ( $V_{out}$ ) can be estimated from the capacitive divider involving  $C_\phi$ ,  $C_{ox}$ ,  $C_d$ , and  $C_g$  as<sup>20</sup>

$$\Delta V_{out} = \frac{Q_{SIG}}{\frac{C_d}{C_{ox}}(C_{ox} + C_\phi + C_g) + (C_\phi + C_g)} \quad [19]$$

If the clock voltage  $\phi_2$  is adjusted to produce the same surface potential variations as the phase clocks  $\phi_1$  and  $\phi_3$ , the nondestructive voltage sensing of the charge signal by the FGA shown in Fig. 20 will not interfere with the complete charge-transfer process.

Another type of FGA is illustrated in Fig. 21. In this FGA the dc biased gates  $N_1$  and  $N_2$  shield the output from the phase clocks  $\phi_1$  and  $\phi_3$ . The operation of this type of FGA has the advantage that it can provide a complete charge-transfer mode for emptying the charge signal from wells associated with the floating gate by adjustment of the d-c voltages  $V_1$ ,  $V_2$ , and  $V_3$ .

In addition to providing a high transfer efficiency and nondestructive and relatively noise-free sensing of the charge signal, the FGA can be constructed to perform a summing operation if a number of floating gates are connected in parallel. Furthermore, a split-gate construction such as illustrated in Fig. 22 can provide not only a sensing and summing operation, but also a positive and negative weighting of the charge signals in the CCD register.

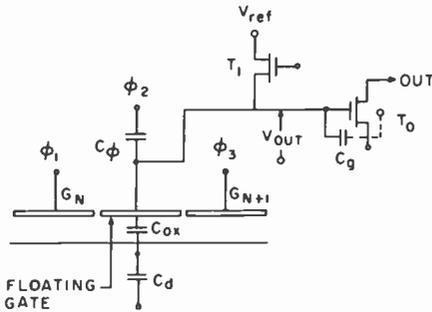


Fig. 20—Floating-gate amplifier (FGA).

### 5.4 Output Considerations

One of the most important properties of charge-coupled devices is that they can transfer the signal charge packets between the potential wells without any capacitive coupling from the clock pulses into the charge signal. Some capacitive coupling, however, is unavoidable at the output. Therefore, the choice of output circuit is often determined by the extent to which we wish to minimize this capacitive feedthrough from the clocks to the output.

The floating diffusion amplifier, FDA, provides a very effective output scheme for the destructive sensing of the output. However, if nondestructive sensing is needed, a suitable floating-gate amplifier circuit is better able to combine the nondestructive sensing with good charge-transfer efficiency.

Linearity is another consideration in the choice of input as well as output circuits. Therefore, the question of linearity must be attacked

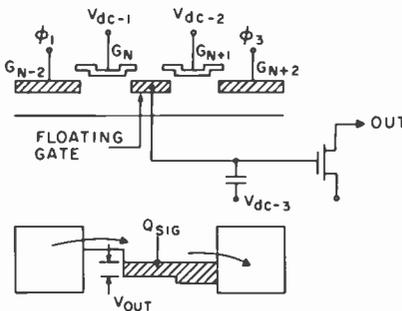


Fig. 21—Floating-gate amplifier in which the floating gate is dc biased and capacitively decoupled from the clocks  $\phi_1$  and  $\phi_3$ . The surface potential under gate  $G_{N+1}$  is adjusted by dc voltages  $V_{dc-1}$ ,  $V_{dc-2}$ , and  $V_{dc-3}$ .

by considering the composite characteristics of the input and the output.

The noise introduced into the output in the process of sensing the charge signal is also of concern, since CCD's are capable of virtually noiseless transfer of charge<sup>4</sup> (especially the BCCD's) and are being considered for applications where the noise introduced at the output stage may be of considerable importance. An example of one such application is the image sensing for a low-light-level TV.<sup>5</sup>

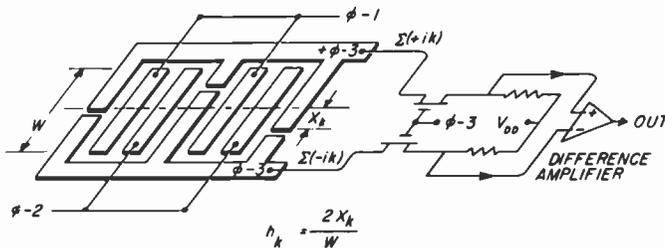


Fig. 22—Three-phase transversal CCD filter employing the split-gate electrode weighting and summing approach.

In the current output scheme the major noise source is associated with the characteristics of the off-chip amplifier and the larger output terminal capacitance. The transfer of charge into the drain diffusion is basically a complete charge transfer process and as such is essentially noiseless.<sup>4</sup>

The major noise associated with the operation of the FDA is due to the periodic resetting of the floating diffusions. The actual noise fluctuations can be treated in very much the same way as the noise introduced by the input circuits. However, the reset noise of the FDA can be virtually eliminated by a use of double-synchronous sampling,<sup>22</sup> in which the output of the on-chip MOS amplifier is sampled twice and then compared, i.e., first after resetting of the floating diffusion and then after the charge signal is introduced into the floating diffusion.

## 6. Summary and Conclusions

In this paper we have tried to present in simple physical terms some of the important concepts useful in the understanding of CCD operation and limitations. This included a description of the transfer and trapping mechanisms for charge packets in the CCD register, a short discussion of buried-channel operation, and a review of the concepts and choices involved in signal input and output.

Even though a CCD can be considered as a special type of MOS device, the most useful description of the CCD is in terms of microscopic physical parameters rather than equivalent-circuit models. To some degree, the same also applies to input and output techniques, especially at the point where conventional MOS components join with the CCD channels. This is because conventional MOS circuits are normally described in terms of currents and voltages (or amplifiers and switches), and their internal device transient time constants, such as the times involved in the forming and collapsing of an MOS channel, are normally much shorter than the time constants of the external circuits. Thus these devices can be thought of in terms of steady-state MOS channels that are readily represented in equivalent circuit format.

On the other hand, in the CCD, transfer of signal charge packets from one low capacitance ( $\sim 155$  pF/mm<sup>2</sup> of CCD channel) storage site to the next involves very small transient currents. Therefore, the operation of CCD's is determined only by the transient channel characteristics rather than steady state. These transient characteristics are not amenable to equivalent-circuit formulation.

It is interesting to note, however, that the theory developed to describe CCD operation is providing insight resulting in new MOS components, such as the concept of charge pumping,<sup>23</sup> the development of the charge-injection device for image sensing,<sup>24</sup> and new design techniques for high-density one-transistor-per-bit MOS RAM's.<sup>25</sup> It is safe to conclude, therefore, that the various charge-coupling principles that are being developed will eventually become accepted as integral parts of MOS circuit technology.

### **Acknowledgment**

It has been more than five years since the authors began their work on charge-coupled devices. Looking back on how our understanding of these devices progressed, we recall with warm feelings of friendship many of our conversations with Al Rose. Al's clear and probing understanding of the fundamental physical phenomena of semiconductor and imaging devices and his generous and sympathetic interest in our questions guided us in the development of our understanding of CCD's, especially in the areas of transfer phenomena,<sup>2,3</sup> noise and imager sensitivity, and resolution.<sup>4-6</sup> In our conversations with Al, probably his most valuable contribution was his ability to reduce the originally complex and confusing formulations of device operation to simply understood physical concepts. It is appropriate, therefore, that this paper on CCD's be dedicated to Al Rose.

## References

- <sup>1</sup> W. S. Boyle and G. E. Smith, "Charge Coupled Semiconductor Devices," *Bell System Tech. J., Briefs*, **49**, No. 4, p. 587, Apr. 1970.
- <sup>2</sup> J. E. Carnes, W. F. Kosonocky, and E. G. Ramberg, "Drift-aiding Fringing Fields in Charge-Coupled Devices," *IEEE J. Solid-State Circuits*, **SC-6**, p. 322, Oct. 1971.
- <sup>3</sup> J. E. Carnes, W. F. Kosonocky, and E. G. Ramberg, "Free Charge Transfer in Charge-Coupled Devices," *IEEE Trans. on Electron Dev.*, **ED-19**, No. 6, p. 798, June 1972.
- <sup>4</sup> J. E. Carnes and W. F. Kosonocky, "Noise Sources in Charge-Coupled Devices," *RCA Review*, **33**, p. 327, June 1972.
- <sup>5</sup> J. E. Carnes and W. F. Kosonocky, "Sensitivity and Resolution of Charge-Coupled Imagers at Low Light Levels," *RCA Review*, **33**, p. 607, Dec. 1972.
- <sup>6</sup> J. E. Carnes, W. F. Kosonocky, and P. A. Levine, "Measurements of Noise in Charge-Coupled Devices," *RCA Review*, **34**, p. 553, Dec. 1973.
- <sup>7</sup> C. K. Kim and M. Lenzlinger, "Charge Transfer in Charge-Coupled Devices," *J. Appl. Phys.*, **42**, p. 3586, Aug. 1971.
- <sup>8</sup> W. F. Kosonocky and J. E. Carnes, "Two-Phase Charge-Coupled Devices with Overlapping Polysilicon and Aluminum Gates," *RCA Review*, **34**, No. 1, p. 164, Mar. 1973.
- <sup>9</sup> F. L. J. Sangster and K. Teer, "Bucket-Brigade Electronics—New Possibilities for Delay, Time-Axis Converter, and Scanning," *IEEE J. Solid-State Circuits*, **SC-4**, p. 131, June 1969.
- <sup>10</sup> J. E. Carnes and W. F. Kosonocky, "Fast-Interface-State Losses in Charge-Coupled Devices," *Appl. Phys. Lett.*, **20**, No. 7, p. 261, Apr. 1972.
- <sup>11</sup> W. F. Kosonocky and J. E. Carnes, "Design and Performance of Two-Phase Charge-Coupled Devices with Overlapping Polysilicon and Aluminum Gates," 1973 International Electron Devices Meeting Tech. Dig., p. 123, Wash. D.C., Dec. 3-5, 1973.
- <sup>12</sup> A. M. Mohsen, M. F. Tompsett, and C. H. Sequin, "Noise Measurements in Charge-Coupled Devices," *IEEE Trans. Electron Dev.*, **ED-22**, No. 5, p. 209, May 1975.
- <sup>13</sup> R. H. Walden, R. H. Krambeck, R. J. Strain, J. McKenna, N. L. Schryer, and G. E. Smith, "The Buried Channel Charge Coupled Device," *Bell System Tech. J.*, **51**, No. 7, p. 1635, Sept. 1972.
- <sup>14</sup> M. F. Tompsett, "Surface Potential Equilibration Method of Setting Charge in Charge-Coupled Devices," *IEEE Trans. Electron Dev.*, **ED-22**, No. 6, p. 305, June 1975.
- <sup>15</sup> S. P. Emmons and D. D. Buss, "Noise Measurements on the Floating Diffusion Input for Charge-Coupled Devices," *J. Appl. Phys.*, **45**, No. 12, p. 5303, Dec. 1974.
- <sup>16</sup> C. H. Sequin and A. M. Mohsen, "Linearity of Electrical Charge Injection into Charge Coupled Devices," 1974 Int'l Electron Devices Meeting Tech. Dig., p. 229, Wash., D.C., Dec. 1974.
- <sup>17</sup> S. P. Emmons, A. F. Tasch, Jr., and J. M. Caywood, "A Low-Noise CCD Input With Reduced Sensitivity to Threshold Voltage," Int'l Electron Devices Meeting Tech. Dig., p. 233, Wash., D.C., Dec. 1974.
- <sup>18</sup> "Stabilized input" is a term first introduced by M. H. White.
- <sup>19</sup> "Fill and spill" is a term first introduced by Emmons, Buss, and Brodersen.
- <sup>20</sup> This is a CCD equivalent to the circuit described by Emmons, et al. in Ref [17].
- <sup>21</sup> W. F. Kosonocky and J. E. Carnes, "Charge-Coupled Digital Circuits," *IEEE J. Solid-State Circuits*, **SC-6**, No. 5, p. 314, Oct. 1971.
- <sup>22</sup> M. H. White, D. R. Lampe, F. C. Alaha, and I. A. Mack, "Characterization of Surface Channel CCD Image Arrays at Low Light Levels," *IEEE J. Solid-State Circuits*, **SC-9**, p. 1, Feb. 1974.
- <sup>23</sup> J. S. Brugler and P. G. A. Jespers, "Charge Pumping in MOS Devices," *IEEE Trans. Electron Dev.*, **ED-16**, No. 3, p. 297, March 1969.
- <sup>24</sup> G. J. Michon and H. K. Burke, "Charge Injection Imaging," Tech. Dig. 1973 ISSCC, Philadelphia, Pa. p. 138, Feb. 1973.
- <sup>25</sup> L. G. Heller, D. P. Spampinato, and Y. L. Yao, "High-Sensitivity Charge-Transfer Sense Amplifier," Tech. Dig. of Int'l Solid State Circuits Conference, p. 112, Phila., Pa., Feb. 1975.

# Tunneling Conductivity in Granular Metal Films

B. Abeles

RCA Laboratories, Princeton, N.J. 08540

**Abstract**—A review is given of electrical transport in granular metals in which conduction is due to tunneling of electrons and holes between isolated metal grains. Some of the conflicting results that have appeared in the literature are resolved. It is shown that the criticism by Hill of the low-field conductivity model of Neugebauer and Webb is unfounded. The interpretation of the field dependence of the high-field conductivity in terms of the Poole-Frenkel effect, which has been used by many workers, is found to be inconsistent with the tunneling model. The confusion arises from the fact that the charging energy was treated as a barrier, rather than as the difference in the Fermi levels of the neighboring grains between which the tunneling takes place. The tunneling model, generalized to the case of a distribution in charging energies and tunnel barrier thicknesses, explains well the temperature and field dependence of the conductivity observed in a wide variety of cermets. In discontinuous metal films the agreement with theory is not as consistent.

## 1. Introduction

Granular metals are two-phase mixtures of metals and insulators. When the volume fraction of the metal  $x$  is less than 0.5, the metal breaks up into isolated grains, and the resistance is thermally activated with activation energies in the range 0.01–0.1 eV. The transport properties of granular metals in which the metal particles are isolated have been of continuing<sup>1–13</sup> interest since the first reported investigations of a system of this kind at the beginning of this century. The early work was done on evaporated ultra-thin films, which, below a critical thickness, are discontinuous and form isolated metal islands

on the substrate. Granular metals are now usually prepared by the co-sputtering of metals and insulators, and are being used in a variety of scientific<sup>3-5</sup> and technological applications.<sup>4-6</sup> The two types of granular metals reviewed in this work are dispersions of metal grains in insulators, which are referred to as "cermets," and ultra-thin metal films, which are referred to as "discontinuous films." A typical electron micrograph of a cermet consisting of isolated Au grains in silica is shown in Fig. 1.

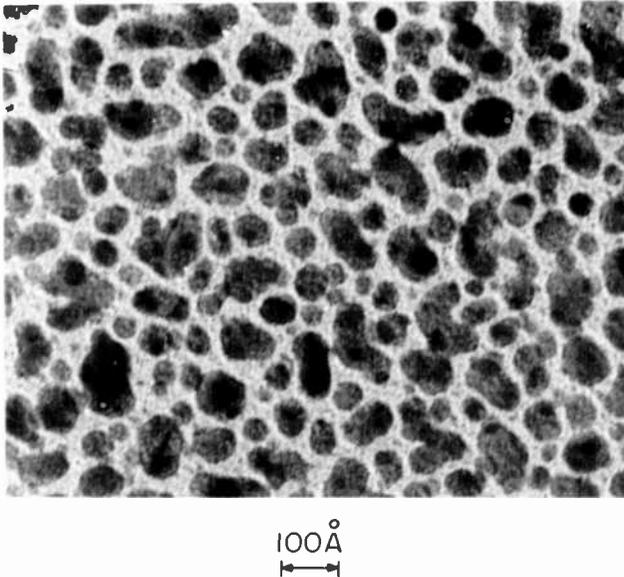


Fig. 1—Electron micrograph of an Au-SiO<sub>2</sub> cermet with 0.4 volume fraction Au. Thickness of film 70 Å. The dark particles are crystalline Au grains.

The explanation for the temperature dependence of the conductivity,  $\sigma$ , in granular metals remained elusive for many years. It was thought at one time that the conduction mechanism is thermionic emission over the barriers separating the metal grains. One would expect in that case  $\sigma \propto \exp(-\varphi/kT)$ , where  $\varphi$  is the barrier height,  $k$  is Boltzmann's constant, and  $T$  is the absolute temperature. However, the values of  $\varphi$  for metal-insulator interfaces are much too large to explain the observed activation energies. Image forces result in barrier height lowering;<sup>14</sup> however, it is unlikely that this results in barriers of much less than 1 eV for barrier thicknesses  $s \geq 10$  Å. It has

also been postulated that the conductivity could be due to emission of electrons from the metal into localized states in the substrate.<sup>15</sup>

A more natural way of explaining the observed low activation energy was suggested by Gorter.<sup>7</sup> He pointed out that in order to transfer an electron from one neutral grain to another neutral grain, a finite energy is required. This charging energy,  $E_c$ , is the electrostatic energy of the positively and negatively charged grains, and is of the order  $e^2/C_G$ , where  $e$  is the electronic charge and  $C_G$  is the capacitance of a

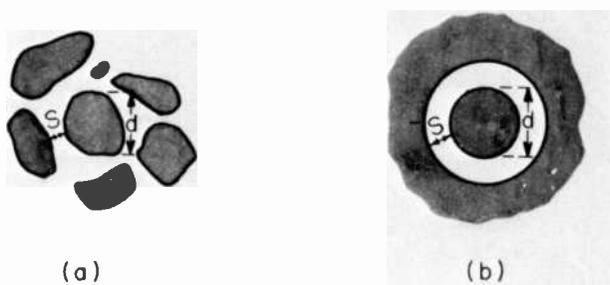


Fig. 2—For the purpose of calculating the charging energy  $E_c^0$ , the geometry in (a) is approximated by that of (b).

grain. The charging energy,  $E_c^0$ , corresponding to a fully dissociated pair of positively and negatively charged grains, can be estimated by replacing the grain and its neighbors by a spherical capacitor, as in Figure 2. The expression for  $E_c^0$  in that case is<sup>3</sup>

$$E_c^0 = \frac{e^2}{d} \frac{1}{\epsilon_1} \frac{4s}{d + 2s}, \quad [1]$$

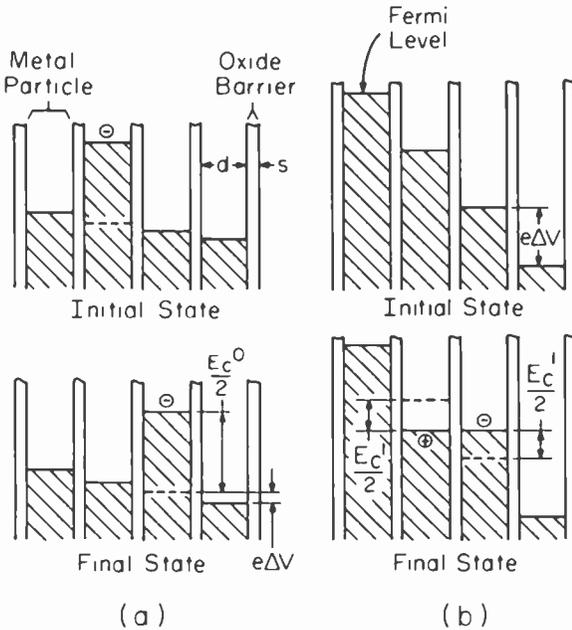
where  $d$  is the diameter of the grain,  $s$  the separation between the grains, and  $\epsilon_1$  is the dielectric constant of the insulator. Using the values of  $d = 10\text{--}100 \text{ \AA}$ ,  $s = 5\text{--}50 \text{ \AA}$ , and  $\epsilon_1 = 2\text{--}10$ , commonly found in granular metals, Eq. [1] yields values of  $E_c^0$  in reasonable agreement with the conductivity activation energy.

A quantitative treatment of the conductivity using the concept of charging energy was given by Neugebauer and Webb.<sup>8</sup> At temperatures at which  $kT \gg E_c^0$ , the thermal equilibrium number density of positively or negatively charged metal grains is given by<sup>16</sup>

$$\frac{\mathfrak{N}}{2} \exp(-E_c^0/2kT), \quad [2]$$

where  $\mathfrak{N}$  is the number density of metal grains. The charged grains act

as carriers that hop by tunneling of electrons or holes from charged grains to neutral grains. When the voltage drop,  $\Delta V$ , between the adjacent grains is small enough so that  $\Delta V \ll kT/e$ , then the thermal equilibrium number of charge carriers is undisturbed and the conductivity is determined by the net drift of the charge carriers due to the applied field. A schematic representation of the hopping of charge is given in Fig. 3a.



**Fig. 3**—Schematic illustration of tunneling in granular metals. The voltage drop between adjacent grains is  $\Delta V$ .  $E_c^0$  is the energy required to generate a pair of fully dissociated positively and negatively charged particles, and  $E_c^1$  is the energy required to generate a pair of charges on neighboring grains. The relative potential levels indicated are as seen by a negative test charge. (a)  $e\Delta V \ll E_c^1$ . Initial state, one grain is charged with an extra electron; final state, the extra electron tunneled to a neighboring grain. (b)  $e\Delta V = E_c^1$ . Initial state, each grain is neutral before tunneling takes place; final state, after tunneling has occurred, a hole is left on one grain and an electron is added to the other grain. The hole and electron will then drift to the electrodes.

As the magnitude of electric field is increased, an additional mechanism, field-induced tunneling, becomes important, and the charge carrier concentration gradually deviates from its thermal equilibrium value. The relative importance of the two generation processes, ther-

mal activation and field induced tunneling, is gauged by the ratio  $kT/e\Delta V$ . The transition between the high-field and the low-field regimes in cermets is observed as a rapid increase in conductivity when  $e\Delta V$  becomes comparable to or greater than  $kT$ . When the strength of electric field is further increased, electrical conductivity becomes highly non-ohmic, a direct manifestation that field-induced tunneling has replaced thermal activation as the dominant process for charge-carrier generation. The field-induced tunneling process, in which the energy,  $E_c^1$ , required for the generation of a pair of positively and negatively charged grains on neighboring grains is provided by the external field, is illustrated in Fig. 3b. Because of the interaction energy between the neighboring charged grains,  $E_c^1 < E_c^0$ .

Sheng et al.<sup>3,13</sup> showed that the conductivity of cermets,  $\sigma$ , can be represented by a single function  $\sigma(T^*, \mathcal{E}^*)$  of reduced field  $\mathcal{E}^* = \mathcal{E}/\mathcal{E}_0$  and reduced temperature  $T^* = kT/C_0$ , where the parameters  $\mathcal{E}_0$  and  $C_0$  depend on the composition of the cermet and on the dielectric constant of the insulator. This scaling property of the conductivity of cermets was shown to be a direct consequence of the fact that in cermets there is a distribution of charging energies and a strong correlation between them. Using an argument based on the structure of the materials it was shown that this correlation can be approximated by the relation  $sE_c^0 = \text{constant}$  for a given composition of the cermet. Based on this assumption, the tunneling theory predicts the correct form for the function  $\sigma(T^*, \mathcal{E}^*)$  and yields expression for  $C_0$  and  $\mathcal{E}_0$  in terms of the parameters of the granular metal—the volume fraction of the metal,  $x$ , the average grain size  $d_0$ , the average thickness  $s_0$ , and height  $\varphi$  of the barriers and the dielectric constant of the insulator  $\epsilon_1$ .

In the case of discontinuous films, the agreement between the tunneling theory and the experimental data is not as consistent as it is for cermets. For instance, in some discontinuous metal films the temperature dependence of the low-field conductivity is characterized by a single activation energy,<sup>8</sup> which is surprising in view of the fact that electron microscopy of such films indicates a wide distribution in  $s$  and  $d$ ;<sup>8,10</sup> one would, in such a case, expect a distribution in activation energies. Another difficulty is that the field dependence of the conductivity has been interpreted in terms of Poole-Frenkel thermionic emission,<sup>8,10</sup> which is inconsistent with the tunneling model. The purpose of this paper is to present a coherent treatment of the electrical conductivity of granular metals in the dielectric regime and to resolve some of the conflicting results that have appeared in the literature.

## 2. Theory of Tunneling Conductivity

The tunneling conductivity in granular metals results primarily from two kinds of tunneling transitions: (1) tunneling of an electron from a neutral grain to a neighboring neutral grain, producing a pair of positively and negatively charged grains (Fig. 3b), and (2) tunneling of the charge from a charged grain to a neighboring neutral grain (Fig. 3a). The first process represents charge-carrier generation, while the second represents charge-carrier transport. In the following, the derivation of the conductivity is carried out in three parts—the calculation of the tunneling rates, the low-field conductivity, and the high-field conductivity.

### 2.1 Tunneling Rates

The energy diagrams depicting the two tunneling processes are shown in Fig. 4. Because the potential of the metal grain is changed by the addition or subtraction of an electron, the transfer of electrons between grains is properly described in terms of the Fermi levels  $\mu_i$  of the individual grains  $i$ , measured with respect to a fixed electrochemical potential. In the case of transfer of an electron from a neutral grain 1 to a neighboring neutral grain 2 (Fig. 4a), the Fermi levels are  $\mu_1 = \varphi_v + \frac{1}{2}E_c^1$  and  $\mu_2 = \varphi_v - \frac{1}{2}E_c^1 + e\Delta V$ , while for the case of transfer of an electron from a charged grain 1 to a neutral grain 2 (Fig. 4b),  $\mu_1 = \varphi_v - \frac{1}{2}E_c^0$  and  $\mu_2 = \varphi_v - \frac{1}{2}E_c^0 + e\Delta V$ , where  $\Delta V$  is the voltage drop between the grains due to the externally applied field and  $\varphi_v$  is the work function of the metal. The energies required to produce a pair of positively and negatively charged grains are given by  $E_c^0$  for fully dissociated charges and by  $E_c^1$  for neighboring charges (to a good approximation<sup>3</sup>  $E_c^1 \simeq \frac{1}{2}E_c^0$ ).

The relation between the initial energy  $E_1$  and the final energy  $E_2$  of the electron is determined as follows. To remove to infinity an electron that initially is on grain 1 at an energy  $E_1$  above the Fermi level requires an energy  $\mu_1 - E_1$ . In placing the electron at an energy  $E_2$  above the Fermi level of grain 2, an amount of energy  $\mu_2 - E_2$  is released. Conservation of energy requires that  $\mu_1 - E_1 = \mu_2 - E_2$ , so that for the two tunneling processes:

*neutral-neutral*

$$E_2 = E_1 + e\Delta V - E_c^1, \tag{3}$$

*charged-neutral*

$$E_2 = E_1 + e\Delta V. \tag{4}$$

The rate of tunneling of electrons from grain 1 to grain 2,  $r^+$ , is given by<sup>3</sup>

$$r^+ = \gamma G \int_{-\infty}^{+\infty} |M_{12}|^2 \nu f(E_1) [1 - f(E_2)] N(E_1) dE_1, \quad [5]$$

where the integration extends over all initial states  $E_1$ . The final states,  $E_2$ , are given by either Eq. [3] or Eq. [4],  $f(E) = [1 + \exp(E/$

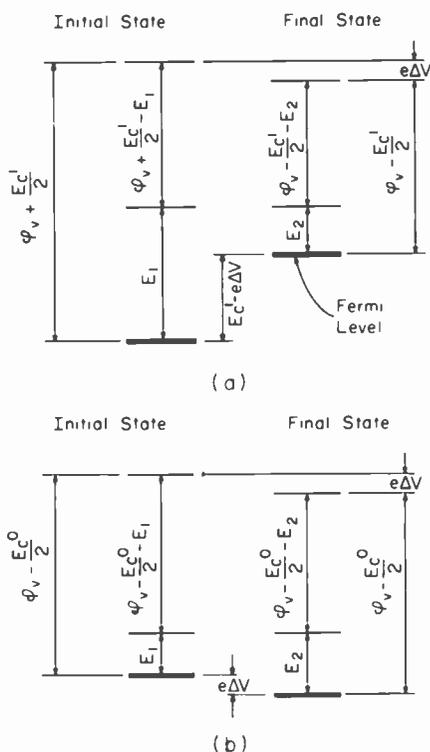


Fig. 4—Schematic energy diagram of transfer of electrons between grains.  $E_1$  and  $E_2$  are the initial and final energies and  $\varphi_v$  is the work function. (a) electron is transferred from neutral grain 1 to neutral grain 2; (b) electron is transferred from charged grain 1 to neutral grain 2.

$kT)^{-1}$  is the Fermi factor,  $G$  is the grain volume,  $N(E)$  is the density of states,  $\nu$  is the frequency at which the electron strikes the barrier,  $M_{12}$  is a matrix element, and  $\gamma$  is a factor that takes into account the fact that tunneling occurs across nonplanar barriers.<sup>3</sup> The tunneling rate  $r^-$ , corresponding to electrons tunneling in the direction opposed

to  $\Delta V$ , is given by the transformation  $\Delta V \rightarrow -\Delta V$  in Eqs. [3] and [4]. The frequency of oscillation for the electron in the grain,  $\nu$ , is analogous to that of nucleons oscillating inside the nucleus. The frequency of such an oscillation is given by  $\nu = \delta/h$  where,

$$\delta = 1/[GN(E_F)]. \quad [6]$$

is the energy level separation due to quantization of electronic motion in irregularly shaped metal grains<sup>17</sup> and  $h$  is Planck's constant.

The integration of Eq. [5] extends over the total kinetic energy of the electron, which is appropriate for irregularly shaped grains in the WKB approximation. The matrix element  $M_{12}$  in the WKB approximation is given by

$$|M_{12}|^2 = \exp\{-2[2m(\varphi - E)/\hbar^2]^{1/2} s\}, \quad [7]$$

where  $m$  is the effective mass of the electron. The dependence of  $M_{12}$  on  $E$  can be neglected<sup>18</sup> for barrier voltages  $e\Delta V \ll \varphi$ . Taking  $|M_{12}|^2$  as a constant, the integration of Eq. [5] yields for the tunneling rates from a neutral to a neutral grain  $r_n^\pm$  and from a charged to neutral grain  $r_c^\pm$ , the expressions

$$r_n^\pm = \frac{\gamma}{h} \exp(-2\chi s) \frac{\pm e\Delta V - E_c^1}{1 - \exp[-(\pm e\Delta V - E_c^1)/kT]}, \quad [8]$$

$$r_c^\pm = \frac{\gamma}{h} \exp(-2\chi s) \frac{\pm e\Delta V}{1 - \exp(\mp e\Delta V/kT)}. \quad [9]$$

where  $\chi = (2m\varphi/\hbar^2)^{1/2}$ .

We now wish to draw attention to some effects that were not taken into account in the derivation of Eqs. [8] and [9]. Dryer et al.<sup>19</sup> points out that the time required for tunneling of an electron between two grains is much shorter than the response time of the dielectric polarization of the insulator. This response time is of the order of a reciprocal phonon frequency ( $10^{-13}$  sec). The tunneling thus proceeds through an intermediate state in which the electron tunnels to a neighboring grain with the polarization of the dielectric remaining unchanged. The transition to the final state in which the dielectric relaxes to the equilibrium polarization is accompanied by the emission of an optical phonon. This process is analogous to the transport of the small polaron. The existence of an intermediate state results in a conductivity activation energy that is larger than one derived on the basis of only electrostatic considerations.

Due to quantization of electronic motion within every metal parti-

cle, phonon-assisted tunneling must be taken into account. Energy levels in irregularly shaped particles<sup>17</sup> are separated on the average by an energy  $\delta$  given by Eq. [6]. Due to this discrete separation between the energy levels, sometimes the closest empty state that the electron can tunnel to is at an energy  $\Delta E$  away, where  $0 < \Delta E < \delta$ . When this happens, there can be no direct tunneling between the grains, even when the condition  $e\Delta V \geq E_c$ <sup>1</sup> is satisfied, and the electron must emit or absorb a phonon during the tunneling process in order to reach the final state. The probability of such a tunneling process (with phonon emission or absorption) is expected to increase with temperature and to approach a high-temperature limit when  $kT$  is of the order of  $\delta$  or larger. Hartman<sup>9</sup> proposed that the conductivity activation energy in granular metals is given by  $\delta$  and that the contribution of  $E_c$  can be ignored. This conclusion is not justified because Hartman overestimated the value of  $\delta$  by 1–2 orders of magnitude in assuming rectangular-parallelepiped-shaped particles. In fact, using Eq. [6] appropriate to irregularly shaped particles, we find that  $\delta \ll E_c$ . Effects due to level splitting  $\delta$  have been observed in the field dependence of the conductivity<sup>3</sup> at temperatures  $kT \leq \delta$ .

## 2.2 Low-Field Conductivity

We first give the derivation of the low-field conductivity,  $\sigma_L$ , for the case of a granular metal characterized by a single charging energy  $E_c$ . An idealized metal system such as this is realized, for instance, by a simple cubic lattice of metal spheres of diameter  $d$  with a lattice constant  $d + s$  embedded in a dielectric matrix. The conductivity is then generalized to the more realistic case of a granular metal with a distribution in  $E_c$  and  $s$ .

According to the model of Neugebauer and Webb, the motion of the charge carriers results from tunneling of charge from a charged grain to a neutral grain (Fig. 3a). The applied field causes a net drift of the charge carriers, and it is assumed that  $e\Delta V \ll kT$ , so that the thermal equilibrium number given by Eq. [2] is undisturbed. Accordingly, the expression for  $\sigma_L$  is given by

$$\sigma_L = \mu_d e n \exp(-E_c^0/2kT), \quad [10]$$

where  $\mu_d$  is the drift mobility of the charge carriers. The net tunneling rate of electrons down-field is  $r_c^+ - r_c^-$ , the drift velocity is given by  $(r_c^+ - r_c^-)(d + s)$ , the voltage drop is  $\Delta V = \mathcal{E}(d + s)$ , where  $\mathcal{E}$  is the average field, and the mobility is  $\mu_d = [(r_c^+ - r_c^-)/\Delta V](d + s)^2$ . Substituting for  $(r_c^+ - r_c^-)/\Delta V$  from Eq. [11] we obtain

$$\mu_d = \frac{\gamma e}{h} \exp(-2\chi s)(d + s)^2, \quad [11]$$

and from Eq. [10],

$$\sigma_l = \frac{\gamma}{h} \frac{e^2}{d + s} \exp\left(-\frac{E_c^0}{2kT} - 2\chi s\right), \quad [11a]$$

where we have used  $\varkappa = (d + s)^{-3}$  for the case of a three-dimensional granular metal. For a two-dimensional granular metal, such as a discontinuous film, the sheet conductivity  $\Sigma$  is given by

$$\Sigma = (d + s)\sigma_l. \quad [12]$$

Apart from the pre-exponential factor, Eq. [11] differs from the result of Neugebauer and Webb in that their conductivity activation energy is twice that of ours.<sup>16,20</sup>

Neugebauer and Webb's model has been criticized<sup>10,11</sup> on the ground that the current contribution due to tunneling between neutral grains had not been taken into account. This objection is invalid since such processes are in fact included in their model. This can be shown clearly by calculating the conductivity from the generation rate. Consider inside the sample a plane of unit area perpendicular to the direction of the macroscopic field. The current density  $j$  is equal to the sum of negative charges that cross the plane along the field direction and positive charges that cross the plane in the opposite direction during unit time. If electrons and holes have a recombination length  $l$ , then  $j$  is equal to twice the rate of generation of charge carriers in a volume of unit area and thickness  $l$ ; therefore,

$$j = 2le[\frac{1}{2}\varkappa \exp(-E_c^0/2kT)]/\tau, \quad [13]$$

where  $\tau$  is the lifetime of the charge excitation and

$$\frac{1}{2}\varkappa \exp(-E_c^0/2kT)/\tau \quad [14]$$

is the generation rate in the low-field limit. The recombination length, defined as the net distance (parallel to the macroscopic field) traversed by the charge excitation during its lifetime  $\tau$ , is given by

$$l = \varepsilon\mu_d\tau. \quad [15]$$

Since the recombination of charge carriers is a bimolecular process,  $\tau$  is inversely proportional to the density of charge carriers. In the low-field limit, the temperature dependence of  $\tau$  and of  $l$  are therefore given by  $\exp(E_c^0/2kT)$ , while at high fields,  $\tau$  is a function of  $\mathcal{E}$  because the charge carriers are field generated.

Substituting in Eq. [13] for  $l$  from Eq. [15] results in a conductivity identical to Eq. [10], which demonstrates that the low-field conductivity calculated from generation rate is equivalent to that of Neugebauer and Webb.

We now wish to make some remarks regarding the calculation of the current density  $j$  carried out by Hill<sup>10</sup> and Kiernan and Stopps.<sup>11</sup> The authors assume that  $j$  is proportional to the difference in the down-field and up-field tunneling rates between neighboring neutral grains. However, the generation of a charge carrier by tunneling between neutral grains constitutes only the first step in the life of a charge carrier; additional tunneling processes are required for the two charges to contribute a net current. In the correct calculation of  $j$  the whole history of a charge excitation must be taken into account. This is accomplished by using the concept of the recombination length, as was done in Eq. [13]. Another point we wish to make is that the authors,<sup>10,11</sup> in calculating the tunneling rates, carried out the integration separately over the components of the kinetic energy of the electron normal to the barrier,  $E_{\perp}$ , and parallel to the barrier,  $E_{\parallel}$ . Such a separation in the integration is permissible only for the case of planar barriers, where transverse momentum is conserved, and is not applicable to the irregularly shaped grains in granular metals; the integration over total kinetic energy as done in Eq. [5] appears to us as a more realistic treatment of the situation.

We now generalize the calculation of the conductivity to the case where there is a distribution of  $E_c$  and  $s$ . A closer look at the structure of granular metals suggests that although  $d$ ,  $s$ , and  $E_c$  are not constant, their variations are correlated. Let us consider the case of cermets. Due to the fact that the metal grains are formed by surface diffusion of sputtered metal atoms and insulator molecules, we can expect the relative volume fractions of metal and dielectric to be uniform when averaged over a volume larger than a few surface diffusion lengths ( $\sim 100$  Å). That is, if we take any 100-Å-sized region and analyze its relative volume fractions of metal and insulator, we would expect that number to be a constant and equal to that of the bulk sample. This condition of uniform composition on the scale of surface diffusion length implies a close relationship between the average grain size and the average distance of separation between the grains within small regions (100 Å in diameter, say) of the sample. Roughly speaking, we can expect the large grains to be further separated than small grains. If it is further assumed that the grains are roughly uniform in size inside small regions of the sample (but can vary from one region to the next), then the grain size  $d$  in each region should be directly proportional to the grain separation  $s$  in that region, and the ratio  $s/d$

is a constant whose magnitude depends only on  $x$ . This simple fact is illustrated for the one-dimensional case in Fig. 5. From Eq. [1] and the condition  $s/d = \text{constant}$ , it follows directly that  $sE_c^0 = \text{constant}$  for a cermet of a given composition. In the case of discontinuous films, a functional relationship between  $E_c^0$  and  $s$  can be derived by requiring the amount of deposited metal, averaged over any small area, to be constant.

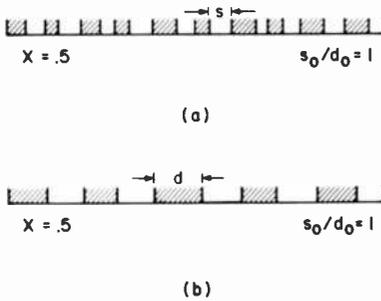


Fig. 5—One-dimensional illustration of the idea that the ratio of metal particle separation  $s$  to metal particle size  $d$  is only a function of  $x$ . Particles are shaded. The average particle size in (b) is about twice that in (a).

We now derive, on the basis that  $sE_c^0 = \text{constant}$ , the low-field conductivity. We assume that the generated positive-negative pairs of charged grains are roughly of the same size, so that each contributes about half of the charging energy  $E_c^0$ . Let us consider all those charge carriers with charging energy  $E_c^{0'}$ . Their number density is proportional to the Boltzmann factor  $\exp(-E_c^{0'}/2kT)$ . When a weak electric field ( $e\Delta V \ll kT$ ) is applied, the charge will drift along a path of largest mobility towards the electrodes. The charge is inhibited from tunneling to a much smaller grain with  $E_c^0 \gg E_c^{0'}$  because the charge has insufficient energy. The charge will, therefore, tunnel to grains with  $E_c^0 \leq E_c^{0'}$ . However, since smaller  $E_c$  is associated with a larger tunneling barrier by the rule  $sE_c^0 = \text{constant}$ , the optimal path, which we will refer to as the percolation path, follows the regions with the least deviations of  $E_c^0$  from  $E_c^{0'}$ . The total low-field conductivity,  $\sigma_L$ , is the sum of the conductivity given by Eq. [11] integrated over all possible percolation paths:

$$\sigma_L \propto \int_0^{\infty} \beta(s) \exp[-2\chi s - (C/2\chi s kT)] ds. \quad [16]$$

where

$$C \equiv \chi s E_c^0 \quad [17]$$

is a constant and  $\beta(s)$  is the density of percolation paths associated with the value  $s$ . The exponent in Eq. [16],  $-2\chi s - (C/2\chi s kT)$ , is a peaked function of  $s$  whose maximum occurs at  $s_m = \sqrt{C/kT}/2\chi$ . By performing the integration by the method of steepest descent, the dominant temperature dependence is given by  $\sigma_L \sim \exp[-2(C/kT)^{1/2}]$ . The pre-exponential factor of the conductivity is not given by the above argument because its exact form depends on more complicated considerations of counting the percolation paths. For purpose of comparing theory with experiment, the temperature dependence of the pre-exponential factor is neglected completely and we write

$$\sigma_L = \sigma_0 \exp[-2\sqrt{C/kT}]. \quad [18]$$

A more detailed derivation of Eq. [18] has been given previously.<sup>3</sup>

### 2.3 High-Field Conductivity

In the high-field regime, where  $e\Delta V \gg kT$  and  $kT \ll E_c$ , the number density of charge excitations deviates appreciably from the thermal equilibrium value given by Eq. [2], and is determined by the generation rate given by Eq. [8].

To illustrate clearly the consequences of the relation  $E_c^0 s = \text{constant}$  on the high-field conductivity, we first give a simple derivation of the conductivity at  $T = 0$ . We make use of the approximation  $E_c^1 \simeq \frac{1}{2}E_c^0$ , which is valid in granular metals over a wide range of compositions.<sup>3</sup> For a certain barrier thickness  $s$ , there is a voltage threshold due to the existence of a charging energy  $E_c^1 = C/2\chi s$  below which there is no tunneling. An equivalent restatement of the threshold condition is that at a given voltage, electrons will only tunnel through barriers with thickness  $\geq C/2\chi e\Delta V$ . Since the tunneling rate depends exponentially on the barrier thickness, the dominant contribution to the conductivity is due to electron tunneling through barriers with thickness  $s \simeq C/2\chi e\Delta V$ . Therefore, the dominant field dependence of the high-field conductivity  $\sigma_H$  is proportional to  $\exp(-\chi C/\chi e\Delta V)$ , or

$$\sigma_H = \sigma_0 \exp(-\varepsilon_0/\varepsilon), \quad [19]$$

where  $\sigma_0$  and  $\varepsilon_0$  are constants.

To determine  $\sigma_H$  at finite temperatures and to calculate  $\sigma_0$  and  $C_0$ , we express the current density  $j_H$  in the form<sup>3</sup>

$$j_H = e\mathcal{N}l_H \int_0^\infty r_n^+(s)D(s)ds. \tag{20}$$

where the generation rate  $r_n^+$  is given by Eq. [8],  $r_n^-$  was neglected,  $D(s)$  is the distribution function for  $s$ , and  $l_H$  is the recombination length given by Eq. [15]. Because of the field dependence of  $\tau$  in the high-field regime,  $l_H$  decreases with increasing field to a limiting value. Thus at sufficiently high fields  $l_H$  can be taken as constant. Assuming  $D(s) = (s/s_0)^2 \exp(-s/s_0)$  and  $\Delta V = w\mathcal{E}$ , where  $w$  is a constant of the order of  $s_0 + d_0$  ( $s_0$  and  $d_0$  are the averages of  $s$  and  $d$ ), integration of Eq. [20] yields<sup>3,13</sup>

$$\sigma^* = \exp(-1/\mathcal{E}^*) F(\mathcal{E}^*, T^*). \tag{21}$$

where  $\sigma^* = \sigma/\sigma_\infty$ ,  $\mathcal{E}^* = \mathcal{E}/\mathcal{E}_0$ ,  $T^* = kT/C_0$ , and

$$F(\mathcal{E}^*, T^*) = \int_{-1/\mathcal{E}^*}^\infty dZ Z \exp(-Z) \times \left\{ 1 - \exp\left[-\frac{\mathcal{E}^*}{Z + (\mathcal{E}^*)^{-1}}\right] \right\}^{-1}. \tag{22}$$

$$C_0 = C \left( 1 + \frac{1}{2\chi s_0} \right) \equiv \chi s E_c^0 \left( 1 + \frac{1}{2\chi s_0} \right). \tag{23}$$

$$\mathcal{E}_0 = C_0/ew. \tag{24}$$

$$\sigma' = e^2 l_H \gamma / [(s_0 + d_0)^2 (2\chi s_0 + 1)^2 h].$$

Since usually  $1/(2\chi s_0) \ll 1$ , it follows that  $C_0 \simeq C$ . In the limit of  $T = 0$ , where  $F = 1$ , Eq. [21] reduces to Eq. [19]. Tabulated values of  $F(\mathcal{E}^*, T^*)$  are given in Ref. [4].

Eq. [21] predicts a scaling behavior for the temperature and the field dependences of the high-field conductivity. In other words, Eq. [21] states that when the high-field conductivity data for granular metals with different compositions are plotted on scales of reduced temperatures and reduced field, they obey a universal behavior that is characteristic of field-induced tunneling. The constants  $C_0$  and  $\mathcal{E}_0$ , whose values depend on the composition and dielectric constant of the granular metal, determine the field and the temperature scales over which this universal behavior is observed. This remarkable result, a direct consequence of the rule  $sE_c^0 = \text{constant}$ , has been verified experimentally and offers an additional test of the theory.

It is instructive to point out the similarities and the differences between Mott's model<sup>21</sup> of variable range hopping in amorphous semiconductors and the tunneling model in granular metals. In Mott's

model the density of charge carriers is assumed to be temperature independent, and the percolation paths for the charge carriers are determined by optimizing the mobility. In the granular-metals model, the charge carriers are thermally activated in the low-field regime and field-generated in the high-field regime; tunneling occurs between nearest neighbors only; and the optimization, when it is used in deriving the temperature dependence of  $\sigma_L$ , is applied to the product of mobility and number density of charge carriers. The differences between charging energies in the granular metal are analogous to the relative displacements of the energy levels for the localized states in Mott's model. However, in the case of the granular metal not only the differences, but also the magnitudes of  $E_c$ , play an important role. This is especially obvious in the high-field regime, where the governing factor for field generation of charge carriers is the value of  $E_c$  rather than the differences in  $E_c$ .

We now wish to comment on some inconsistencies that have appeared in the literature in the treatment of the field dependence of the conductivity. Some authors<sup>7,8,10,22</sup> proposed that the field dependence of the conductivity results from a Poole-Frenkel effect; the authors assumed that the electrons are thermally emitted across a barrier given by the charging energy and that this barrier is lowered by the applied electric field. The authors treated  $E_c$  as a barrier and not as the difference in the Fermi levels of the neighboring grains between which the tunneling takes place. In the tunneling model<sup>3,13</sup> the barrier  $\varphi$  is due to the insulator, and the contribution of  $E_c$  to  $\varphi$  is small as long as  $\varphi \gg E_c$ . The effect of  $E_c$  is treated correctly as the difference in the Fermi levels  $\mu_1$  and  $\mu_2$  of the grains 1 and 2 measured with respect to a fixed electrochemical potential (Figure 4).

Thermionic emission is operative only when thermal excitation over the barrier  $\varphi$  exceeds tunneling through the barrier, i.e., when  $\varphi/kT < 2\chi s$ . In granular metals this condition is usually not met, and transport in most cases is due to tunneling. Nonetheless, if thermionic emission were operative, an additional condition must be satisfied in order for the Poole-Frenkel field-dependence to apply. The condition is that the electric field must be larger than a certain minimum value. In the following we derive the value of this minimum field.

The low-field conductivity in the case of thermionic emission is determined from the Richardson-Dushman equation, and is given by<sup>10</sup>

$$\frac{e(s+d)}{kT} i_0 \exp(-\varphi/kT), \quad [25a]$$

where  $i_0 = 4\pi me(kT)^2/h^3$  is the Richardson-Dushman constant.

When an image potential is included, the energy  $W$  required to move an electron a distance  $y$  from the grain is given by

$$W = \varphi - \left[ e^2 / \epsilon_l \left( \frac{d}{2} + y \right) \right] - ye \mathcal{E}', \quad [25b]$$

where  $\mathcal{E}'$  is the local field, which can be expressed in terms of the average field  $\mathcal{E}$  by,

$$\mathcal{E}' = \gamma' \mathcal{E}, \quad [26]$$

where  $\gamma' > 1$  is a field-enhancement factor. If we make the simplifying assumption that  $\mathcal{E}'$  is constant, Eq. [25b] exhibits a maximum,  $W_{max}$ , at  $y_{max} = (e/\epsilon_l \mathcal{E})^{1/2} - 1/2d$ . The thermionic emission current,  $i_{TH}$ , is given by

$$i_{TH} = i_0 \exp(-W_{max}/kT) \simeq i_0 \exp(-\varphi + \beta \sqrt{\mathcal{E}'})/kT, \quad [27]$$

where

$$\beta = 2(e^3/\epsilon_l)^{1/2}, \quad [28]$$

we have neglected in  $W_{max}$  the term  $(d/2)e\mathcal{E}'$ , and we assumed  $e\Delta V \gg kT$ . The field dependence of  $i_{TH}$  expressed in terms of the average field  $\mathcal{E}$  is given by

$$i_{TH} \propto \exp(\beta' \sqrt{\mathcal{E}}/kT), \quad [29]$$

where

$$\beta' = \gamma' \beta. \quad [30]$$

In order for the Poole-Frenkel field dependence given by Eq. [29] to apply to granular metals, it is required that  $y_{max} < 1/2d + s$ , i.e., the maximum in the barrier potential must fall between two grains. To satisfy this condition, the field must be strong enough so that

$$\mathcal{E} > e/\epsilon_l (d + s)^2 \gamma'. \quad [31]$$

Eq. [31] is used later in the interpretation of the field dependence of the conductivity of discontinuous films.

### 3. Comparison of Theory and Experiment

#### 3.1 Cermets

Results of measurements of conductivity on a variety of cermets by different workers<sup>3</sup> agree well with the predictions of the tunneling

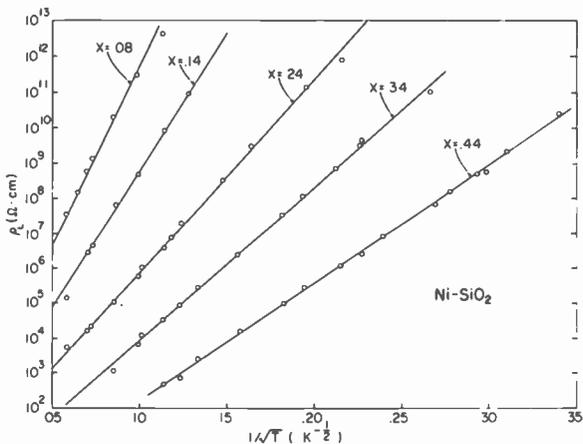


Fig. 6—Electrical resistivity in the low-field ohmic regime  $\rho_L$ , as a function of  $1/\sqrt{T}$ , for several Ni-SiO<sub>2</sub> films after Abeles et al.<sup>3</sup> The volume fractions  $x$  of Ni are indicated in the figure. The points are the experimental data, the straight lines represent the relation  $\log \rho_L \sim 1/\sqrt{T}$ .

theory. This is demonstrated by the examples in Figs. 6 and 7. The low-field resistivity  $\rho_L$  in Fig. 6 exhibits the characteristic temperature dependence  $\log \rho_L \sim 1/\sqrt{T}$ , in agreement with Eq. [18], and the field dependence of the high-field resistivity  $\log \rho_H \sim 1/\delta$  is in agreement with Eq. [19]. Due to the wide range of resistivity, temperature,

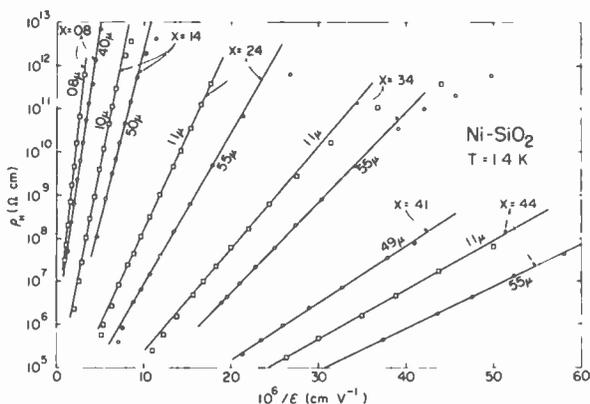
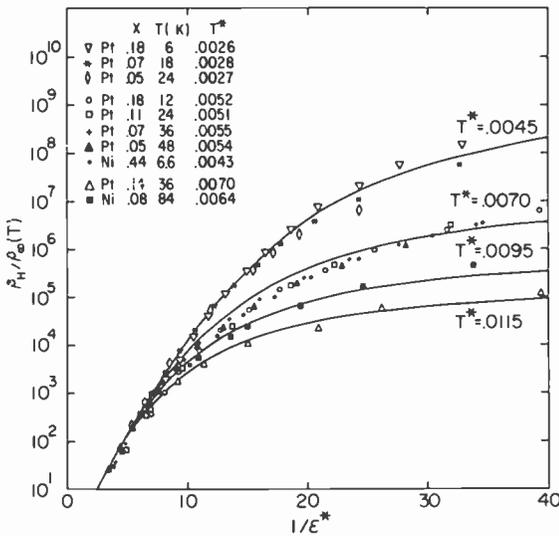


Fig. 7—The high-field electrical resistivity  $\rho_H$  measured at 1.4°K as a function of  $1/\delta$  for several Ni-SiO<sub>2</sub> films after Abeles et al.<sup>3</sup> The thicknesses of the films and the volume fraction  $x$  of Ni are indicated in the figure. The points are the experimental data, the solid lines represent the relation  $\log \rho_H \sim 1/\delta$ .

and field values, the dependence of the resistivity on  $T$  and  $\mathcal{E}$  can readily be distinguished from the Mott  $\ln \rho_L \sim 1/T^{1/4}$  or the  $\ln \rho_L \sim 1/T$  type dependence at low fields and the Poole-Frenkel behavior, Eq. [29], at high fields.

The theory further predicts that the reduced resistivity  $\rho^* \equiv \rho_H(\mathcal{E}, T)/\rho_\infty(T)$ , plotted as a function of reduced field  $\mathcal{E}^* \equiv \mathcal{E}/\mathcal{E}_0$ , and reduced temperature,  $T^* \equiv kT/C_0$  (or  $T^* \equiv kT/C$ , since  $C_0 \simeq C$ ), is a universal function independent of the composition of the granular metals. The validity of this scaling behavior is demonstrated by the



**Fig. 8**—Reduced resistivity  $\rho_H/\rho_\infty(T)$ , as a function of reduced field  $\mathcal{E}^* = \mathcal{E}/\mathcal{E}_0$  and reduced temperature  $T^* = kT/C_0$ , for several Ni-SiO<sub>2</sub> and Pt-SiO<sub>2</sub> films after Abeles et al.<sup>3</sup> The volume fraction of the metal  $x$ , the temperature  $T$ , and reduced temperature  $T^* = kT/C$  of the samples are indicated in the figure. The points are experimental, the solid curves were computed from Eq. (2) for the values of  $T^*$  indicated next to each curve.

fact that the field dependences of granular metals with widely different compositions measured at widely different temperatures but at approximately the same reduced temperature follow a single curve. This is shown for several Ni-SiO<sub>2</sub> and Pt-SiO<sub>2</sub> films in Fig. 8. The values of  $T^*$  obtained by fitting Eq. [21] to the experimental data are somewhat larger than the values of  $T^* = kT/C$  obtained from low-field measurements. The values of the parameters  $\mathcal{E}_0$ ,  $C$ , and  $C_0$  that yield the best fit to the data for the Ni-SiO<sub>2</sub> films are listed in Table

1. The theoretical prediction that  $C_0 \approx C$  is borne out experimentally within a factor of two, as is seen from Table 1 and Fig. 8.

Other authors have interpreted the field dependence of the conductivity of granular metals in terms of the Poole-Frenkel effect given by Eq. [29]. For instance, Minn and Pinguet<sup>23</sup> and Hauser<sup>24</sup> found a linear relation of  $\log \sigma$  with  $\sqrt{\mathcal{E}}$  over a range of several orders of magnitude in  $\sigma$ . However, this fit is fortuitous, since an unreasonably large value of the dielectric constant  $\epsilon_1$  is required to obtain quantitative agreement with Eq. [28]. For instance, Hauser used  $\epsilon_1 \approx 100$  to fit the data for Sn-SnO<sub>2</sub>. It should be pointed out that although the dielectric constant of granular metals  $\epsilon_{GM}$  can be high<sup>4</sup> (in the case of the Sn-SnO<sub>2</sub> films the value  $\epsilon_{GM} = 82$  was reported), it is the dielectric constant of the insulator, which is in the range<sup>4</sup> 2-10,

Table 1—Parameters of Ni-SiO<sub>2</sub>

x vol. fraction Ni	$\mathcal{E}_0$ (10 <sup>6</sup> V/cm)	$C_0$ (eV)	$C$ (eV)	$w$ (Å)	$s_0 + d_0$ (Å)
0.44	0.22	0.085	0.13	60	50
0.34	0.48		0.22	45	49
0.24	1.0	0.30	0.35	44	35
0.14	1.8		0.68	38	25
0.08	3.4	0.90	1.1	34	20

that must be used in Eq. [28]. Thus, in the case of cermets, we are led to reject the Poole-Frenkel model on the ground that it leads to inconsistencies in the dielectric constant of the insulator and, as has been pointed out in the previous section, that it lacks theoretical justification. Moreover, the data of Hauser and Minn and Pinguet can be fitted well by the tunneling theory. This is shown in Figs. 9 and 10. The values of the parameters  $\mathcal{E}_0$ ,  $C$ , and  $C_0$  that yield the best fit to the data are listed in Table 2. We note that  $C > C_0$ , a trend that is also found in the Ni-SiO<sub>2</sub> and Pt-SiO<sub>2</sub> systems (Fig. 8 and Table 1).

We now wish to relate the transport parameters  $\mathcal{E}_0$ ,  $C$ , and  $C_0$  to the material and structural parameters of the granular metal. The quantities  $C_0$  and  $\mathcal{E}_0$  are related through Eq. [24], and the parameter  $w$  is expected to be of the order  $w \approx s_0 + d_0$ . The values of  $w$  ( $= C_0/e\mathcal{E}_0$ ) in Tables 1 and 2 are in reasonable agreement with  $d_0 + s_0$  determined from electron micrographs. The parameter  $w$  can also be deduced from the value of the field at which the conductivity goes from the ohmic (low-field) to the non-ohmic (high-field) behavior. This is expected to occur when  $e\Delta V \equiv e\mathcal{E}w = kT$ . Fig. 11 shows the resistivity  $\rho$  ( $\mathcal{E}$ ,  $T$ ) at constant temperature, normalized by the zero-

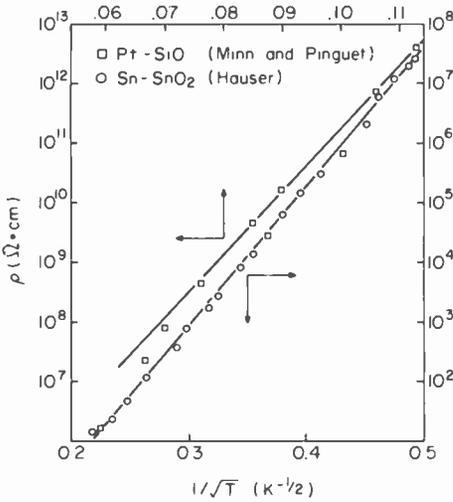


Fig. 9—Electrical resistivity in the low-field ohmic regime  $\rho_L$ , as a function of  $1/\sqrt{T}$ , of Pt-SiO (Minn and Pinguet<sup>23</sup>) and Sn-SnO<sub>2</sub> (Hauser<sup>24</sup>). The points are the experimental data; the straight lines represent the relation  $\log \rho_L \sim 1/\sqrt{T}$ .

field resistivity  $\rho(O, T)$ , as a function of the field  $\mathcal{E}$ , normalized in the form  $e\mathcal{E}w/kT (\equiv \mathcal{E}^*/T^*)$ . It can be seen that, over a wide range of compositions and reduced temperatures, the transition from the low-field ohmic behavior to the non-ohmic high-field behavior can be described by the relation

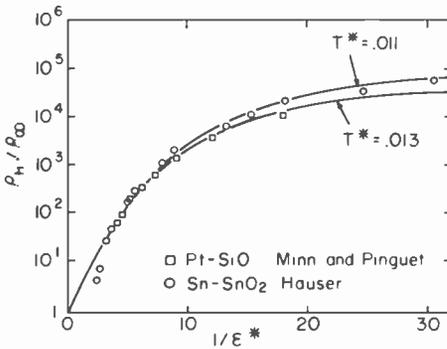


Fig. 10—Reduced resistivity  $\rho_H/\rho_\infty(T)$ , as a function of reduced field  $\mathcal{E}^* = \mathcal{E}/\mathcal{E}_0$  for Pt-SiO (Minn and Pinguet<sup>23</sup>), at 78°K and Sn-SnO<sub>2</sub> (Hauser<sup>24</sup>) at 4.2°K. The points are experimental; the solid curves were computed from Eq. [21] for the indicated value of  $T^*$ .

Table 2—Parameters of Pt-SiO and Sn-SnO<sub>2</sub>

	vol. fraction metal	$\epsilon_0$ 10 <sup>6</sup> V/cm	$C_0$ (eV)	$C$ (eV)	$w$ Å
Pt-SiO <sup>23</sup>	0.04 <sup>a</sup>	0.96	0.52	1.3	54
Sn-SnO <sub>2</sub> <sup>24</sup>	0.46 <sup>b</sup>	0.10	0.033	0.062	30

<sup>a</sup> Calculated from mole percent of SiO (92%) assuming that the density of the insulator SiO is 2.13 and of Pt is 21.4 gm cm<sup>-3</sup>.

<sup>b</sup> Estimated from the room-temperature resistivity of sample #44 (1 ohm-cm) and Fig. 7, Ref [24].

$$\rho(\epsilon, T) / \rho(0, T) = \exp(-\alpha e \epsilon w / kT), \quad [32]$$

where  $\alpha$  is a constant in the range 1–2. Thus, the microscopic granular metal structural constant  $w$  can be deduced from macroscopic measurements of temperature and field dependence of the electrical conductivity.

The constant  $C$  defined by Eq. [17] depends only on the composition  $x$ , the dielectric constant  $\epsilon_1$ , and  $\chi$ . Thus, in a granular metal for a definite value of  $x$  the constant  $C$  is expected to be independent of grain size, and the constituent metal and insulator will affect the value of  $C$  only insofar as the barrier height and  $\epsilon_1$  are affected. The invariance of  $C$  with respect to grain size has been demonstrated ex-

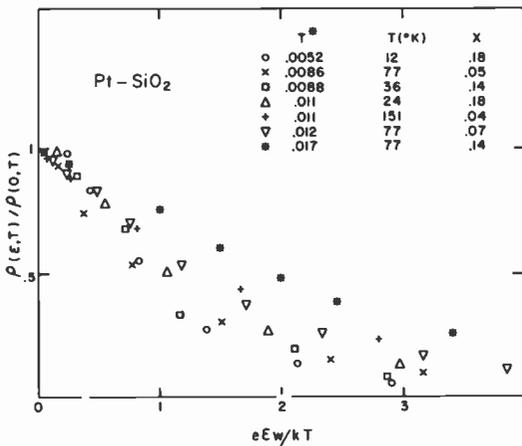


Fig. 11—Normalized resistivity  $\rho(\epsilon, T) / \rho(0, T)$ , as a function of electric field  $\epsilon$  (normalized by  $kT/ew$ ), for several Pt-SiO<sub>2</sub> films after Abeles et al.<sup>3</sup> The volume fraction  $x$  of Pt, the temperature, and the reduced temperatures  $T^* = kT/C$  of the samples are indicated in the figure.

perimentally in the  $W-Al_2O_3$  system.<sup>5</sup> In these experiments, it was found that on annealing the  $W-Al_2O_3$  samples at high temperatures, the values of  $d_0$  increased by a factor of 2–5 depending on the composition  $x$ . At the same time, in the dielectric regime,  $\sigma_0$  was found to decrease by several orders of magnitude, yet the values of  $C$  remained unchanged. Thus,  $C$  is a useful materials characterization parameter because, unlike resistivity, it is insensitive to the grain size and the method of preparation. In Fig. 12,  $C$  is plotted as a function of  $x$  for

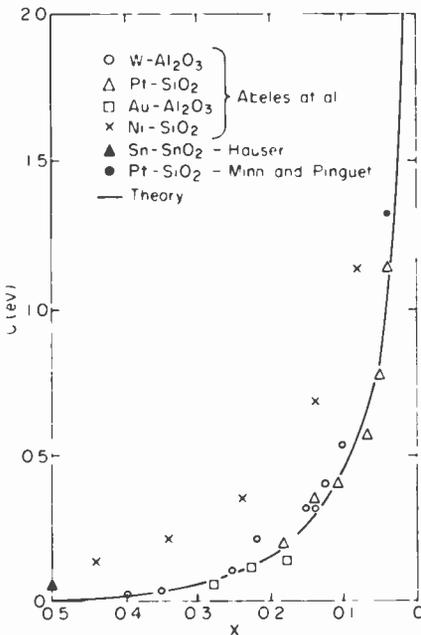


Fig. 12—The parameter  $C$  as a function of volume fraction of the metal  $x$ , after Abeles et al.,<sup>3</sup> Pinguet and Minn,<sup>23</sup> and Hauser.<sup>24</sup> The theoretical curve was computed from Eqs. [33] and [34] with  $\eta = 1$  eV.

several granular metal systems. With the exception of  $Ni-SiO_2$ , we note that  $C(x)$  is insensitive to the constituent metals and insulators. In the case of  $Ni-SiO_2$  the displacement of the  $C(x)$  curve to higher values of  $x$  has been ascribed to the presence of  $NiO$  in the films.<sup>3</sup> The theoretical curve for  $C(x)$  was calculated assuming a model in which the granular metal is replaced by a simple cubic lattice of metal spheres of constant diameter  $d$  with a lattice constant  $s + d$ . The expression for  $C$  is<sup>3</sup>

$$C = \eta \frac{(s/d)^2}{[1/2 + (s/d)]}. \quad [33]$$

where

$$s/d = (\pi/6x)^{1/3} - 1, \quad [34]$$

$$\eta = 2\chi e^2 / \epsilon_1. \quad [35]$$

The solid curve in Fig. 12 represents the best fit of Eqs. [33] and [34] with a value of  $\eta = 1$  eV. We can calculate a theoretical value of  $\eta$  from Eq. [35]. For instance, using  $\chi = 0.5 \text{ \AA}^{-1}$ , which corresponds to  $\varphi = 1$  eV and a free-electron mass, and using  $\epsilon_1 = 8.5$ , which is the value measured in Au-SiO<sub>2</sub>,<sup>4</sup> we obtain  $\eta = 1.6$  eV, which is somewhat larger than the experimental value. Such a discrepancy could be due to the effect that sharp protrusions on the metal grains have on tunneling.<sup>3</sup>

### 3.2 Discontinuous Metal Films

Measurements by Neugebauer and Webb<sup>8</sup> on discontinuous Au, Pt, and Ni films indicate a temperature dependence of the low-field conductivity given by the relation  $\log \sigma_L \sim -1/T$ , characteristic of a single activation process. Fig. 13 shows the low-field conductivity of an Ni film after Neugebauer and Webb,<sup>8</sup> demonstrating the linearity of  $\log \sigma_L$  with  $1/T$  over a range of six orders of magnitude. It should be pointed out, however, that there are also results for discontinuous films for which the temperature dependence of the conductivity deviates from the  $\log \sigma_L \sim -1/T$  behavior, indicating a distribution of activation energies.<sup>8,15,25</sup>

The field dependence of the conductivity in discontinuous films has customarily been fitted by the Poole-Frenkel relation, Eq. [29]. An example of such a fit is given in Fig. 14 for an Ni film and Table 3 summarizes the results of several workers.<sup>8,10,26</sup> The values of  $\beta'$  are given in column 3, the temperatures at which these values were determined in column 2, and the activation energies, obtained from the slope of  $\log \sigma_L$  vs.  $-1/T$ , are given in column 5.

The interpretation of the field dependence of the conductivity in terms of the Poole-Frenkel thermionic emission model raises a number of problems. If we assume conduction by thermionic emission through the glass with  $\epsilon_1 = 4$  and we take  $d = 15 \text{ \AA}$  and  $s = 25 \text{ \AA}$  (the values reported for the Pt film<sup>10</sup>), we calculate from Eq. [31]  $\gamma' \mathcal{E} > 2.3 \times 10^5 \text{ V/cm}$ , and from Eq. [28],  $\beta = 3.8 \times 10^{-4} \text{ eV V}^{-1/2} \text{ cm}^{1/2}$ . The value  $2.3 \times 10^5 \text{ V/cm}$  is orders of magnitude higher than the fields at which the films were measured, and  $\gamma' = 10^2\text{--}10^3$  is required to satisfy

Table 3—Parameters of Discontinuous Metal Films

metal	T °K	$\beta'$ eV V <sup>-1/2</sup> cm <sup>1/2</sup>	w (Å)	activation energy eV
Pt <sup>10</sup>	300	$1.6 \times 10^{-4}$	(a)	0.11
Au <sup>26</sup>	300	$1.1 \times 10^{-4}$	25	0.24
Au <sup>26</sup>	300	$3.7 \times 10^{-4}$	350	0.07
Ni <sup>8</sup>	118	$3.7 \times 10^{-4}$	400	0.10

<sup>a</sup> Author does not provide the field dependence of  $\sigma$  over sufficiently wide range of fields to be able to fit Eq. [29].

Eq. [31]. Such high values of  $\gamma'$  imply extremely inhomogeneous samples in which the voltage drop is across a small number of high-impedance gaps. Moreover, such an interpretation leads to an inconsistency; the assumption of a large  $\gamma'$  results, according to Eq. [30], in values of  $\beta$  that are 2-3 orders of magnitude smaller than the theoretical value given by Eq. [28].

There are also difficulties in interpreting the data in terms of the tunneling model. The observed temperature dependence of the low-field conductivity is in some cases<sup>8</sup> characteristic of a single activation process. This is difficult to reconcile with the fact that the films

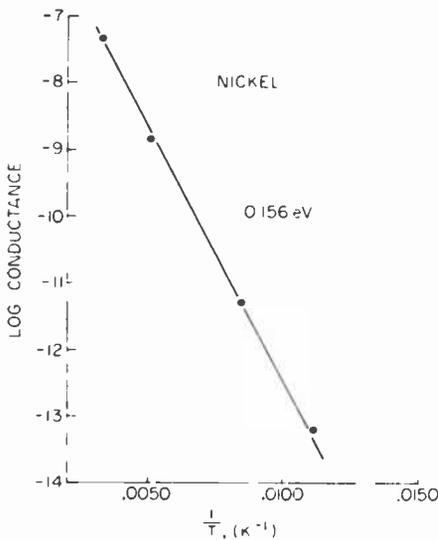


Fig. 13—Low-field resistivity  $\rho_L$  as a function of  $1/T$  for a discontinuous Ni film; after Neugebauer and Webb;<sup>8</sup> straight line represents relation  $\log \rho_L \sim 1/T$ . The value of the activation energy is indicated in the figure.

possess a distribution of grain sizes and hence a distribution of charging energies. Since data on the field dependence of the conductivity in discontinuous films is in most cases available only up to fields where the onset of the non-ohmic behavior is just beginning, it is not possible for us to test the validity of Eq. [21], which applies only at high fields, i.e.,  $e\Delta V \gg kT$ . In view of this we have used the empirical

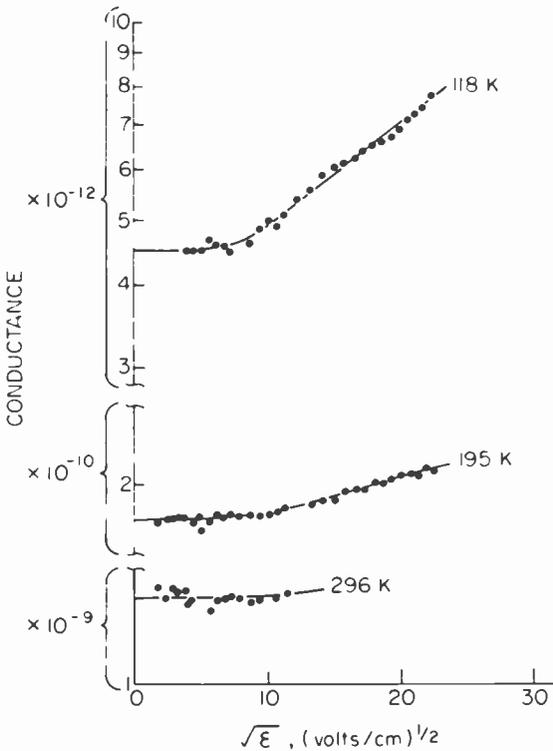


Fig. 14—Conductivity  $\sigma$  of a discontinuous Ni film as a function of  $\sqrt{\mathcal{E}}$ ; after Neugebauer and Webb;<sup>8</sup> straight line represents the relation  $\log \sigma \sim \sqrt{\mathcal{E}}$ .

relation given by Eq. [32], which was found to describe the transition from the low field to the high field behavior in the case of cermets (Fig. 11). Using  $\alpha = 2$ , which, according to Fig. 11, is appropriate for films whose field dependence is measured at high reduced temperatures, we have fitted Eq. [32] to the experimental data with  $w$  as an adjustable parameter. The values of  $w$  obtained in this manner are given in Table 3. With the exception of one of the Au films ( $w = 25 \text{ \AA}$ ) the values of  $w$  are a factor of 5–10 times larger than typical values

of  $s_0 + d_0$  reported for discontinuous films. We note that also in the case of cermets (Table 1 and also Ref. [3])  $w$  tends to be larger than  $s_0 + d_0$ ; however, there the discrepancy is never larger than a factor of 2.

In order to establish more firmly the validity, or the lack of it, of the tunneling model in the case of discontinuous films, data over a wider range of fields and temperatures are required. Unfortunately such data are difficult to obtain because the conductivity of discontinuous films is by necessity measured in the plane of the film, and consequently the resistance is very high. In cermets the situation is much more favorable, because the conductivity can be measured normal to the film, so that the resistance is much lower, and high fields can be readily established by applying only several volts across the sample. This perhaps explains the fact that discontinuous films, which have been studied since the beginning of this century, are not as well understood as cermets, which were introduced only in the past decade, mainly as a result of the previous work on discontinuous films.

### Acknowledgments

I am grateful to Ping Sheng and Albert Rose for many helpful discussions. This work is dedicated to Albert Rose who so generously and patiently shared his great physical insight with us.

### References:

- <sup>1</sup> F. G. Swann, "The Electrical Resistance of Thin Metallic Films, and a Theory of The Mechanism of Conduction in Such Films," *Phil. Mag.*, **28**, p. 467 (1914).
- <sup>2</sup> Reviews of transport in thin films are given by T. J. Coultts in *Electrical Conduction in Thin Metal Films* (American Elsevir, 1974); and Z. H. Meiksin, "Discontinuous and Cermet Films," *Physics of Thin Films*, Vol. 8, Academic Press Inc., San Francisco, 1975.
- <sup>3</sup> B. Abeles, P. Sheng, M. D. Coultts, and Y. Arie, "Structural and Electrical Properties of Granular Metal Films," *Advances in Physics*, **24**, p. 407 (1975).
- <sup>4</sup> B. Abeles, *Applied Solid State Science*, Vol. 6, edited by R. Wolfe (Academic Press, New York) in press.
- <sup>5</sup> B. Abeles, H. L. Pinch, and J. Gittleman, "Percolation Conductivity in  $W-Al_2O_3$  Granular Metal Films," *Phys. Rev. Lett.*, **35**, p. 247 (1975).
- <sup>6</sup> C. R. Wronski, B. Abeles, and A. Rose, "Granular Metal-Semiconductor Vidicon," *Appl. Phys. Lett.*, **27**, p. 91 (1975).
- <sup>7</sup> C. J. Gorter, "A Possible Explanation of the Increase of the Electrical Resistance of Thin Metal Films at Low Temperatures and Small Field Strengths," *Physica*, **17**, p. 777 (1951).
- <sup>8</sup> C. A. Neugebauer and M. B. Webb, "Electrical Conduction Mechanism in Ultra Thin, Evaporated Metal Films," *J. Appl. Phys.*, **33**, p. 74 (1962).
- <sup>9</sup> T. E. Hartman, "Electrical Conduction in Discontinuous Thin Metal Films," *J. Appl. Phys.*, **34**, p. 943 (1963).
- <sup>10</sup> R. M. Hill, "Electrical Conduction in Ultra Thin Metal Films, I. Theoretical," *Proc. Roy. Soc. A*, **309**, p. 377 (1969); and R. M. Hill, "Electrical Conduction in Ultra Thin Metal Films, II. Experimental," *Proc. Roy. Soc. A* **309**, p. 397 (1969).

- <sup>11</sup> R. Kiernan and D. W. Stops, "Tunneling Between Metallic Islands on a Dielectric Substrate," *Nature*, **224**, p. 907 (1969); A. Bar and R. D. Finney, "Tunneling of Electrons in Discontinuous Metal Films," *Thin Solid Films*, Vol. 24, p. 511, (1974).
- <sup>12</sup> P. Sheng and B. Abeles, "Voltage—Induced Tunneling Conduction in Granular Metals at Low Temperatures," *Phys. Rev. Lett.*, **28**, p. 34 (1972).
- <sup>13</sup> P. Sheng, B. Abeles, and Y. Arie, "Hoping Conductivity in Granular Metals," *Phys. Rev. Lett.*, **31**, p. 44 (1973).
- <sup>14</sup> J. G. Simmons in *Tunneling Phenomena in Solids* edited by E. Burstein and S. Lundquist (Plenum Press, New York, 1969), p. 135.
- <sup>15</sup> A. A. Milgram and Chih-Shun Lu, "Field Effect and Electrical Conduction Mechanism in Discontinuous Thin Metal Films," *J. Appl. Phys.*, **37**, p. 4773 (1966); and P. J. Dobson and B. J. Hopkins, "Electrical Conductance of Thin Ultra High-Vacuum-Evaporated Films of Tungsten on Glass," *J. Appl. Phys.*, **39**, p. 3074 (1968).
- <sup>16</sup> We note that the activation energy for the number density of fully dissociated charge carriers is  $E_c^0/2$  rather than  $E_c^0$ . The reason for this is that the dissociated positive and negative charge carriers are statistically independent particles and the energy of each charge excitation is  $E_c^0/2$  above the ground state (neutral grain). An analogous situation arises in intrinsic semiconductors when the density of states for holes and electrons are equal. In that case, the Fermi level lies in the middle of the band gap and the activation energy for the densities of holes and electrons is given by half the band gap.
- <sup>17</sup> R. Kubo, "Electronic Properties of Metallic Fine Particles," *J. Phys. Soc. of Japan*, **17**, p. 975 (1962); and R. Kubo, "Electrons in Small Metallic Particles," *Volume Jubilaire en l'Honneur d'Alfred Kastler*, La Societe Francaise de Physique, p. 325 (1969).
- <sup>18</sup> This is borne out experimentally by the fact that in ordinary (non-granular) metal-insulator-metal tunnel junctions, in which  $\varphi$  was estimated to be  $\approx 1$  eV, the tunneling conductance is only weakly voltage dependent up to voltages of 0.5 V, W. F. Brinkman, R. C. Dynes, and J. M. Rowell, "Tunneling Conductance of Asymmetrical Barriers," *J. Appl. Phys.*, **41**, p. 1915 (1970).
- <sup>19</sup> J. Dryer, D. Gore, and R. Speiser, "Electrode and Substrate Considerations in Discontinuous Thin Film Resistance," *J. Vac. Sci. Technol.*, **10**, p. 316 (1973).
- <sup>20</sup> N. C. Miller, P. Hardiman, and A. Shirn, "Transport Properties, Microstructure and Conduction Model of Co-Sputtered Au-SiO<sub>2</sub> Cermet Films," *J. Appl. Phys.*, **41**, p. 1850 (1970), measured conductivity activation energies in Au-SiO<sub>2</sub> cermets which were about one half as large as those they computed from the expression of Neugebauer and Webb.<sup>8</sup> As was pointed out in Ref. [16], the activation energy is  $E_c/2$  and not  $E_c$ , which brings the theoretical values into good agreement with the experimental ones.
- <sup>21</sup> N. F. Mott, "Conduction in Non-Crystalline Materials, III. Localized States in a Pseudogap and Near Conduction and Valance Bands," *Phil. Mag.*, **19**, p. 835 (1969).
- <sup>22</sup> J. E. Morris, "Calculation of Activation Energy in Discontinuous Thin Metal Films," *J. Appl. Phys.*, **39**, p. 6107 (1968).
- <sup>23</sup> S. S. Minn and J. Pinguet, "Étude des Résistances Électriques des Couches Minces de Cermet Pt-SiO à Faibles Teneurs en Métal," *C. R. Acad. So. Paris*, **71**, p. 915 (1970).
- <sup>24</sup> J. J. Hauser, "Electrical and Structural Properties of Amorphous Metal-Metal Oxide Systems," *Phys. Rev.*, **B7**, p. 4099 (1973).
- <sup>25</sup> B. T. Boiko, L. S. Palatnik, and A. N. Synelnikov, "Electric Conductivity and Structure of Discontinuous Metal Films on Dielectric," *Thin Solid Films*, **7**, p. 305 (1971).
- <sup>26</sup> L. A. Weitzenkamp and N. M. Bashara, "Conduction in Discontinuous Metal Films," *Trans. Met. Soc. AIME*, **236**, p. 351 (1966).

# Pattern Meaningfulness and Pattern Detectability

Joseph J. Mezrich

RCA Laboratories, Princeton, N.J. 08540

**Abstract**—The object-superiority effect is a recent finding which supports the notion that perceptual processing for recognition in brief exposures is enhanced by pattern meaningfulness. Since this result has potential significance for the modeling of visual pattern recognition, it is important to examine whether this type of effect can be demonstrated in a situation that does not involve brief exposures to determine whether the meaningfulness attribute simply affects the durability or retrievability of the stimulus in memory, and not perceptual processing.

Geometric patterns composed of thin straight lines are portrayed in static random dot arrays by deleting dots to construct the lines. Only one of these patterns represents a meaningful object. It is found that the ability to detect lines embedded in random dot arrays is significantly enhanced if the lines are part of an object pattern. This is an object-superiority type of effect, without the requirement of a brief exposure. An additional demonstration is presented which emphasizes the judgemental aspect of visual processing. Results are interpreted in terms of perceptual decision making.

## Introduction

Consider the following experimental scenario. An observer, while fixating a point, is presented with a brief exposure of one of the four line segments shown in Fig. 1. Presented along with this line segment are eight other line segments, such as those shown in Fig. 2. The arrangement of these distractor lines can be as shown in this figure, or can be

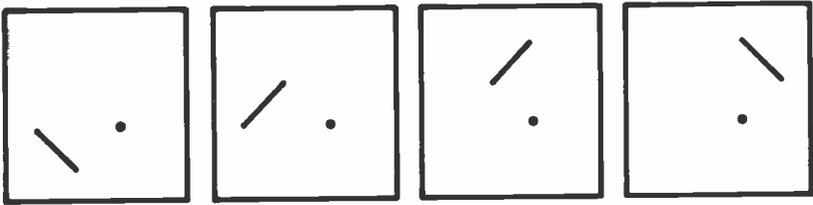
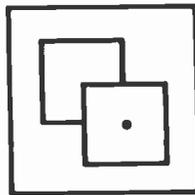


Fig. 1—Four test-line segments.

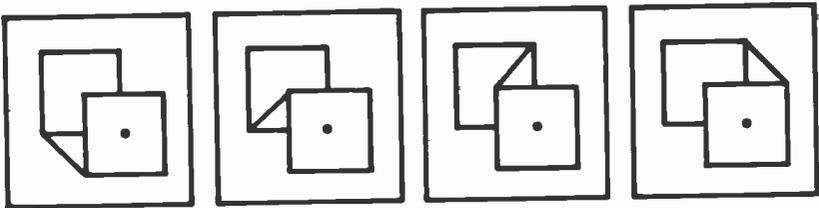


Fig. 2—One arrangement of distractor-line segments.

derived from a rigid translation of the individual lines. In particular, one possible arrangement of these distractors is shown in Fig. 3a, which, when presented with one of the test lines of Fig. 1, forms one of the unitary patterns, or identifiable objects, shown in Fig. 3b. In any other presented arrangement of the distractor lines, a single



(a)



(b)

Fig. 3—(a) Distractor-line arrangement that yields an object when presented with a test line, and (b) object patterns resulting from presentation of line arrangement in Fig. 3a with one of test lines of Fig. 1.

object pattern composed of the test and distractors is not produced. The observer's task is simply to report which of the four line segments of Fig. 1 was presented. Is report accuracy enhanced if the distractors are arranged so that the test line is part of one of the meaningful patterns shown in Fig. 3b?

An affirmative answer to this question has recently been reported by Weisstein and Harris.<sup>1</sup> They refer to this as the object-superiority effect.

On the face of it, this result says some quite significant things about visual information processing. For example, it is probably the most direct counterexample to the popular belief that visual form detection or recognition is a hierarchical process, the first stage being a local feature analysis. (A feature is usually taken to be an element in the visual field, such as an edge or a line, that can be identified with the class of trigger stimuli for single cells in the visual cortex reported, e.g., by Hubel and Wiesel.<sup>2</sup>) The object-superiority effect suggests that some global-holistic process is involved in pattern perception at a very early stage, and this global analysis is not preceded by local feature analysis. The reader may refer to the Weisstein and Harris<sup>1</sup> paper for further discussion of this point.

For a variety of reasons, such as restricting eye movements and limiting attention, a brief-flash paradigm was used. Perhaps the major reason for using this procedure is that it is a convenient way to degrade stimulus information so that errors are made. The occurrence of errors in a systematic fashion tells us something about how the observer handles the stimulus. The brief-flash technique is a classical device for examining human information processing, which was popular even in the nineteenth century (e.g., Pillsbury<sup>3</sup>).

Unfortunately, the use of brief exposures necessarily confounds the perceptual processing of the stimulus with the durability of the stimulus in short-term memory once it has been processed. It might be the case that in the Weisstein and Harris experiment all test-line and distractor-line arrangements are processed with equal efficiency; however, making the test line part of a meaningful figure makes it easier to retain in short-term memory, so that a correct response can be made. The object-superiority effect might, therefore, be solely attributable to the greater ease with which meaningful patterns can be held in memory, which is hardly a surprising result. The best protection workers using a brief-flash paradigm have against artifacts of this sort is to use a forced-choice response procedure rather than ask the observer to freely report what he saw. A forced-choice procedure was used by Weisstein and Harris. However, forced-choice response cannot completely eliminate memory as a potential source of ob-

tained effects. This general point has also been made by Smith and Spoehr<sup>4</sup> in a related context.

### Object Superiority in Steadily Viewed Pictures

Can one demonstrate an object-superiority effect in a situation that does not involve brief exposures, so that memory-based explanations

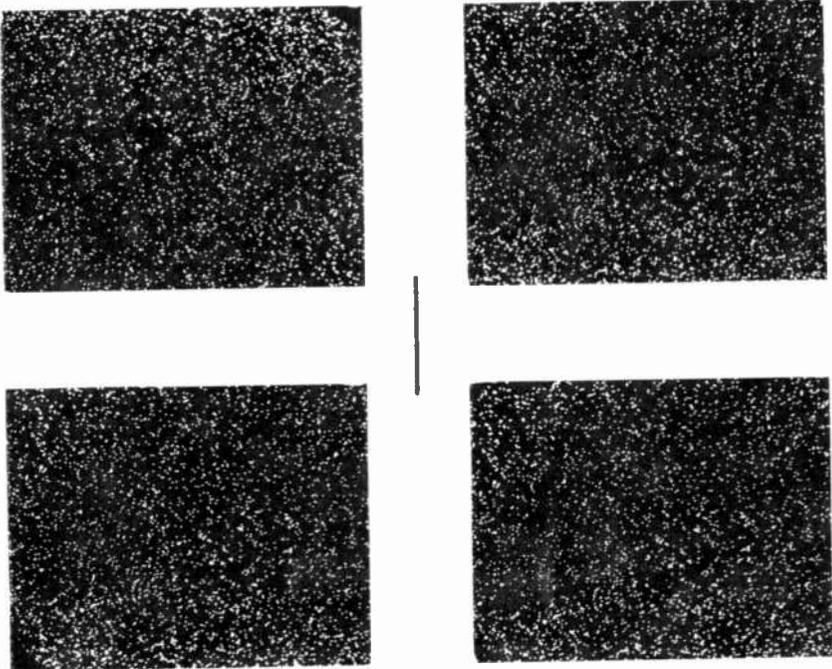


Fig. 4—Random needles in the haystack (see text for details).

can be disregarded? The type of test pattern introduced by Rose (see, e.g. Ref [5], pp. 12-14) to demonstrate signal detection properties of vision seems well suited for this purpose. An example of how such patterns can be used is shown in Fig. 4.

Consider the following "problem": A black line, like the one in the center of this figure, can be hidden in these dot pictures. After casual inspection, indicate which of these dot pictures you would feel most confident in reporting as having at least two black lines, each line resembling the one in the center of the figure. The lines may be at any orientation or location in the picture.

Fifty-one people were asked this question with a larger version of the figure. About half were shown the figure as oriented here, the remainder being shown the figure after 180° rotation. Responses did not noticeably depend on which orientation was used. Thirty-three out of fifty-one, or 65% of the people asked, chose the upper left dot picture in Fig. 4. There are, however, three correct answers to this question.

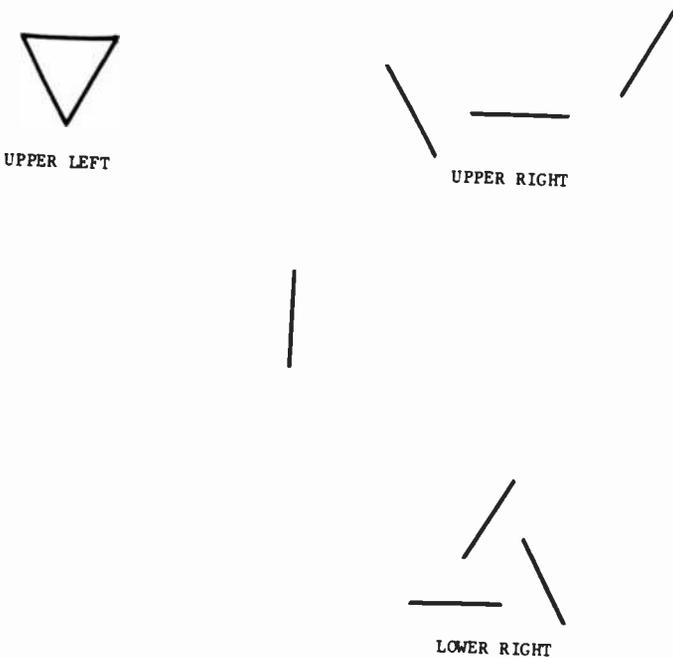


Fig. 5—Line transparencies used in producing the dot pictures of Fig. 4.

The lower left dot picture in this figure contains no intended target lines. This picture was produced by photographing a TV monitor supplied with a random train of brief, constant-amplitude pulses. Each white dot in the picture is the photographic record of a pulse displayed on the monitor during the exposure time of the film. A positive transparency of this photograph was made and used to produce the remaining three pictures.

Fig. 5 shows the line pattern transparencies that were overlaid on the master random dot transparency to produce the other dot pictures. The test line is displayed in the center of this figure. Three

such lines were used to construct the equilateral triangle of the upper left transparency. The upper-right and lower-right transparencies were derived from the triangle by rigidly translating its sides to form meaningless line arrays. The reader will find the three lines in each of the dot pictures of Fig. 4 if he allows for the fact that the line transparencies were off-center and slightly rotated when the pictures were produced.

So, three out of the four pictures in Fig. 4 contain at least two test lines. No one selected the dot pattern without intentionally embedded lines in response to our question. As was pointed out, 65% of those asked were most confident that the dot picture containing the triangle contained at least two lines. The tendency to select this pattern, rather than the other two correct alternatives, was statistically significant ( $\chi^2 = 21.2$ ,  $df = 1$ ,  $p < 0.001$ ). The responses of the eighteen people who did not select the triangle pattern were not evenly split among the other two patterns. Fifteen selected the upper right dot picture, and three chose the lower right picture. This difference is significant ( $\chi^2 = 6.7$ ,  $df = 1$ ,  $p < 0.01$ ), and is most likely due to the fact that the  $45^\circ$  line in the lower right picture fell on regions of fairly low dot density, essentially eliminating one of the opportunities for line detection in this picture. This type of line masking does not seem particularly evident for the eight other lines in Fig. 4.

If we assume that the dot picture containing the triangle was not overwhelmingly chosen over the other correct alternatives simply because of where the lines happened to fall for the particular dot background used, then we have a result that looks very much like a confirmation of the Weisstein and Harris object-superiority effect in a static picture.

It might be argued that our result is simply a consequence of line proximity, and the fact that a line arrangement can be interpreted as an object is irrelevant for visual detection. Further tests with a greater variety of line arrangements are in order to adequately respond to this. However, it is worth pointing out that when asked why they selected a particular picture, those who chose the upper left dot pattern often claimed they saw a triangle or the letter V (one side of the triangle not detected); i.e., perception of an object was reported. Those who selected one of the other pictures reported seeing separate lines, rather than an object. The assertion that the opportunity for an observer to perceptually construct a unitary image from a stimulus does affect visual processing will be made more convincing by a demonstration presented in the next section.

In any event, our results are not in conflict with those of Weisstein

and Harris, and furthermore our results avoid short-term memory-based explanations for a possible object-superiority effect. The use of these dot patterns to illustrate an effect of meaningfulness on visual detection also provides a convenient framework for speculating on how the fact that a line is potentially a member of an object can affect its being processed.

### Decision Making in Pattern Perception: A Source of Meaningfulness Effects

We have presented observers with what may be called noisy line segments in these pictures. The observer inspects a picture for evidence of lines, where this evidence will be a blackened region whose geometry resembles that of the prototype line in the center of Fig. 4. Denote what the observer considers as evidence for a line as  $e$ , the probability that this evidence corresponds to a true signal line as  $P(e/s)$ , and the probability that this evidence is derived from random fluctuations in the black noise background as  $P(e/n)$ . That is, there is a distribution of shapes that can be taken as evidence for presence of an embedded test line ranging from evidential shapes that in no way correspond to the test line,  $P(e/s) = 0$ , to those that can only correspond to a true line,  $P(e/s) = 1$ . The  $P(e/s) = 1$  situation can be considered absent in the dot pictures used here. So, a considered evidential shape evokes for the observer some impression,  $P(e/s)$ , or equivalently some likelihood ratio that a true line is being considered,  $1(e) = P(e/s)/P(e/n)$ , which he acts on. If  $1(e)$  is very high, the observer will certainly decide that  $e$  represents a line; and if  $1(e)$  is very low, he will reject the hypothesis that a line is seen. This is equivalent to postulating that the observer brings to the task a decision criterion,  $k$ , such that  $1(e) \geq k$  will cause him to accept the hypothesis that  $e$  is a line, and  $1(e) < k$  will cause him to reject this hypothesis. There is a considerable body of evidence that experimental subjects do, in fact, use such decision rules in a variety of situations (e.g., Green and Swets<sup>6</sup>). That is, the observer can be modeled as an intelligent statistical decision maker.

The decision criterion,  $k$ , is set by the observer, and reflects his tolerance for false alarms. For example, a low value of  $k$  leads to a high false-alarm rate (i.e., noise being called signal) since  $1(e)$  will frequently exceed the decision threshold; whereas a high value of  $k$  will lead to few false alarms, but will also lead to frequent false rejections (i.e., failure to consider a noisy signal as different from noise) since  $1(e)$  rarely exceeds  $k$ . Indeed, one of the reasons Rose introduced this

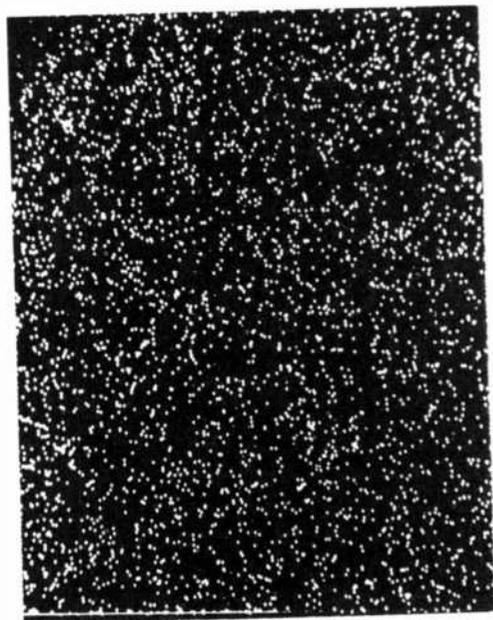
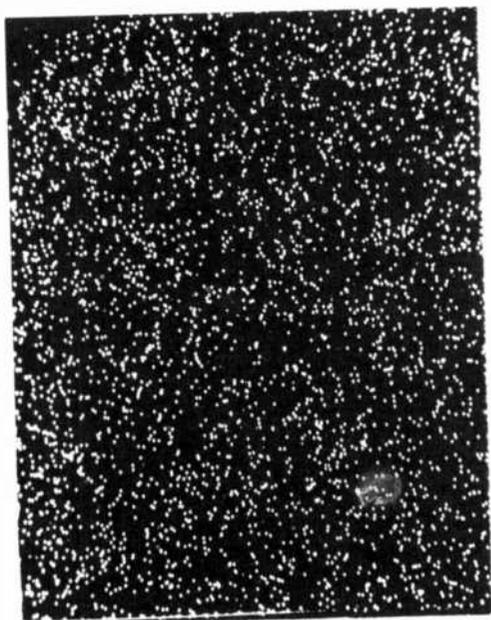


Fig. 6—Which picture contains a black triangle?

type of test pattern was to demonstrate some of these signal-detection aspects of vision, e.g., the occurrence of false alarms.

It is important to note that the observer is essentially free to set any value of  $k$  that he desires. By asking people to select the dot picture they were most confident in reporting as having at least two lines, rather than simply asking for a report of any occurrence of lines, it was hoped that a relatively high value of  $k$  would be induced. Thus the observer, intolerant of false alarms, scans the four dot pictures for lines and finds, among other things, what might be considered a triangle, or the letter V, in the upper left picture. Noise may seem to conspire to make itself look like a line, but a figure or a letter is unlikely to be a false alarm. Hence the meaningful line array may have more detectable line segments, since this property minimizes false rejections.

The point is, the individual line segments of the potentially meaningful pattern in the upper left dot picture are not treated with any more efficiency for detection than are the line segments hidden in the other dot pictures.  $P(e/s)$  or  $1(e)$  are not dramatically different for the various black lines nestled among the dots in these pictures. However, the possibility that a candidate line segment will be deemed a false alarm can depend on whether or not it is part of a meaningful pattern.

The same point can be made about the object-superiority effect of Weisstein and Harris. All line segments may be processed equivalently in their brief exposures, regardless of arrangement. However, once a processed array of line segments reaches the appropriate cortical center a decision must be made on this degraded stimulus, according to tolerances for false alarms and false rejections, as to what was presented. It is in this interpretation stage of perception that the source of the object-superiority effect may lie.

The decision making that an observer engages in to arrive at his percept can allow elements in the visual field to be included in the percept as signal, or be rejected as noise. The likelihood that the meaningfulness of the stimulus being processed affects these perceptual judgements is more convincingly demonstrated by the dot patterns in Fig. 6. For this figure we pose the following question: Which picture contains a black triangle? The most frequently selected dot picture is the one on the left, even though the correct answer is the picture on the right. These two pictures were produced in the same way as were those of Fig. 4. A different noise transparency than was used in the previous demonstration was overlaid with two of the line transparencies of Fig. 5. The triangle transparency (upper left in Fig. 5) was used to produce the picture on the right in Fig. 6, the line

transparency on the lower right in Fig. 5 was used to produce the pattern that is easily misperceived as a relatively large triangle in the picture on the left in Fig. 6.

Inspection of the left picture in this figure will reveal the line segments that were placed in this random dot array. Noticing these lines will make it obvious that perceiving a triangle here requires an observer to make a number of false alarms in connecting the triangle, and to make a number of false rejections in disregarding the line portions that extend beyond the vertices of the imaginary figure. The fact that one can either perceive a triangle here, or perceive the line segments that were actually put in this picture, emphasizes the significance of the decision-making aspect of perception. The fact that a triangle is spontaneously perceived, even after being told a triangle really is not being presented, suggests that meaningfulness strongly affects this decision making process.

Tangential to this discussion, but still worth noting, is another effect that can be obtained with the patterns of Fig. 6. The reader is requested to observe the triangles in these dot figures and then, while looking away from the page, rotate it by  $90^\circ$ . Now, look for the triangles. What typically happens is that the dot pictures at first look free of the patterns that were previously quite apparent. After quite a few seconds the triangles pop into view. This slowing down of the time to perceive an image is similar to what occurs when viewing a random-dot stereogram (see, e.g., Julesz<sup>7</sup>). In this type of demonstration, a random-dot array is presented to each eye, each array looking very much like our dot pictures with no intentionally embedded lines. A subset of the dots in each array is correlated at some non-zero binocular disparity, so that under stereoscopic viewing a figure in depth is seen. The time required to perceive a figure is often remarkably long, and when it does emerge it doesn't come in bits and pieces but as a unit. Since this perceptual delay effect has previously only been obtained with random-dot stereograms, it has been thought of as being attributable to the operation of binocular disparity detectors in the visual cortex and could simply be an artifact of transient, incorrect binocular eye coordination for detecting the appropriate disparity in the stereogram. However, this certainly cannot be the major source of the delay phenomenon, since being told what the figure is can dramatically reduce the delay, even though this clue gives no information about binocular disparity. The fact that we can induce a delay effect without involving stereopsis suggests that this sort of phenomenon is associated with more fundamental aspects of visual processing.

### Concluding Note

Visual perception involves passive transduction of information and active decisions on the resulting neural message. The focus here has been the use of variations on Rose's dot pictures to expose the judgmental phase of visual processing. Our results for the substantive issue addressed, i.e., whether the fact that a set of distinct elements in the visual field may be perceptually coalesced to form an object affects the processing of these elements, are preliminary. Nevertheless, the methodological value of using such dot patterns to examine the properties of perceptual decisions seems well demonstrated.

### References:

- <sup>1</sup> N. Weisstein and C. S. Harris, "Visual Detection of Line Segments: An Object Superiority Effect," *Science*, 186, p. 752 (1974).
- <sup>2</sup> D. H. Hubel and T. N. Wiesel, *J. Physiol. (London)*, 160, p. 106 (1962).
- <sup>3</sup> W. B. Pillsbury, "A Study in Apperception," *Amer. J. Psychol.*, 8, p. 315, (1897).
- <sup>4</sup> E. E. Smith and K. T. Spoehr, "The Perception of Printed English: A Theoretical Perspective," in *Human Information Processing: Tutorials in Performance and Cognition*, ed. by B. Kantowitz, Halsted Press, New York (1974).
- <sup>5</sup> A. Rose, *Vision, Human and Electronic*, Plenum Press, New York (1973).
- <sup>6</sup> D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, John Wiley and Sons, New York (1966).
- <sup>7</sup> B. Julesz, *Foundations of Cyclopean Perception*, University of Chicago Press, Chicago (1971).

# An X-Ray-Sensitive Fiber-Optic Intensifier Screen for Topography

R. W. Smith

RCA Laboratories, Princeton, N.J. 08540

**Abstract**—The time required to obtain an x-ray topograph can be reduced many orders of magnitude by replacing the photographic film detector with a television system. An efficient high-resolution x-ray-sensitive intensifier screen to couple directly to an ISIT camera is a crucial element necessary for this time reduction. This paper describes preliminary work on fabricating a high-resolution screen composed of an array of fiber optic elements made with an x-ray-sensitive phosphor, CsI:Na.

## Introduction

The objective of this work is to fabricate an x-ray-sensitive fiber optic transducer with sufficient response and resolution to be useful in an all-electronic high-speed x-ray topographic system. Most x-ray topographic systems in current use employ photographic film for the detecting medium. Photographic film with high resolution has inherently poor sensitivity to x-rays. It is possible to obtain several orders of magnitude improvement in speed by replacing the film with an image-intensifier TV system. A sensitive high-resolution x-ray-to-visible-light transducer has been the missing component for this system. This paper describes the preliminary results of work aimed at making a transducer with an array of x-ray-sensitive fiber-optic elements.

Meieran<sup>1</sup> and Reifsnider and Brown<sup>2</sup> describe video display of x-ray images and the latter paper also describes a fiber-optic x-ray in-

tensifying faceplate. The x-ray-sensitive element in Reifsnider and Brown's faceplate is glass. Since it is unlikely that any activated glass will soon equal the x-ray response of the good inorganic phosphors, our approach has been to make the center core element a solid optically clear inorganic, i.e., CsI:Na.

### Experimental Approach

It is feasible to consider making the central core element in a fiber-optic array out of something other than glass, since empty core glass matrices are commercially available. These matrices, called glass capillary array plates, can be made from various types of glass, with pore diameters from about 1  $\mu\text{m}$  up, several cm in diameter, and with arbitrary thickness. The working temperature of the glass (600–700°C) essentially determines the choice of the phosphor that can be fused into the capillaries. CsI:Na is widely used<sup>3</sup> in x-ray scintillation detectors, and its fluorescent emission spectrum is a good match for the response characteristic of our intensifier photocathode. It can be repeatedly melted and recrystallized without losing its efficiency. CsI melts at 621°C into a liquid that wets glass and flows into the glass capillaries. A preliminary experiment quickly established that a chip of CsI could be melted on a glass plate and formed into an array of x-ray-sensitive elements. Translating that observation into a uniform large-area faceplate has not been straightforward.

Since glass is very pliable at the melting point of CsI, there is a problem in supporting the plate during processing. Also, the way the molten CsI wets the glass, flows over the surface and into the capillaries, and is finally recrystallized, is extremely sensitive to temperature gradients. Furthermore, although it may be possible to minimize temperature gradients at any particular temperature, the final response of the recrystallized phosphor in a capillary is dependent on transient temperature gradients that arise throughout the cycle from initial melting to recrystallization.

During the early stages of this work the basic approach was to process the *whole* plate by baking it in a Pt dish charged with CsI. The CsI was melted, wet the glass, filled the capillaries, and was recrystallized, all by various schemes designed to control the temperature in the plate by the rate at which heat was supplied to the dish from an arrangement of heaters. This approach was relatively successful for plates 12 mm in diameter, with 25  $\mu\text{m}$  capillaries, and 0.5 mm thick. It is extremely difficult to process 25 mm diameter plates, 0.1 to 0.25 mm thick and with 10  $\mu\text{m}$  capillaries.

Recently we have completely changed our processing scheme. A

uniform layer of CsI:Na is initially evaporated on the glass plate. The CsI is then manipulated into the capillaries by mechanically scanning the plate past a highly localized heat source. The local processing is monitored with a microscope. With this scheme, it is much easier to control the over-all operation step-by-step and also to avoid distorting the plate. We are currently exploring two different methods for applying the localized heat. In the first method, the glass is heated with a CO<sub>2</sub> laser beam. In the second, the heat is applied to the CsI layer by means of a fine jet of hot gas.

## Results

Fig. 1 is the schematic of the x-ray image intensifier microscope used to evaluate the samples. A Radelin Super High Speed STF-2 intensifier screen can be inserted for internal comparison of x-ray response.

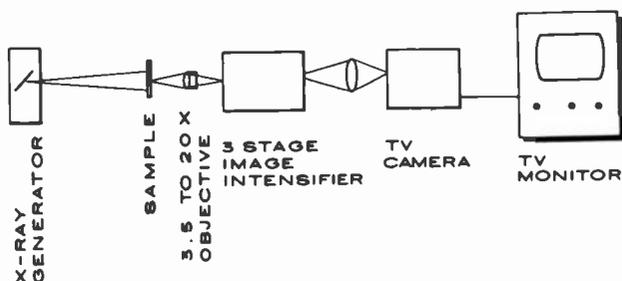
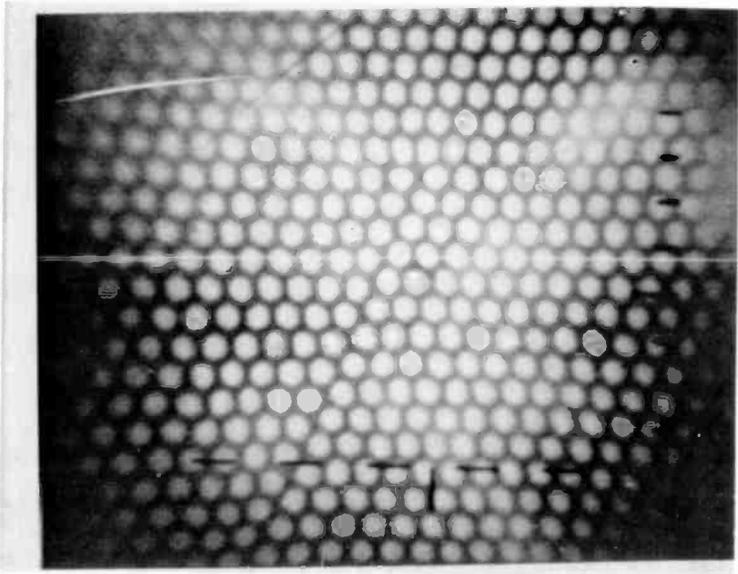


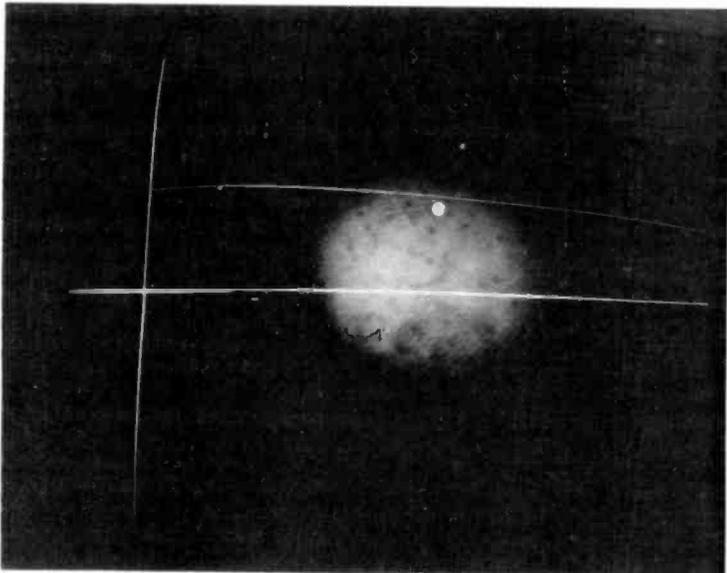
Fig. 1—Schematic for x-ray image-intensifier microscope used to evaluate CsI:Na transducers.

Fig. 2a is a picture taken from the monitor with a 25  $\mu\text{m}$  diameter capillary array and Fig. 2b is a similar picture for a 10  $\mu\text{m}$  diameter capillary array. The input microscope objective limits and distorts the field of view in both cases. The x-ray response from each plate is equal to or greater than the output from the Radelin standard, and the output from the CsI:Na is far from being optimized.

Fig. 3 is a picture from the TV monitor of the Laue x-ray transmission diffraction pattern for sapphire obtained with an x-ray generator. For this demonstration, the CsI:Na fiber optic transducer was directly coupled to the faceplate of an RCA C21159 I-SIT pickup tube. The output from the I-SIT camera goes directly to the TV monitor. We assume that, to a first approximation, the x-ray intensity in a Laue x-ray transmission diffraction spot is about the same intensity



(a)



(b)

Fig. 2—Photographs from TV monitor of x-ray image-intensifier microscope of (a) 25- $\mu\text{m}$  diameter capillary CsI:Na transducer, and (b) 10- $\mu\text{m}$  diameter capillary array transducer.

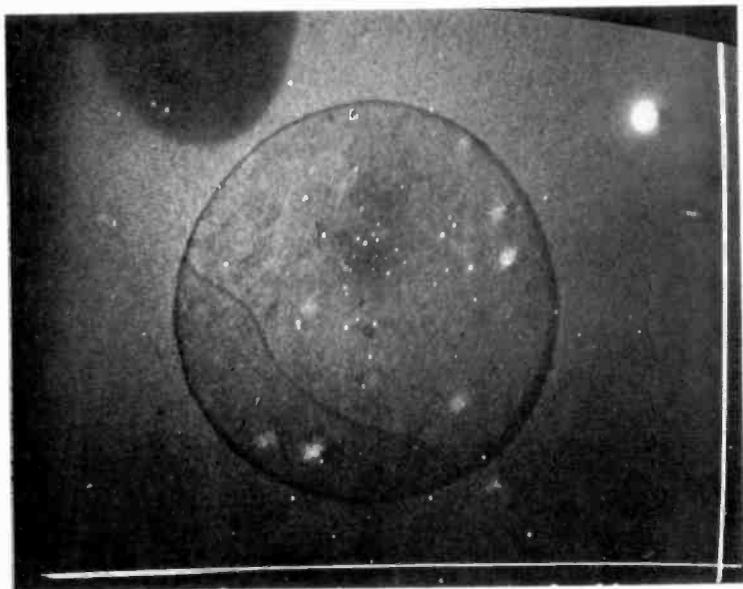


Fig. 3—Laue x-ray transmission diffraction pattern for sapphire. Circular outline in center of picture is the CsI:Na transducer. Main diffraction spots are along the inside edge of the transducer.

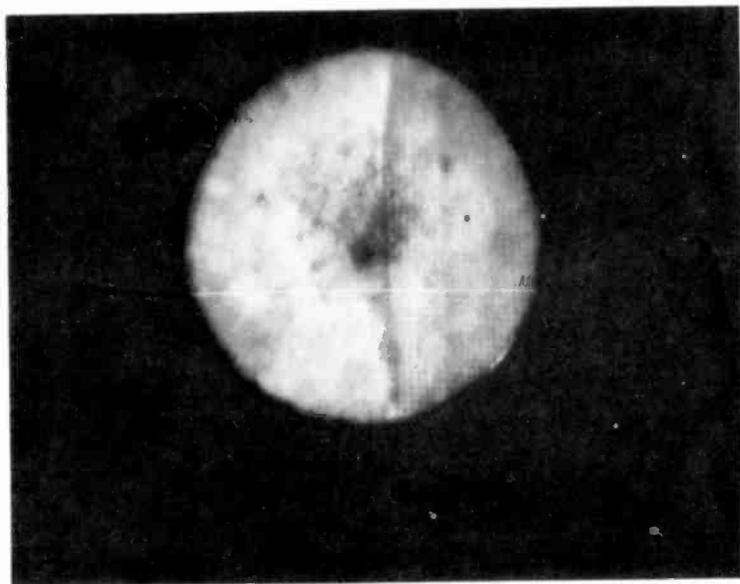


Fig. 4—X-ray shadowgraph of 100 X 100 mesh, 0.4 mm tungsten wire screen on one half of CsI:Na transducer.

that would be obtained in the case of a transmission topographic pattern. Since the Laue diffraction can be seen directly on the TV monitor and the TV frame time is several orders of magnitude faster than the minutes of time required to take the same diffraction pattern on film, this confirms the anticipated improvement in speed. There are a number of spurious spots in the diffraction picture due to nonuniformities in the fiber-optic structure and also to the crude coupling to the I-SIT faceplate in this demonstration.

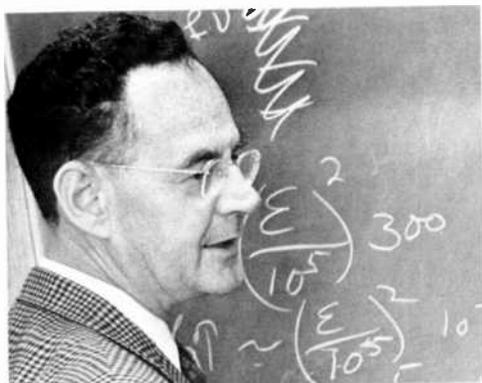
Fig. 4 shows a shadow-graph of a  $100 \times 100$  mesh, 0.4 mm tungsten wire screen. This is a rough indication of the resolution for this transducer which was made from a capillary array plate with  $25 \mu\text{m}$  pores.

### Acknowledgments

I would like to acknowledge the suggestions and encouragement from Arnold Moore during this work. A. D. Cope suggested the use of glass channel plate. I also acknowledge the technical assistance of Harvey Saltzman.

### References:

- <sup>1</sup> E. S. Meieran, "Video Display of X-Ray Images: I. X-Ray Topographs," *J. Electrochem. Soc.*, **118**, p. 619 (1971); "Part II. Laue Patterns," *J. Electrochem. Soc.*, **118**, p. 628 (1971).
- <sup>2</sup> K. L. Riefsnyder and J. J. Brown, Jr., "A High-Gain Light-Intensifying Device Incorporating a Fiber-Optic X-Ray Intensifying Faceplate," *J. Electrochem. Soc.*, **120**, p. 282 (1973).
- <sup>3</sup> J. Menefee, Y. Cho, and C. Swinehart, "Sodium Activated Cesium Iodide as a Gamma Ray and Charged Particle Detector," *IEEE Trans. Nucl. Sci.*, **NS-14**, p. 464 (1967).



**Albert Rose**, a Fellow of the Technical Staff of RCA Laboratories at the David Sarnoff Research Center, Princeton, N. J., received an A.B. degree in 1931 and a Ph.D. degree in Physics in 1935 from Cornell University. Dr. Rose joined RCA in 1935 in Harrison, N. J., as a member of the tube research group which was then engaged in an effort to improve the sensitivity of the Iconoscope, the pickup tube of early television systems. This research effort culminated in the initial development in 1937 of the Orthicon. When RCA and NBC introduced television as a new public service at the New York World's Fair in April 1939, cameras employing the new Orthicon tube were a vital part of the system.

Dr. Rose transferred to RCA Laboratories, Princeton, N. J., on its establishment in 1942, bringing with him an experimental model of a tube that later evolved into the image orthicon. At Princeton he was joined by two RCA scientists from Camden, Dr. P. K. Weimer and Dr. H. B. Law, who had developed what was essentially an electrostatic counterpart of Rose's tube. The three proceeded to combine their efforts to perfect a new pickup tube with increased sensitivity, greater contrast, improved picture quality, and decreased size and weight. For the first quarter century of television, this tube, the Image Orthicon, was adopted throughout the world as the major source of live television broadcasts.

In 1942 Dr. Rose also became interested in the relative limitations in the light sensitivity of photographic film, television pickup tubes, and the human eye, and developed a theory of visual sensitivity based on the absolute scale of the noise limitations due to the quantum nature of light, described in his book, *Vision: Human and Electronic*. His studies in the field of solids led to his widely recognized work on the theory and analysis of photoconductivity and related phenomena in insulators and semiconductors. Much of this work is contained in his monograph, *Concepts in Photoconductivity and Allied Problems*. In May 1955, Dr. Rose was appointed to head the research activity of the newly-established Laboratories RCA Ltd., Zurich, Switzerland. On his return to Princeton in 1957, he resumed research relating to the study and application of photoelectric phenomena, and in 1959 was named a Fellow of the Technical Staff.

In 1959 Dr. Rose appeared as a guest lecturer on the subject of electron optics on NBC's educational science program, Continental Classroom, and in 1961 he presented a series of lectures on photoconductivity at the International Summer Course in Solid-State Physics held at the University of Ghent, Belgium, under the auspices of NATO. He has lectured at summer schools at the University of Illinois (1953), California Institute of Technology (1961), University of Rochester (1963), the Massachusetts Institute of Technology (1963), Polytechnic Institute of Mexico City (1968), the Hebrew University in Jerusalem (1968), and at the University of Puerto Rico (1969). During the academic year 1961-62, he was a visiting lecturer at Princeton University. In 1967 he was the Mary Upson Distinguished Visiting Professor at Cornell University and in 1975 was appointed a Fairchild Distinguished Scholar at the California Institute of Technology.

Dr. Rose retired from RCA in July 1975, and is currently a Visiting Professor at the California Institute of Technology, Pasadena, Calif.

Dr. Rose has received the Television Broadcasters Association Award and Morris Liebmann Award of the Institute of Radio Engineers (now IEEE), the Journal Award of the Society of Motion Picture Engineers, and in March 1958 was one of the two recipients of the first David Sarnoff Outstanding Achievement Awards in Science and Engineering. In October 1958, he received the David Sarnoff Gold Metal Award of the Society of Motion Picture and Television Engineers for basic contributions to the development of the Orthicon, Image Orthicon, and Vidicon television pickup tubes.

Dr. Rose is a Fellow of the American Physical Society and the IEEE, a member of Phi Beta Kappa and the Societe Suisse de Physique, and in 1975 was elected to the National Academy of Engineering.

## Recent Papers by RCA Authors

Listing is alphabetical by name of primary author. For copies of reprints, the reader should contact the publication directly.

- R. C. Alig, "The Nephelauxetic Effect in Divalent Rare Earth Ions," *RCA Review*, Vol. 36, No. 1, p. 125, March 1975.
- R. C. Alig and A. T. Fiory, "A Paraelectric Defect in Sodalite," *J. Phys. Chem. Solids*, Vol. 36, p. 695, 1975.
- G. A. Alphonse, R. C. Alig, D. L. Staebler, and W. Phillips, "Time-Dependent Characteristics of Photo-Induced Space-Charge Field and Phase Holograms in Lithium Niobate and Other Photoreflexive Media," *RCA Review*, Vol. 36, No. 2, p. 218, June 1975.
- K. Ametani and T. Takahashi, "Composition of Single Crystal of Iron Sulfide Grown by Normal Freezing," *Bull. Chem. Soc. Japan*, Vol. 48, No. 2, p. 721, Feb. 1975.
- C. H. Anderson, P. Call, J. Stott, W. Hayes, "Correlation between Superhyperfine Constants and 4f Crystal Field Parameters of  $Tm^{2+}$  in  $CaF_2$  and  $BaF_2$ ," *Phys. Rev. B*, Vol. 11, No. 9, pp. 3305-3310, May 1975.
- D. Baeriswyl and W. Czaja, "Bose-Einstein Condensation in Exciton Systems and Liquid Helium," *RCA Review*, Vol. 36, No. 1, pp. 5-39, March 1975.
- E. Belohoubak and E. Denlinger, "Loss Considerations for Microstrip Resonators," *IEEE Trans. Microwave Theory and Technique*, p. 522, June 1975.
- K. R. Bube, "Laser-Induced Microcrack in Thick-Film Resistors—a Problem and Solutions," *Amer. Ceramic Soc.*, Vol. 54, No. 5, May 1975.
- J. K. Butler, H. Kressel, and I. Ladany, "Internal Optical Losses in Very Thin CW Heterojunction Laser Diodes," *J. Quantum Electronics*, Vol. QE-11, No. 7, July 1975.
- R. L. Camisa, B. F. Hitch, S. Yuan, and M. Ettenberg, "Microwave Integrated-Circuit MIS Varactor Phase Shifter," *RCA Review*, Vol. 36, No. 2, p. 296, June 1975.
- C. R. Carlson and E. L. Peskin, "One-Dimensional Particle Velocity Probability Densities Measured in Turbulent Gas-Particle Duct Flows," *Int'l. J. Multiphase Flow*, Vol. 2, p. 67, 1975.
- D. E. Carlson and C. E. Tracy, "Injection of Ions into Glass from a Glow Discharge," *J. Appl. Phys.*, Vol. 46, No. 4, p. 1575, April 1975.
- D. J. Channin, J. M. Hammer, and M. T. Duffy, "Scattering in ZnO-Sapphire-Optical Waveguides," *Appl. Optics*, Vol. 14, No. 4, p. 923, April 1975.
- D. J. Channin, "Triode Optical Gate: A New Liquid Crystal Electro-Optic Device," *Appl. Phys. Lett.*, Vol. 26, No. 11, p. 603, June 1, 1975.
- W. R. Curtice, J. J. Risko, and P. T. Ho, "A Study of Second-Harmonic-Extraction Trapatt Amplifiers for X-Band Operation," *RCA Review*, Vol. 36, No. 2, p. 274, June 1975.
- B. J. Curtis and H. R. Brunner, "The Growth of Thin Films of Lithium Niobate by Chemical Vapour Deposition," *Mat. Res. Bull.*, Vol. 10, p. 515, 1975.
- R. S. Crandall, "Hot Electrons," *Phys. Rev. B*, Vol. 12, No. 1, July 1, 1975.
- J. M. Cusack, S. M. Perlow, and B. S. Perlman, "Automatic Load Contour Mapping for Microwave Power Transistors," *IEEE Trans. Microwave Theory and Techniques*, Vol. MTT-22, No. 12.
- R. H. Dean, L. S. Napoli, and S. G. Liu, "Silicon Solar Cells for Highly Concentrated Sunlight," *RCA Review*, Vol. 36, No. 2, p. 324, June 1975.
- R. H. Dean and R. J. Matarese, "Submicrometer Self-Aligned Dual-Gate GaAs FET," *IEEE Trans. Electron Devices*, 1975.
- E. C. Douglas, "High Light Level Blooming in the Silicon Vidicon," *IEEE Trans. Electron Devices*, Vol. ED-22, No. 5, p. 224, May 1975.
- I. Drukler, R. L. Camisa, S. T. Jolly, H. C. Huang, and S. Y. Narayan, "Medium-Power GaAs Field-Effect Transistors," *Electronics Lett.*, Vol. 11, No. 5, March 6, 1975.
- R. E. Enstrom and D. G. Fisher, "The Effect of Lattice Parameter Mismatch in NEA GaAs Photocathodes Grown on GaP/ $In_xGa_{1-x}P$  Substrates," *J. Appl. Phys.*, Vol. 46, No. 5, p. 1976, May 1975.
- M. Ettenberg and H. F. Lockwood, "A Comparison of  $In_xGa_{1-x}As:Ge$  and  $GaAs:Si$  Prepared by Liquid Phase Epitaxy," *Inst. Phys. Conf. Ser.*, No. 24, Chapter 3, 1975.
- M. Ettenberg and C. J. Nuese, "Reduced Degradation in  $In_xGa_{1-x}As$  Electroluminescent Diodes," *J. Appl. Phys.*, Vol. 46, No. 5, p. 2137, May 1975.
- W. H. Fonger and C. W. Struck, "Temperature Dependences of  $Cr^{+3}$  Radiative and Nonradiative Transitions in Ruby and Emerald," *Phys. Rev. B*, Vol. 11, No. 9, p. 3251, May 1975.

- B. W. Faughnan, R. S. Crandall, and Philip M. Heyman, "Electrochromism in  $WO_3$  Amorphous Films," *RCA Review*, Vol. 36, No. 1, p. 177, March 1975.
- B. Goldstein and D. J. Szostak, "Preferential Evaporation of In from  $Ga_xIn_{1-x}As$ ," *Appl. Phys. Lett.*, Vol. 26, No. 12, June 15, 1975.
- A. M. Goodman, K. W. Hang, and J. W. Breece, "The Relative Dielectric Constant of Two Semiconductor Passivation Glasses," *J. Electrochem. Soc.*, Vol. 122, No. 6, p. 823, June 1975.
- J. J. Hanak, "Electroluminescence in  $ZnS:Mn_x:Cu$ , PF-Sputtered Films," *Proc. 6th Internl. Vacuum Congr. 1974, Japan. J. Appl. Phys. Suppl. 2, Part 1*, p. 809, 1974.
- J. M. Hammer, "Electro-Optic Switching and Modulation in Bulk and Waveguide Configurations," *Proc. Soc. Photo-Optical Instru. Eng.*, Vol. 53, p. 60, 1975.
- J. M. Hammer, D. J. Channin, M. T. Duffy, and C. C. Neil, "High Speed Electrooptic Waveguide Grating Modulator Using Epitaxial  $ZnO$ ," *IEEE J. Quantum Electronics*, Vol. 11, No. 4, p. 138, April 1975.
- G. Harbeke and E. Tosatti, "Optical Properties of Layer Structure Compounds," *RCA Review*, Vol. 36, No. 1, p. 40, March 1975.
- K. G. Herrqvist and R. Carbonetta, "Lasers for Information Scan and Control," *Proc. 18th Tech. Meeting*, Vol. 53, Laser Recording and Information Handling, San Diego, Ca., Aug. 21-22, 1974.
- H. Fujita, K. Ametani, and Michiko Inoue, "Optical Properties of Polystyrene Suspension in Water (Latex)," *RCA Review*, Vol. 36, No. 1, p. 108, March 1975.
- S. T. Hsu and J. H. Scott, Jr., "Mobility of Current Carriers in Silicon-on-Sapphire (SOS) Films," *RCA Review*, Vol. 36, No. 2, p. 240, June 1975.
- M. Inoue, "Lattice Green's Function for the Body Centered Cubic Lattice," *J. Math. Phys.*, Vol. 16, No. 4, p. 809, April 1975.
- W. Kern, "Chemical Vapor Deposition Techniques for Glass Passivation of Semiconductor Devices," *NAECON '75 Record*, p. 93.
- H. Kressel and I. Ladany, "Reliability Aspects and Facet Damage in High-Power Emission from (Al-Ga)As CW Laser Diodes at Room Temperature," *RCA Review*, Vol. 36, No. 2, p. 231, June 1975.
- I. Ladany and H. Kressel, "(AlGa)As Double Heterojunction CW Lasers: The Effect of Device Fabrication Parameters on Reliability," *Inst. Phys. Conf. Ser. No. 24*, Chapter 4, 1975.
- H. W. Lehmann and R. Widmer, "Preferred Orientation in RF Sputtered  $ZnO$  Films," *Japan. J. Appl. Phys.*, Supplement 2, Part 1, p. 741, 1974.
- R. W. Klopfenstein and C. W. Wu, "Computer Solution of One-Dimensional Poisson's Equation," *IEEE Trans. Electron Devices*, Vol. Ed.-22, No. 6, p. 329, June 1975.
- A. R. Moore, "Short-Circuit Capacitance of Illuminated Solar," *Appl. Phys. Lett.*, Vol. 27, No. 1, p. 26, July 1975.
- L. H. Meray, "Evaporator Facility for Deposition of Multielement Thin-Film Patterns," *J. Vacuum Science and Tech.*, Vol. 12, No. 3, p. 709, May/June 1975.
- D. Meyerhofer, "Field Induced Distortions of a Liquid Crystal with Various Surface Alignments," *Phys. Lett.*, Volume 51A, No. 7, p. 407, April 21, 1975.
- R. Mezrich, K. F. Etzold, D. Vilkomerson, "Ultrasonovision: An Interferometric Means for Measurement and Visualization of Ultrasonic Wavefront Interactions with Biologic Tissue," *Proc. Soc. Photo-Optical Instru. Engrs.*, Vol. 47, Meeting Aug. 1-2, 1974.
- A. Miller and M. T. Duffy, "Orientation of Crystals by Extinction Between Crossed Linear Polarizers," *Appl. Optics*, Vol. 14, No. 4, p. 809, April 1975.
- C. J. Nuese, G. H. Olsen, "Room Temperature Heterojunction Laser Diodes of  $In_xGa_{1-x}As/In_yGa_{1-y}P$  with Emission Wavelength between 0.9 and 1.15  $\mu m$ ," *Appl. Phys. Lett.*, Vol. 26, No. 9, p. 528, May 1, 1975.
- R. M. Moore, J. T. Fischer, and F. Koziol, Jr., "Vacuum-Deposited CdSe Films Grown Under Excess Se Flux," *Thin Solid Films*, Vol. 26, p. 363, 1975.
- G. H. Olsen, M. S. Abrahams, C. J. Buicocchi, and T. J. Zamerowski, "Reduction of Dislocation Densities in Heteroepitaxial III-V VPE Semiconductors," *J. Appl. Phys.*, Vol. 46, No. 4, p. 1643, April 1975.
- J. I. Pankove, "Blue Anti-Stokes Electroluminescence in GaN," *Phys. Rev. Lett.*, Vol. 34, No. 13, p. 809, March 31, 1975.
- J. I. Pankove, S. Bloom, and G. Harbeke, "Optical Properties of GaN," *RCA Review*, Vol. 36, No. 1, p. 163, March 1975.
- J. I. Pankove and E. R. Levin, "Scanning Electron Microscopy Studies of GaN," *J. Appl. Phys.*, Vol. 46, No. 4, p. 1647, April 1975.
- W. Phillips and J. M. Hammer, "Formation of Lithium Niobate-Tantalate Waveguides," *J. Electronic Materials*, Vol. 4, No. 3, p. 549, 1975.
- B. Robinson and C. Lakhani, "Dynamic Price Models for New-Product Planning," *Management Science*, Vol. 21, No. 10, p. 1113, June 1975.
- J. R. Sandercock, "Some Recent Developments in Brillouin Scattering," *RCA Review*, Vol. 36, No. 1, p. 89, March 1975.

- P. Scheng, E. B. Priestley, and P. J. Wojtowicz, "Comment on "Elastic Continuum Theory Cutoffs and Order in Nematics and Solids,"" *J. Chem. Phys.*, Vol. 63, No. 2, p. 1041, July 15, 1975.
- H. S. Sommers, Jr., and J. K. Butler, "Coupled Lateral Modes in Narrow Stripe Injection Lasers," *J. Appl. Phys.*, Vol. 46, No. 5, p. 2319, May 1975.
- H. S. Sommers, Jr., "Internal Dynamics of Narrow Striped Lasers," *J. Appl. Phys.*, Vol. 46, No. 4, p. 1844, April 1975.
- F. Sterzer, "The Conversion Efficiency of Ideal Shockley P-N Junction Photovoltaic Converters in Concentrated Sunlight," *RCA Review*, Vol. 36, No. 2, p. 316, June 1975.
- L. C. Upadhyayula, S. Y. Narayan, and E. C. Douglas, "Fabrication of 3-Terminal Transferred-Electron Logic Devices by Proton Bombardment for Device Isolation," *Electronics Lett.*, Vol. 11, No. 10, May 15, 1975.
- L. J. Vieland, A. W. Wicklund, and J. G. White, "Structure and Properties of  $Nb_3SnH_x$ ," *Phys. Rev. B*, Vol. 11, No. 9, p. 3311, May 1975.
- H. A. Weakliem, W. J. Burke, and D. Reifeld, "Optical Properties of  $SrTiO_3$  and  $LiNbO_3$ ," *RCA Review*, Vol. 36, No. 1, p. 149, March 1975.
- R. K. Wehner and E. F. Steigmeier, "Coupled Lattice Modes in Light Scattering," *RCA Review*, Vol. 36, No. 1, p. 70, March 1975.
- R. Williams, "Interfacial Free Energies between Polymers and Aqueous Electrolyte Solutions," *J. Phys. Chem.*, Vol. 79, No. 13, p. 1274, 1975.
- P. J. Wojtowicz, "Two Dimensional Phase Transitions of Mobile Ions at  $SiO_2$ -Si Interfaces," *RCA Review*, Vol. 36, p. 132, March 1975.
- C. R. Wronski, B. Abeles, A. Rose, "Granular Metal-Semiconductor Vidicon," *Appl. Phys. Lett.*, Vol. 27, No. 2, p. 91, July 15, 1975.
- C. P. Wu, E. C. Douglas, and C. W. Mueller, "Limitations of the CV Technique for Ion-Implanted Profiles," *IEEE Trans. Electron Devices*, Vol. ED-22, No. 6, p. 319, June 1975.

## Patents Issued to RCA Inventors Second Quarter 1975

April

- A. A. Ahmed Stabilization of Quiescent Collector Potential of Current-Mode Biased Transistors (3,878,471)
- A. A. Ahmed Biasing Circuit for Differential Amplifier (3,876,955)
- A. F. Arnold Method of Electroless Plating (3,877,981)
- A. F. Arnold and A. Z. Miller Method of Bonding Metals Together (3,875,652)
- A. D. Arsem Range Tracking Circuit (3,879,730)
- A. R. Balaban Digital Synchronizing System (3,878,335)
- A. R. Balaban Digital Synchronizing System (3,878,336)
- S. Berkman and P. M. Britt Method of Growing Single Crystals of Compounds (3,877,883)
- C. D. Boltz, Jr. FM Demodulator (3,878,470)
- R. J. Bosselaers Sawtooth Waveform Generator (3,879,683)
- D. E. Carlson Corona Discharge Method of Depleting Mobile Ions from a Glass (3,879,183)
- R. H. Dean High Frequency Device Assembly (3,878,436)
- W. F. W. Dietz Boost Regulator with High Voltage Protection (3,881,135)
- G. B. Dodson, 3rd Liquid Level Limit Indicator (3,878,541)
- E. C. Douglas, C. P. Wu, and C. W. Mueller Method of Making a Semiconductor Device (3,880,676)
- G. J. Dusheck, Jr. Energy Peak/Time Averaging Seismic Intrusion Detector (3,879,720)
- N. Feldstein and H. B. Law Method of Depositing a Pattern of Metal Plated Areas on an Insulating Substrate (3,878,007)
- N. Feldstein Master Matrix for Making Multiple Copies (3,878,061)
- N. Feldstein Method for Making a Photomask (3,877,810)
- W. V. Fitzgerald, Jr. Voltage Supply System (3,878,326)
- R. H. Godfrey Apertured-Mask Cathode-Ray Tube Having Half-Tone Array of Heat-Absorbing Areas on Target Surface (3,878,427)
- B. Golant and N. R. Landry Distributed Transmission Line Filter (3,879,690)
- P. E. Haferl Raster Centering Circuit (3,881,134)
- F. P. Helman and P. H. Robinson Method of Treating Semiconductor Devices to Improve Lifetime (RE28,386)
- H. Khajezadeh Semiconductor Temperature Sensor (3,881,181)
- H. W. Kuzminski and F. M. Sohn Cathode Ray Tube Having Shadow Mask and Screen with Tailored Heat Transfer Properties (3,878,428)
- E. Lachocki DC to Polyphase Inverter Utilizing a Plurality of Switching Device and a Transfer Having a Plurality of Primary and Feedback Windings Connected in Circuit with the Switching Device (3,879,650)
- N. F. Maxemchuk and H. E. White Data Processor for a Loop Data Communications System (3,879,710)
- A. Mayer Method of Treating Semiconductor Devices (RF28,385)
- E. P. McGrogan, Jr. Delta Modulator Utilizing a Duty Cycle Circuit (3,879,663)
- P. Nyul Modularized Laser Diode Assembly (3,878,556)
- M. A. Pollnsky Method of Achieving Semiconductor Substrates having Similar Surface Resistivity (3,876,472)
- G. D. Pyles Tape Cartridge Changer with Selective Cartridge and Track Programming Controls (3,879,758)
- W. R. Roach Deformable Mirror Light Valve and Method of Making the Same (3,877,791)
- O. H. Schade, Jr. Protective Diode Network for MOS Devices (3,879,640)
- J. Stark, Jr. Television Degaussing System with Saddle-Type Coils Adjacent CRT Cone (3,879,633)

- J. A. Van Raalte and V. Christlano** Predetermined Thickness Profiles Through Electroplating (3,880,725)  
**P. K. Weimer** Signal Processing Circuits for Charge-Transfer Image-Sensing Arrays (3,876,952)  
**J. A. Weisbecker** Read/Write Circuits for Reliable Digital Recording (3,881,183)  
**H. E. White and N. F. Maxemchuk** Data Loop Communication System (3,879,582)

## May

- A. D. Arsem** Amplifier Blanking Circuit (3,882,407)  
**T. V. Bolger** Doppler Correlation Radar Exhibiting Reduced Time Side Lobes (3,882,493)  
**T. V. Bolger** Doppler Correlation Radar Providing Combined Target Detection (3,882,494)  
**T. V. Bolger** Doppler Correlation Radar Providing Coarse-Range Detection Resolution (3,882,495)  
**A. Brown, Jr., and K. M. Schlesler** Absolute Valve Circuit Employing Opposite Conductivity Type (3,882,327)  
**J. F. Corboy, G. W. Cullen, and H. Pastal** Dual Growth Rate Method of Depositing Epitaxial Crystalline Layers (3,885,061)  
**G. W. Cullen, S. R. Bolin, J. F. Corboy, J. E. Creamer, and A. J. Wasielewski** Modified Czochralski-Grown Magnesium Aluminate Spinel and Method of Making Same (3,883,313)  
**A. G. F. Dingwall** Self-Biased Complementary Transistor Amplifier (3,886,464)  
**W. W. Evans and D. E. Christensen** On-Off System for Television Receivers (3,886,307)  
**R. E. Fernsler** Fail-Safe High Voltage Protection Circuit (3,885,201)  
**C. J. Hall** Dynamic Convergence Circuit (3,882,350)  
**W. J. Hannan and J. R. Frattarola** Process of Producing Double-Sided Holographic Replicas (3,882,207)  
**D. S. Jacobson** Semiconductor Package Having Means to Tune Out Output Capacitance (3,886,505)  
**J. D. Knox** Optical Beam Scanning System (3,882,273)  
**H. Kressel and H. F. Lockwood** Efficiency Light Emitting Diode (3,883,888)  
**M. A. Leedom** Rotary Stylus Cleaner (3,881,734)  
**M. A. Leedom** Video Playback System Tracking Arm and Pickup Assembly (3,882,267)  
**S. A. Lipps** Image Display Employing Filter Coated Phosphor Particles (3,886,394)  
**S. Liu** Magnetically Tunable Ferrite Stripline Trappat Mode Oscillator and Amplifier Circuits (3,882,420)  
**J. D. Mazgy and H. M. Kleinman** Cross Over Detector and Rectifier (3,882,328)  
**R. W. Nosker and L. P. Fox** Trapezoidal Smooth Grooves for Video Disc (3,882,214)  
**C. B. Oakley and H. G. Schwarz** Control Apparatus for a Two-Way Cable Television System (3,886,454)  
**D. H. Pritchard and A. C. Schroeder** High Voltage Regulator (3,885,198)  
**M. R. Royce** Europium-Activated Alkaline-Earth Pyrophosphate Phosphors (3,882,041)  
**H. G. Seer, Jr.** Short Circuit Protection Apparatus for a Regulated Power Supply (3,886,410)  
**A. H. Sommer** Method of Making a Multialkali Electron Emissive Layer (3,884,539)  
**H. Sorkin** Fabrication of Liquid Crystal Devices (3,885,860)  
**S. A. Steckler** VBE Voltage Voltage Source Temperature Compensation Network (3,886,435)  
**G. A. Swartz** Varacter Tuned Impatt Diode Microwave Oscillator (3,882,419)  
**D. L. Thoman** Hermetic Seal and Method (3,884,007)  
**C. R. Thompson and R. J. Flint** Redundant Signal Processing Error Reduction Technique (3,883,891)  
**J. K. Unger** Method for Conditioning Transmission Lines Utilizing Adjustable Equalizers and a Recording Technique (3,883,703)

- H. A. Wittlinger** Temperature Sensing Circuit (3,882,728)
- R. E. Wilson** Pulse Position Discriminator (3,886,460)
- C. F. Wheatley, Jr.** Bias Circuitry for Stacked Transistor Power Amplifier Stages (3,886,466)

**June**

- R. S. Baker** Photographic Method for Printing Viewing-Screen Structure Including Treatment of Exposed Coating with Ammonium Compound (3,887,371)
- W. Bohringer** Start-Up Control Circuit for SCR Deflection (3,891,892)
- A. G. F. Dingwall** Method of Making a Compact Guard-Banded MOS Integrated Circuit Device Using Framelike Diffusion-Masking Structures (3,888,706)
- M. Ettenberg** Method for Depositing on a Substrate a Plurality of Epitaxial Layers in Succession (3,890,194)
- R. B. Goyer** Correlator and Control System for Vehicular Collision Avoidance (3,887,916)
- E. H. Griffin** Analog Echo Protection Circuit for DME (3,889,260)
- W. E. Ham and D. W. Flatley** Stabilized Semiconductor Devices and Method of Making Same (3,890,632)
- J. J. Hamak and P. D. Southgate** Energizing Technique for Electroluminescent Devices (3,889,151)
- W. A. Harmening** Equipoise Mechanism (3,889,551)
- H. Huang** Impatt Diode (3,890,630)
- W. F. Kosonocky** Charge-Coupled Circuits (3,890,633)
- I. Ladany and V. M. Cannuli** Deposition of Epitaxial Layer from the Liquid Phase (3,891,478)
- A. J. Leidich** Bias Circuitry for Stacked Transistor Power Amplifier Stages (3,887,880)
- A. L. Limberg** Transistor Biasing Arrangement (3,891,935)
- J. S. Radovsky** Current Mirror (3,887,879)
- W. L. Ross** Raster-Scan Display System Having Improved Means for Reading Out Stored Game-Score Information (3,889,253)
- O. H. Schade, Jr.** Transistor Series Amplifier (3,887,878)
- J. Stark, Jr.** High Voltage Protection Circuit (3,891,891)
- L. A. Torrington and F. R. Stave** Apparatus for Inhibiting a Plurality of Records from Being Disposed on a Turntable (3,888,493)
- J. P. White and P. J. Delpriore** Method of Shaping Semiconductor Workpiece (3,888,053)
- W. C. Wilkinson, Jr.** Scanning Antenna (3,887,924)
- H. A. Wittlinger** Ground Fault Detection (3,891,895)
- L. Woontner** Cathode Ray Tube Having a Luminescent Screen Including a Two Component White-Emitting Phosphoric Mixture (3,891,886)

## AUTHORS

**Benjamin Abeles** received the M.Sc. degree in physics at the Charles University in Prague (1945-49) and was instructor of physics at that university in 1949. He received the Ph.D. degree in Physics in 1956 from the Hebrew University in Jerusalem. From 1949 to 1951 he was employed as instructor in physics and as meteorologist by the Israeli Meteorological Service. From 1951 to 1952 he was a member of the Geophysical Group of the Weizmann Institute of Science, Rehovoth, Israel, engaged in research on seismic detectors. From 1952 to 1956 he was a member of the Solid-State Physics group of the Weizmann Institute working on galvanomagnetic effects in semimetals and semiconductors. In 1956 he became a Member of the Technical Staff of the RCA Laboratories. From 1969 to 1972 he served as Head of the General Research Group and in 1972 he was named Fellow of the Technical Staff.



In the area of solid-state physics his research interests have included galvanomagnetic effects in semimetals and semiconductors, thermoelectric devices, thermal conductivity, the use of microwave phonons to study superconducting tunneling, the generation of phonons by electromagnetic waves, the enhancement of the transition temperature and fluctuation phenomena in superconductors, transport, optical and magnetic properties of granular metals, and the physics of sputtering. Dr. Abeles received the David Sarnoff Outstanding Achievement Award in Science, as well as three RCA Achievement Awards for outstanding research. He is a fellow of the American Physical Society, a member of Sigma Xi, and was elected member of the executive committee of the Division of Solid State Physics (1971-73).

**Bruno K. Binggell** graduated in electrical engineering at the Kantonales Technikum Burgdorf, Switzerland, in 1948. During 1948-1950 he was engaged in color television research at the A.B. Dumont Laboratories, Passaic N.J., and from 1950 to 1954 in television receiver design at Aldepa Co., Zürich. During 1954-1955 he designed electronic equipment for air traffic control at Radio Schweiz A.G. Bern. In 1955, he joined the RCA Industry Service Laboratory at New York City, followed by a transfer to the then newly established Laboratories RCA Ltd., Zurich, Switzerland.



**Richard H. Bube** received the Sc.B. degree from Brown University in 1946 and the Ph.D. degree in Physics from Princeton University in 1950. From 1948 to 1962 he was associated with the Physical and Chemical Research Laboratory of the David Sarnoff Research Center as a Senior Staff member involved in research on luminescence and photoconductivity of solids. In 1962 he joined the faculty of Stanford University, where since 1964, he has been Professor of Materials Science and Electrical Engineering, and as of September 1975 is Chairman of the Department of Materials Science and Engineering. Dr. Bube is a Fellow of the American Physical Society.



ety, the American Association for the Advancement of Science, and the American Scientific Affiliation, and a member of Sigma Xi, and the American Society for Engineering Education. Dr. Bube is Associate Editor of Annual Review of Materials Science, a member of the editorial advisory board of Solid-State Electronics, and the editor of the Journal of the American Scientific Affiliation.

**James E. Carnes** received the B.S. degree from Pennsylvania State University, University Park, in 1961, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, N.J., in 1967 and 1970, respectively. His Ph.D. dissertation was an investigation on photoinduced currents and charge transport in polyvinylcarbazole, an organic polymer. He was in the U.S. Navy from 1961 to 1965. During the summers of 1966 and 1967 he investigated metallic contacts and dc electroluminescence in strontium titanate at RCA Laboratories, David Sarnoff Research Center, Princeton, N.J., which he joined as a member of the technical staff in 1969. Since that time he has studied electrical breakdown, conduction, and interface properties of various insulating films on silicon and is currently involved in the investigation of charge-coupled devices. Dr. Carnes is a member of the American Physical Society, Tau Beta Pi, Phi Kappa Phi, and IEEE.



**David Eger** obtained his B.Sc. degree (1970) and M.Sc. degree (1972) from The Hebrew University of Jerusalem, Israel. He is at present studying for his Ph.D. degree at the same university under supervision of Professors A. Many and Y. Goldstein. His work is concerned mainly with the electric properties of the ZnO surface.



**Richard S. Crandall** received a B.S. degree in physics from Cornell University in 1960, and M.S. in physics in 1962, and a Ph.D. in physics from the University of Illinois. He joined the RCA Laboratories in 1964 and has worked in the fields of electron and phonon transport in insulators and semiconductors, electron surface states on liquids, and is now engaged in studies of electrochromism. Dr. Crandall received an RCA Achievement Award for work on the electron-phonon interaction and spent the year from 1971 to 1972 at the Zurich Laboratories of RCA. He is a member of the American Physical Society and the IEEE.



**Yehuda Goldstein** received the M.Sc. in 1959 and the Ph.D. in 1962 from the Hebrew University of Jerusalem. He was a member of the Technical Staff of RCA Laboratories, Princeton, N.J., from 1962 to 1965. In 1965 he became a member of the Academic Staff of the Hebrew University and in 1971 an Associate Professor. His present research activities include electron-phonon interactions in superconductors and normal metal at low temperatures, granular metal films, and semiconductor surfaces. In addition, he was a Visiting Lecturer at Bar-Ilan University, Israel (1965-1966) and at Tel-Aviv University (1966-1967). His sabbatical during the year 1969-1970 was spent at RCA Laboratories, Princeton, N.J.



**Joseph J. Mezrich** received the B.S. in Electrical Engineering from the Polytechnic Institute of Brooklyn in 1967. The M.S. and E.E. degrees were received from the Massachusetts Institute of Technology in Electrical Engineering in 1969 and 1970, respectively. He received the M.A. in Statistics in 1974 and the Ph.D. in Psychology in 1975 from the University of Michigan. Since joining RCA Laboratories in 1974, Dr. Mezrich has worked in visual perception and in choice behavior. Society memberships include the Association for Research in Vision and Ophthalmology, The American Association for the Advancement of Science, Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and Sigma Pi Sigma.



**Alvin Malcolm Goodman** received a B.S.E. in 1952 from the Drexel Institute of Technology, Philadelphia, Pa. He did graduate work at Princeton University, receiving an M.A. in 1955 and a Ph.D. in 1958. From June 1956 to January 1957 he served as Research Assistant at Princeton University, and then as Assistant Professor of Electrical Engineering at the Case Institute of Technology up to June 1959. He performed research at RCA Laboratories as a summer employee in 1954, 1955, and 1958. He has been a Member of the Technical Staff since June 1959. During the year 1970-71, Dr. Goodman engaged in postdoctoral studies at the Swiss Federal Institute of Technology. Dr. Goodman has specialized in solid-state physics. His thesis subject was "Dember Effect and Trap Levels in Silver Chloride," and he has worked extensively in the areas of photoconductivity, metal-semiconductor contacts, metal-insulator contacts, insulator properties, and tunnel diodes. He has received two RCA Laboratories Achievement Awards for work on metal-semiconductor contacts (1963) and MNOS (metal-nitride-oxide-silicon) memory devices (1969). Since January 1973, Dr. Goodman has served as a Member of the Editorial Board of the Review of Scientific Instruments. He is a member of the American Physical Society, Sigma Xi and the Institute of Electrical and Electronics Engineers, and is listed in American Men of Science.



**Helmut Kiess** attended the University of Stuttgart from 1953 to 1957 and the University of Darmstadt from 1958 to 1960. He received the doctor's degree from the University of Karlsruhe in 1963. Dr. Kiess joined the Laboratories RCA Ltd., Zürich, Switzerland in 1963 and since then has worked on the physics of photoconductors and insulators. During the period from July 1970 to April 1971 he was at RCA Laboratories in Princeton.



**Walter F. Kosonocky** received the B.S. and M.S. degrees in electrical engineering from Newark College of Engineering, Newark, N.J., in 1955 and 1957, respectively, and the Sc.D. degree in engineering from Columbia University, New York, N.Y., in 1965. Since June 1955 he has been employed at RCA Laboratories, Princeton, N.J., where he has conducted research on application of new phenomena and new devices for information processing systems. This work has included ferrite memory systems, parametric digital devices, tunnel-diode circuits, tunnel-diode and transistor circuits, pattern-recognition systems, applications of lasers for digital sys-



tems (including a study of saturable absorbers for Q-switched lasers and semiconductor laser digital devices), optical hologram memory systems and page composition, optically controlled *p*-MOS circuits, and a liquid-crystal image converter. Since 1970 he has been working on the development and applications of charge-coupled devices. Dr. Kosonocky is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and a senior member of IEEE.

**Murray A. Lampert** writes "I grew up in Brooklyn, New York, during its glory period, when Ebbe's Field housed the immortal Babe Herman and the quite mortal, but speedy, Buddy Young. I majored in esoterica at Harvard until Pearl Harbor brought me into electronics. I taught the latter to assorted and sundry armed forces personnel, all of whom ultimately learned more about the subject than I knew. After brief diversions into optical design work and 'high-energy' (100 Mev) physics, I returned to electronics in 1949 at ITT, joining an R and D group in microwave and gaseous electronics. In 1952 I began a lengthy career at RCA Laboratories, largely in the areas of semiconductor and insulator physics. My work in these areas was strongly influenced by Al Rose, a stout-hearted, strong-minded chap who never let a mathematical formalism stand between him and comprehension. In 1966, in a strangely unrealistic effort to flee reality, I made a miscalculation of epic proportions and settled into a Grove of Academe. Three coronaries and one lengthy heart operation later I retired from that bizarre arena. At which point the thought occurred to me that possibly we dwell in a hazardous, uncertain world. And so currently, to explore further the mysteries of this familiar, unknown world I am studying biology and psychology—areas richly dense with unanswered and unanswerable questions. I also do a modest amount of consulting on the more comfortable terrain of solid-state physics."



**Abraham Many** received the M.Sc. in 1945 and the Ph.D. in 1950 from the Hebrew University of Jerusalem. In 1951 he became a Member of the academic staff of the Hebrew University. He later became Head of the Semiconductor Division, the Racah Institute of Physics, and in 1966 a Professor of Physics. His present research activities include work on semiconductor surfaces, electron-phonon interactions, and electronic processes in molecular crystals. Dr. Many has, during sabbatical leaves, been a research associate at General Telephone and Electronics (1957-1958) and at RCA Laboratories (1962-1963). He was a visiting professor at The Technion, Israel Institute of Technology, Haifa (1966-1973); at Tel-Aviv University (1969); at Chalmers University, Goteburg, Sweden (summer 1969); and at the University of Wisconsin, Milwaukee (summers of 1971-1975). He is a member of the Editorial Board of several journals and of the Commission on Semiconductors of the Internal Union of Pure and Applied Physics.



**Arnold R. Moore** received his B.S. in Chemistry from the Polytechnic Institute of New York in 1942, with thesis work in chemical kinetics. Subsequently, he worked on gas discharge and photoelectric surface problems with the manufacturing divisions of RCA in Harrison, N.J., and Lancaster, Pa. In 1945 he returned to graduate school at Cornell University, where he received the Ph.D. degree in Experimental Physics, minor in Theoretical Physics and Mathematics, in 1949 under National Research Council Fellowship. His work at Cornell consisted chiefly of experimental research into the motion of electrons in alkali and silver halides, along with the growth



of the required single crystals. He joined RCA Laboratories in 1949, working in the fields of transistor physics, optical absorption in semiconductors and semiconductor alloys, magnetic susceptibility, acoustoelectric effect in cadmium sulfide and gallium arsenide, and most recently photovoltaic solar converters. In 1970-71 he was a visiting professor at Brown University where he taught a graduate course in physics of solid-state devices and participated in research on photovoltaic cells. Since 1971 he has been Head of a group at RCA Laboratories working on insulators and photo devices. He has received three RCA Achievement Awards. Dr. Moore is a Fellow of the American Physical Society and a member of Sigma Xi.

**Roland W. Smith** received the B.S. degree from Western Kentucky State University, Bowling Green, Kentucky, in 1939 and the M.S. degree from Northwestern University, Evanston, Illinois, in 1942. From 1942 to 1947, he was a Research Associate at Northwestern. He joined RCA Laboratories in 1947 as a member of the Technical Staff. Since that time he has worked on research connected with insulators, photoconductors, and the acoustoelectric effect. He has received several RCA Achievement Awards and is a member of the American Physical Society and Sigma Xi.



**Fritz Stöckmann** received his degree Dr. rer.nat. from the University in Göttingen (Germany) in 1942. Until 1952 he worked as an assistant of R. W. Pohl in Göttingen. From 1952 to 1959 he was chief assistant at the Technische Hochschule Darmstadt, and has been a full professor at the University of Karlsruhe since 1959. As a Fulbright fellow he spent 6 months with K. Lark-Herovitz at Purdue University in 1954/1955. He has been a consultant for RCA Laboratories, Zürich, since early 1955. His research has been mainly concerned with the mechanisms of photoconductivity. Dr. Stöckmann is a member of the German and the European Physical Societies.



**Richard Williams** received the A.B. in chemistry from Miami University and his Ph.D. in physical chemistry from Harvard University. He joined RCA Laboratories, Princeton, N.J., in 1958. He has worked in numerous areas, including semiconductor-electrolyte interfaces, liquid crystals, internal photoemission, and insulator physics. He was group leader in Insulator Physics from 1967 to 1970 and in Quantum Electronics from 1970 to 1972; in 1972, he was made a Fellow. He has received several achievement awards and shared in the David Sarnoff Team Award in Science in 1969. He was a visiting scientist at the RCA Zurich laboratory in 1963 and has been a Fulbright Lecturer in Sao Carlos, Brazil, a summer school lecturer at Instituto Politécnico Nacional in Mexico, and a visiting lecturer at Princeton University.



**Paul K. Weimer** received his A.B. degree from Manchester College in 1936, his M.A. degree in physics from the University of Kansas in 1938, and his Ph.D. degree in physics from Ohio State University in 1942. From 1937 to 1939, he taught physics and mathematics at Tabor College, Hillsboro, Kansas. Since 1942, Dr. Weimer has been engaged in research at RCA Laboratories, Princeton, N.J., where he is a Fellow and member of the technical staff. At the Laboratories he has participated in the basic development of various types of television tubes and solid-state devices. From 1959 to 1960, Dr. Weimer was granted an RCA fellowship for study abroad, which was spent at the Laboratoire de Physique, Ecole Normale Supérieure, Paris, France, working in the field of semiconductors. He was recipient of a television Broadcasters' Award in 1946, the IRE Vladimir K. Zworykin Television Prize in 1959, and the 1966 IEEE Morris N. Liebmann Prize Award. He received the RCA David Sarnoff Outstanding Achievement Award in Science in 1963. In 1968, Dr. Weimer was awarded an honorary degree of Doctor of Science by Manchester College. He is a Fellow of the Institute of Electrical Engineers and a member of the American Physical Society.



**Christopher R. Wronski** obtained his B.Sc. in 1960 and Ph.D. in 1963 in Physics at Imperial College, London University. His thesis was concerned with electron diffraction and study of the melting phenomena in thin films. From 1963 to 1966 he was with 3M Research Laboratories in St. Paul, Minnesota, working in the field of photoconductivity in thin films. Since joining RCA Laboratories in Princeton, N.J., in 1966, he has been engaged in work on television pick-up tubes and cold cathodes; photoconductivity of insulators; metal insulator, and granular metal semiconductor barriers and heterojunctions. Dr. Wronski is a member of the American Physical Society.







The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that every entry, no matter how small, should be recorded to ensure the integrity of the financial statements. This includes not only sales and purchases but also expenses and income. The text suggests that a systematic approach to record-keeping is essential for identifying trends and making informed decisions.

In the second section, the author addresses the challenges of managing cash flow. It is noted that many businesses struggle with timing their payments and receipts. The text provides practical advice on how to negotiate better terms with suppliers and customers to improve liquidity. It also discusses the importance of having a reserve fund to cover unexpected expenses or downturns in business.

The third part of the document focuses on the role of technology in modern accounting. It highlights how software solutions can streamline the recording and analysis of financial data. The text mentions various types of accounting software and their benefits, such as automated calculations and real-time reporting. It also touches upon the security of digital records and the importance of regular backups.

Finally, the document concludes with a section on the ethical responsibilities of accountants. It stresses that accuracy and honesty are paramount in the profession. The text discusses the potential consequences of unethical behavior, such as fraud or misstatement, and provides guidance on how to handle difficult situations. It encourages accountants to act in the best interests of their clients and the public.