DEVELOPMENT OF A SPECTROMETER FOR THE ELECTROREFLECTANCE SPECTROSCOPY OF METALS

by

James Daniel Rancourt

A Dissertation Submitted to the Faculty of the

COMMITTEE ON OPTICAL SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 7 4

THE UNIVERSITY OF ARIZONA

GRADUATE COLLEGE

I hereby recommend that this dissertation prepared under my direction by JAMES DANIEL RANCOURT,

entitled DEVELOPMENT OF A SPECTROMETER FOR THE ELECTROREFLECTANCE SPECTROSCOPY OF METALS

be accepted as fulfilling the dissertation requirement of the degree of DOCTOR OF PHILOSOPHY

B.O. Seraphin — Dissertation Director

4-18-74 — Date

After inspection of the final copy of the dissertation, the following members of the Final Examination Committee concur in its approval and recommend its acceptance:*

Murray Sargent III	April 18, 1974
Roland V. Shack	18 April 1974
Robert R. Shannon	18 April 74
Angelo M. Sturmolis	18 April 1974

*This approval and acceptance is contingent on the candidate's adequate performance and defense of this dissertation at the final oral examination. The inclusion of this sheet bound into the library copy of the dissertation is evidence of satisfactory performance at the final examination.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: James Daniel Roncourt

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under grant number GH-34576 and monitored by Dr. R. Silberglitt.

I wish to thank my advisor, Professor Bernhard O. Seraphin for suggesting this problem and for his encouragement and guidance through the project.

It is impossible to acknowledge here individually all of the many people who contributed ideas and suggestions during the course of this work, and to them all I say sincerely, "Thank you." In particular, I want to mention Mr. Michael Kottke who shared the lab with me, Mr. David Rock, who made the thermoreflectance measurements on the bismuth sample, and Mr. L. N. Nelson, who made the reflectance measurements on the sample films.

Thanks are extended to Dr. Arlon Hunt for many hours of fruitful discussion and for his many constructive comments during the writing of this dissertation.

I wish to extend my thanks to Visiting Professors J. L. Deiss, A. Donnadieu, and R. Garrigos for providing stimulating discussions on all aspects of this work.

A special note of thanks is due to Professor Robert Shannon for his support in the early stages of these studies.

Finally, I wish to thank my wife Jeannine for her patient understanding and support.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS--Continued

LIST OF ILLUSTRATIONS--Continued

LIST OF TABLES

ABSTRACT

This study involves two phases: (a) the theoretical study of the penetration of a static electric field into a metal film; and (b) the construction of a spatial modulation reflectometer.

An equation is derived for the interaction of a strong electric field with the free electrons in a metal film. The penetration depth reduces to the Thomas-Fermi screening length for films with a large number of free electrons. When the number of carriers in the film is near the number of fixed charges which cause the electric field, then it is shown that the field will penetrate into the film to a depth on the order of that probed by a light wave. The use of strong electric fields generated by the polarization of a ferroelectric ceramic substrate is considered for impressing an electric field on a metal film.

A reflectometer utilizing the principles of spatial modulation and all-reflective optics was built and tested. It measures the differential reflectance of the two halves of a sample film deposited on a ferroelectric substrate. By applying different static electric fields on each half of the film, it is possible to maintain equal free electron densities, but different field strengths, in the sample halves. This could lead to unambiguous detection of the band structure information without the detection of the large free electron effects which are normally superimposed on such spectra.

No structure was observed in the electroreflectance spectrum of bismuth. The cause of this was likely due to the shielding of the electric field by the film which poled the ceramic.

Suggestions for improvements to the apparatus are given, and directions for further work are proposed.

CHAPTER 1

INTRODUCTION

The spectroscopy of non-interacting atoms and molecules has provided experimental information of extraordinary wealth and quality. When the quantum theory of atoms and molecules was devised, it was applied to assess numerous families of spectral lines, each classified with respect to shift, splitting, and state of polarization in magnetic and electric fields. This theory related spectroscopy and the electronic structure with dramatic success by interpreting emission and absorption of light as transitions of an electron between energy states of the system of which they were a part.

Radiation incident on a solid induces similar transitions, and, in principle, proper interpretation of the optical spectra of solids should reveal characteristic parameters of their electronic structure. The interpretation of the static spectrum of solids is not so straightforward as that for atoms and molecules, however. Instead of being sharply localized on an energy scale as is the case for atoms and molecules, the initial and final states of an electronic transition in a solid are arranged in quasi-continuous energy bands extending throughout the three-dimensional Brillouin zone (BZ) in momentum space. Discontinuities in the energy profile of the density

of these states in certain parts of the zone introduce characteristic features into the optical spectrum. However, the contributions of these transitions to the shape of the spectrum must be added to those produced by all other energetically possible and allowed transitions in other parts of the zone. These latter transitions have little intrinsic information about the band structure since the initial and the final states of the transition could be buried in the middle of some allowed band; in a static measurement, there is no way of differentiating this pair of states from any other pair with a similar energy separation. The sum of all allowed transitions produces an optical spectrum which lacks the contrast seen in the lines of atomic and molecular spectroscopy.

The task, then, is to separate these contributions to the optical spectrum and to extract only those which supply data characteristic of the band structure of the material. This is accomplished by the methods of modulation spectroscopy which utilize synchronous detection techniques to mask out the unwanted transitions; only band-edge to band-edge transitions in certain parts of the BZ are thereby observed. A spectrum produced by this technique shows no response over large ranges of photon energy, and the structure that does appear has a spectral width much smaller than that which appears in the usual static reflectance spectrum. This is interpreted as meaning that the modulation which is applied to the sample affects only limited areas of the BZ. This structure can be associated with features of the BZ

known as critical points (CP), and this gives the information required by the theorist to perfect his theories and adjust his semi-empirical computer calculations.

A large amount of the credit for the rapid progress made in the study of the band structure of dielectrics and semiconductors in recent years is due to the analysis of the data obtained from modulation spectroscopy, and, in particular, from the specific technique of electroreflectance (Seraphin, 1972a, 1972b). Electroreflectance (ER) spectroscopy has established itself as a powerful diagnostic tool in the exploration of the band structure of these materials. Unfortunately, electroreflectance has not been unambiguously applicable to the study of the metallic band structure because the method requires the penetration of an electric field into the material. For metals and good conductors this electric field is screened out by free electrons within one atomic layer of the surface. This thin layer of material cannot be expected to be representative of the metallic structure; therefore, to a great extent the results represent surface interfacial effects combined with the free carrier effects.

Since ER has been instrumental in the rapid progress of the study of the band structure of dielectrics and semiconductors, and since the band structure of conductors should be qualitatively similar, a technique for the measurement of the ER spectra of metals should prove to be a very useful tool. In order for such a technique to succeed, it is necessary to devise a means whereby an electric field can

be made to penetrate significantly deeper into a metal; then the band structure information could be obtained from the ER spectrum.

This study has therefore focused on developing a method and the equipment with which to observe the ER spectrum of metals. In particular, an equation which describes the penetration of an electric field into a metal film has been derived and solved numerically. Examination of the result leads to a sample configuration in which the free electron effects are eliminated. The band structure of a layer of material which interacts with the probing light is then influenced by the perturbing electric field without the complicating effects of the free electrons. The overlap of the field and the light profiles leads to an observable change in the reflectance of the affected film as compared to a film with a negligible field. A spectrometer was built which compares the reflectance of these two film halves in the manner of spatial modulation. We thereby try to observe unambiguous interband transition spectra. Finally, we discuss the problems encountered and suggest directions for future efforts.

The importance of this work derives from the fact that very little data is available on the band structure of metals. Other types of studies, such as DeHaas-van Alphen techniques, probe the metallic band structure only in the vicinity of the Fermi level. Optical studies such as those discussed in this work offer an opportunity to probe to much deeper energies into the band structure. The superior

diagnostic capabilities of electroreflectance would thereby yield valuable information for the theorists and extend our knowledge of the band structure of metals and conductors.

CHAPTER 2

PRINCIPLES OF MODULATION SPECTROSCOPY

In this chapter, we briefly review the band theory of solids as it applies to modulation spectroscopy. The connection between the experimentally observed reflectance spectrum and variation of the dielectric constant (dielectric function) is examined. Reflectance and transmission spectroscopy are compared and it is found that modulated reflectance measurements are more easily obtained than modulated transmission ones. Modulation of the reflectance is associated with the differential of the dielectric function and the latter is discussed in terms of critical points. Finally, electroreflectance spectroscopy, a particular technique of modulation spectroscopy, is explained in terms of the band structure.

Insulators, Semiconductors, Semimetals, and Metals

When 10^{23} atoms are brought together into a cube 1 cm on a side, it is a well-known fact that due to the Pauli principle the energy levels of the atoms are going to change. In effect the levels divide into quasicontinuous bands instead of producing sharp lines. These bands consist of allowed and forbidden regions on an energy scale. At 0 K, the electrons of the atoms completely fill these levels up to an energy called the Fermi level. At higher temperatures, some of the

electrons may occupy higher levels and vacate some lower lying ones. Whether they do or not depends on there being enough energy available to make an otherwise allowed transition energetically possible.

The complete description of a band structure of a crystal requires a four dimensional space, where three of the axes are momentum coordinates and the fourth represents energy. A projection of the surfaces in this space onto the energy axis gives rise to the common band structure diagram shown schematically in Fig. 2.1. There is no abscissa in this diagram; the regions on the energy scale have been spread out horizontally for ease of visualization. This figure also shows the various locations at which a Fermi level could be found in the band structure of different materials. There is only one Fermi level for a given crystal, of course. In the case marked (a) the Fermi level falls in a forbidden gap where the gap width is around 5 electron volts (eV). Thermal energy added to the material cannot cause a transition across the gap since kT for room temperature is only of the order of 1/40 eV. The conduction of an electric current results from an excess of electrons with momentum in one direction over those in the opposite direction. In the case of level (a), all states in the lower band are filled so an imbalance in momentum cannot occur. The upper band has no electrons and cannot conduct without carriers. This material is therefore an insulator. A photon in the visible region of the spectrum has an energy on the order of 3 eV and it does not have enough energy to raise the electron to a higher band. Since

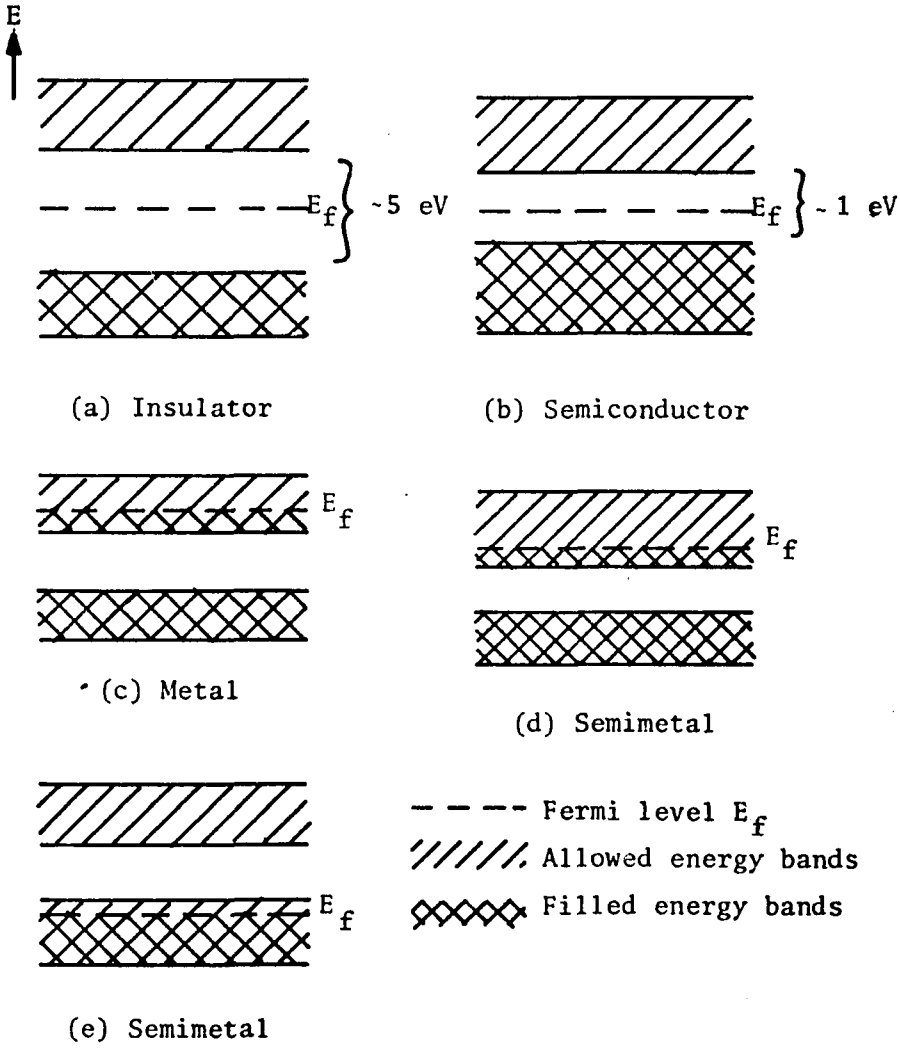


Fig. 2.1. Fermi Levels and the Band Structure at 0 K

the photon is not absorbed, the material is transparent in this spectral region. An example of such a material is diamond.

Next, we consider case (b) where once again the Fermi level is in a gap, but this time the gap is narrower, say of the order of 1 eV. Some electrons can be thermally excited into the upper level where they can change momentum states and give rise to an electric current. We can now state that the major difference between an insulator and a semiconductor is that the width of the forbidden gap is smaller for a semiconductor so that some electrons have enough thermal energy to cross it and occupy states in the conduction band.

Case (c) has the Fermi level in the middle of a band so that many electrons and many states are readily available and conduction can take place. In case (d) the Fermi level is within a band, but there are few electrons available to carry out the conduction process though many states are accessible. A similar but reversed situation is true with the level in (e) where there are many electrons available but few states are accessible. Materials with the Fermi level as in case (c) are considered to be metals while those with levels as in (d) and (e) are called semimetals.

Interband transitions occur when an electron acquires or loses enough energy to cross an energy gap so that it ends up in a band different from the one it started out in. This is generally the type of transition in which we are interested in modulated spectroscopy. Unfortunately for band structure studies, interband transitions are

swamped by intraband transitions which are much more favorable energetically. Furthermore, since we are interested in studying materials which have states available for conduction, there will be conduction if an electric field is applied with a potential source such as a battery. We will deal with this problem later.

Band Structure Analysis from Modulation Spectroscopy

Before we discuss modulation spectroscopy (MS), we must first take a look in more detail at the elementary optical absorption process. The electromagnetic field of the incident light wave stimulates the transition of an electron from an initial state with energy E_i and wave vector $\vec{k}$ into a final state of energy E_f and wave vector $\vec{k}'$. The interaction between the electrons and the electromagnetic field of frequency ω and polarization $\vec{e}$ is described as a time-dependent perturbation to the Hamiltonian of the system. The number of transitions between initial and final states per unit time and volume is given by

$$\frac{dN}{dt} \sim \int |\vec{e}\cdot\vec{M}_{if}|^2 \, \delta(E_f - E_i - \hbar\omega) \, d\vec{k}, \tag{2.1}$$

where M_{if} is the matrix element of the momentum operator. The dot product $\vec{e}\cdot\vec{M}_{if}$ is given with respect to the wave functions ψ of the initial and the final states:

$$\vec{e}\cdot\vec{M}_{if} = \int \psi_f^* (\vec{k}', \vec{r}) \vec{e}\cdot\nabla\psi_i (\vec{k}, \vec{r}) \, d\vec{r}.$$

The probability of finding the electron in the excited state is proportional to the square of the matrix element of the perturbation.

The delta function in Eq. (2.1) makes the absorption process a probe for energy differences in the band structure. It stipulates that no transitions can occur unless the photon energy matches the energy difference between the initial and the final states. We have made the implicit assumption that E_i and E_f are not changed by the transition. This is the so-called one-electron approximation and it holds well in covalently bound materials with high mobility of the carriers, and wave functions that spread through correspondingly large regions of the crystal.

The matrix element offers a similar diagnostic potential with respect to the momentum vector $\vec{k}$. Transitions between states in the BZ are classified into two groups. "Direct" transitions involve no change in momentum thereby connecting states located vertically above each other in the conventional $(E,\vec{k})$ diagram. Transitions classed as "indirect" involve a change in momentum; the initial and final states of such a process can lie anywhere in the BZ. The momentum of an optical photon ($k = 2\pi/\lambda \sim 4\pi \times 10^6 m^{-1}$) is small on the scale of the BZ ($k = 2\pi/\text{lattice constant} \sim 5\pi \times 10^9 m^{-1}$) and is usually ignored. Such a photon cannot induce an indirect transition without the emission or the absorption of a phonon. This type of transition is expected to have a much lower probability of occurrence than a direct transition. This probability is given by the matrix element in Eq. (2.1).

This concept relates simply the optical absorption process to the band structure in a straightforward manner. The analysis can rely on a simple counting throughout the BZ of states that meet the requirements of energy and momentum conservation. The density-of-states function obtained by this counting process for all of the energies of the spectral range then establishes a significant band structure parameter which correlates to the macroscopic quantity ε_2, the imaginary part of the complex dielectric constant $\tilde{\varepsilon}$. On the basis of the assumptions described above, an experimentally determined ε_2 spectrum can then be expected to provide a replica of the energy profile of the joint density-of-states function (JDS) in the band structure.

The correlation between ε_2 and band structure parameters is established by assuming that absorption is the sum of the transitions between electronic states of the same $\vec{k}$ vector and separated by the energy difference $E_f - E_i = \Delta E$:

$$\varepsilon_2(\omega) \sim \int_{BZ} \frac{2}{(2\pi)^3} |\vec{e}\cdot\vec{M}_{if}|^2 \delta(\Delta E - \hbar\omega) d^3k.$$

If we further assume that the matrix element is the same for all direct transitions, regardless of their location in the BZ, i.e., $|\vec{e}\cdot\vec{M}_{if}|^2$ is not a function of $\vec{k}$, we can define the joint density of states function $J(\Delta E)$ as

$$J(\Delta E) = \int_{BZ} \frac{2d^3k}{(2\pi)^3} \delta(\Delta E - h\omega). \qquad (2.2)$$

Since all pairs of states that meet the requirements of energy and momentum conservation are admitted to the integral with the same weight, the joint density of states function represents the simple counting of states in the BZ on which the analysis relies. The spectral profile of an experimentally determined ε_2 is expected to reflect the features of this sum and, therefore, the features of the energy structure. This is the working hypothesis of band analysis.

A transformation of the volume integral in Eq. (2.2) into a surface integral over the surface ΔE=constant shows the properties of the JDS function more clearly. This gives

$$J(\Delta E) = \frac{2}{(2\pi)^3} \oint_{\Delta E=\text{constant}} \frac{ds}{|\nabla_k(\Delta E)|}, \tag{2.3}$$

where ds is the surface element on the isoenergetic surface ΔE=constant and $\nabla_k(\Delta E)$ is the gradient of the separation of initial and final states with respect to $\vec{k}$. Energy conservation requires that $\hbar\omega$ be exactly the vertical separation of a pair of states. If the energy bands are nearly parallel at some given point, the energetic separation changes little over a finite $\vec{k}$ interval. States are spaced equidistantly on the $\vec{k}$ coordinate, and such an interval of near parallelism offers a large number of allowed transitions to photons of this energy range. Hence, the area makes a large contribution to the absorption, as described by the small value of the gradient $\nabla_k(\Delta E)$ in the denominator of the integrand in Eq. (2.3). If on the other hand, the bands spread apart,

few transitions are resonant with the light, and the corresponding absorption is small (large $\nabla_k(\Delta E)$).

Points for which $\nabla_k(\Delta E) = 0$ are particularly important and they are called "critical points" (CP). This condition can be satisfied in two different ways. First, the energy bands can be parallel but not horizontal. This can occur anywhere in the BZ and such CP's are therefore called "general" CP's. Second, both gradients can vanish separately so that their difference is also zero. This is a necessary condition at the center and certain other points of the BZ for reasons deriving from the symmetry of crystal structure. Furthermore, a BZ connects with other identical zones which surround it in momentum space. Since the energy bands must connect smoothly to the corresponding bands in the next zone, the slopes at the zone edges are zero. Such a CP is called a "symmetry" CP.

Critical points introduce slope discontinuities into the JDS function, and, subsequently, into ϵ_2. These slope discontinuities correlate to the band separation at points of high symmetry in the BZ, such as the center or the edges.

A classification of symmetry critical points into four groups can be made. A Taylor series expansion about some critical point located at k_0 in $(E,\vec{k})$ space gives (Seraphin, 1972b, p. 180; Wooten, 1972, p. 119)

$$E_f - E_i = \Delta E \simeq \Delta E_0 + \sum_{j=1}^{3} \alpha_j (k_j - k_o)^2,$$

where

$$\alpha_j = 1/2 \frac{\partial^2}{\partial k_j^2} (\Delta E(\vec{k})).$$

From the definition of a CP, the first derivative is identically zero in this expansion. Since we are considering only what happens in the immediate vicinity of a CP, we choose to ignore the third and higher order terms.

The band separation can either be at an extremum in which case all derivatives (α's) are positive (absolute minimum) or negative (absolute maximum). Since ΔE increases or decreases in all three k directions at such a point, this kind of CP is called "parabolic." When one or two of the derivatives are of one sign, the CP is called "hyperbolic." The critical points are normally labelled as M_j, where j = 0, 1, 2, or 3. The value of j is obtained by counting the number of negative derivatives in the expansion. Integration of the Taylor series expansion given above, gives the functional dependence of the JDS function for a small region about the CP. The expected slope discontinuities are of a square root nature and are plotted in Fig. 2.2 which is reproduced from Seraphin, 1972b, Fig. 4.3.

Optical Constants, Dielectric Function, and Reflectance

In this section we show how the reflectance is related to the band structure by the dielectric function.

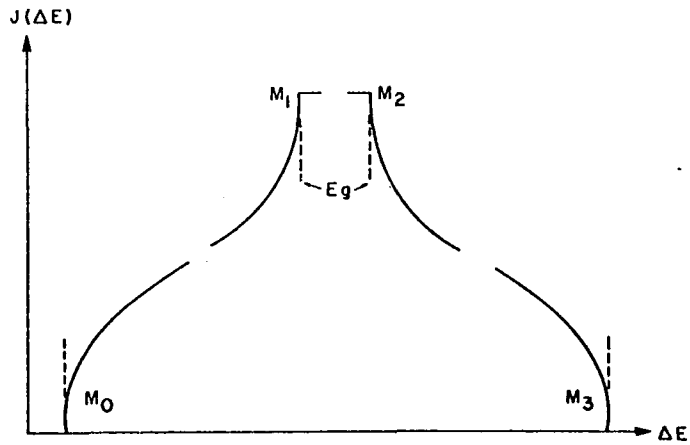


Fig. 2.2. Joint Density-of-States Function J(ΔE) Near Critical Points of M_0, M_1, M_2, and M_3 Types

On a macroscopic basis the effects due to optical transitions are summed up in a parameter called the dielectric function, $\tilde{\varepsilon}$. In general, this function is complex and it is related to the complex index of refraction $\tilde{N}$:

$$\tilde{\varepsilon} = \varepsilon_1 + i\varepsilon_2 = \tilde{N}^2 = (n + ik)^2$$

where ε_1 is the real part of $\tilde{\varepsilon}$, ε_2 is the imaginary part of $\tilde{\varepsilon}$, n is the index of refraction, and k is the extinction index. We immediately see that ε_1 and ε_2 are related to n and k by

$$\varepsilon_1 = n^2 - k^2 \text{ and } \varepsilon_2 = 2nk. \tag{2.4}$$

The reverse relation can be shown to be

$$n = \{1/2[(\varepsilon_1^2 + \varepsilon_2^2)^{1/2} + \varepsilon_1]\}^{1/2}$$

and

$$k = \{1/2[(\varepsilon_1^2 + \varepsilon_2^2)^{1/2} - \varepsilon_1]\}^{1/2}.$$

Absorption measurements are the most direct to determine ε_2. In fact, ε_2 is shown to be proportional to the absorption of energy from the light wave by differentiating the time averaged Poynting vector (Stone, 1963, p. 384):

$$\delta W = \varepsilon_2 \frac{\omega}{8\pi} |E_0|^2 \delta z$$

where δW is the energy lost by a plane wave in traversing a thin slab δz thick, E_0 is the electric field strength of the plane wave, and $\omega = 2\pi c/\lambda$. The spectrally dependent features of ε_2 thus reflect the features of the energy band structure in which we are interested.

These measurements are accurate in spectral regions in which enough light can penetrate a bulk sample. Unfortunately, these transparent regions are of little use to band structure analysis since the very fact that little absorption takes place implies that there are few interband transitions in this spectral region. In general, reflectance measurements are much more practical when we are studying spectral ranges where the photon energy is sufficient to lift electrons between band states.

The complete description of the reflectance process involves two functions as given by the complex reflectivity

$$r(\omega) = \rho(\omega)e^{i\phi(\omega)} = \frac{n'-n+ik}{n'+n-ik},$$

where the primed quantities refer to the incident medium and the unprimed quantities are for the absorbing medium, $\rho(\omega) = R(\omega)^{\frac{1}{2}}$, $R(\omega)$ is the reflectance, and $\phi(\omega)$ is the phase shift introduced in the light wave by the reflection. Reflectance depends on both parts of the dielectric function. This can be seen by substituting the dielectric function into the Fresnel reflection equation for an interface (normal incidence):

$$R = |r|^2 = \frac{(n'-n)^2 + k^2}{(n'+n)^2 + k^2}.$$

The result is

$$R = \frac{\varepsilon' - \{2\varepsilon'[\varepsilon_1 + (\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}}]\}^{\frac{1}{2}} + (\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}}}{\varepsilon' + \{2\varepsilon'[\varepsilon_1 + (\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}}]\}^{\frac{1}{2}} + (\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}}}$$

or,

$$R = \frac{(\varepsilon_1 - \varepsilon')^2 + \varepsilon_2^2}{[\varepsilon' + \{2\varepsilon'[\varepsilon_1 + (\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}}]\} + (\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}}]^2}$$

where $\varepsilon' = (n')^2$, and n' is the index of refraction of the nonabsorbing incident medium. A similar relation can be written for the phase shift:

$$\phi = \tan^{-1} \frac{2n'k}{n^2 - n'^2 + k^2} = \tan^{-1} \frac{\{2\varepsilon^1[(\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}} - \varepsilon_1]\}^{\frac{1}{2}}}{(\varepsilon_1^2 + \varepsilon_2^2)^{\frac{1}{2}} - \varepsilon'}.$$

These two results can be inverted to give the dielectric function in terms of the two observable quantities R and ϕ (Weiss, 1971, Table I):

$$\varepsilon_1 = \frac{(1 - R)^2 - 4R\sin^2\phi}{(1 + R - 2R^{\frac{1}{2}}\cos\phi)^2}, \qquad (n' = 1)$$

and

$$\varepsilon_2 = \frac{4R^{\frac{1}{2}}(1 - R)\sin\phi}{(1 + R - 2R^{\frac{1}{2}}\cos\phi)^2}.$$

The phase shift upon reflection ϕ is a rather difficult quantity to measure experimentally; fortunately, R and ϕ are not independent functions. As a consequence of the causality principle, they are related by Kramers-Kronig (KK) relations (Wooten, 1972, p. 182; Stern, 1963):

$$\ln \rho(\omega_0) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\phi(\omega)}{\omega - \omega_0} d\omega$$

and

$$\phi(\omega_0) = -\frac{2\omega_0}{\pi} P \int_0^\infty \frac{\ln \rho(\omega)}{\omega^2-\omega_0^2} \, d\omega, \tag{2.5}$$

where P indicates that the Cauchy principal value of the integral is used. A similar relation holds between the real and the imaginary parts of the dielectric function:

$$\varepsilon_1(\omega_0) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\omega\varepsilon_2(\omega)}{\omega^2-\omega_0^2} \, d\omega$$

and

$$\varepsilon_2(\omega_0) = \frac{2\omega_0}{\pi} P \int_0^\infty \frac{\varepsilon_1(\omega) - \varepsilon_0}{\omega^2-\omega_0^2} \, d\omega + \frac{\sigma_0}{\omega_0\varepsilon_0}$$

where σ_0 is the d.c. conductivity and ε_0 is the permittivity of free space, which has the MKS value of 8.85×10^{-12} $C^2/(N \cdot m^2)$. Kramers-Kronig analysis can be cast into a form of the ubiquitous Fourier analysis, as shown by Sceats and Morris (1972) and then later by Peterson and Knight (1973).

If the reflectance of a material can be measured over a sufficiently wide range of energy, ϕ does not have to be determined experimentally but can be evaluated by using Eq. (2.5). Since the integral has to be taken from zero to infinite energy, extrapolation of the reflectance curve is necessary. This is fraught with difficulties since oscillator models are postulated to account for the

behavior of the reflectance at the energy extremes. These are generally adjusted empirically until a reasonable result is obtained for the correlated function. This was the way most optical constants were obtained from optical measurements. From these, correlations with band models were attempted; the results of these studies were rather indefinite (e.g., Ehrenreich and Philipp, 1962). In the next section, we see how the methods of modulation spectroscopy enabled the observation of hidden or weak structure in the optical response. With this new information more precise correlation to the band structure was possible.

Modulation Spectroscopy

The basic objective of modulation spectroscopy as it applies to band theory consists of separating the critical point contribution to the optical spectra from that due to the non-critical point background, and subsequently identifying the position of the critical point in the BZ.

The separation can be demonstrated on the basis of the square root dependence of the absorptive part of the dielectric function near a CP of photon energy ΔE_0 (Wooten, 1972, sec. 5.3):

$$\varepsilon_2 \propto \vec{e}\cdot\vec{M}_{if}\,(E - \Delta E_0)^{\frac{1}{2}} + \text{constant} \tag{2.6}$$

where $E = \hbar\omega$ is the incident photon energy. Differentiation of this with respect to some parameter x of the electronic structure gives

$$\frac{d\varepsilon_2}{dx} \propto \frac{\vec{e}\cdot\vec{M}_{if}}{2}\,\frac{1}{(E - \Delta E_0)^{\frac{1}{2}}}\,\frac{d(E - \Delta E_0)}{dx} + \frac{d(\vec{e}\cdot\vec{M}_{if})}{dx}\,(E - \Delta E_0)^{\frac{1}{2}}. \tag{2.7}$$

Note that the constant in Eq. (2.6) that takes in all of the non-critical point effects disappears upon taking the derivative. Thus, MS eliminates the static background response and allows the measurement of the locations and the shapes of critical points on an energy scale. In an oversimplified manner this describes the advantage of modulated spectroscopy over static spectroscopy. The various terms in Eq. (2.7) define several classes of modulation spectroscopy.

a. Modulation of the incident photon energy (dE/dx) leads to wavelength modulation. This is an "external" modulation in that the parameters intrinsic to the material under study are unchanged, and the modulation acts externally to the sample. This technique enhances the resolution of a static reflectance measurement and allows the observation of structure that is too weak to be resolved in a standard spectral measurement. It adds no new information that was not already in the static spectrum. The interpretation of the resulting spectra is straightforward since it involves only the theory of the optical constants. Therefore, it is subject to the same problems encountered in analyzing the static spectrum as previously mentioned.

The interpretation of spectra resulting from internal modulation requires an additional knowledge beyond that needed to analyze static or externally modulated spectra. The applied modulating parameter that perturbs the optical constants makes the analysis more complicated, but conversely, the diagnostic capability of internal

modulation is much greater than that of external modulation. The following two classes are of the internal modulation type.

b. If the modulation affects either the position of the critical point ΔE_0 on the spectral scale or the separation of the bands, we have "gap" modulation ($d\Delta E_0/dx$). Typical examples of this effect are piezo- and thermoreflection.

c. A modulation mechanism could also alter the probability of a given transition $[d(\vec{e}\cdot\vec{M}_{if})/dx]$. Electroreflectance is an example of such a mechanism and is the one of interest here. It is discussed in a later section.

We now consider the effect of differentiation on the macroscopic observable, R, and see how it can be related to the band structure through ε_2.

When the Fresnel equation for reflectance at an interface is differentiated, the result has the functional form

$$\frac{\Delta R}{R} = \alpha(\varepsilon_1,\varepsilon_2)\Delta\varepsilon_1 + \beta(\varepsilon_1,\varepsilon_2)\Delta\varepsilon_2, \qquad (2.8)$$

where $\Delta\tilde{\varepsilon} = \Delta\varepsilon_1 + i\Delta\varepsilon_2$. The coefficients α and β, known as the Seraphin coefficients (Cardona, 1969), are functions of photon energy through their relation with ε_1 and ε_2. Their sign and relative magnitude determine the outcome of an analysis of the modulated spectrum in the different spectral regions. The line shape of a reflectance response can be discussed only after its composition is analyzed in terms of $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$. In a $\Delta\varepsilon_2$ dominated region, for which $|\beta| >> |\alpha|$, an absorptive line shape is preserved in the reflectance response. If, however, the KK relation must be applied to such an absorptive line

shape in a $\Delta\varepsilon_1$ dominated region ($|\alpha| >> |\beta|$), the reflectance response consists of an up-down sequence of two peaks of opposite sign. Fractional admixture will result in reflectance responses with line shapes anywhere between these two extremes. Similarly, the sign, magnitude, and spectral position of a response with respect to the correlated CP is strongly influenced by the size and sign of the two coefficients. From this we see that it is not sufficient to consider only one part of the complex differential dielectric function for the description of a modulated reflectance spectrum.

In order to separate out the two components $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$ from the $\Delta R/R$ spectrum [Eq. (2.8)], a second equation is necessary. In principle, the differential KK relation could be used:

$$\Delta\varepsilon_1(\omega_0) = \frac{2}{\pi}\int_0^\infty \frac{\omega\Delta\varepsilon_2(\omega)}{\omega^2 - \omega_0^2}\, d\omega.$$

Substitution of this into Eq. (2.8) gives

$$\frac{\Delta R(\omega_0)}{R} = \alpha(\varepsilon_1,\varepsilon_2)\frac{2}{\pi}\int_0^\infty \frac{\omega\Delta\varepsilon_2(\omega)}{\omega^2 - \omega_0^2}\, d\omega + \beta(\varepsilon_1,\varepsilon_2)\Delta\varepsilon_2(\omega_0).$$

The mathematical difficulties encountered in attempting to solve this for $\Delta\varepsilon_2$ in terms of $\Delta R/R$ are formidable, so an alternative approach is used.

The wavelength dependence of $\Delta R/R$ can be obtained from experiment. A second function can be found through the dispersion relation that connects the real and the imaginary parts of the complex reflection coefficient previously given as

$$r = R^{\frac{1}{2}} e^{i\phi}.$$

Differentiation of this and using the Fresnel relations gives (Seraphin, 1972a, p. 14)

$$\Delta\varepsilon_1 = \frac{r}{2}\frac{\Delta R}{R} - \delta\Delta\phi \tag{2.9a}$$

$$\Delta\varepsilon_2 = \frac{\delta}{2}\frac{\Delta R}{R} + r\Delta\phi, \tag{2.9b}$$

where

$$r = (n/n_0)(n^2 - 3k^2 - n_0^2)$$

and

$$\delta = (k/n_0)(3n^2 - k^2 - n_0^2).$$

If $\Delta R/R$ is known for a sufficiently large range of photon energies $\hbar\omega$, the KK dispersion relation gives for the differential $\Delta\phi$

$$\Delta\phi(\omega_0) = \frac{\omega_0}{\pi}\int_0^\infty \left[\frac{\Delta R}{R}(\omega_0) - \frac{\Delta R}{R}(\omega)\right]\frac{d\omega}{\omega_0^2 - \omega^2}, \tag{2.10}$$

thus allowing the calculation of $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$ through Eqs. (2.9a) and (2.9b).

Contributions to the integral given in Eq. (2.10) are restricted to spectral regions in which the modulation affects ε_2. This is of considerable practical value, since extrapolation outside the region of measurement is less critical. In most cases, the denominator in the integrand increases so rapidly that even the neighboring structure of a spin-orbit split transition can be ignored. Also, since modulation spectra are generally zero except for some relatively narrow structure, there is usually no problem encountered in the extrapolation of the integrand as is the case with static analysis as previously mentioned.

Electroreflectance

In electroreflection, the electric field not only produces interband effects that are common to all other types of internal modulation, but it also modifies the transition probability matrix element and the JDS function. The symmetry-breaking effect on the matrix element is found in all vectorial modulation techniques. In a cubic crystal, for example, the optical isotropy is destroyed under the action of the electric field or the uniaxial stress of piezomodulation, a preferred direction is established, and the tensorial character of the dielectric function becomes important.

Inserting the potential energy term (-Ex) into the one-electron Hamiltonian of the unperturbed crystal breaks up the invariance to lattice translation in the direction x of the electric field E. As a consequence, the component k_x of a wave vector is no longer a valid quantum number. Block functions are the eigenfunctions of the wave equation for a perfectly periodic potential. Using these functions we can build the field-perturbed wave functions from linear combinations of the unperturbed Bloch functions with wave vectors parallel to the field direction. Aspnes, Handler and Blossey (1968) have worked out the theory of this situation and show that for the small fields which are typically present in such an experiment the field-perturbed function can be obtained by convoluting the zero field function with an Airy function.

Aspnes (1973) gives a good phenomenological description of the physical model of ER and its effect on the optical transitions. Figure 2.3 is taken from this work and shows the change in the transitions which are allowed in the two cases. In the first case, the bands are approximately parabolas, and the transitions are strictly vertical since there cannot be a large momentum change. In such a transition we have what is known as a first derivative modulation process since shifts induced in the band spacing by the modulation process are small on the scale of the energy gap.

In the case where the modulation is produced by varying an electric field, the electron accelerates, and momentum is no longer a valid quantum number in the field direction. The one-electron Bloch functions become mixed and this is equivalent to spreading the formerly sharp vertical transitions over a finite range of initial and final energy and momentum states, as shown in the bottom half of the figure. This smears out the structure in the unperturbed dielectric function and gives a more complicated difference spectrum. This difference is approximated by higher order derivatives, and the third derivative is characteristic of these spectra. A rigorous theoretical derivation of this is also given in this paper by Aspnes.

The third-derivative characteristic of ER spectra provides optimum sensitivity for the detection of weak critical points. This is certainly true for the present study since the electric field is greatly attenuated by the electrons in the metal. Typically, it has

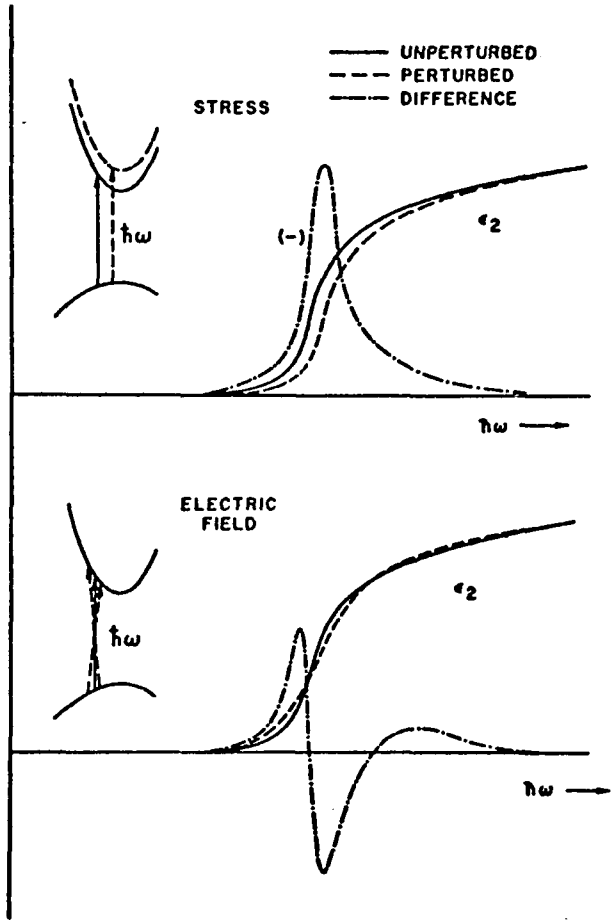


Fig. 2.3. First and Third Derivative Modulation

Top: a schematic diagram of the change in the imaginary part of the dielectric function expected for a first derivative modulation process where the lattice periodicity is preserved. Bottom: similar diagram for electric field modulation where lattice periodicity is not preserved. The effect of the perturbation on the energy band structure and the optical transition is shown at the left in each case (from Aspnes, 1973, Fig. 1.).

been possible to observe structure in ER spectra as small as a few parts per million in $\Delta R/R$.

From this type of analysis it is possible to draw some conclusions about the type of critical point one expects to see in an experiment. Unfortunately, line shape analysis still is not completely understood and is just on the verge of producing results. Insufficient knowledge of the effective field is probably the most important cause for the failure of line shape interpretation in providing one, and only one, answer. Unless the modulating voltage variation about the D.C. operating level is known, the relationship between electrical and optical modulation cannot be established. Aspnes and Frova (1969) have shown that the shape of the lines can take on rather different appearances when we apply an inhomogeneous field rather than a homogeneous one. Figure 2.4 is a figure reproduced from their paper. This graphically shows the error that could arise from using the uniform perturbation approximation indiscriminately. As we will see later in Chapter 4, the experimental configuration discussed in this paper creates a particularly simple field profile from which it should be possible to calculate the exact effect produced by the inhomogeneous field.

In this chapter we have tried to show, albeit in a qualitative manner, the process of modulation spectroscopy. We have touched upon possibilities of its potential as a probe of the band structure of solids. It is interesting to compare the capabilities of modulated

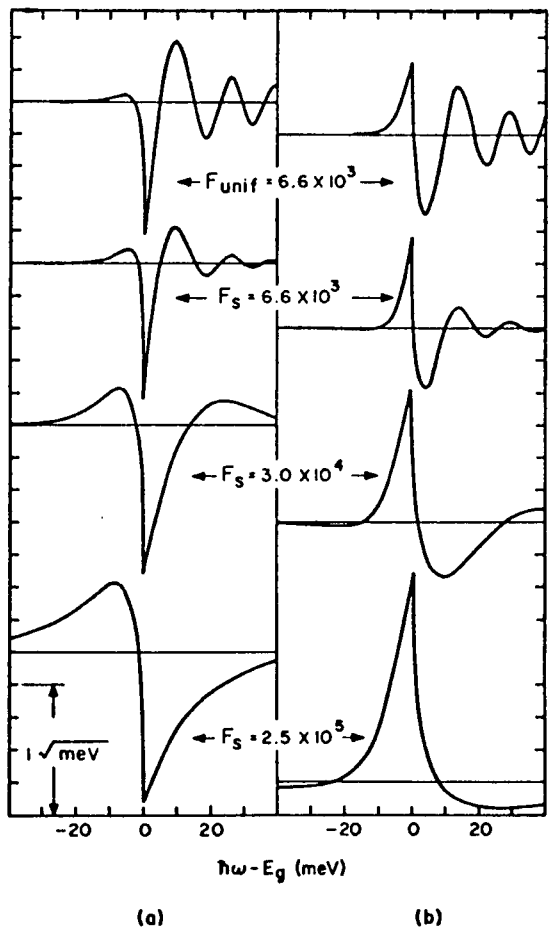


Fig. 2.4. Homogeneous and Inhomogeneous Modulation

Changes in the (a) real and (b) imaginary part of the dielectric function at the fundamental edge of intrinsic germanium, as induced by a homogeneous electric field (top row) and space-charge fields of a value F_s (from Aspnes and Frova, 1969).

spectroscopy with other ways of studying solids, in particular, metallic conductors. Most of these other methods, such as the deHaas-vanAlphen techniques, probe only the region around the Fermi surface. They cannot get down deeper into the band structure. Until the advent of modulated spectroscopy, theorists had only this data to use to check their calculations of the entire structure of a given material. Unfortunately, the Fermi surface is rather insensitive to the details of a given model of the band structure so that the agreement of a particular model with experimental Fermi surface data is not an exacting test of the model's accuracy.

Hopefully, the methods developed here will provide more data for a deeper understanding of these materials.

CHAPTER 3

ELECTROREFLECTANCE SPECTROSCOPY OF METALS

In this chapter we review the work done by others in their attempts to measure the electroreflectance of metals. The experimental and the analytical difficulties are discussed, and this gives us a vantage point from which to view the work reported in subsequent chapters.

Electroreflectance (ER) has established itself as a powerful diagnostic tool in the exploration of the band structure of dielectrics and semiconductors. Unfortunately, for all its potential, it has not been applied unambiguously to metals. The reason for this is that the method relies on the penetration of a perturbing electric field into the material. For metals and good conductors in general, this modulating field is screened out within the Thomas-Fermi screening length (Kittel, 1966, p. 235). On the other hand, the light wave might penetrate into the material to a depth an order of magnitude or more larger than the screening length; thus, most of the material interacting with the probing light wave is in a field-free region. In copper, for example, the field penetration depth of 0.5 Å should have little effect on the band structure of the material with which the light wave is interacting to a depth of about 100 Å (Born and Wolf, 1970). Thus, most reliable data we have to date on the band structure of metals come from studies of the Fermi surface.

Early Work

Electrolytic ER

Considering what has just been said, one would not expect to be able to observe any electroreflectance of a metal sample, especially one with a very small screening length. In 1966, Feinleib (1966) published results showing that indeed he did observe an effect with his apparatus. Briefly, his method consisted of a metal sample immersed in an aqueous potassium chloride electrolyte. The sample served as one electrode while a second electrode, made of platinum, was situated remotely. The modulating voltage was applied to the electrode and the light which was reflected from the sample-electrolyte interface was directed into a monochromator and analyzed using standard modulation spectroscopy techniques. The dipole layer that was formed at the interface gave rise to a large potential gradient at the sample surface.

In his paper Feinleib himself expresses some surprise at his results. Identical results were obtained by P. Grant (Seraphin, 1972a, p. 145) and by B. J. Parsons (Seraphin, 1972a, p. 145) but these remain unpublished. Upon and following the publication of Feinleib's results, several criticisms were advanced concerning the analysis. Feinleib (1966) stated that there might be an added complication since the method measured the reflection from a thin film of metal in intimate contact with another material, the electrolyte.

The index of the electrolyte in the neighborhood of the surface might be changed by the same process that changes the metal film reflectance. On the other hand, he goes on to point out that if the change in the index of the electrolyte is ignored, one could obtain reasonable results from calculations based on the bulk optical constants. Another effect which clouds the picture is that no precautions were taken to avoid the formation of surface films or of soluble metal-chloride complexes at the interface and thus the entire results are put into doubt (McIntyre, 1973a, p. 121).

Basing their calculations on a simplified model of the metal-electrolyte interface, Prostak and Hansen (1967) argued that the observed effect could not be introduced by a change in the index of the space charge layer of the electrolyte. They in turn proposed a so-called rigid-shift theory which states that the ER effect observed by Feinleib and others can be accounted for by a rigid shift to lower photon energies of the entire dielectric function of the 0.5 Å thick metal surface layer. This amounts to changing the optical constants by an amount proportional to their first derivative. This theory seems to work well for gold, but for silver and copper the results are not at all in agreement with carefully done experiments. When this model was refined by Hansen and Prostak (1968), they found that including free electron effects shifted the curve to higher energies. They then proposed that the Fermi level in the conduction band is modulated by the field-induced free electron concentration

change but that other bands involving bound electronic states are unaffected by the field. This theory predicts a result similar to the rigid-shift, but now the translation along the energy axis is not uniform. This latter model has also proved unsatisfactory since the predictions for copper and silver do not agree with the experimental data (Parsons, 1969). It is also physically unrealistic because it assumes a low-frequency modulation of the Fermi level of the metals (McIntyre, 1973a, p. 127).

Stedman (1968) agrees with Prostak and Hansen (1967) in that the modulation of the ionic double layer is only of secondary importance. She later (Stedman, 1970) suggested that possibly compression in the dipole layer gave rise to a sort of piezomodulation effect, but this has been discounted on the basis that it is a quadratic effect and that it should be negligible when operating with a bias close to the point of zero charge (McIntyre, 1973a, p. 124).

Ferroelectric ER

Ishibashi and Stadler (1969) investigated the reflectance modulation of evaporated gold films. The modulation was obtained by reversing the polarization of the ferroelectric substrate. The metal film was affected and a response similar to that obtained by Feinleib was observed. Though the authors claim to show that they have eliminated the possibility that what they observed might be due to modulated piezoreflectance effects, not everyone is convinced that the analysis is as simple as they make it out to be (Seraphin, 1972a, p. 145).

Thus, though some effect has been observed and has been attributed to true electroreflectance, many doubts remain about this early work.

Recent Work on the ER of Metals

Recently, two new theories have been put forth that claim to explain properly the physics of the electrolytic measurements of ER. Both are based on free electron effects and are similar in some ways.

Cheyssac et al. (1973) suggest that the observed ER response is correlated with the surface itself. They assume that the electric field gives rise to excess free electrons ΔN, which in turn give rise to a surface current under the action of the electric field of the incident light wave. By modifying the usual boundary conditions used for solving Maxwell's equations for a light wave at an interface, they arrive at a fractional change in the reflectivity

$$\frac{\Delta R}{R} = F \frac{\Delta N}{N}$$

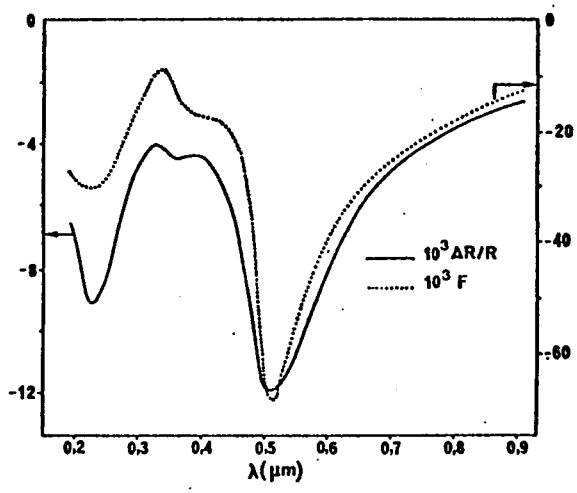
where N is the unperturbed free electron concentration,

$$F = -4n_E\rho \frac{n_E^2 + k^2 - (\rho+n)^2}{[n_E^2 + k^2 - (\rho+n)^2]^2 + 4k^2(\rho+n)^2},$$

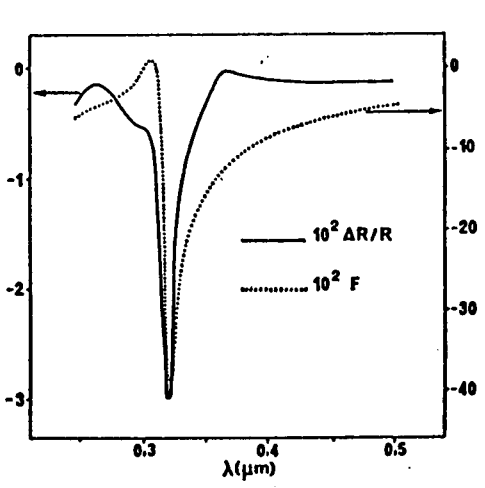
n_E is the index of refraction of the electrolyte, n and k are the optical constants of the metal, $\rho = (\mu_0/\varepsilon_0)^{\frac{1}{2}}\sigma$, and σ is the surface conductivity of the metal and is related to the bulk value, σ_0, by $\sigma = \sigma_0 d$, d being the penetration depth of the modulating electric field.

Figures 3.1(a,b,c) are reproductions of their comparison between the theoretical and experimental ER spectra of bulk gold, silver, and copper samples (Garrigos, Kofman, and Richard, 1973). The agreement between the calculated and the observed spectra shows that their simple theory can account reasonably well for the observed results. Extension of the theory to include more sophisticated expressions for parameters such as conductivity (McIrvine, 1966a,b) might lead to even better agreement. Of course, more data with which to compare this result would be useful in determining the suitability of the model in the general case.

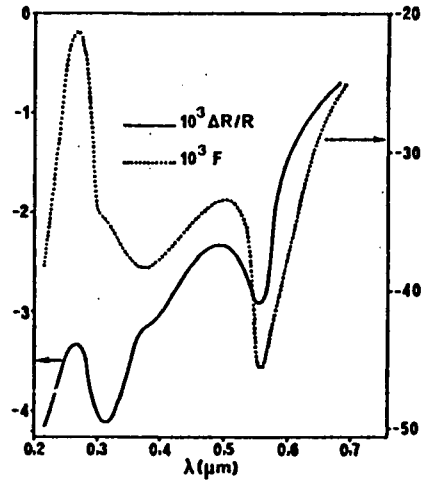
In the model proposed by McIntyre and Aspnes, (MA) (McIntyre, 1973a,b) a transition layer is assumed between the electrolyte and the metal surface. The electrolyte and the metal both have their respective bulk optical constants, and these connect smoothly from one to the other through the transition region. The only requirement on the thickness of the region is that it be much less than the wavelength of the light. A linear approximation is made for the ratios of Fresnel coefficients for a three phase system given in terms of (d/λ) where only the first order terms are kept. From this model, and on the basis of experimental results of the ratio of the reflectance of the s and p polarized components at an angle of incidence of 45°, it can be argued that the modulation takes place in the optical constants of the metal and not in that of the electrolyte (McIntyre, 1973a). It is not necessary to assume any specific model for the form of the field-induced



Gold



Silver



Copper

Fig. 3.1. Comparison of Experimental (ΔR/R) and Calculated (F) ER Spectra of Gold, Silver, and Copper

(From Garrigos, Kofman, and Richard, 1973)

shifts to draw this conclusion. The physical reason for this is that 45° is very near Brewster's angle for the experimental conditions considered, and the equations simplify a great deal to give this result.

The dielectric constant of the metal can be resolved into two components--the free-electron intraband transition term, and the interband transition contribution

$$\varepsilon_{tot} = \varepsilon_f + \varepsilon_b$$

where ε is the complex dielectric function, and tot, f, and b refer to the total, free electron, and bound electron parts, respectively. If we set the interband part, ε_b, equal to zero, in other words we assume that the only factor contributing to the optical constant of metal is that due to the free electrons, we get the interesting result that the model predicts no ER effect, to first order in d/λ, (McIntyre, 1973a, p. 131) when the index of the incident medium is unity.

It is assumed that interband transitions do exist in addition to the free electron effects, but are not affected by the externally applied field; thus,

$$\varepsilon_b = \text{constant.}$$

This now modifies the results derived for the first case; McIntyre and Aspnes show that the results of this theory tend to follow the general trend of the experimental data (see Fig. 3.2) for gold, silver, and copper.

McIntyre and Aspnes argue that field assisted interband transitions would result in either a lowering of the Fermi level or a

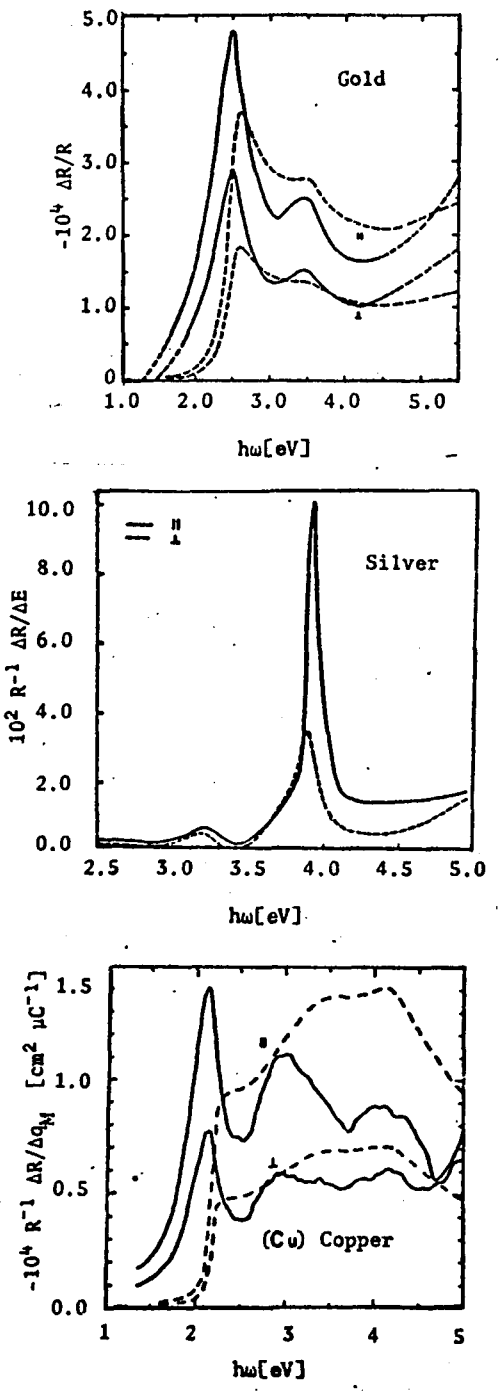


Fig. 3.2. Comparison of Calculated Curves (Dashed) and Experimental Curves (Solid) of the Electrolytic ER of Gold, Silver (from McIntyre, 1973a), and Copper (from McIntyre, 1973b)

raising of the energy levels of the lower level states in the surface atom layer. In either case, this would result in a large positive peak in the ER spectrum of silver near 4 eV. They conclude that the absence of such a peak in the experimental spectrum, which is everywhere negative, indicated that ε_b is not modulated to any great extent by the external electric field. On the other hand, there is some evidence that some such transitions have been seen in gold (McIntyre, 1973b, p. 678).

It should be noted that nowhere in their discussion do MA consider a specific shape for the spatial profile of the dielectric function in the transition layer.

Comparing the theories of MA and Cheyssac et al., we note that the MA model assumes a three phase system, while the other only assumes a two phase interface. While the MA model is well-suited to the case where electrolytic experiments are being analyzed, we must remember that similar results have been obtained in air by Ishibashi and Stadler (1969). Another difference between the two theories is that in the one of Cheyssac et al., there is an adjustable parameter, namely, the thickness of the surface layer in which the electron density was changed by the modulating field. No such parameter exists in the MA model. The result of adjusting this thickness value is of the same order of magnitude as the Thomas-Fermi length.

In summary, it is apparent that there are still many different ideas as to what is occurring in the ER of metals. Up to now, results

have been obtained on three materials with large free electron concentrations. Some theories seem to work remarkably well with the results measured on some materials, while others agree more poorly with all three. There is obviously a dire need for more results to give the theoreticians more data to work with.

CHAPTER 4

PENETRATION OF AN ELECTRIC FIELD INTO A THIN METAL FILM

In this chapter we consider the effect of an externally applied electric field on the charge distribution within a thin conducting metal film. An equation is derived which gives the profile of the electric field under various boundary conditions. With these results we evaluate classes of conductive materials in order to determine which might be most suitable in terms of field penetration. Further discussion includes the band structure of the selected sample material and the structure which might be observed in its ER spectrum.

Field Equation Derivation

Several different approaches are possible to derive the equilibrium distribution of carriers in a film which in turn gives us the electric field. Among these are (a) the minimization of the total energy of the electron-ion core system of charges, and (b) the balancing of the forces on a charged test particle. Unfortunately, though the physics of these approaches is correct, the mathematics of both soon proves to be rather formidable.

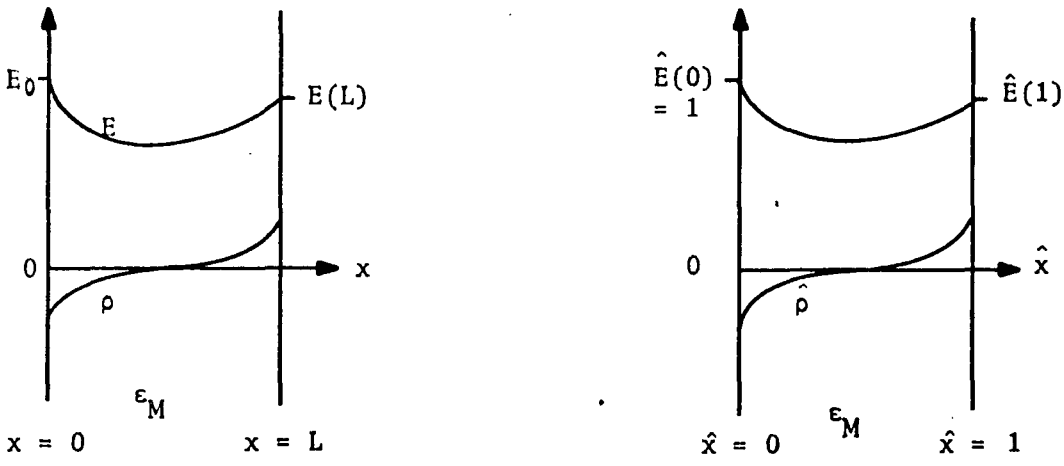
Professor Rolf Enderlein of Humboldt University in East Berlin suggested the following method which is elegantly simple in comparison to the other methods. Rather than balancing forces, we evaluate electron currents and set their sum equal to zero for the steady state solution. With this approach, the mathematical obstacles are neatly circumvented. The rationalized MKS system of electrical units is used in expressing all derivations and equations.

Let us assume that we have a film as shown in Fig. 4.1, with an arbitrary electric field E which has yet to be defined. We will develop the normalized and the unnormalized field equations in parallel. Since the equation is nonlinear, it will be seen that though the normalized form can be more easily solved numerically, its results are more difficult to interpret. We have also plotted qualitatively in Fig. 4.1 the charge distribution for the arbitrary E given in the figure.

We further assume that the charge carriers are electrons and are mobile while the ionic cores, consisting of the atomic nuclei and the tightly-bound inner electrons, are rigidly fixed at the lattice sites.

The charge carriers in the film will have in general a nonuniform distribution so that diffusion occurs. This can be described by Fick's law, (Kittel, 1958, p. 155):

$$\vec{J}_D = -D \text{ grad } n,$$



(a) unnormalized coordinates

(b) normalized coordinates

Fig. 4.1. Geometry for Derivation of the Field Equation

where $\vec{j}_D$ is the particle diffusion current density, $n = n(x,y,z)$ is the number density of the particles, and D is the diffusion constant. Since the particles are charged, Fick's law can be converted to electric current density by multiplying each side by the electric charge on each particle.

We now make the assumption that our film is very large in extent in both y and z directions in comparison to the x direction; we can therefore ignore edge effects and the problem is reduced to a one dimensional one. Thus we can drop the vectorial nature of Fick's law and consider the diffusion to take place only along the x coordinate. We can write

$$j = ej_D = -De\frac{dn}{dx} = D\frac{d\rho(x)}{dx}$$

where $-\rho(x)$ is the charge distribution of the negative carriers, and j is the electric current density.

Since we expect that some sort of nonzero electric field will be present in the final solution, we must consider the effect of this on the charged particles. Ohm's law describes the effect:

$$j_E(x) = \sigma_c E(x) = \mu_e \rho(x) E(x)$$

where j_E is the electric current density, σ_c is the electrical conductivity, $-\mu_e$ is the electron mobility $= -\mu e$, μ is the mobility given as drift velocity per unit force, and $-\rho(x)$ is the electron density function. Now we can write the total current density as

$$j_{tot} = \mu_e \rho(x) E(x) + D \frac{d\rho(x)}{dx} .$$

In the steady state situation, an equilibrium must exist between the two currents so that the total current is zero:

$$j_{tot} = 0 = \mu_e \rho(x) E(x) + D \frac{d\rho(x)}{dx} .$$

The macroscopic form of one of Maxwell's equations relates the electric field and charge density:

$$\nabla \cdot \vec{D} = \rho_{tot} \tag{4.1}$$

where $\vec{D} = \varepsilon_M \vec{E}$, ε_M is the dielectric constant in the film, and ρ_{tot} is the total charge density in the film. ε_M is not the usual tabulated dielectric function since in our approach we account explicitly for the polarization due to the free electrons. Thus, the dielectric function we have to use is that which arises from the polarizability of the inner electrons and the nucleus. Since the binding between these electrons and the nucleus can be expected to be tight, the relative dielectric constant will be approximately unity, so that $\varepsilon_M \simeq \varepsilon_0$, where ε_0 is the dielectric constant of free space.

The total charge in the film can be broken down into two parts: (a) that due to the free electron carriers, and (b) that due to the fixed ionic cores and their tightly bound inner electrons. We can rewrite Eq. (4.1) as

$$\frac{dE}{dx} = \frac{\rho_{tot}}{\varepsilon_0} = \frac{\rho_0 - \rho(x)}{\varepsilon_0} ,$$

where ρ_0 is the charge density due to the fixed ionic cores and includes the tightly bound electrons. Differentiating this once with respect to x we get

$$\frac{d^2E}{dx^2} = -\frac{1}{\varepsilon_0}\frac{d\rho(x)}{dx}.$$

ρ_0 is constant since we have assumed a uniform lattice throughout our film.

Combining these results we get

$$\mu_e E(x)\left(-\varepsilon_0\frac{dE}{dx} + \rho_0\right) - D\varepsilon_0\frac{d^2E}{dx^2} = 0$$

or

$$\frac{d^2E}{dx^2} + \frac{\mu_e}{D}E\frac{dE}{dx} - \frac{\mu_e\rho_0}{D\varepsilon_0}E = 0. \qquad (4.2)$$

For a classical gas having a Boltzmann distribution μ and D are related by an equation first deduced by Einstein (Feynman, Leighton, and Sands, 1963, p. 43-9):

$$\frac{\mu}{D} = \frac{1}{k_B T}$$

where k_B is the Boltzmann constant and T is the absolute temperature.

Electrons in a metal behave like a degenerate Fermi gas and therefore are described by Fermi statistics. An expression will be derived for μ/D for this case.

Under equilibrium conditions, a distribution of particles in a one dimensional box can be expressed as [Feynman, Leighton, and Sands, 1963, Eq. (43.36)].

$$D\frac{dn}{dx} = n\mu F \tag{4.3}$$

where F is the force on the particles and n = n(x) is the particle concentration. This is obtained by equating the diffusion current and the current due to the force on the particles. This differs from the previous result in that the electrostatic nature of the particles is not considered since Fermi particles do not necessarily have to be charged. Rewriting Eq. (4.3) we have

$$\frac{\mu}{D} = \frac{(dn/dx)}{nF}. \tag{4.4}$$

The Fermi energy for a free electron gas is given by [Kittel, 1971, Eq. (7.21)],

$$E_F = E_F(x) = \frac{\hbar^2}{2m}\left(\frac{3\pi^2 n(x)}{V}\right)^{2/3}$$

where V is the volume of the box, m is the mass of the particle, and $\hbar \equiv h/2\pi$. Note that E_F is a function of position through its dependence on n. Taking the natural logarithm of both sides of this equation makes differentiation simple:

$$\ln E_F = \frac{2}{3}\ln(n) + \text{constant},$$

and,

$$\frac{1}{E_F}\frac{dE_F}{dx} = \frac{2}{3n}\frac{dn}{dx}$$

or,

$$\frac{1}{n}\frac{dn}{dx} = \frac{3}{2E_F}\frac{dE_F}{dx} \quad . \tag{4.5}$$

Another expression can be obtained for $1/n \cdot dn/dx$ by using the Fermi distribution function in its low temperature, high density limit (Huang, 1963, p. 227):

$$n = \frac{n_o}{\exp\{[eU(x)-E_F(x)]/k_BT\} + 1} \quad ,$$

where $U(x) = \int E(x)dx$ and n_0 is the total number of particles in the box and k_B is the Boltzmann constant. Again taking the logarithm before differentiating, we can write directly

$$\frac{1}{n}\frac{dn}{dx} = \frac{eE(x) - (dE_F/dx)}{k_BT\,[\exp\{[eU(x) - E_F(x)]/k_BT\} + 1]} \quad .$$

Using Eq.(4.4),and rearranging, we get

$$\frac{dE_F}{dx} = \frac{2E_F}{3}\,\frac{eE(x) - (dE_F/dx)}{k_BT\,[\exp\{[eU(x) - E_F(x)]/k_BT\} + 1]}$$

and,

$$\frac{dE_F}{dx} = \frac{eE(x)}{1 + \{[3k_BT\,(\exp\{[eU(x) - E_F(x)]/k_BT\} + 1)]/2E_F\}} \quad .$$

The second term in the denominator can be shown to be negligibly small in comparison to unity: for $T \approx 300$ K, $k_BT \approx 1/40$ eV, $E_F \sim 10$ eV, and the exponential term will have a value close to zero when $eU(x) < E_F(x)$, so we can write

$$\frac{3(1)}{2(40)(10)} \simeq \frac{3}{800} .$$

Therefore, $dE_F/dx \simeq eE(x)$. Combining this with Eqs. (4.4) and (4.5), we get the result

$$\frac{\mu}{D} = \frac{3}{2E_F}$$

after identifying the force F in Eq. (4.3) as eE(x).

We can now eliminate the ratio μ_e/D from Eq. (4.2) and get

$$\frac{d^2E}{dx^2} + \frac{3}{2}\frac{e}{E_F} E \frac{dE}{dx} - \frac{3e\rho_0}{2E_F\varepsilon_0} = 0. \tag{4.6}$$

A length (Kittel, 1971, p. 279) can be defined which is called the Thomas-Fermi screening length or the Debye length depending whether we are talking about a metal (Fermi distribution) or a semiconductor (Boltzmann distribution) (Jackson, 1962, p. 342):

$$d^2 = \frac{2\varepsilon_0 E_F}{3e\rho_0} .$$

Using this in Eq. (4.6), we have

$$\frac{d^2E}{dx^2} + \frac{\varepsilon_0}{\rho_0 d^2} E \frac{dE}{dx} - \frac{1}{d^2} E = 0. \tag{4.7}$$

The screening distance d is a function of x through its dependence on E_F. We now put in this dependence explicitly:

$$E_F(x) = \frac{\hbar^2}{2m}\left(\frac{3\pi n(x)}{V}\right)^{2/3} = \frac{\hbar^2}{2m}\left(\frac{3\pi}{V}\right)^{2/3}\left(\frac{\rho(x)}{e}\right)^{2/3}$$

$$= E_F^0\left(\frac{\rho(x)}{\rho_0}\right)^{2/3},$$

where E_F^0 is the unperturbed Fermi level given by

$$\frac{\hbar^2}{2m}\left(\frac{3\pi\rho_0}{Ve}\right)^{2/3}.$$

Using $\rho(x) = \rho_0 - \varepsilon_0 E'$, we find

$$E_F(x) = E_F^0\left(\frac{\rho_0 - \varepsilon_0 E'(x)}{\rho_0}\right)^{2/3} = E_F^0\left(1 - \frac{\varepsilon_0}{\rho_0}E'(x)\right)^{2/3}.$$

Substituting this into the expression for the screening distance, we get

$$d^2 = \frac{d_0^2}{E_F^0}E_F(x) = d_0^2\left(1 - \frac{\varepsilon_0}{\rho_0}E'(x)\right)^{2/3},$$

where d_0 is the unperturbed screening distance given by $(2\varepsilon_0 E_F^0/3e\rho_0)^{\frac{1}{2}}$. Putting this into Eq. (4.7) we have

$$E'' + \frac{1}{d^2}\frac{\varepsilon_0}{\rho_0}EE' - \frac{E}{d^2} = E'' - \frac{E}{d^2}\left(1 - \frac{\varepsilon_0}{\rho_0}E'\right)$$

$$= E'' - \frac{E}{d_0^2}\left(1 - \frac{\varepsilon_0}{\rho_0}E'\right)^{1/3} = 0. \qquad (4.8)$$

This equation describes the electric field within a thin sheet of material under the assumptions that we made of the problem's being limited to one dimension.

We can now write the corresponding equations for the electric potential U, and for the charge density, ρ. Since $\vec{E} = -\nabla U$ and $\nabla \cdot \vec{E}$

$= \rho/\varepsilon_0$, we have

$$\frac{d^3U}{dx^3} - \frac{1}{d_0^2}\frac{dU}{dx}\left(1 + \frac{\varepsilon_0}{\rho_0}\frac{d^2U}{dx^2}\right)^{1/3} = 0 \qquad (4.9)$$

and

$$\frac{d\rho}{dx} - \frac{1}{d_0^2}\left[1 - \frac{\rho(x)}{\rho_0}\right]^{1/3}\int\rho(x)dx = 0 \,. \qquad (4.10)$$

The three labeled equations (4.8) to (4.10) all represent the same physics stated in different forms. The next task is to solve one or more of these for particular boundary conditions which obtain from the experimental configuration.

Since the potential equation is a third order differential equation, three constants of integration are required to get a complete solution. The electric field equation requires two, while the charge distribution requires only one. In the latter case, however, there is an integral appearing twice and this integral also requires a constant of integration, so we need to specify a minimum of two boundary conditions in any case for a complete solution.

The equation for the electric field appears to be the most tractable one of the three: (a) it is only of the second order and it therefore requires only two boundary conditions; (b) it is not a mixed integral-differential equation; (c) as we will see later, its boundary conditions are known from consideration of the experimental arrangement; and (d) it can be normalized easily for mathematical

convenience. A further reason for concentrating on the field equation as opposed to the other two is that the band structure changes are usually discussed in terms of the action of an electric field on the optical transitions. From this point of view, it is the most important of the three forms. We therefore confine our attention for the moment on the equation given in terms of the electric field.

If we write Eq. (4.7) in the following way, we can get some insight as to the origin of each term

$$\frac{d^2E}{dx^2} - \left(\frac{\rho_0}{\epsilon_0} - \frac{dE}{dx} \right) \frac{3e}{2E_F} E = 0. \tag{4.11}$$

The first term is proportional to the diffusion current. The current due to the electric field comes in as the factor $(3e/2E_F)E$. This factor is modified by a coefficient made up of two terms. The first, ρ_0/ϵ_0, is the ionic core restoring force. The second has the same sign as the diffusion term and is due to the electric repulsion between the charge carriers.

To study the behavior of this equation from a mathematical point of view, we can normalize it as follows. The circumflex over variables indicates normalized quantities:

$$\hat{E}(x) = \frac{E(x)}{E_0}, \quad \frac{d\hat{E}}{dx} = \frac{1}{E_0} \frac{dE}{dx}, \quad \frac{d^2\hat{E}}{dx^2} = \frac{1}{E_0} \cdot \frac{d^2E}{dx^2},$$

where E_0 is defined as the value of the electric field at $x = 0$.

Next, we normalize the spatial coordinate x to values between zero and 1.0:

$$\hat{x} = \frac{x}{L}, \quad \frac{d}{dx} = \frac{d\hat{x}}{dx}\frac{d}{d\hat{x}} = \frac{1}{L}\frac{d}{d\hat{x}}, \quad \frac{d^2}{dx^2} = \frac{1}{L^2}\frac{d^2}{d\hat{x}^2},$$

where L is the thickness of the metal film. We also define the following quantities

$$\hat{\alpha} = \frac{L}{d}, \quad \hat{\beta} = \frac{\sigma}{\rho_0 L},$$

$$\alpha = 1/d, \text{ and } \beta = \varepsilon_0/\rho_0$$

where σ is the surface charge which generates the external electric field and is equal to $\varepsilon_0 E_0$. With these definitions we can write, after appropriate algebra,

$$\frac{d^2\hat{E}}{d\hat{x}^2} - \hat{\alpha}^2\hat{E}\left(1 - \hat{\beta}\frac{d\hat{E}}{d\hat{x}}\right)^{1/3} = 0$$

or simply,

$$\hat{E}'' - \hat{\alpha}^2\hat{E}\,(1 - \beta\hat{E}')^{1/3} = 0 \qquad (4.12)$$

where the primes indicate differentiation with respect to the normalized x. Equation (4.12) can be reduced to

$$\frac{1}{2}\alpha^2 E^2 = \int \frac{p\,dp}{(1-\beta p)^{1/3}}$$

by the substitution p = dE/dx and dp/dx = (dp/dE)/(dE/dx) = p(dp/dE). If we look at the equation in which the variation of the screening

length with x is ignored [Eq. (4.7)], the result corresponding to the above is

$$\frac{1}{2}\alpha^2E^2 = \frac{1}{\beta^2}(1 - \beta E' - \ln|1 - \beta E'|) + \text{constant}.$$

This equation is a first order transcendental equation which can be shown to belong to a class that has no solution in closed form (Stavroudis, 1973). Because of the apparent mathematical obstacles, it was thought that either linearization of the equation or numerical integration would prove to be more fruitful than further attempts at direct integration. Before looking at the numerical solution, we will consider some approximations that linearize the equation.

A few words are in order here on the physical significance of the ratio $\hat{\beta}$. The denominator is the total number of atoms that have given up an electron to the free electron gas per unit area of the sheet. The numerator, on the other hand, is the surface charge density necessary to fully compensate the external field. Thus, we see that the value of $\hat{\beta}$ can range from zero for the case where there is no applied field, to one where the field is so large that all available free charges in the sheet can only partially compensate it. The parameter $\hat{\alpha}$ is the thickness of the film expressed in units of the screening length d.

Two Sets of Boundary Conditions

In general, an electrostatic problem with specified Neumann boundary conditions over a closed surface results in a unique stable solution (Jackson, 1962, pp. 5, 16-17). The Neumann boundary conditions specify the electric field (normal derivative of the potential)

everywhere on the surface (corresponding to a given surface charge density); the closed surface may, in whole or in part, be at infinity. We can specify $\vec{E}$ on two surfaces, which reduces to specifying E(x) at two points for our particular one-dimensional case. Alternatively, we could specify E and E' at any point. This is consistent with what is to be expected of a second order ordinary differential equation, i.e., there will be two constants of integration.

The first condition is easily specified: at $x = \hat{x} = 0$, the left side of the sheet, we can specify an electric field, E_0, which might be due, for instance, to a specified and fixed external surface charge density σ. To simplify this conceptually, we can suggest two ways to obtain this experimentally: (1) a ferroelectric material with a bound surface charge density σ is used as the substrate of our thin sheet; (2) our sheet is self-supporting and it is suspended between the two plates of a plane-parallel capacitor in a vacuum. In the latter case, the voltage on the capacitor can be adjusted to give the desired electric field E_0 in the vacuum space. At this point, we introduce a very useful result of Gauss' law, which states that

$$\oint_S \vec{E} \cdot \hat{n} da = 0$$

if there is no net charge inside the surface S (Jackson, 1962, p. 5). Another way of writing this for our particular case is

$$\int_A^B E(x)dx = E(B) - E(A) = 0$$

if there is no net free charge between A and B. Applying this to the case of the vacuum capacitor just mentioned, the vacuum space does not

attenuate the electric field and can be made arbitrarily small. This is very convenient in considering an isolated sample as we now do.

If we consider the sample to be completely electrically insulated by surrounding it on both sides with a perfect vacuum, for example, and then place it between the plates of a capacitor, there will be no net electric charge in the sheet. Since there is charge neutrality within the film, the electric films at both surfaces of the film are equal:

$$\hat{E}(0) = \hat{E}(1) = 1$$

or,

$$E(0) = E(L) = E_0.$$

A second boundary condition which is realizable consists of connecting the right surface of the sheet to ground potential so that charge may flow into the sheet. We now have a plane parallel capacitor with the sheet now forming one of the plates. In such a case, the current flow through the ground wire will correspond to that which will give a surface charge, σ, equal to that which gives rise to the electric field E_0. In such a case, we have a zero electric field at the right surface

$$\hat{E}(1) = E(L) = 0,$$

and we specify the potential to be zero at $\hat{x} = 1$. This condition will have to be modified later for the case where all of the free electrons are depleted by a strong negative electric field.

In either case, it now remains for us to discover how the charge, whether it be the intrinsic free electrons or these electrons

plus some charge from the outside, distribute themselves as a function of x.

Approximate Solutions

We now look at various approximations which can be made. We will consider both the normalized and unnormalized forms of the equation in this section; it is shown that care should be exercised in interpreting and in applying the normalized results directly to the physical situation because of the type of normalization used and the fact that the basic equation is nonlinear.

Negligible Diffusion Term

Using Eq. (4.12), we let $|\hat{E}''| << |\hat{\alpha}^2\hat{E}(\hat{\beta}\hat{E}' - 1)^{1/3}|$ so that

$$\hat{\beta}\hat{E}\hat{E}' - \hat{E} = (\hat{\beta}\hat{E}' - 1)\hat{E} = 0.$$

This leads to two solutions

$$\hat{E} = 0 \text{ and } (\hat{\beta}\hat{E}' - 1) = 0.$$

Integrating the second, we get for the pair of solutions

$$\text{(a)} \quad \hat{E}(\hat{x}) = 0, \text{ and (b)} \quad \hat{E}(\hat{x}) = \hat{x}/\hat{\beta} + c.$$

We will now determine the regions over which each is valid. Besides the boundary conditions, we have one other fact which will enter the discussion: there is no limit on the negative charge density attainable, but the positive charge density has a maximum value given by the ion core density, ρ_0. Taking first the case of the insulated

sample with a positive field, we find that at $\hat{x} = 0$ the (b) solution does not have a proper slope. Solution (a) leads to a step in $\hat{E}$ which requires an infinite negative charge density. This is compatible with our assumptions. On the right boundary, this solution would not be acceptable because it would require an infinite positive charge density to bring the field back up to the boundary value. The constant in solution (b) can be evaluated to give

$$\hat{E}(\hat{x}) = 1 + (\hat{x} - 1)/\hat{\beta}$$

for the right side of the interval. The dividing point between the two solutions is given by

$$1 + (\hat{x} - 1)/\hat{\beta} = 0 \quad \text{or} \quad \hat{x} = 1 - \hat{\beta}.$$

The solution for positive field and grounded sample is given by $\hat{E}(\hat{x}) = 0$ over the entire region. The results are different when the applied field is negative. Using arguments similar to those above, we find a reversed result for the insulated sample. The result for the grounded sample is different in that the two solutions are required. Near the left boundary we use solution (b):

$$\hat{E}(\hat{x}) = \hat{x}/\hat{\beta} - 1 \qquad 0 \leq \hat{x} \leq \hat{\beta}$$

For the rest of the interval we use $\hat{E}(\hat{x}) = 0$. The field and the charge profiles of these solutions are given in Fig. 4.2 and are summarized in Table 4.1. Repeating this for the unnormalized equation, we have

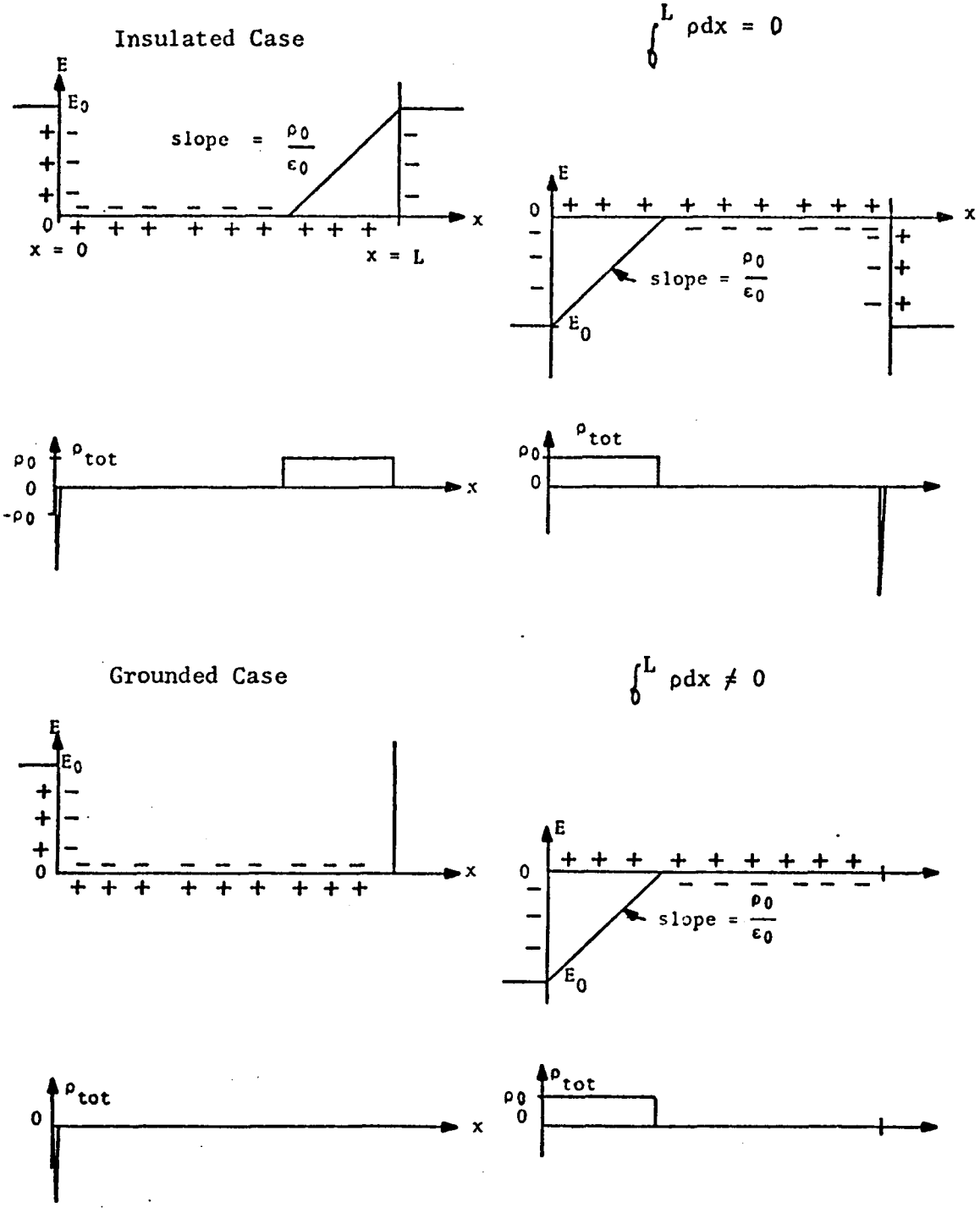


Fig. 4.2. Solutions of Field Equation When Diffusion is Ignored

$$\mu\rho(x)E(x) = 0,$$

and using

$$\rho(x) = \rho_0 - \varepsilon_0 \frac{dE}{dx}$$

we get the differential equation

$$\mu E(x) \left(\rho_0 - \varepsilon_0 \frac{dE}{dx} \right) = 0$$

with solutions

$$E(x) = 0$$

and

$$E(x) = \frac{\rho_0}{\varepsilon_0} x + E_0.$$

At this point we can see one of the difficulties that can arise in trying to apply the normalized results directly to the real situation. The slope of the field profile, $d\hat{E}/d\hat{x}$, appears to be a function of σ, the applied field, since $\hat{\beta} = \sigma/\rho_0 L$ in the normalized case. On the other hand, the derivation using the unnormalized equation shows no such dependence--only material parameters enter into the result.

The results obtained from this approximation give us a qualitative feel for the way the charges will distribute themselves under the influence of the external field. Inclusion of the diffusion term will tend to smooth out the results, but it should not change the overall trend of the separation of the charges.

Table 4.1. Characteristics of Solutions when Diffusion is Ignored

Solution	$\hat{E}$	$\hat{E}'$	$\rho_{tot} = \rho_0 - \rho(x)$	$\rho(x)$
(a)	0	0	0	ρ_0
(b)	$\frac{\hat{x}}{\hat{\beta}} + c$	$1/\hat{\beta}$	ρ_0	0

Negligible $\beta E'$ Term

The normalized form of the field equation is (Eq. 4.12)

$$\hat{E}'' - (1-\hat{\beta}\hat{E}')^{1/3}\hat{\alpha}^2\hat{E} = 0.$$

The product $\hat{\beta}\hat{E}'$ derives from the mutual electrostatic repulsion of the charged carriers, and we now assume that this term is negligible with respect to unity. This approximation can be effected in several ways. The applied field can be made small so that $\hat{\alpha}$ and $\hat{\beta}$ are small. Conversely, the gradient $\hat{E}'$ can be made small, since

$$\hat{E}' = \frac{\rho_{tot}}{\varepsilon_0} = \frac{\rho_0 - \rho(x)}{\varepsilon_0},$$

and this means that $\rho(x) \sim \rho_0$. This can be interpreted another way by rewriting the condition:

$$\hat{\beta}\hat{E}' = \left(\frac{\sigma\varepsilon_0}{\rho_0 L}\right)\left(\frac{\rho_0 - \rho(\hat{x})}{\varepsilon_0}\right) = \left(\frac{\sigma}{L}\right)\left(1 - \frac{\rho(\hat{x})}{\rho_0}\right) \ll 1.$$

If the field is of such a magnitude that only a very small percentage of the available charge is required to compensate it, then the ratio $\rho(x)/\rho_0$ will be small and our assumption will be valid. This case will later be seen to obtain for the case of bulk metals.

We will treat the normalized form here and, with this assumption, it reduces to

$$\frac{d^2\hat{E}}{d\hat{x}^2} - \hat{\alpha}^2\hat{E} = 0.$$

The general solution of this is the sum of two particular solutions (Nelson, Folley and Coral, 1960, Chapter 4):

$$\hat{E} = a_1 e^{\hat{\alpha}\hat{x}} + a_2 e^{-\hat{\alpha}\hat{x}}.$$

Boundary conditions for the insulated sample give us the result

$$\hat{E} = \frac{\cosh \hat{\alpha}(\tfrac{1}{2}-\hat{x})}{\cosh \tfrac{1}{2}\hat{\alpha}}$$

while the set for the grounded sample gives

$$\hat{E} = \frac{\sinh\hat{\alpha}(1-\hat{x})}{\sinh\hat{\alpha}} .$$

Plots of these curves along with the first derivative and the first integral are given in Fig. 4.3 for various values of the parameter $\hat{\alpha}$.

Bearing in mind the cautions in the interpretation of the results of the normalized curves indicated below, we convert these solutions to the unnormalized form and calculate the first derivative

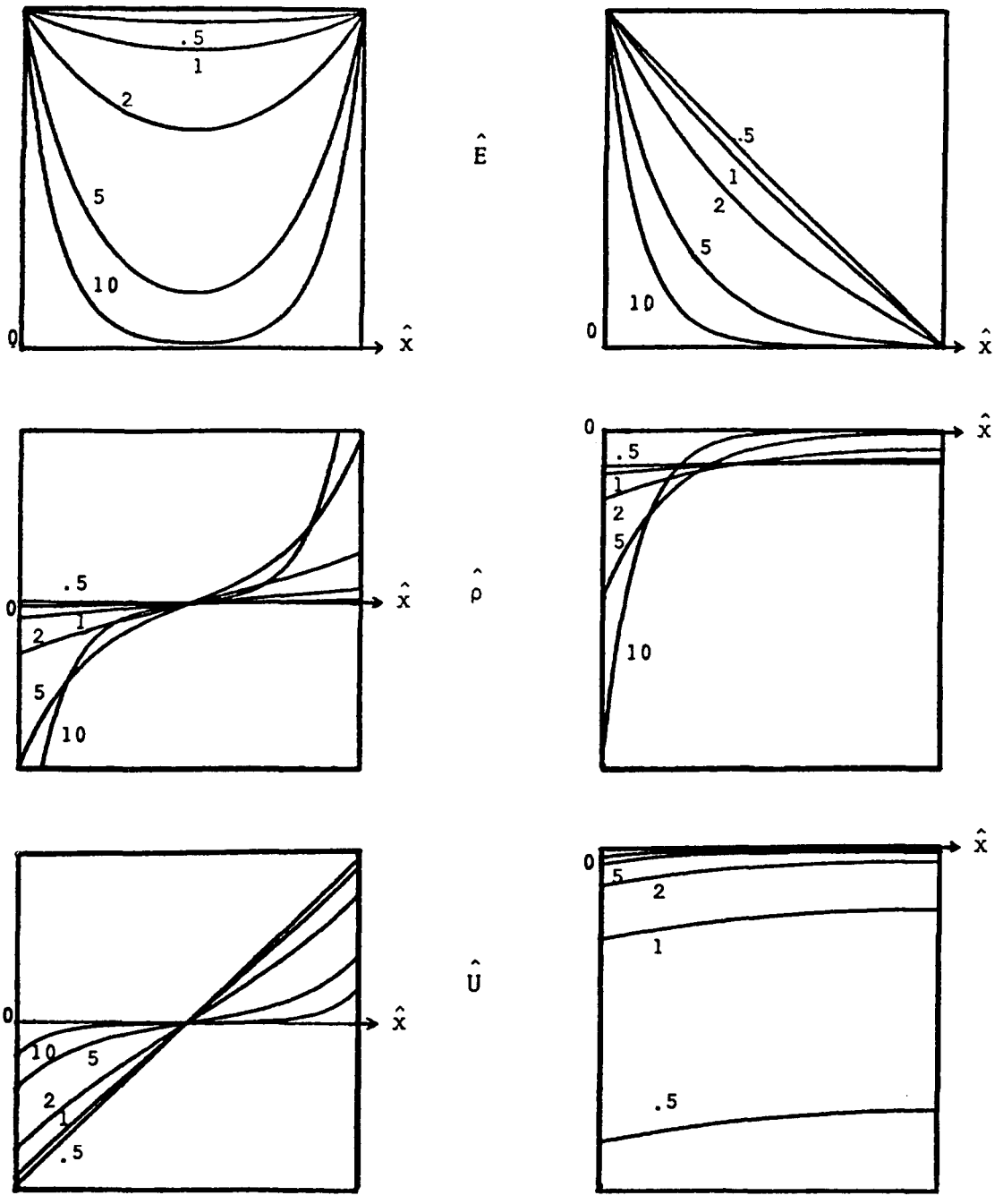


Fig. 4.3. Field, Charge, and Potential Profiles When βE' Term Can Be Neglected

Insulated case $E = E_0 \dfrac{\cosh (L-2x)/2d}{\cosh L/2d}$ $\qquad E' = \dfrac{-E_0}{d} \dfrac{\sinh (L-2x)/2d}{\cosh (L/2d)}$

Grounded case $E = E_0 \dfrac{\sinh (L-x)/d}{\sinh (L/d)}$ $\qquad E' = \dfrac{-E_0}{d} \dfrac{\cosh (L-x)/d}{\sinh (L/d)}$

The slopes at the origin for the two boundary value cases are given by

$$\text{Insulated case} \qquad E'(0) = -\frac{E_o}{d} \tanh \frac{L}{2d} \tag{4.13}$$

$$\text{Grounded case} \qquad E'(0) = -\frac{E_o}{d} \operatorname{ctnh} \frac{L}{d} .$$

These slopes are plotted in Fig. 4.4 as a function of $\alpha = L/d$. When the sample is sufficiently thick, the boundary condition at the second surface does not affect the curve near the first surface. This is of practical value when we discuss the possible experimental configurations. The asymptotic value which both cases approach in the limit can be compared to the slope obtained in the first approximation where diffusion was ignored. There, the slope was given by ρ_0/ε_0 and was therefore independent of the applied field. Under the present assumption, however, the starting slope is now dependent on the size of the field at the origin. We will return to this point in the discussion of the numerical solution of the equation where no physical approximations are made.

Figure 4.4 shows an interesting result as $\alpha \to 0$. In the case of the isolated film, the slope of the field goes to zero, i.e., the field does not get attenuated, which is hardly surprising. On the other hand,

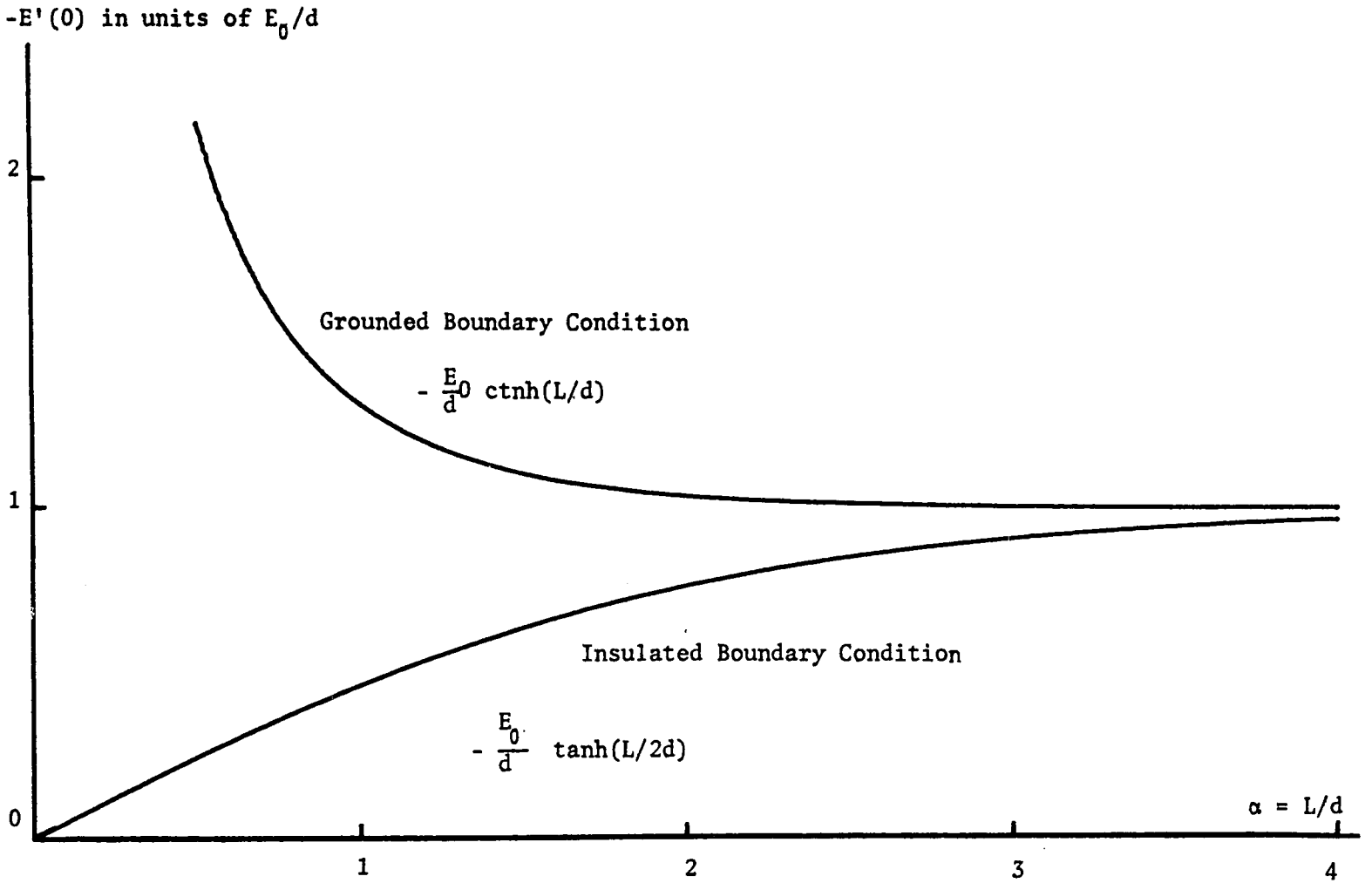


Fig. 4.4. Slope at Origin as a Function of Film Thickness

in the case of the grounded sample the slope goes to infinity. The interpretation here is that as the film is removed, for example, the free charges which originally came from external sources remain behind in an infinitely thin sheet. Of course, such a limit cannot be reached in practice with a thin self-supporting film, but the concept has validity when we consider the case of a film deposited on a ferroelectric substrate, as we will do later.

In each of the approximations used above, important aspects of the problem were ignored in the attempt to arrive at a mathematically tractable solution. Let us look at the validity of these approximations.

In the first case we let $|\hat{E}''| << |\hat{\alpha}^2\hat{E}(\hat{\beta}\hat{E}'-1)^{1/3}|$. We have no way of assuring a priori that $\beta E'$ will not be on the order of unity, thus requiring rather large values of $\hat{\alpha}^2\hat{E}$ to make the assumption valid. As a matter of fact, in the process of generating a solution with the assumption of no diffusion, we set $\hat{\beta}\hat{E}'$ equal to 1! Thus we should not emplasize the results based on this assumption.

In the second approximation we let $|\hat{\beta}\hat{E}'| << 1$. $\hat{E}'$ takes on its steepest value of $x = 0$, so we can easily determine some limits on β from values of $E'(0)$. In the case of the insulated sample, the worst case occurs for values of L/d which will make the hyperbolic tangent in Eq.(4.13) approximately one. A film with a thickness of at least three screening lengths satisfies this condition. This yields $dE/dx = E_0/d$ and the assumption, in physical units, becomes

$$\frac{\varepsilon_0}{\rho_0}\frac{E_0}{d} << 1.$$

The largest fields which one can expect to get in a laboratory would not exceed 10^8V/cm. A highly conductive metal, e.g., copper, has $\rho_0 \sim 10^{23}$ carriers per cubic centimeter and d ~ 0.5Å. This gives

$$\frac{\epsilon_0}{\rho_0}\frac{E_0}{d} \simeq 0.1.$$

Thus, if a field of 10^8 V/cm is attainable, the approximation which neglects the nonlinear term is almost valid. By the same token, a field of 10^6 V/cm will reduce the contribution of this term to 0.1% or less at most. In the case of a material with significantly lower carrier density, this term is again seen to be important through the reduction of ρ_0 in the denominator.

For very thin samples the factor corresponding to the field gradient becomes smaller since the hyperbolic tangent is zero when the argument is zero [Eq.(4.13)]. This is another situation where the nonlinear term might be neglected. Unfortunately, the hyperbolic tangent function becomes negligible only for values of L which are much less than the screening length d (see Fig. 4.4).

In the case of the boundary conditions for the grounded sample, the thinner the sample is made, the more significant this term becomes. In a reasonably thick sample, the same conditions as the insulated case obtain.

In summary we state that none of the approximations are very satisfactory in providing a solution.

Numerical Solution of Equation

Given a well-behaved differential equation, it is possible to use the fundamental definition of a derivative and go from a continuous function to a discretely sampled one, the sampling depending, of course, on the steepness of each differential involved. In particular, for our equation we could start at $\hat{x} = 0$ with a value for $\hat{E}$ and $\hat{E}'$, calculate $\hat{E}''$, and from this calculate new values for $\hat{E}'$ and $\hat{E}$ at $\hat{x} = \Delta x$. This process is repeated as often as necessary to proceed across some given interval. The size of the step increment Δx taken must be such that the true value for $\hat{E}$ is accurately approximated by the calculated value. This is the basis for numerical solutions of differential equations. As we have already seen in the preceding sections, the behavior of the equation is sensitive to all of the terms and therefore rather small steps in x will have to be used to achieve the desired accuracy. Fortunately, we know from physical considerations that all of the derivatives are well behaved, so we do not have to worry about any discontinuities.

The boundary conditions, however, do pose a problem. As previously mentioned, there are two ways of specifying them for our equation: (1) the field E is given at two points; (2) the field and its derivative are given at some point. This latter case would be ideal, for if we knew E and E' at x = 0, we could simply apply the algorithm given above to calculate E from x = 0 to x = L. Alas, we know the field at two points, x = 0 and x = L, and we have no a priori knowledge of

the exact starting slopes. Therefore, we must adjust E' (0) until the curve passes through the point E(L). Since the procedure is a rather lengthy one, the only way of finding the solution is to program a computer to select a starting E'(0), evaluate the solution and then adjust E'(0) appropriately. After a certain number of iterations a satisfactory solution should result. Due to the steep curves and the large numbers which are sometimes added to smaller ones, it is necessary to have a large number of significant digits, especially when evaluating some curves which have large values for ρ_0, α, and β.

The program which was written to accomplish this is listed in Appendix A along with a short user's description and a block diagram of the logic flow.

Results and Discussion of Numerical Solution

Plots of the numerical integration of the normalized equation are given in Fig. 4.5. Along with the solution to the electric field equation are plots of the electric potential and the charge distribution.

There are several interesting features of these solutions which merit comparison with the results obtained in the approximations previously tried. First, there is an obvious asymmetry of the curves which was previously seen in the first approximation. The mathematical origin of this is the nonlinear term that disappeared when we ignored β. The physical explanation for this comes from the fact that the positive carriers (the ionic cores) are not mobile, while the negative carriers (the free electrons) are capable of moving from one part of the sample to another under the influence of the field.

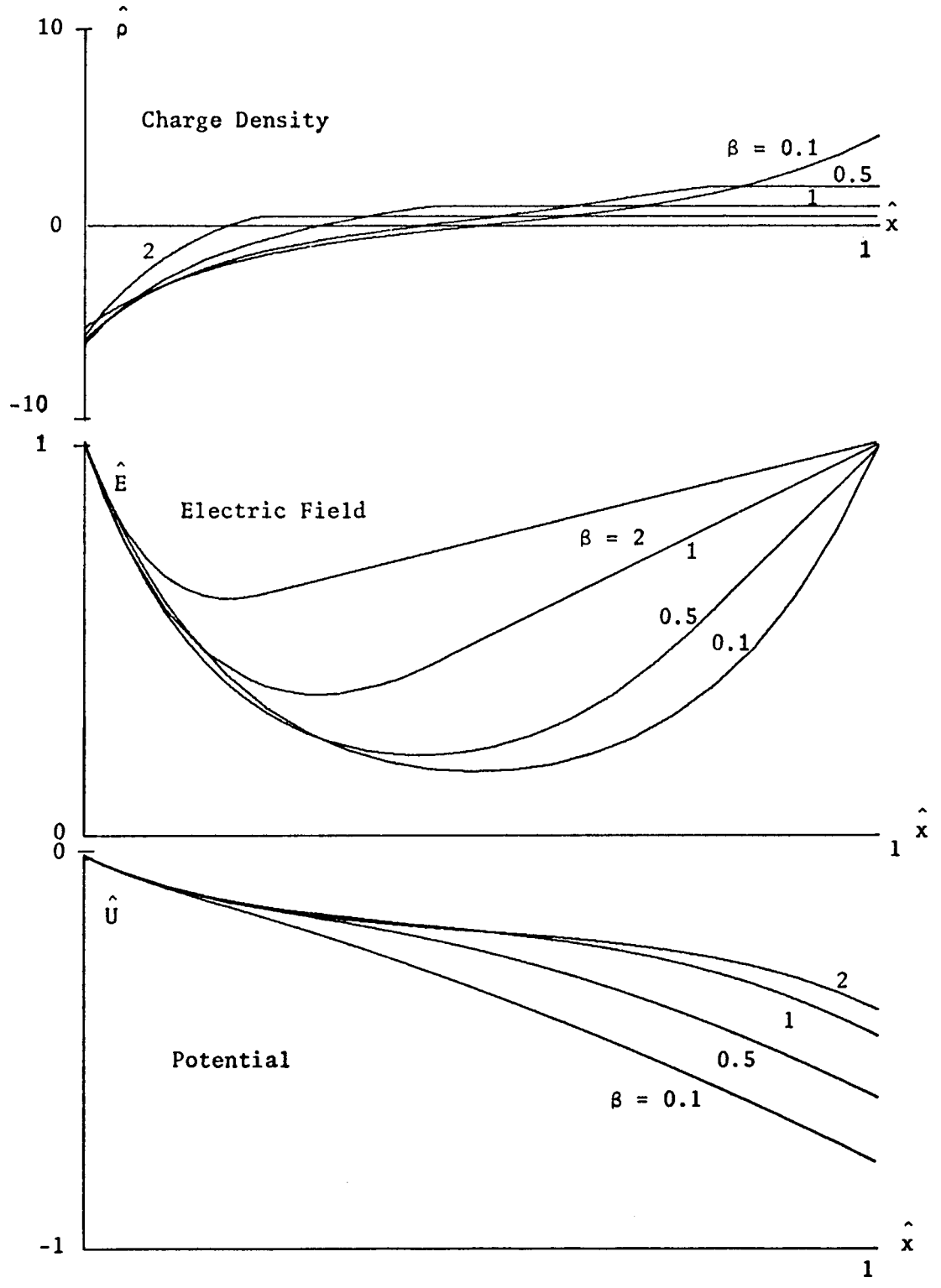


Fig. 4.5. Potential, Field, and Charge Density for the Normalized Equation ($\alpha = 5$)

A second feature that becomes apparent for certain ranges of parameters is that the E curve is very close to a straight line.

On one side of the sample the electric field profile initially decays almost exponentially, while on the other side, the field initially decreases almost linearly. The linear behavior is advantageous if line shape analysis is to be done on the resulting spectra. The shape of the spectral ER line can be theoretically evaluated for such a simple case of inhomogeneous modulation. The electric field is known quite accurately so that the change in reflectance can be put on an absolute basis (Seraphin, 1967, discussion).

In the discussion above we have been considering light incident on a film with a positive electric field as shown in Fig. 4.6(a). A change of the polarity of the field then presents the exponentially decreasing side of the field curve to the light [Fig. 4.6(b)]. If the light is somehow allowed to probe the film from the substrate side as shown in (c) of the same figure, the result is the same relative orientation of the light and the field vectors as was the case in (a). In other words, reversing the polarity of the field will not simply reverse the polarity of the linear region. The only way to get the light to penetrate into a region where a negative field exists over a depth comparable to that probed by the light is to make the sample so thin that the field does not decrease very much throughout the thickness of the film. Since ER is a quadratic effect, the response is independent of the sign of the field. This minimizes the problem caused by the inability to reverse the relative polarities in the linear region.

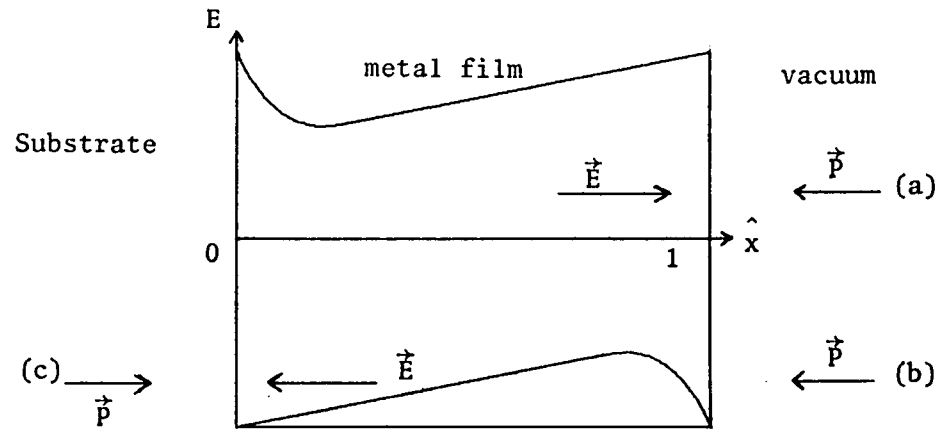
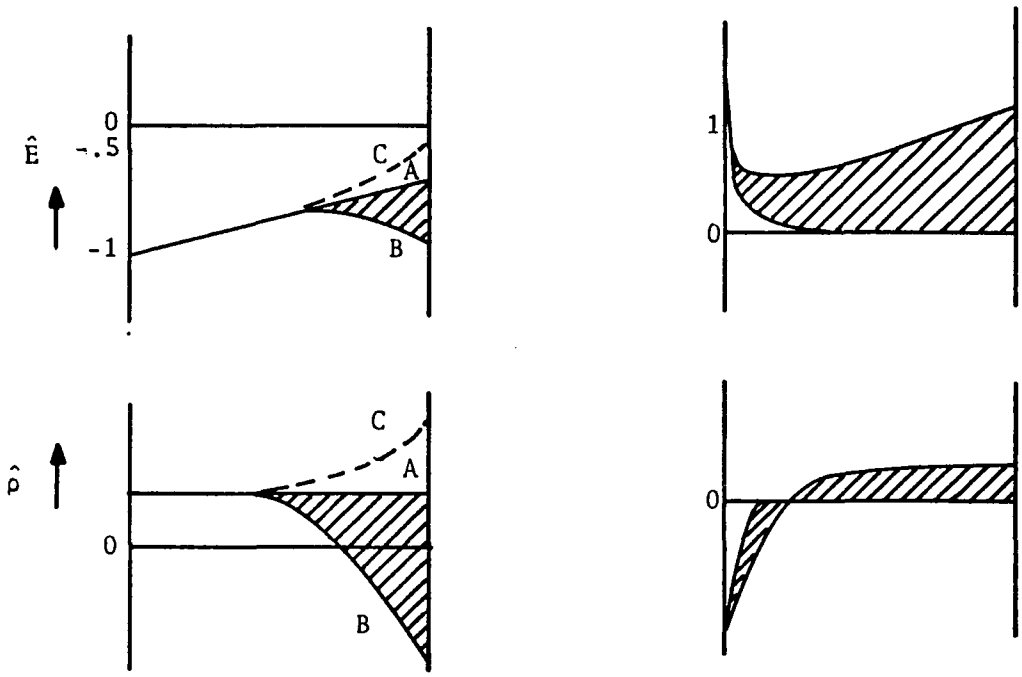


Fig. 4.6. Orientation of the Incident Light Vector $\vec{P}$ With Respect to the Electric Field Vector $\vec{E}$

In our earlier discussion on the results of the two approximations we noted that a different type of slope was obtained for each case. It is now seen that they both appear in the numerical solution. In the region where we have depleted all of the electrons we have the same result as when we ignored diffusion. In the region where we have a large charge density, we have a result similar to the other case where we assumed the field term to dominate the electrostatic repulsion term.

Not all mathematical solutions of the field equation are valid physical solutions. This may most easily be illustrated by an example. Let us assume values for $\hat{\alpha}$ and $\hat{\beta}$ of 5 and 2, respectively. This means that the number of external fixed charges which generate the field is twice as large as the number of free electrons in the metal film. Gauss' law predicts that the field will decrease by a factor of two through the thickness of this film. This result is shown in Fig. 4.7 and is labeled A. Curve B is that which we get when the film is insulated. Curve C, which is dashed, is a possible mathematical solution. This latter result is unphysical, however, because the normalized charge density cannot rise above 0.5. If it did, this would mean that we are taking out of the film more negative charges than are available from the free electrons. This could only come about by removing inner core electrons, and this is contrary to the working hypothesis that these electrons are tightly bound in comparison to the fields we are applying externally. Thus the shaded region is the only one in which physical results are possible for negative fields.



(a) negative field

(b) positive field

Fig. 4.7. Physically Allowed Solutions to the Field Equation

This is not the case with positive fields, since in this case we can easily get external negative charges to flow into the film in the grounded case, and they can find a place in the band structure as long as there is no gap in the conduction band.

Elimination of Free Electron Effects

One of the most serious objections to the results obtained in past studies of the electroreflectance of metals is that the free electron effects may dominate as discussed in Chapter 3. The equations developed here are consistent with these views. The solutions to the field equation and the corresponding charge density plots, however, point to a way to overcome these objections. Figure 4.8 is a plot of the electric field profile, the charge distribution, and the light field profile. If we now vary the electric field to the extent shown by the three curves, we find that the charge density in the region sampled by the light does not change. Therefore, the interaction between the light wave and the solid will be independent of any free electron effects.

The physical explanation for this effect lies in the fact that we have a large region in which the free electrons are completely depleted. The slope of the field in the normalized case is $1/\hat{\beta}$, while in the unnormalized case it is ρ_0/ε_0. It is significant to note here that this applies only to the case where the sample is electrically insulated.

When the sample is connected to ground, in general, there is no way of eliminating the free electron effects. As the electric field

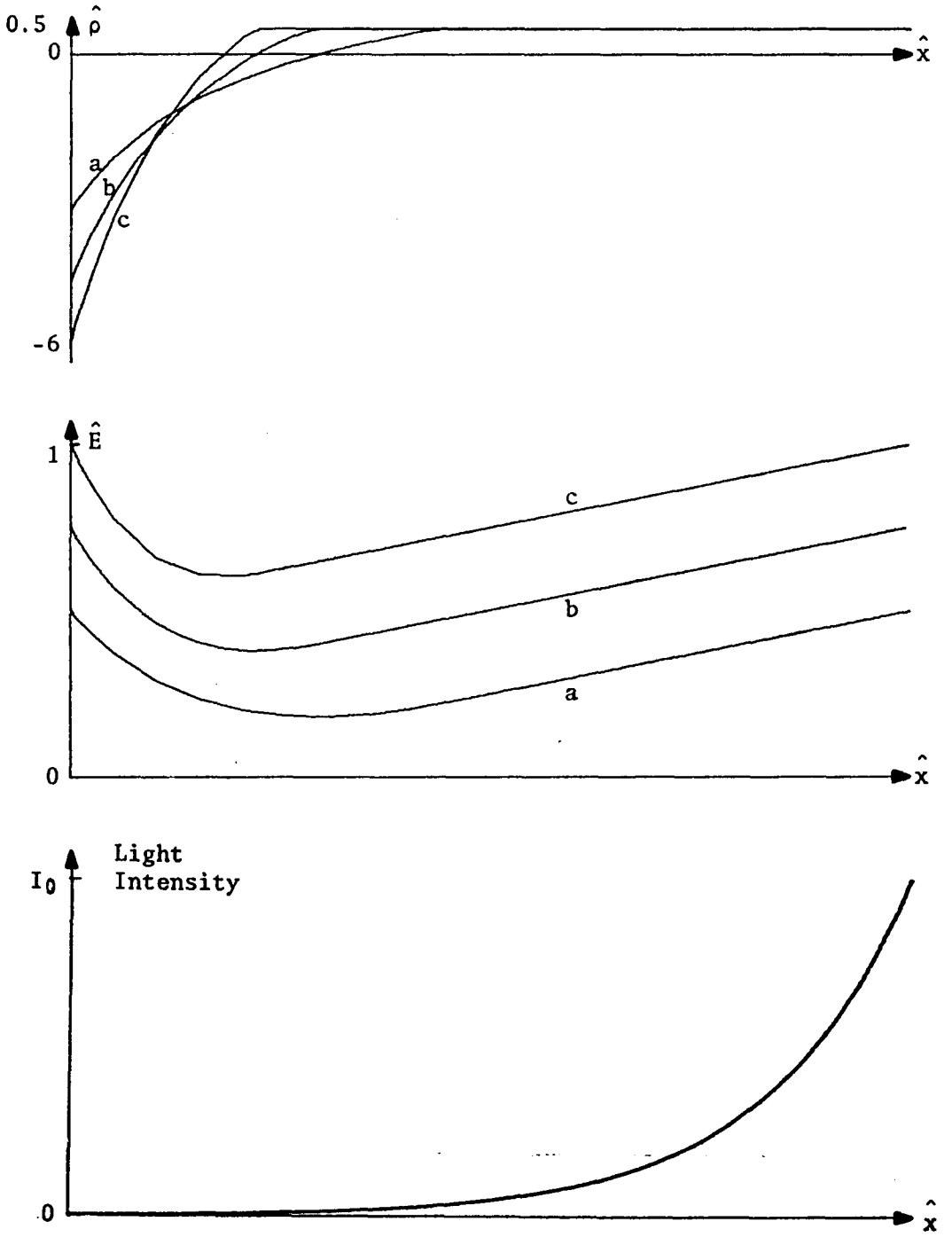


Fig. 4.8. Profiles of the Electric Field, the Charge Distribution, and the Light Wave in an Insulated Sample

is varied, charges flow into and out of the film. The light meanwhile probes the tail of the electric field profile from the other side, where for all practical purposes, the electrons are always at their bulk value (see Fig. 4.9). If the sample is such that the two profiles overlap, the varying charge distribution will also reach into the region probed by the light, thus giving rise to free electron effects. Furthermore, this will happen only if the sample is made exceedingly thin (100Å for the case of copper), in which case it is doubtful whether we are actually observing the band structure of the metal or whether we are seeing something else, such as some interface reaction between the substrate and the first few molecules, or even the electroreflectance of the substrate itself. This discussion is valid for the general case of metals with large free electron densities; in the particular case of bismuth, it is shown in another section that it is indeed possible to obtain ER because of its low free electron concentration.

Real Materials

In this section we consider real metals and the possibility of their being used as sample materials. Table 4.2 lists the physical parameters of some metals whose band structure has been calculated by various investigators. Since we would like the slope of the field to be as small as possible in the region where the field is nearly linear for the insulated case, we should consider metals with low carrier densities. This leads to those materials in the semimetal group: bismuth, antimony, and arsenic. Of these, bismuth is the best candidate, both

Table 4.2. Optical and Physical Constants of Selected Materials

	Ge	Bi	Sb	Cs	Fe	Ag	Au	Cu
n(a)	4.5	1.79	2.6	0.264	1.51	0.18	0.47	0.44
k	1.7	3.46	4.19	1.123	1.63	20.6	6.02	7.4
λ(μm)	0.6	0.6	0.6	0.578	0.589	0.589	0.589	0.65
$t_{1/2}$(Å)	195	96	26	283	199	15.8	54.0	48.5
E_F(eV)	--	9.97(b)	11.04(c)	1.58(d)	2.8(e)	5.5(d)	5.5(d)	5.5(d)
n_0 (cm^{-3})	2.4E13(d)	2.75E17(f)	5.54E19(f)	0.91E22(d)	1.7E22(e)	5.76E22(d)	5.90E22(d)	8.50E22(d)
d(Å)	2442	366	27.1	.80	0.78	0.59	0.58	0.55
α(m^{-1})	4.095E6	2.735E7	3.69E8	1.240E10	1.283E10	1.695E10	1.724E10	1.818E10
β(m^2/V)	2.305E-12	2.011E-16	9.98E-19	6.507E-21	3.254E-21	9.603E-22	9.375E-22	6.507E-22

$t_{1/2}$ = depth at which light intensity is reduced to ½ surface intensity

$= (\lambda_0 \ln 2)/(4\pi k)$

d = screening length $= \left(\frac{2\varepsilon_0 E_F}{3e\rho_0}\right)^{1/2} = \left(\frac{\varepsilon_0 k_B T}{e\rho_0}\right)^{1/2}$ for Ge

E_F = fermi energy $\alpha = 1/d$

n_0 = free electron concentration $\beta = \varepsilon_0/\rho_0 = \varepsilon_0/n_0 e$

(a) Optical constants from Ge to Fe are from AIP (1972); those from Ag to Cu are from Chemical Rubber Company (1961)

(b) From Golin (1968)

(c) From Falicov and Lin (1966)

(d) From Kittel (1971), p. 248

(e) From Sze (1969)

(f) From Dresselhaus (1971)

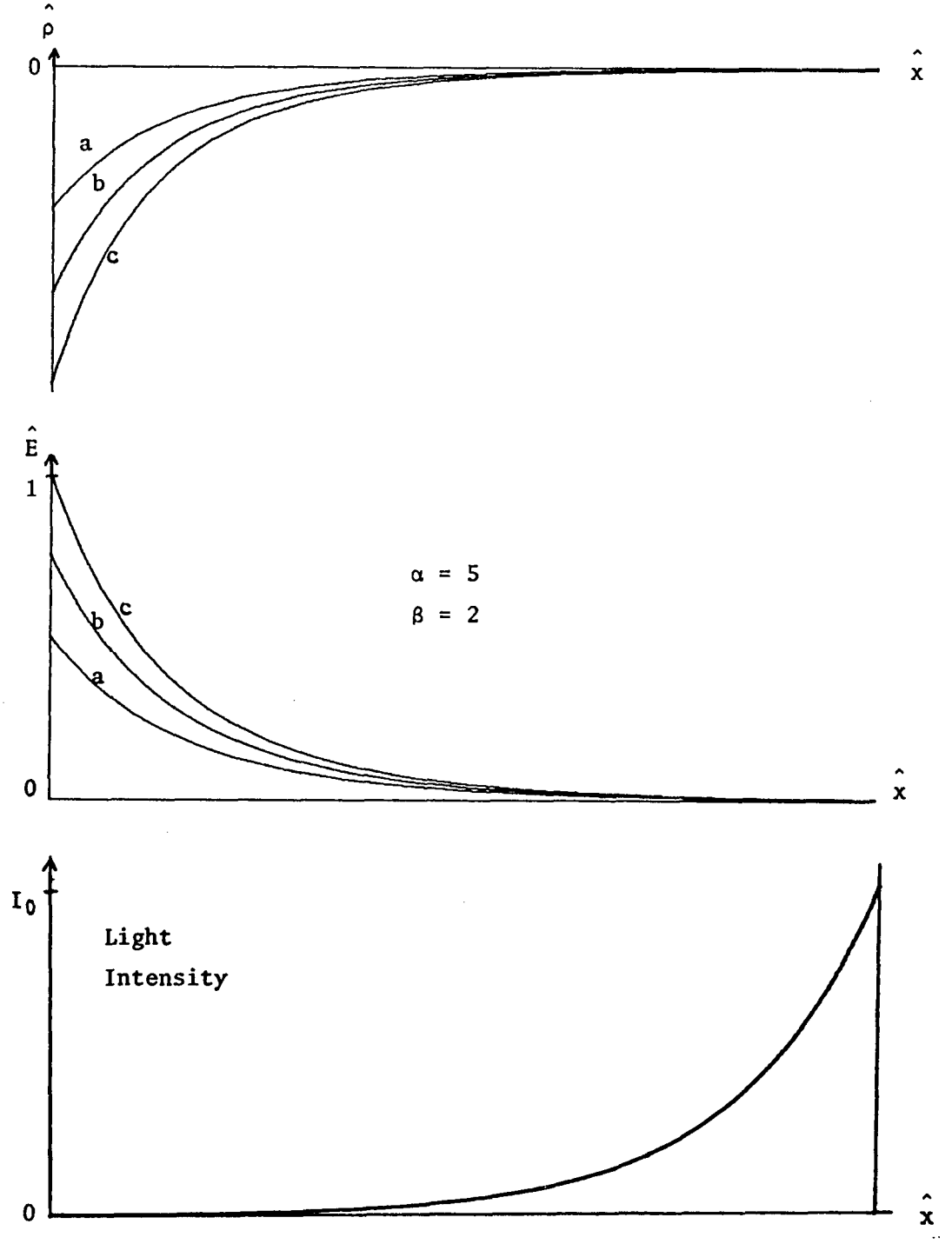


Fig. 4.9. Profiles of the Electric Field, the Charge Distribution, and the Light Wave for a Grounded Sample

from the theoretical and from a practical standpoint. It has the lowest free electron density and its band structure has been calculated (Golin, 1968). Experimentally, it is easy to evaporate and it forms good thin film samples. In the linear field region of an isolated sample we calculate for dE/dx the value of 5.4×10^{11} V/cm^2. If the applied field is 10^6 V/cm, we can immediately get a rough estimate of the penetration depth of the field by assuming that the linear portion extends all the way to zero field. We then find that the profile reaches to a depth on the order of 200 Å. This is twice as deep as the depth at which the light is attenuated to one half of its surface value. Since the field never actually gets to a zero value in the insulated case, our assumption is pessimistic, so the result is on the conservative side. Thus, it would seem that the semimetals, and bismuth in particular, would be appropriate materials with which to start.

The numerical solution was evaluated for several thicknesses of a bismuth film for both the insulated and the grounded cases. The results for a 1000 Å thick Bi film are shown in Fig. 4.10. It is seen that the electric field on the right hand side for the insulated case remains relatively large in the 200 Å thick region probed by the light wave.

The linear approximation suggested above can be applied to the true metals as well; in this case the penetration depth in gold is on the order of 0.1 Å. Thus it is seen that it will be very difficult to get an electric field inside such a material. The Thomas-Fermi distance for the noble metals is about 0.5 Å, not in bad agreement with the rough approximation. Diffusion effects contribute to make the region wider.

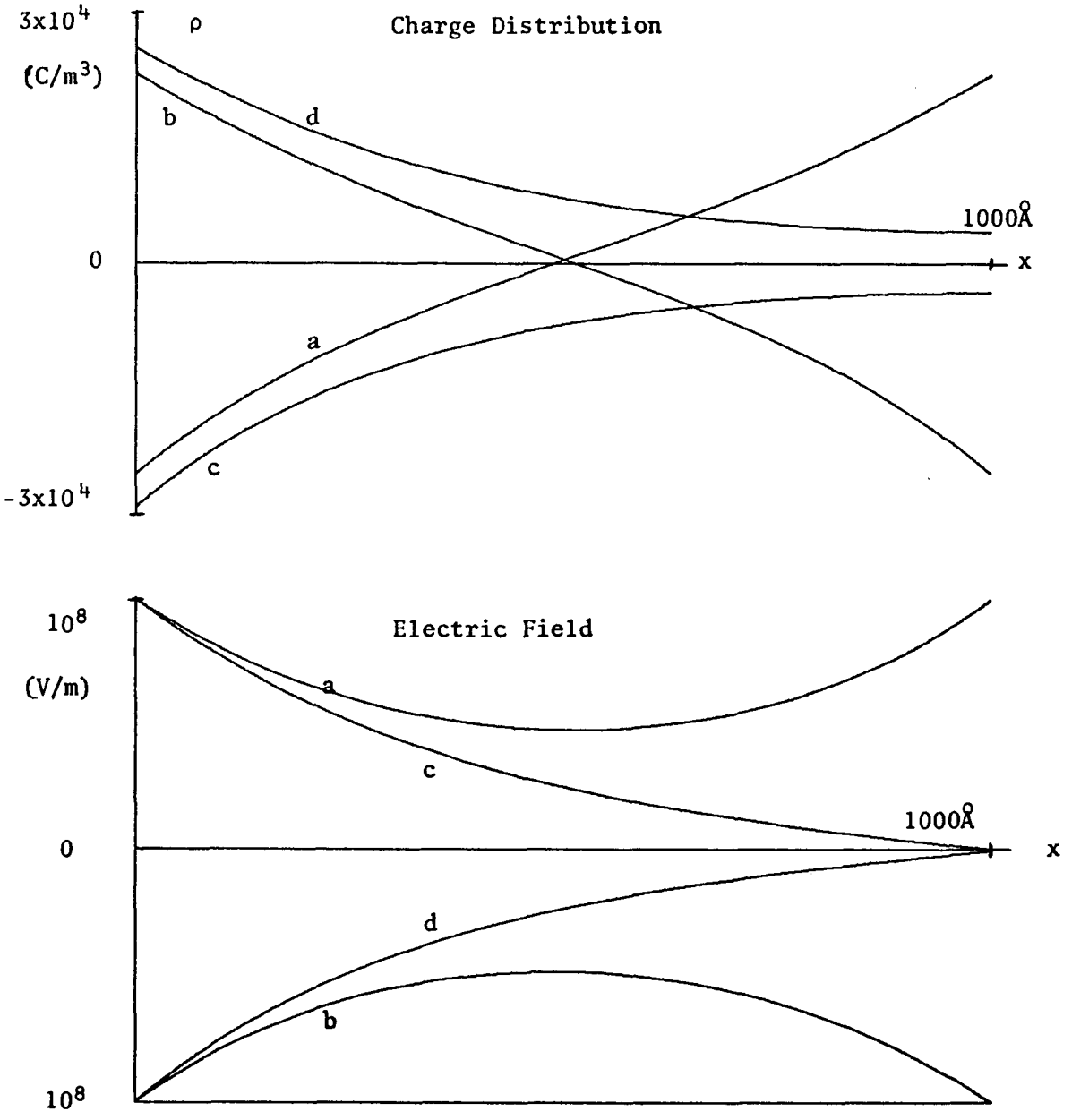


Fig. 4.10. Field Profiles in a Bismuth Film

Numerical integration using the parameters for gold confirms this; in fact, the field drops to 10% of its initial value within 1.5Å, and to 1% within 3Å. On the opposite side of the sample, the behavior of the field is the same, since we are in a situation where β is negligible and we have the symmetric solution given by the hyperbolic cosine.

ER of Bismuth

It should be possible to observe ER spectra for 500Å thick bismuth films even though they are not insulated. In this method, the free electron effects would be eliminated by ensuring that both halves of the sample on the ceramic are fully depleted of free electrons. It would only be possible to do this with one polarity of the electric field; but since ER is a quadratic effect, this should not be objectionable.

In terms of the normalized equation, the proposal is to make $\hat{\beta}$ equal to unity on one half, while on the other half it would be made closer to 45. In the first case, the field would just deplete the free electrons, while in the second, a large field would exist with the same charge density in the metal. Comparison of the two reflectances would lead to seeing the change introduced in band structure by the large differences in the electric field.

In terms of real numbers, a fully polarized ceramic's field of 3×10^{10} V/m would be attenuated to 2.975×10^{10} V/m at the free side of the film. On the other half, the ceramic needs to be polarized to only $0.25 \mu C/cm^2$; this will give rise to a field of about 2.5×10^{8} V/m which

will just be completely attenuated by the charges in the bismuth film. This is shown in Fig. 4.11.

The large difference in the electric field between the two sample halves is of the order of magnitude previously discussed for observing interband transitions in metals. It would be necessary to eliminate the poling electrode for this experiment since the equations show that very thin layers of metals with large free electron densities can obliterate the field. It would not be possible to use the bismuth directly as the poling electrode since this would produce different strains in each half, and this would interfere with the ER response.

Cardona and Greenaway (1964) measured the static reflectance of bismuth and their static spectrum is reproduced in Fig. 4.12. The band structure of bismuth was calculated by Golin (1968) and his results are reproduced in Fig. 4.13. As a result of these studies, it becomes easier to estimate where to look for structure in the ER spectra. It is expected that features in the static spectrum will be reflected in the modulated ER spectrum, although the inverse is definitely not true. On this basis, it is therefore predicted that strong structure should be found in the near infrared region around 0.75 and 1.0 micrometers, and in the violet at about 4000 Å.

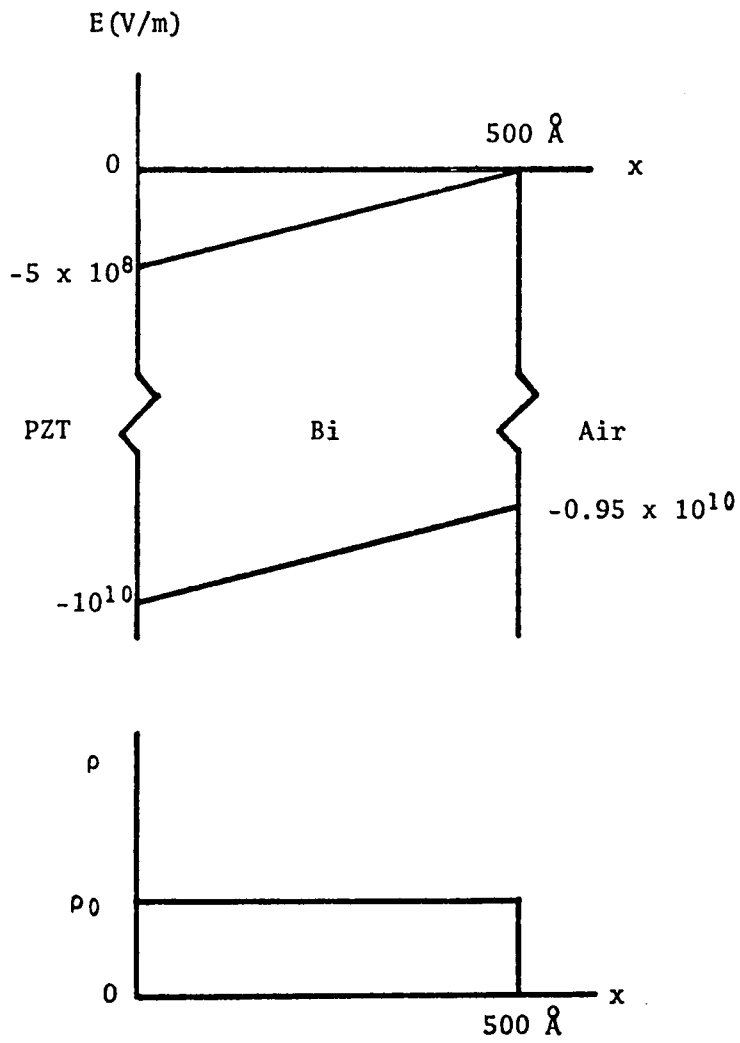


Fig. 4.11. Field Profile and Charge Distribution for a Fully Depleted Bismuth Film

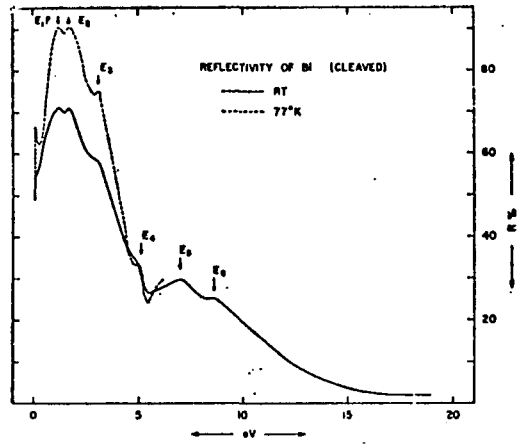


Fig. 4.12. Reflectance of a Cleaved Single Crystal of Bismuth

(From Cardona and Greenaway, 1964, Fig. 6)

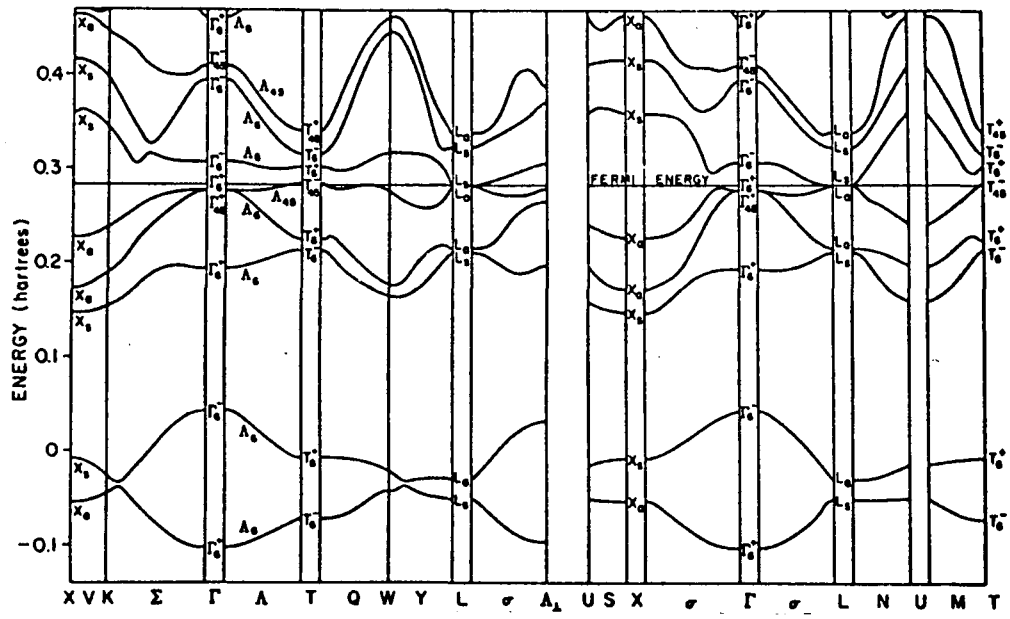


Fig. 4.13. The Calculated Band Structure of Bismuth

(From M. S. Dresselhaus, 1971, Fig. 1)

CHAPTER 5

EXPERIMENTAL APPROACH TO THE ER OF METALS

In this chapter, the experimental aspects of the measurement of the electroreflectance of metals are considered. The feasibility of constructing various sample configurations which attempt to approximate the models presented in Chapter 4 are discussed. The techniques of preparing the samples are then covered, followed by a description of the construction of the instrument which measures small differences in reflectance (reflectometer) and its electronic and optical systems. Finally, the results obtained with this apparatus are compared to published results of other similar apparatus.

Material Choice and Sample Configuration

The results obtained in Chapter 4 are used as a guide in selecting materials and in designing the sample. The choice of material and the experimental arrangement are interrelated since some materials are incompatible with the preparation of some sample configurations. Therefore, these two aspects of the ER of metals have to be considered together.

Actually, a choice of two materials is necessary. The first material used should be one whose ER response is relatively large and well known. This sample will not be a metal since metallic ER responses

are still rather indefinite as discussed in Chapter 3. This material is used to calibrate and to "fine tune" the apparatus, since the wavelengths at which a response should be seen are known. The second material is a metal whose ER spectrum is unknown; its selection is such as to simplify the observation of the spectrum.

We have previously seen that the difficulty involved in observing the ER spectrum of a metal is in a sense proportional to its free carrier density ρ_0. It is prudent, therefore, to select materials with low carrier concentrations. A semiconductor is adequate for the calibration phase, while a semimetal is a suitable starting point for the metal studies. With these thoughts we now consider the experimental arrangement.

The difficulties encountered with the electrolytic method were discussed in Chapter 3. This method was not considered for this study because of the indefinite results of the extensive studies already made by other researchers. A popular technique, called the "Seraphin method" (Cardona 1969, p. 206), has been used extensively for the measurement of the ER of semiconductors. This uses a field effect sandwich structure and was first described by Seraphin (1964 and 1967). Unfortunately, two drawbacks, which stem from the fact that the grounded boundary conditions hold in this method, prevent the use of the technique for most metals: (1) The results of the numerical integrations show that the electric field is shielded within a few Ångstroms of the surface. (2) The free electron effects are not eliminated. The two effects can be

overcome by completely insulating the sample from its conductive surroundings.

A self-supporting thin metal film (Hunter, Angel, and Tousey, 1965) placed in an electric field in a perfect vacuum would be a close approximation to the mathematical model (Fig. 5.1). Unfortunately, such a film is in unstable equilibrium in the electric field, since any unevenness in the film would allow charge to accumulate at that point. This would cause large forces to act at those points and thereby cause even larger distortions. The result is an avalanche effect which would lead to a rupturing force on the film.

Replacement of the vacuum by a solid insulator produces a configuration which encapsulates the metal as shown in Fig. 5.2. This structure which resembles a "ravioli" is reminiscent of the Seraphin method, except that now the sample is electrically "floating." The electrodes are placed directly on the insulator surfaces; this structure can be constructed in a straightforward way using standard thin film evaporation techniques. The film thicknesses are made thin enough to preclude optical interference effects in the wavelength region of interest.

Many benefits accrue from such a structure, although there are some drawbacks as well. The worst of the drawbacks is the fact that the field is now limited by the physical parameters of the insulating film. Thin films of SiO_2 and Al_2O_3 are about equal in this regard in that the dielectric breakdown strength of each is in the range of 10^6

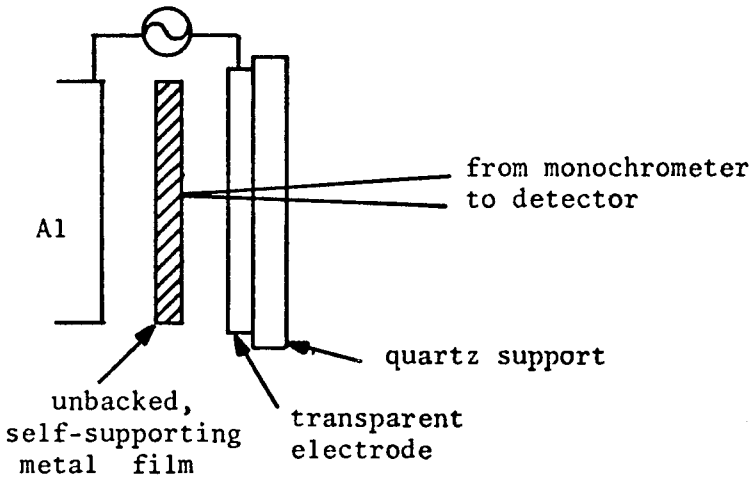


Fig. 5.1. Self-Supporting Film ER Configuration

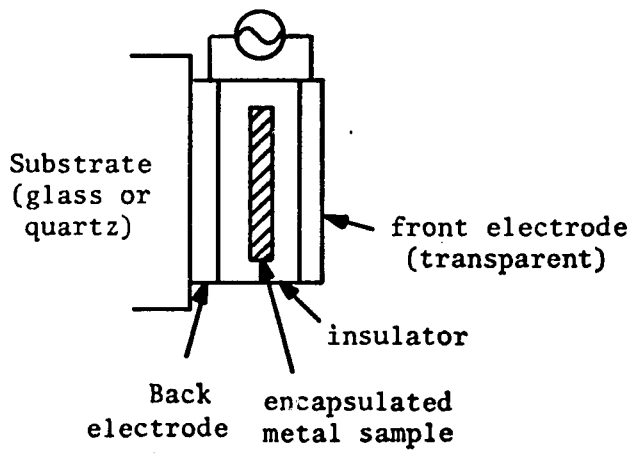


Fig. 5.2. Encapsulated Metal Film Configuration

to 10^7 V/cm. SiO_2 is by far the easier to evaporate and is the choice for that reason.

This configuration also closely approximates the mathematical model with the insulated boundary conditions discussed in the previous chapter. It is easy to construct and since it can tolerate elevated temperatures, this allows the use of germanium as a sample material. A thin film of Ge has to be heated to at least 450°C to convert it from the amorphous phase to a polycrystalline one. The ER response of Ge has been extensively studied and its spectrum is well known (Seraphin and Hess, 1965). Such a material would be very useful to fine-tune and to calibrate the apparatus before making measurements on a material whose ER response is totally unknown. Insulator breakdown would be readily apparent with this structure since the modulating voltage can be monitored with an oscilloscope. When this occurs, the metal film is no longer electrically isolated and the field in the metal is difficult to calculate, since the boundary conditions are uncertain.

A third configuration involves the deposition of the sample film on a ferroelectric substrate. This is similar to the approach taken by Ishibashi and Stadler (1969); in this case, however, the polarization of the ceramic is not changed during the measuring process. The basic structure of this type of sample is given in Fig. 5.3. This approach is an attempt to take advantage of the surface charge produced through the polarization of the ferroelectric. The ceramic PZT-5H has a polarization charge density of about 30 μCoulomb/cm^2. Let us assume

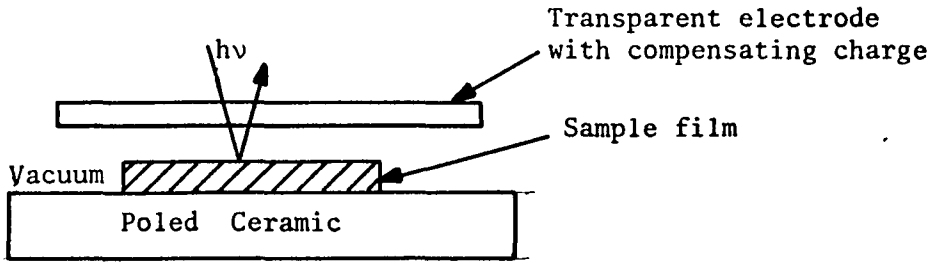


Fig. 5.3. Ferroelectric Substrate and Sample Film

for the moment that we can instantaneously polarize an unpoled ceramic (e.g., by using a noncontacting electrode) and also keep it insulated. An electrical surface charge of 30 $\mu C/cm^2$ would appear at the surface and give rise to an electric field just outside the ceramic of σ/ε_0, or about 3×10^8 V/cm (Reitz and Milford, 1967). This is one to two orders of magnitude larger than that possible with the ravioli structure discussed previously. The ER response scales as the square of the applied electric field in the low field regime. The present approach would then produce a response 10^2 to 10^4 times larger than what is possible with the ravioli configuration.

The 30 $\mu C/cm^2$ of charge on the surface of the ceramic tend to attract an equal and opposite charge to the surface. A 1000 Å thick film of a semimetal such as bismuth has a free electron density of about $10^{17}/cm^3$. If all of these electrons are concentrated at the surface of the film which is closest to the ferroelectric substrate (clearly an unlikely situation when Coulomb repulsion, diffusion, and the ion cores left behind in the film are considered), the electric field is not completely compensated by this layer of charge, and the field penetrates into the film. A 1000 Å high column of Bi film above a square centimeter of ceramic can supply at most $10^{17}/cm^3 \times 1cm^2 \times 1.6 \times 10^{-19}$ C x 1000Å ≈ 0.2 μCoulomb of charge. This is only 0.6% of the charge required to shield the field, so a very strong electric field penetrates through this sample. The same calculation made for a noble metal film which is 10 times thinner than the previous one and which has a free electron density of about $10^{23}/cm^3$, gives $10^{23}/cm^3 \times 1cm^2 \times 1.6 \times$

10^{-19}C x 100Å ≈ 2 x 10^{-2} Coulomb of charge. It now takes only 0.1% of the charges in this rather thin film to completely shield the field. It is therefore apparent that it will be very difficult to observe an effect for the noble metals with this type of experiment.

As discussed above, the ferroelectric ceramic substrate approach provides the largest electric field. The next aspect to consider is the feasibility of constructing such a sample. One source of difficulty is that a ferroelectric material is also piezoelectric so that the process of polarizing the substrate will produce large lateral surface strains. This was ignored in the discussion above, but since unambiguous results are desired, differential strains in the film must be avoided. The solution to this is to deposit the film after the substrate has been poled.

With one half of the substrate poled and the other half unpoled and by sampling alternate halves, the change in the reflectance of the sample film under the influence of the electric field can then be measured. The details of an apparatus which measures this will be given below.

One configuration which was tried for the polarization of the ceramic was the evaporation of a very thin film of chromium in two sections on the substrate. The choice of Cr was based on the fact that it forms a conductive film when its thickness is on the order of 20 Å. It was hoped that this thickness of film would not be large enough to shield the field. The Cr film was used as an electrode to pole one half of the substrate and then a sample film was evaporated on both

halves. The boundary conditions of this configuration are those of the grounded film; this is the less desirable of the two cases.

Sample Preparation

Two types of samples were prepared in this study and the preparation of these is described in this section.

Sample and Oven Construction

The samples deposited on the ceramic substrate were prepared in a vacuum chamber. An oven is needed to hold the ceramic and to heat it to at least its Curie temperature. The particular ceramic used is known as PZT-5H and has a Curie temperature of 193°C. It is necessary to make electrical contact with the heated ceramic in order to measure the film resistances and to apply the poling voltage. These requirements are met by the oven outlined in cross-section in Fig. 5.4.

The ceramic substrates were approximately 1 x 2 x 0.2 cm in size so that two areas about 1 cm^2 served as the poled and the unpoled halves. The side of the ceramic onto which the film was to be deposited was polished, while the opposite side of the slab had a silver film fired into it. Small conductive pads were placed in the four corners of the polished side in order to make contact between the evaporated film and the oven contacts. These pads were made in either of two ways; they were, a. painted on with silver paint; b. made of relatively thick evaporated Cr films. External contact is made to these through the use of small stainless steel set screws which are part of the oven assembly.

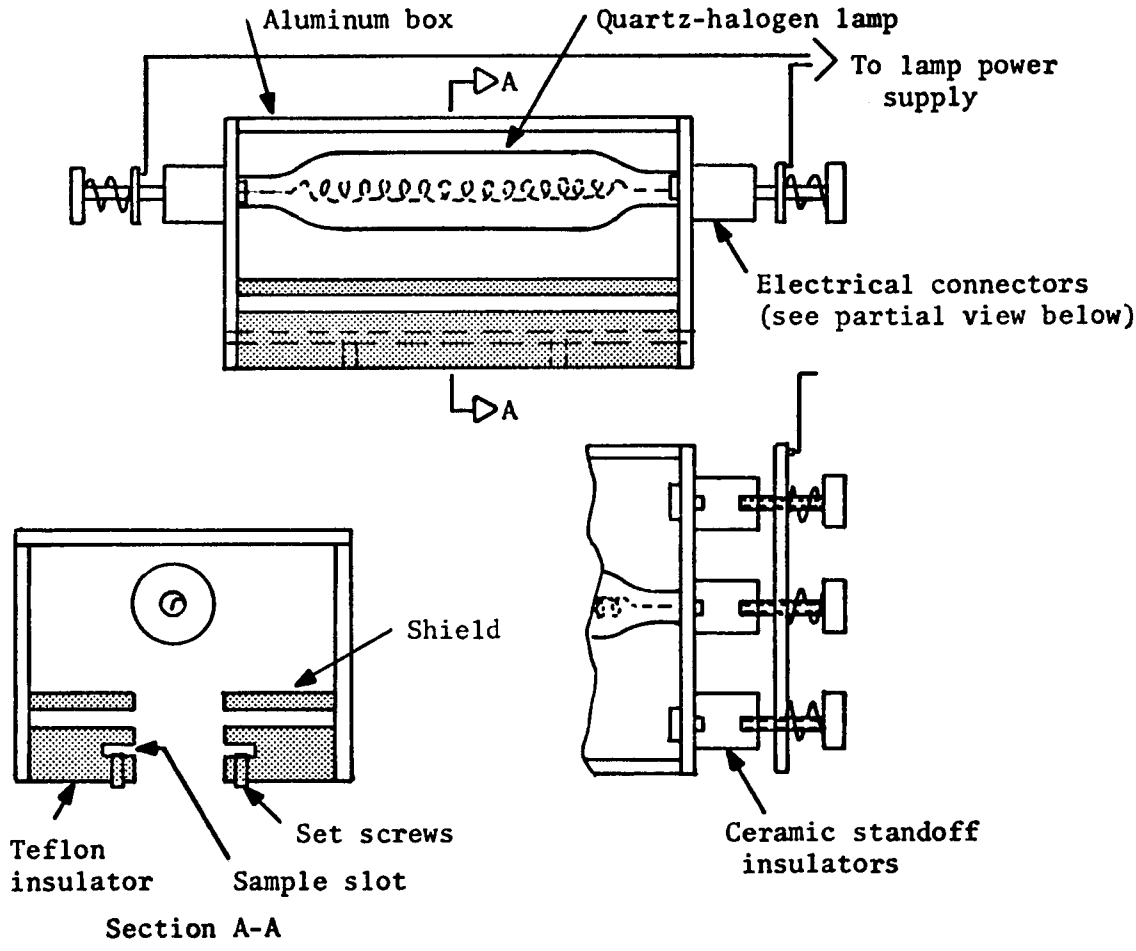


Fig. 5.4. Oven Construction

An iron-constantan thermocouple is wedged in a space between the teflon holder and the ceramic to measure the temperature. It is completely shielded from the direct rays of the heater lamp and is at what is one of the cooler points. We are thereby assured that when the indicated temperature is above the Curie point, no part of the ceramic is below the Curie temperature. The thermocouple and meter circuit was checked at 0°C and at 100°C and was found to be accurate to within a few degrees; this is sufficient for this use.

The lamp used as a heater was a 45 watt quartz iodine unit which was intended for operation at 6.5 volts and 6.6 amps. A variable autotransformer operating from the power lines supplied the primary of a 6.3 volt filament transformer with a variable voltage. In this way the heater power could be varied so that a maximum temperature of about 230°C (indicated) could be reached.

The poling voltage could be transmitted into the vacuum chamber through either of two connectors. One was a standard high voltage feedthrough for operations such as glow discharges, so that several kilovolts could be used if necessary. The other feedthrough was a multipin ceramic header through which up to about 300 volts per pin could be passed into the vacuum chamber. The voltage was applied while the sample temperature was above the Curie temperature; the heater power was then turned off while the high potential remained on until the sample cooled to room temperature.

If a film is used to pole the ceramic, it has to be conductive yet be as thin as possible. The reason for the latter is that we want

to minimize any effects this film might have on the sample film. During the evaporation process the resistance of the deposited film is monitored with an ohmmeter and the run is terminated when a resistance of several megohms is reached. It is necessary to monitor the resistance for a while after the evaporation is completed to ascertain that the film remains sufficiently conductive after it has cooled.

The ceramic substrates were first cleaned with demineralized water and detergent, followed by successive washes in acetone, absolute ethanol, and finally distilled water. No glow discharge was used. This process, though crude by some thin film technique standards, proved adequate for our needs. As an example of the adhesion obtained between the chromium film and substrate, it was sometimes necessary to resort to a polishing compound to remove a film from the ceramic after we were through with a particular sample.

The other sample is as shown in Fig. 5.2. The substrate is a 1 x 3 inch microscope slide; it is cleaned in the same manner as the ceramic substrate. The bottom electrode was an aluminum film, although any good conductor is suitable. Next, a layer of SiO_2 was evaporated so as to cover most of the Al layer except for a small area which was used to make electrical contact. A mask was then used to limit the extent of the bismuth sample film; in fact, five sample areas of about 0.5 by 1.5 cm were made on one slide. An overcoat of SiO_2 was then evaporated, followed by the top electrode. This latter film was a transparent layer of Cr evaporated through the mask used previously

to define the active areas, except that it had slightly enlarged apertures so that electrical contact could be made without covering any part of the active sample.

Material Evaporation

All materials were evaporated by resistively heating them in a vacuum of about 10^{-5} torr until the desired thickness was deposited on the substrate. Initially, alumina-coated tantalum boats were used for the evaporation of Cr, but this proved to be unsatisfactory, since the temperature of the Cr could not be raised high enough for it to evaporate at a reasonable rate. Later, tungsten rods with a chromium electroplate were obtained from commercial sources and subsequent chromium evaporations with these sources were very successful.

Germanium was easily evaporated from tungsten dimple boats, although the two materials alloyed and after a few runs the boat had to be replaced. Due to its lower melting and boiling points, bismuth was even easier to evaporate than germanium and it did not alloy with the boat.

No monitor equipment was used to control the evaporation of the Bi or Ge. The substrate was always cold and since we desired films which were just opaque, we could judge when to stop the evaporation by watching the film deposited on the bell jar. A witness substrate was usually placed where part of the supporting structure would cast a shadow on it. After the chamber was opened, this test slide was placed in another chamber where an overcoat of aluminum was evaporated onto

the entire slide. The step formed by the edge of the shadow could then be measured with a Fizeau interferometer.

The evaporation of silicon dioxide was done in a third chamber where it was possible to monitor optically the film thickness. This system, which consisted of a bell jar eighteen inches in diameter and a 6" diffusion pump, had an ion gauge pressure measuring system along with a needle valve which allowed us to maintain manually a residual pressure in the chamber of about 10^{-4} torr. This pressure is required in order to get good films of SiO_2, rather than a mixture of SiO and SiO_2 (Hahn, 1973). The monitor system measured the reflectance from a sample slide and displayed it on a strip chart recorder. The reflectance of the monitor slide is not only a function of the index of refraction of the incident medium and of the glass but also of the index and the thickness of the film deposited on it (Heavens, 1965). The value of the extrema which appear at a given wavelength as a function of film thickness is related to the difference between the indices of the substrate and the film. When the two indices are nearly equal the extrema are weak and broad. This is the case with silicon and glass ($n = 1.45$ and 1.52, respectively). The monitor slide was therefore precoated with a quarter wave optical thickness film of zinc sulphide ($n = 2.3$). This formed a two layer interference film stack which was much more sensitive to the thickness of the silicon dioxide film. This made the point at which to stop the evaporation much easier to determine. A wavelength of 5461 Å was used, thus the quarter wave layers had an optical thickness of 1365 Å.

Ferroelectric Studies

The original samples of ferroelectric ceramic were obtained through the courtesy of Dr. Cecil Land of Sandia Corp. Later, samples of the same composition, known as PZT-5H, were obtained commercially from the Vernitron Corp. of Cleveland, Ohio. The manufacturer gives 50 V/mil (20kV/cm) as the room temperature dielectric breakdown strength for this material. This drops to about 35 V/mil (14kV/cm) at 100°C. Jaffe, Cook, and Jaffe (1971) state that it is generally necessary to use a voltage close to the breakdown voltage in order to pole a ferroelectric material. Since the samples used in this study were about 2mm thick, this required the application of a potential of about 4 kilovolts at room temperature, and about 3 kV at 100°C. No data was available for higher temperatures, and in particular, at the Curie temperature. An attempt was made to apply about 2kV at a ceramic temperature of about 200°C; the results were catastrophic and the sample was completely ruined. The currents which flowed in the film when the breakdown occurred were very apparent since the film was vaporized in a fern-like pattern. The capacitance of the structure was approximately

$$C = \frac{A\varepsilon}{d} \approx \frac{(1\text{cm}^2)(3000\varepsilon_0)}{2\text{mm}} \approx 1.4 \text{ nanofarad.}$$

The charge stored by this capacitor at 2000 volts flowed through the point of breakdown, and since the circuit had very little resistance

or inductance, the power density in the arc was enough to melt parts of the ceramic.

The question was now raised as to what voltage was required to polarize fully the ceramic at the Curie temperature. It was decided to try to measure the polarization charge during thermal depoling. A Keithley model 610B electrometer which could measure up to 10μ coulomb full scale with less than 10^{-14} coulomb/second drift was used. Since the expected charge was

$$q = A\sigma \simeq (1\ cm^2)\ (30\mu C/cm^2) = 30\mu\ Coulomb,$$

it appeared as though it would not be possible to measure the entire charge. A test of the electrometer showed that it indeed did have negligible drift. Thus it was decided that it would be possible to zero the instrument by first disconnecting the charge source and discharging the integrating capacitor in the electrometer. If this was done as it approached full scale, the charge source could then be reconnected and the measurement continued. The instrument effectively grounded the front surface of the ceramic; during the time it took to discharge the electrometer capacitor the potential built up on the surface of the ceramic, and it was desirable to rezero as quickly as possible. It was generally possible to do this in 2 to 10 seconds, while the overall discharge operation took on the order of 30 minutes, thus making negligible the error from leakage due to a potential appearing on the surface. In this way, we were able to measure reasonably reproducibly the polarization charge in spite of the fact that the literature states otherwise (Fatuzzo and Merz, 1967).

A second way was used to ascertain that the ceramic was polarized. The capacitance of the plane parallel capacitor formed by the sample configuration was measured with a bridge. The values of the dielectric constant could then be calculated for both the poled and the unpoled sample halves. These values agreed well with those supplied by the manufacturer of commercial pieces as shown in Table 5.1. It is not surprising to see a difference between the two values in view of the fact that these ceramics are formed by hot pressing in special atmospheres and therefore variations in the exact process used could result in the differences. Another source of error is the fact that our configuration is not an ideal plane parallel capicator.

Figure 5.5 shows the results obtained when the polarization surface charge was measured as a function of the poling voltage. In every case, the maximum temperature was well above the Curie temperature before the voltage was applied and the heater power turned off.

Table 5.1. Relative Dielectric Constant of PZT-5H

	poled	unpoled	ratio
average of measured values	3803	2537	.666
commercially available values	3400	2040	.60

Not all of the data points at 300 V are shown since experimental difficulties caused the discarded points to fall well below what they should have been. The cause of this problem was apparently a poor

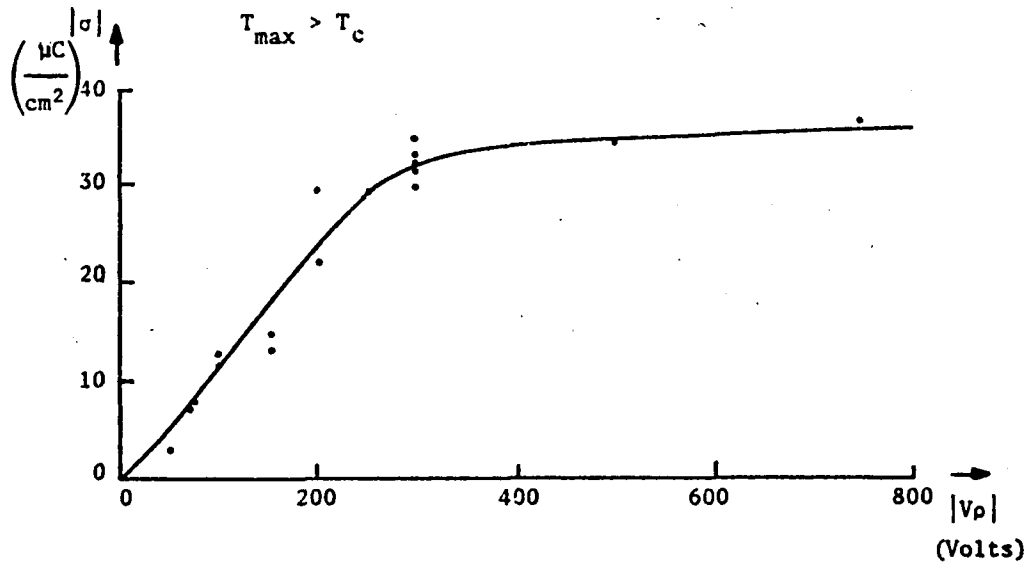


Fig. 5.5. Polarization of 2mm Thick Sample of PZT-5H vs. Poling Voltage

contact, most likely that between the pads on the ceramic and the contact screws in the Teflon holder. The point of this study was to determine the minimum poling voltage for which we could obtain a satisfactory polarization; the answer can be obtained from the curve and is about 300 volts.

The curve in Fig. 5.6 gives the result of an experiment which varied the maximum temperature of the sample at which point the poling voltage of 300 volts was applied. In all of these cases, the heater power was turned off promptly after the voltage was applied, and the voltage was removed as soon as the sample reached room temperature. Undoubtedly, time enters into these curves, i.e., if the heater had been left on for a longer period after the voltage was applied, it is likely that a larger polarization would have been measured. Though such a study could prove to be interesting, it was not the subject of the present project, so once the pertinent answers had been obtained, the subject was not pursued further.

The stress effects which arise from the piezoelectric effect which is always present in ferroelectric materials could be seen plainly on the surface of the ceramic. Since only certain discrete regions were actually polarized, the boundaries between the poled and the unpoled areas were marked by a rise or swelling in the surface topography, so that the poled areas were elevated above the unpoled regions. Since material is conserved, there is a compensating constriction in the plane of the surface. This gives rise to surface

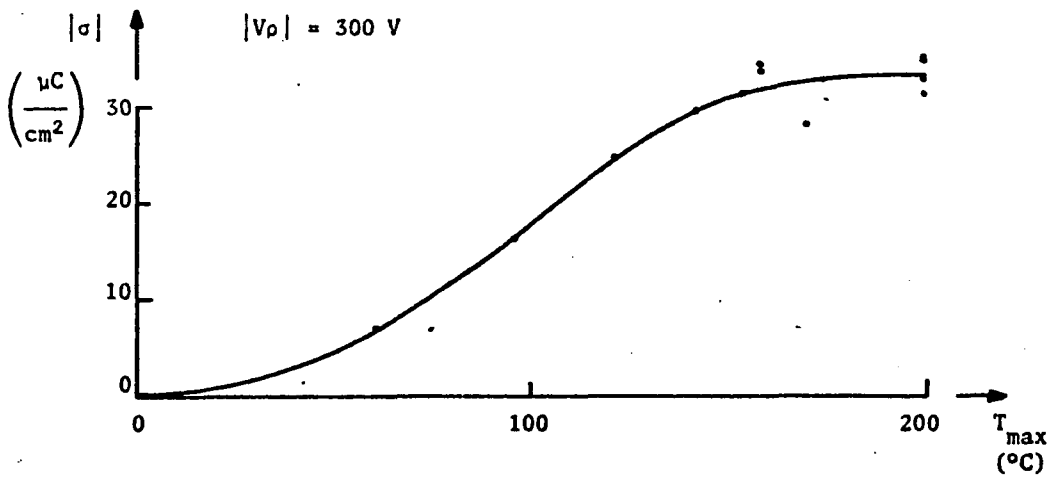


Fig. 5.6. Polarization of PZT-5H vs. Poling Temperature

strains and ultimately to the piezoreflectance effect which can confuse the study of the electroreflectance effect.

Optical Configuration

In this experiment we are comparing the difference in reflectance between two states of the sample, one with and one without an electric field. For the case of the ceramic-based sample, this has to be done in a spatially sensitive manner since the polarization of the substrate is not going to be changed during the measurement process. This type of measurement is known as "spatial modulation" (McNatt and Handler, 1969) and there are two basic ways in which this can be accomplished: 1) move the light beam from one side of the stationary sample to the other half; 2) keep the light path fixed and move the sample laterally in the beam so that the two halves are exposed alternately.

Moving Beam

The moving beam approach was tried initially by using a chopper wheel situated immediately next to the exit slit of a monochromator. Hummel, Dove, and Holbrook (1970) tried a similar system for their compositional modulation studies, although they later changed their system to use a vibrating mirror to move the light beam (Hummel, Holbrook, and Andrews, 1973). It alternately opened the top and bottom halves of the slit; the slit was imaged on the sample so the result

was to illuminate sequentially the two halves of the sample. The reflected beams were then directed to the photocathode of a photomultiplier (PMT).

The basic difficulty with this approach is that the two beams could not be precisely recombined on the photocathode surface. The result of this is that while the intensity difference between the two beams can be balanced to a null at any given wavelength, this zero balance changes as a function of wavelength. The null sought here is on the order of one part in 10^5 in the measurement of $\Delta R/R$. It was found impossible to adjust the system so that the zero point did not change for even a limited spectral range of a few hundred Ångstroms. Results of studies (Birth and DeWitt, 1971; Ballik, 1971; and Sommer, 1973) of the spectral as well as the spatial (areal) response characteristics of photocathodes show that the variations in these parameters within a particular unit can be on the order of tens of percent; the null sought here is many orders of magnitude smaller than that. An adjustable vee-shaped mirror, a ground quartz diffuser, and a mask over the photocathode were all tried in an unsuccessful attempt to minimize the effects of the nonuniformities. As long as the two beams are not coaxial, there will always be a keystone shape in the beams as they are recombined; it is not possible to make them coaxial with all-reflective optics.

An all-reflective optical system is desirable in order to have the capability to operate over a large spectral range, from the near

infrared to the near ultraviolet. The simplest way of covering this wide region of the spectrum is to use mirrors. No problems are thereby incurred from surface reflection losses, absorption in the elements, or variations in the focal point with wavelength.

Another problem was associated with the use of the chopper wheel to open the upper and the lower halves of the monochromator slit. It was desired to minimize the number of transmissive elements in the optical system, so the wheel was an aluminum disk with slots machined in it. Even though great care was used, the hole spacings were not uniform, nor were the slots of the optimum shape. These problems combined to add a significant amount of jitter to the system. It is possible that a quartz disk with a photographically formed pattern would have been more successful in this regard, but the time and effort required to produce such a chopper did not seem to be warranted, especially in view of the other problems associated with this particular approach.

Moving Sample

In the second way of accomplishing spatial modulation, the optical path is kept constant and the sample is moved back and forth in the beam so that each half is sampled alternately. This motion can be obtained in either of two ways: a. The sample is mounted on a structure which is driven in a nonresonant manner. Such a system includes Beaglehole's method (1968) as well as an arrangement where the sample is attached to a loudspeaker coil driven by a hifi amplifier.

These have the advantage that the modulation frequency can be easily changed and optimized to avoid natural resonances of the supports.
b. The sample is part of a mechanical resonator and its oscillations provide the necessary reciprocating motion. This approach has the advantage that a much larger amplitude than that possible with a loudspeaker type of arrangement can be easily obtained. This in turn means that a larger surface area of the sample is used, thereby averaging out effects from dust and other imperfections on the sample.

The reflectometer described here used the approach of a moving sample mounted on a resonant mechanical oscillator. The ceramic substrate was glued to one tine of a tuning fork and the amplitude (peak to peak) obtained was about 0.5 cm at about 70 Hz.

Ideally with this arrangement there would be no motion of the light beam at the photomultiplier. However, if the sample is mounted so that its surface is not exactly parallel to the plane of vibration, a translation of the beam at the photocathode can occur. Similarly, if the tip of the tuning fork tine moves in an arc rather than stays in a plane, the beam will have an angular deviation and a consequent motion at the photocathode. This problem could be overcome to a certain extent if the pupil of the system could be positioned at the sample. Unfortunately, this is not possible becuase a sharp image of the slit is needed at the sample to allow discrimination between the two halves of the sample. The only other alternative, and that which was used, is to ascertain that the beam at the PMT is not an image of the sample.

The moving sample optical arrangement always has the light path reflecting from the same parts of each mirror, thus eliminating spurious signals which could arise if the reflectances of the mirrors were not uniform. Scattered light should be less of a problem with the moving sample arrangement since the light entering the reflectometer box is not modulated while that with the chopper wheel arrangement is. In the moving beam reflectometer any light which is scattered also has the proper modulation impressed on it for detection by the lock-in amplifier. In the former case, scattered light is not modulated (in general) and is therefore not detected. With the lock-in amplifier configuration used in this experiment, constant light levels or even wavelength independent modulated signals are tolerable so long as their level is not so great as to overload the amplifier circuits or exceed the offset capability of the lock-in amplifier.

The optical system will be described in more detail later. Suffice it to say here that with the moving sample and resonant oscillator system we were able to achieve the goal: to get a null balance which would stay constant as a function of wavelength over a significant range of photon energies.

Present Apparatus

The ER measurement system, as presently constructed, can be discussed as separate units such as optical, electrical, and mechanical. Figure 5.7 shows the inter-relationship between the optical and the electrical components. Each of these will be discussed in detail.

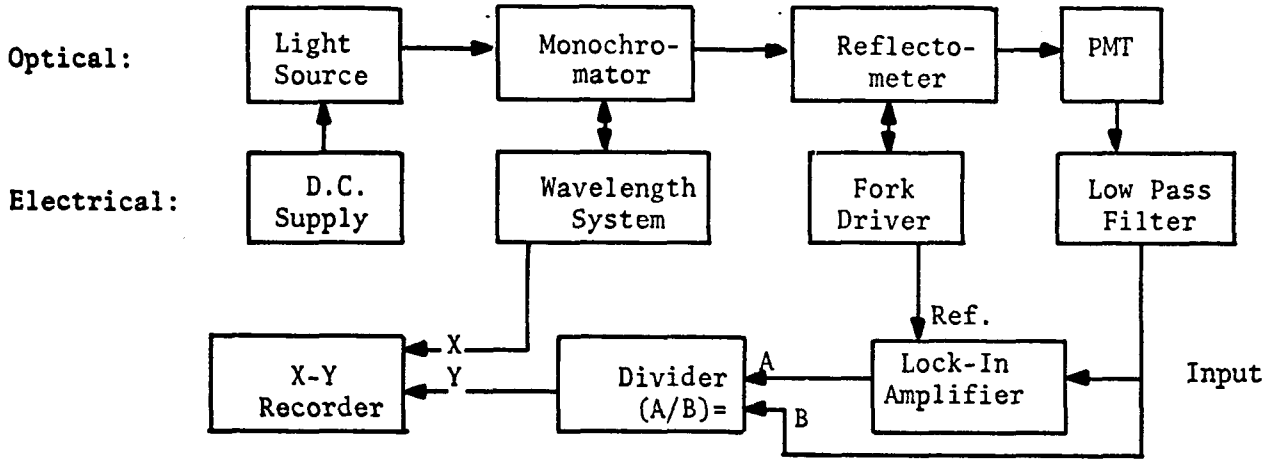


Fig. 5.7. Interrelationship of the Reflectometer Elements

Optical

Two basic light sources are used which provide coverage of the optical spectrum from the near infrared to the near ultraviolet (UV). For the infrared and the visible portions of the spectrum, a 100 watt quartz-halogen lamp is used. The power source can be either a regulated DC supply or a set of storage batteries. The batteries are electrically quieter than the regulated DC source hence these are used when an accurate spectral scan is made. In the UV, a 100 watt high pressure xenon or mercury-xenon arc source is used. The arc source has a much higher brightness than the tungsten source; consequently, we get more light energy onto the sample and this tends to increase the signal to noise (S/N) ratio. Unfortunately, the noise introduced by the arc fluctuations and the radiated spike noise from the lamp power supply tend to offset the gains made in the S/N ratio. It was therefore decided to use the arc source only in spectral regions where the filament source was not bright enough.

The monochromator does not have to have a very high resolution for the ER spectroscopy of interband transitions. The structure in such a spectrum is usually on the order of hundreds of angstroms wide, so that it is obvious that the resolution requirements are very modest. The monochromator used in this apparatus was a Jarrell-Ash f/3.6 Ebert grating instrument. Its nominal resolution with a pair of 100 micrometer slits was about 5 Å. Normally, a pair of 250 μm slits was used in order to increase the light throughput. The manufacturer gave no data on the

instrument's resolution when these latter slits are used, so the resolution was assumed to increase linearly and was estimated to be about 15 Å.

The optical arrangement is shown in Fig. 5.8. The location of the chopper wheel C is given for reference purposes, but it is no longer used. There are only two components which are not reflective: 1) the quartz lens Q is a collimator that focuses the source onto the entrance slit of the monochromator; and 2) the ground quartz diffuser D further reduces the spatial sensitivity of the photocathode. The lens does not form an especially sharp image of the filament on the entrance slit so chromatic aberration in this element is negligible. It will only alter the overall amount of light entering the monochromator, thereby changing only the average intensity of light reaching the detector. The quartz diffuser accounts for some light loss, but it does not change the observed signal.

The mirror M is a weak toric and could just as easily have been a spherical mirror. Its purpose is to image the exit slit of the monochromator onto the sample. The slit has an illuminated height of about 2.5 cm which is imaged onto the sample at approximately unit magnification. Since the active area of the sample is about 0.6 cm, a mask must be used to cut out the unused portion of the slit. In practice a pair of masks is necessary. One is placed at the exit slit of the monochromator and it cuts out the majority of light which would miss the sample. Due to the aberrations incurred by using the focusing mirror off axis,

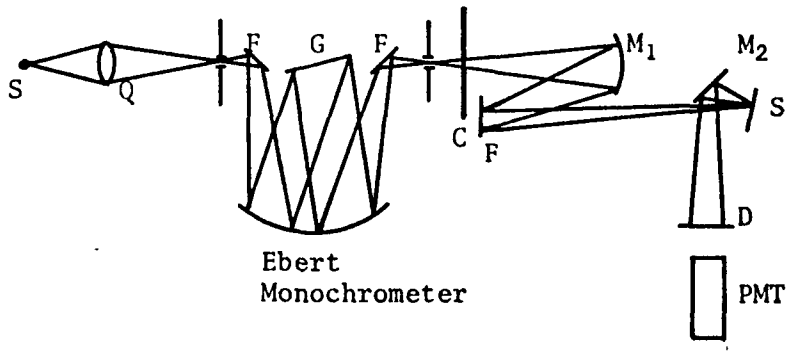


Fig. 5.8. Optical Layout of Reflectometer

the image on the sample is astigmatic. The tangential image is used, and thus the vertical extent of the image is not very well defined. A second mask is therefore placed immediately in front of the sample to define the vertical extent. This mask also serves to cut out any light which might be modulated by reflection from the tuning fork.

Studies were made of various ways of improving the optical efficiency of the system. The wasting of about three quarters of the light by masks appeared to be an area where a great deal of improvement could be made. An optical system which demagnified the exit slit by a factor of four seemed to be what was necessary, and the system might be able to eliminate the aberrations of the focusing mirror as well. Such a system would reduce the f/number of the beam from f/3.6 to f/0.9, and this was the problem. A fast system is more susceptible to focusing errors, and, if used off-axis it would require expensive optics. Furthermore, this created the problem of collecting the fast light cone reflected from the sample. The most likely system which resulted from these considerations required an ellipse whose cost turned out to be prohibitive (see Fig. 5.9a). Another system which uses a folding flat and a spherical mirror is shown in Fig. 5.9b. The only other solution was to try to find a light source with a smaller filament and the same power input, i.e., a source with higher brightness. Arc lamps were not considered because of the noise they introduce. Of the convenient incandescent sources, the quartz halogen one has about the highest brightness available. Thus the only alternative was to use a similar lamp with a lower power rating and a correspondingly shorter

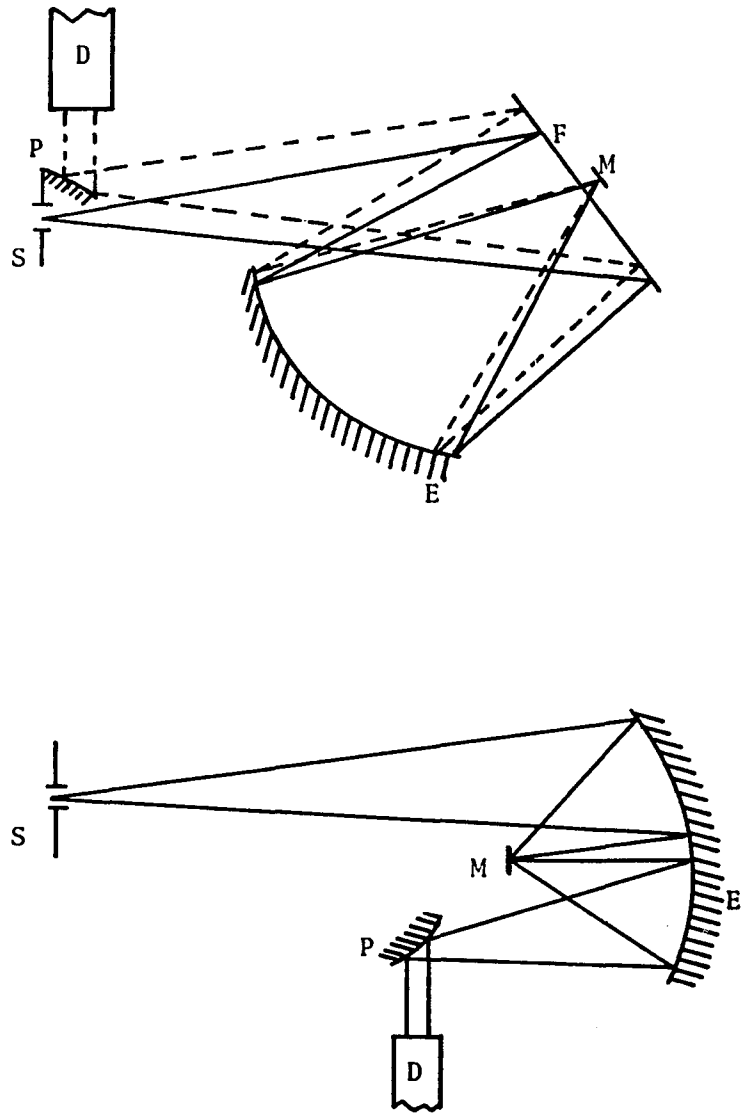


Fig. 5.9. High Efficiency Optical System for Reflectometer

S: monochrometer slit; D: detector; P: convex folding and collimating mirror; F: folding flat with hole; E: elliptical f/0.9 focusing mirror; M: metal sample.

filament and operate it above its rated power. This approach cannot yield a great increase in overall efficiency, but it seemed to be the best course of action, short of finding a continuously tunable laser which would cover the range of our interest.

Several types of detectors were considered and tested. The RCA 7265 phototube was the workhorse detector for the visible portion of the spectrum. We have already discussed the problems associated with the non-uniformity of the photocathode surface. A photomultiplier does have an advantage over almost all other detectors in that its gain is easily and continuously variable by the use of feedback circuitry. In this way, the D. C. output of the PMT can be kept constant, so that the A. C. output is proportional to $\Delta R/R$. This ratio, which is the quantity of interest (see Chapter 2), can also be obtained by using an analog divider circuit or module to divide a voltage proportional to ΔR by one proportional to R. The analog divider is not necessary with the use of a variable gain PMT circuit and the range of light intensity over which the circuit is linear is larger than that which is attainable with the analog divider. The use of the variable PMT supply effectively puts the division process ahead of the lock-in amplifier without introducing excessive noise. Such a system was not generally used in this experiment, and a plan is given in a later chapter for the implementation of such a system with off-the-shelf items at one half the price of a comparable commercial system.

The usual precautions were taken when using and handling the photomultiplier. It was magnetically shielded with mu-metal, was handled only in very subdued incandescent lighting, and was never exposed to strong light. Fluorescent lighting is especially detrimental since the ultraviolet light from these lamps causes an increase in the tube noise due to phosphorescence in the glass envelope of the photo tube (Radio Corporation of America, 1972). The electron transit time of the tube is small enough (about 50ns when the tube is operated at 800 to 1200 volts) so that the changes caused by varying the voltage did not affect the phase setting of the lock-in amplifier.

A second detector, the PIN diode, was tried in the visible and near infrared portions of the spectrum. The response of these diodes extends out to 1.1 micrometer in the IR and out to 0.35 μm in the UV. The noise figure (NEP) of these detectors is about 4×10^{-13} W/Hz$^{\frac{1}{2}}$ for a 1 cm^2 active area detector at 20°C, and their dynamic range extends from 10^{-14} watt to 10^{-2} watt of incident light flux. Furthermore, they can be exposed to room light with or without power applied to them with no detrimental effect. Unfortunately, they appear to have the same bad spatial spectral sensitivity characteristics as the photocathodes. Also, their gain cannot be adjusted as easily as a photomultiplier can, hence the divider circuit had to be used in this case.

A third detector used was a lead sulphide photoconductive cell operated at room temperature. Its spectral range is given in the literature as about from 1.0 to 2.5 μm for operation at 300 K (Wolfe, 1965, p. 474). In terms of range, it is seen that the two solid state detectors

complement each other very nicely. One serious drawback with the PbS detector is that it is extremely temperature sensitive and provision must be made to compensate in some way for the thermal drift. Though little use was made of this detector in this part of the experiment, thought was given to ways of surmounting this problem and these are given in another section. When the apparatus was used with this detector, the final folding flat which directed the beam to the detector was replaced by a short radius cylindrical mirror. This had the effect of focusing the beam in the vertical direction while not appreciably changing the horizontal size. Of course, the aberrations introduced by this were rather atrocious by diffraction limited standards; however, this helped spread the light over the entire detector so that the spot size was well matched to the detector size of about 4 mm diameter. At this point in the optical system, the optics only serve as an energy collection system; no image is formed; and aberrations are inconsequential.

Electronics

Figure 5.7 outlines the electronic circuitry involved in the system. Another mode is possible in which the photomultiplier is supplied by a variable output high voltage supply which eliminates the need for a divider circuit.

The basic element in attempting to detect the very small signals encountered in electroreflectance is the phase sensitive detector, also known as a lock-in amplifier. In a mathematical sense, such an

instrument can be described as a Fourier analyzer which is sensitive to both frequency and phase. In contrast to a tuned amplifier which responds to all components at a given frequency, regardless of phase, the lock-in amplifier needs a reference signal from which to obtain the phase and frequency information. The particular unit used in this experiment was an Ithaco, Inc., model 391. It operates on a heterodyne principle which makes it totally insensitive to any harmonics of the basic signal. This eliminates the need for a "front end" bandpass filter as used in the Princeton Applied Research Corp. model 124 and thereby increases its dynamic range and its capability to reject unwanted components in the input signal without overloading. Munroe (1973) presents a good description of the circuits and their functions in heterodyne lock-in amplifiers.

Despite the large dynamic range of the lock-in amplifier, it was necessary to add a low pass filter between the photomultiplier and amplifier input. The need for this comes from the fact that a very large signal is produced by the strip of bare ceramic between the two sample halves. This signal appears at twice the frequency of the desired modulation signal and can be as much as five orders of magnitude larger. This problem was solved with an eight pole Chebyschev low pass filter with a response as shown in Fig. 5.10. The filter was ordered with a cutoff at 80 Hz; the normal operating frequency of the tuning fork was around 70 Hz. Thus, the 140 Hz component was attenuated by a factor of greater than 1000, which then allowed the lock-in amplifier

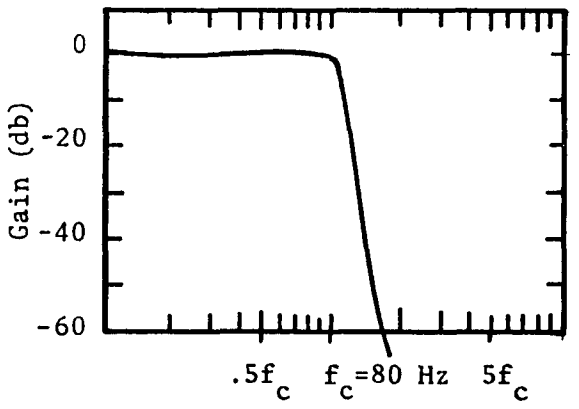


Fig. 5.10. Attenuation of 8 Pole Chebyschev Filter

to operate with a minimum of extraneous noise signals. The cutoff frequency of 80 Hz was chosen to avoid large phase changes with frequency which occur near the cutoff point of this kind of filter. Since the lock-in amplifier has a broadband response and can track a signal which is drifting in frequency, a frequency dependent phase shift introduced by the filter could affect the output.

The output of the analog divider is a voltage which is proportional to the ratio of the two input voltages. It is used here to eliminate the common factor in the modulated signal and the average reflectance and thereby gives the ratio $\Delta R/R$. The common factor includes such items as the lamp intensity and its spectral dependence, the monochromator response, and the detector spectral sensitivity. Any constant component added to either the numerator or the denominator will cause difficulties; for example, if the D.C. offset feature of the lock-in amplifier is used to adjust finely the zero setting, this will produce a wavelength dependent result even if $\Delta R/R$ is constant:

$$\text{divider output} = \frac{I(\lambda)\Delta R + C}{I(\lambda)R} = \frac{\Delta R}{R} + \frac{C}{I(\lambda)R} = \frac{\Delta R}{R} + f(\lambda)$$

where $I(\lambda)$ is the system response as a function of wavelength. On this basis, it is preferable to place the divider in the circuit between the output of the PMT and the lock-in amplifier. This location is undesirable for other reasons, however. This is where the signal is the smallest; any noise introduced here could obliterate the signal. A second problem is the limited dynamic range of the divider. To keep

the S/N ratio at an acceptable level, a large numerator voltage is necessary (1 to 10 volts). On the other hand, whenever the denominator voltage drops below the numerator's, the divider output saturates. These requirements can be met by placing appropriate operational amplifiers with adjustable gains in this part of the circuit, but unless they are extremely quiet (and therefore expensive), they add too much noise to the signal. It was finally decided to do the division after the phase sensitive detection.

The voltage to be used for the denominator of the divider is the average D. C. value of the PMT output. This output cannot be used directly by the divider because of an impedance mismatch. A field effect transistor input operational amplifier was added between the PMT and the denominator input of the divider. This had a twofold function: 1. the gain is adjustable through the selection of different feedback resistors so the input of the divider can be kept within the required range; and 2. it serves as a buffer amplifier to match the impedance of the PMT to that of the divider. The circuitry for this is shown in Fig. 5.11. Figure 5.12 shows a plot of the linearity of the divider circuit.

A time lag occurs in the signal which goes through the ΔR branch of the processing electronics. The output time constant of the lock-in amplifier was often set at 1.25 or 4 seconds, and this introduced a substantial time delay before the ΔR signal reached the divider as compared to the R signal which had no delay. Therefore, it was necessary to introduce a similar time delay for the R signal; this was accomplished

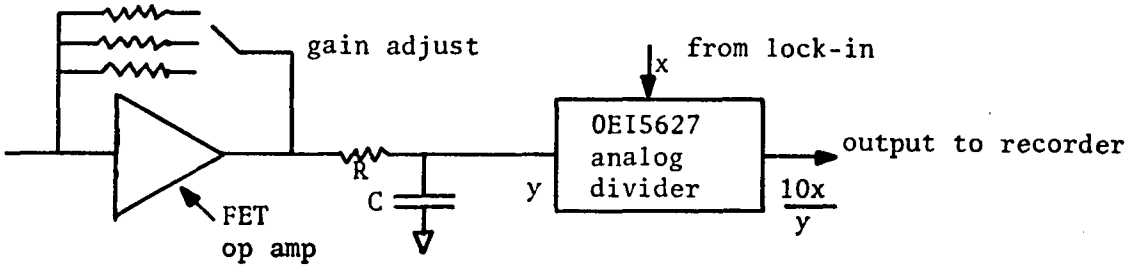


Fig. 5.11. Analog Divider and Buffer Amplifier

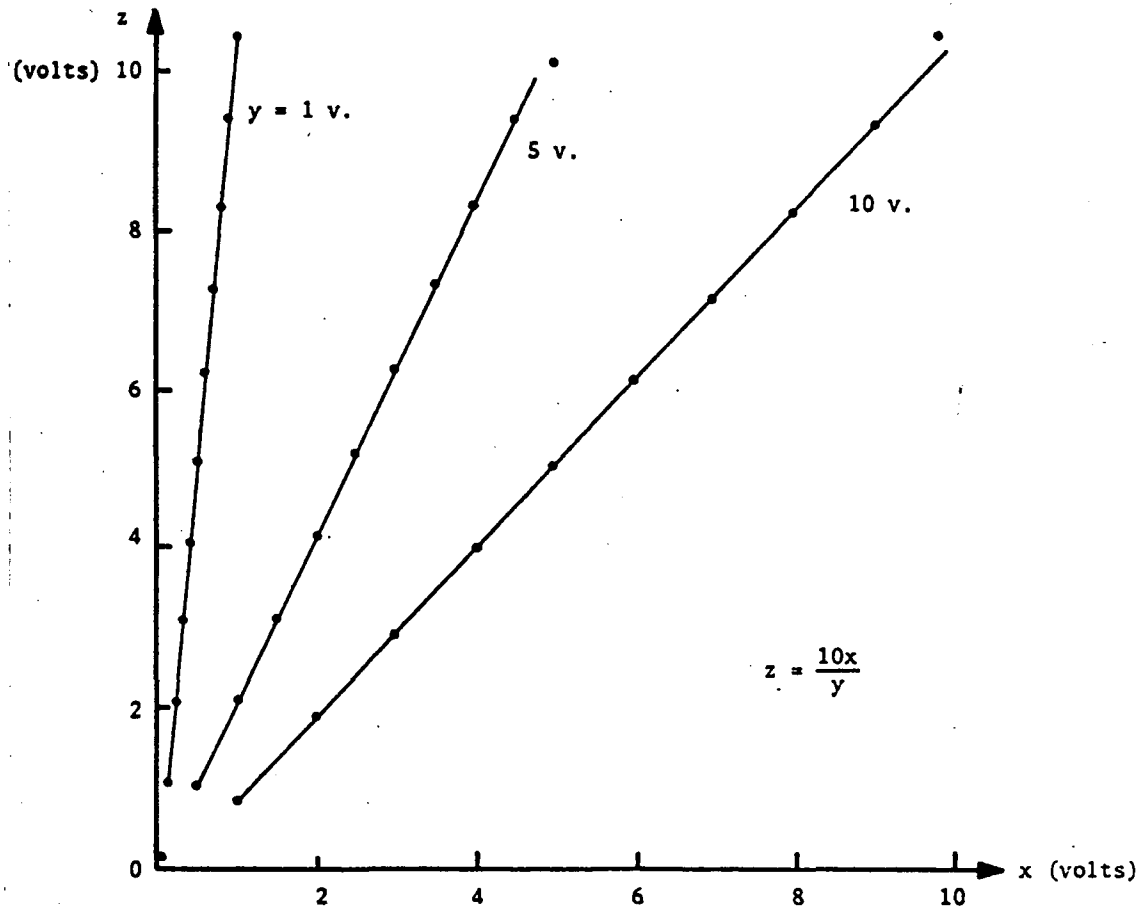


Fig. 5.12. Linearity of Analog Divider

with the RC circuitry shown in Fig. 5.11. Without the delay, what should be a constant divider output would show transient voltage swings whenever an abrupt change occurred in both R and ΔR, such as could be caused by fluctuations in the lamp intensity or gain changes.

The tuning fork had 4 drive coils already incorporated in its mount. There was no feedback mechanism for an oscillator, and the use of an external oscillator to drive the fork proved unsatisfactory due to the fork's very high Q. A feedback mechanism was obtained by mounting a recording head from a tape recorder such that a tine of the tuning fork induced a signal in it. The position of the head had to be offset from the center position of the tine; a centered position produced a feedback signal at twice the required frequency. The pickup coil was connected to an amplifier as shown in Fig. 5.13. The output of the amplifier was fed to the coils which were connected in series-parallel. A phase shifting circuit was added between the two operational amplifier stages to fine tune the circuit. The output power of the 741 op amp is sufficient to drive the fork. The oscillator is self-starting and produces approximately a ±15 volt square wave at equilibrium. Part of the output is attenuated to produce a four volt peak to peak reference signal for the phase sensitive lock-in amplifier. When initially constructed, the circuitry tended to oscillate at several kilohertz; this was presumably due to some mechanical resonance or possibly even fringing magnetic fields linking the driving coil and the feedback pickup. The problem was remedied by adding some bypass filter capacitors to provide an AC short circuit to ground for the higher frequencies.

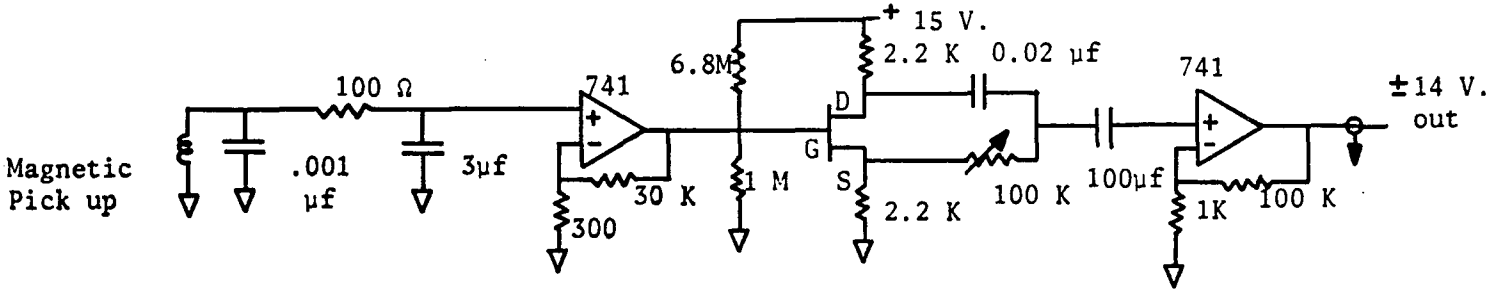


Fig. 5.13. Tuning Fork Driver Circuit

Mechanical

The tuning fork is the basic resonating element in the system; the electronics simply add energy to overcome the mechanical losses in the system. The ceramic sample is attached to one tine with water soluble adhesive, thus making a strong yet easily removable bond. The mass of a ceramic substrate was about 3 grams. Attaching such a mass to only one tine would unbalance the two tines; the resulting system would still oscillate, but it would be in a complicated mode involving the supporting structure as well. Some of the oscillating energy would thus go into making the entire experimental setup oscillate and this would give rise to spurious signals. The first step to bring the two tines into tune again consisted of mounting a second ceramic substrate on the other tine. While this did bring the two halves of the fork closer in tune, it was not close enough and one tine still would execute a large amplitude oscillation while the other remained almost at rest. The solution to this problem consisted of attaching small aluminum masses to the tines. These could be adjusted along the length of the tines to fine tune the mechanical oscillator so that almost all of the energy stayed in the fork and very little could be felt in the supporting structure.

The balance between the reflected light intensities of the two halves of the sample was accomplished by moving the sample perpendicularly in the light beam. The tuning fork and its mount were bolted to a translation stage. The micrometer screw on the stage was used to position

the fork assembly so that each half was exposed to the light beam for an equal amount of time (assuming equal reflectances for each side of the sample). At that position, the lock-in amplifier had a zero output.

The entire experiment bench was supported on inflated inner tubes to minimize any external vibrations from reaching the apparatus.

Originally, a DC motor was used to drive the wavelength shaft of the monochromator, and a toothed timing belt connected the monochromator shaft to a potentiometer. Since the shaft rotation was proportional to the wavelength, the output of the potentiometer served to drive the x axis of a Hewlett Packard 7000AM flatbed x-y recorder.

The DC motor drive system was scrapped after a great deal of work had been expended on it. A mechanical clutch which allowed the manual setting of the wavelength as well as prevented damage when the wavelength limits were reached gave a great deal of difficulty. Later, the DC motor started to produce noise spikes in the output cable of the photomultiplier. The decision to use a variable speed D. C. motor to turn the wavelength drive of the monochromator had been based upon the possible use of an arc lamp which would have sharp intense lines in its spectrum. It was thought desirable to scan through these lines slowly due to the large time constants in both the divider and the lock-in circuits. Conversely, in those sections of the spectrum where there was little structure, the scanning speed could be increased. Eventually the decision was made to use a tungsten filamant lamp; then, it became

possible to continue the work without the variable speed feature of the D. C. motor because of the smooth spectrum of such a lamp.

A synchronous motor and gear train built specifically for the monochromator was obtained and with this arrangement, the elaborate x-y recorder is no longer necessary since a recorder with a time axis serves the purpose as well. The scanning speed is not continuously variable, but it can be adjusted in steps by changing gear sizes or by using different motors which are commercially available.

CHAPTER 6

RESULTS AND DISCUSSION

The results obtained with the reflectometer described in the previous chapter will now be discussed. The performance of the instrument is analyzed and suggestions are given for further improvements. Alternative sample configurations are also discussed.

Reflectometer Sensitivity and Stability

The sensitivity and the stability of the reflectometer were tested by using an opaque aluminum film deposited on a glass substrate. Stability is defined for our purposes here as the change in the zero point as a function of wavelength when both halves of the sample are equal. Under these conditions, there should be no modulation detected at any wavelength.

Room was available on the tuning fork to mount the aluminum film on a glass substrate sample adjacent to the ceramic substrate. The aluminum had been evaporated without a mask so it covered the entire substrate. This continuous film eliminated the contribution at the second harmonic frequency. A difficulty arose with this arrangement in that there was no gap to help in achieving a balance of the intensity reflected during each half cycle. Actually, a balance should be obtained for any position of the beam on this particular sample. This was not the case and dust and surface imperfections caused a signal to be observed at

some settings. A position was finally found at which there was a satisfactory balance for a given wavelength; a spectral scan was then made from 0.4 to 0.725 micrometer.

Except for what appeared to be a transient near the start of the scan, the noise level was about 3×10^{-5} peak to peak in $\Delta R/R$, while the overall flatness of the spectrum seemed to be better than this by a factor of about two. This performance is better than the arrangement reported by Holbrook and Hummel (1973) by about two orders of magnitude.

Subsequently, attempts were made to obtain the ER spectrum of bismuth. The resulting spectrum was basically a straight line with a slope. The change in level from one end of the scanned range to the other was on the order of 10^{-4} in $\Delta R/R$. No structure in the form of spectral lines was observed. One immediate conclusion can be drawn from this, however; if any sharp ER effect exists for this sample in this spectral region, it has to produce a response considerably smaller than 10^{-4} in $\Delta R/R$.

It was suspected that surface imperfections were probably the cause of the observed response, although there were no visible ones on the surface of the Bi film. The film was several weeks old and imperceptible deterioration could certainly have taken place even though the samples were handled with care. A repetition of the stability run with the aluminum film (which also had no visible defects) some few days after it was prepared also produced a similarly tilted spectrum. Since no portion of the Al film could be found which produced a null balance over

an extended wavelength range, and since no change was made in the instrumentation, it was concluded that imperceptible degradations of the films caused the observed response. It is therefore necessary to use only freshly deposited metal films for spatial modulation since unequal surface ageing and tarnishing as well as microscopic dirt and dust contamination give rise to spurious responses.

A modulated thermoreflectance (TR) spectrum of a Bi sample prepared in a manner similar to the ER samples on the ceramic substrates showed no structure in the spectral region where the ER spectrum was measured. Thermoreflectance structure was detected near 2400 and 2800 Å. These TR measurements were carried out at 136 K on a separate reflectometer (Kottke, 1974).

Somewhat surprisingly, no structure was observed in those areas where there are steep slopes in the static reflectance curve (see Fig. 4.12). The static curves were measured for the Bi films prepared in this laboratory and these differed somewhat from that obtained by Cardona and Greenaway (1964) (CG); ours were rather flat with no significant structure in the range from 3500 Å to 2.5 micrometers. The discrepancy between the two results is most likely due to the fact that CG's results were obtained on a cleaved single crystal sample, while the crystalline state of the thin film samples used in these measurements is unknown.

This is probably one reason why no structure was seen in either modulation spectroscopy spectrum. Future work will have to consider

this discrepancy. An absorbing layer sometimes formed on the surface of our Bi films after they had been exposed to air for a while. This tarnish could be removed by gently wiping the film with a bare finger (at the expense of leaving a streaked surface). The tarnish film did not affect the shape of the static reflectance curve other than to lower it uniformly over the spectral range which was examined.

Another source of difficulty which may be responsible for the negative results obtained in the ER measurement is the poling electrode. The samples used a thin Cr film to polarize the ceramic. As previously discussed in Chapter 4, an extremely thin film can shield the electric field if the material has a large free electron density. The electrons in the chromium film could well have prevented the electric field from penetrating beyond the poling film.

There is a need to determine why no spectrally dependent response was observed in the ER studies of bismuth. The most important parameter that should be verified is the shape of the electric field profiles. A sample configuration and a material with a known ER response have to be integrated into a geometry which will allow an experiment to measure the electric field distribution.

Problems Encountered with Spatial Modulation

As previously mentioned, one of the most serious problems encountered was the change in the balance of the reflected light intensities from the two sample halves as a function of wavelength when the

two halves are supposed to be identical. Several sources were found to be the cause of this. One point should be emphasized here. In order to be detected a signal at the output of the photomultiplier must have two qualities: 1) it must have a modulation with the right frequency and phase; and 2) it must be a function of wavelength. These requirements on what is a valid signal make the experiment a great deal simpler, so long as the size of the undesired components do not cause an overloading of the electronic circuits.

Photocathode Nonuniformities

It is a well known fact that photomultiplier photocathodes have a sensitivity which varies with the position of a beam of light on its surface. A second, but not so well known fact, is that the spectral response varies from point to point on the photocathode as previously mentioned. Thus, one must be very careful to ascertain that the same area is always used when two beams are compared. In particular, great care must be exercised in this experiment since we are trying to detect such small signals.

Scattered Light

Scattered light is a problem with most optical systems. It can be especially severe in this experiment because of the small size of the signals being measured. In the alternating beam configuration, all light coming out of the monochromator is chopped. Any light which is scattered

inside the box containing the optics is suitably modulated for detection by the lock-in amplifier which processes the output of the photomultiplier. This problem was eliminated to a great extent by switching over to the moving sample constant optical path configuration. This did not completely eliminate the scattered light, though. It was discovered that there was still a residual signal even after the new arrangement was installed; worse, it was wavelength dependent. Had it not been a function of wavelength, it could have been nulled out when the balance was made between the two sample halves. The problem was finally localized to light reflected by the reddish tine of the tuning fork. Though the exit slit of the monochromator had been masked to keep the image from overfilling the sample, aberrations in the optical system allowed enough light to spill over the edge of the sample to cause a spurious signal. This signal was large enough to cause a shift in the spectral baseline on the order of 10^{-4} in $\Delta R/R$. Once it was discovered, this problem was easily remedied with suitable masks near the sample plane.

Sample Nonuniformities

Since two distinct areas are being compared, they should differ only in that one half is under the influence of an electric field, while the other is not. Any surface defects could give rise to a response, especially if the response of these nonuniformities is a spectrally dependent function. To ascertain that both halves are identical is by no

means an easy task, and this is expected to be the limiting aspect of spatial modulation.

Considerations for Future Work

Calibration Phase

The first aspect which must be examined in the next phase of this research is the establishment of the validity of the equations developed in Chapter 4. This seems to be most easily accomplished through the use of germanium, whose ER response is well known from studies using the field effect structure. It would be expected that Ge would respond in a similar manner if it were used in the ravioli structure, where it is electrically floating instead of being grounded.

A temperature of at least 450°C is needed to crystallize a Ge film, and this is attainable with the materials used in forming the ravioli structure. It is not possible to use a ferroelectric substrate since none is available which has a Curie temperature above 450°C. It is necessary for the substrate to maintain its original poled condition during the film deposition and all subsequent processing, otherwise there will be different strains induced by the substrate in the two sample halves and these could give rise to spurious results. This structure could be made in two halves only one of which would have a field applied to it. In this way, the spatial modulation spectrometer could still be used and thereby be calibrated at the same time.

Some modifications might have to be made to the equations in order to take into account the hole nature of conduction process in the semiconductor. The equations would be verified by comparing the observed spectral line shape with that calculated from inhomogeneous modulation theory using the inhomogeneous field profiles obtained from the equations given in Chapter 4.

Sample Preparation

It appears that there are two viable sample types: (a) the ravioli structure, and (b) the negatively charged surface of a ceramic substrate. The former utilizes known thin film technology and no problems are foreseen in constructing such samples. Such a sample meshes nicely with another modulation spectrometer available in the laboratory (Kottke, 1974) so the thermo- and the electro-reflectance spectra can be measured on the same specimen. While this other spectrometer does not have the capability to make spatial modulation measurements, it does allow the sample to be cooled down to 77 K. This second spectrometer is not a necessity since standard ER can be measured with the apparatus described in Chapter 5, although the measurements could then only be done at room temperature. The same sample can be configured to be suitable for both types of measurement by evaporating the sample film onto two areas on a common substrate.

Another sample configuration which was proposed during the course of this work can be shown to be impractical for a variety of reasons.

This suggestion would have the metal film deposited on the ceramic and subsequently covered with an insulating overcoating to try to preserve the charge so that the insulated boundary conditions obtained (Fig. 6.1). The charge on the surface tends to leak off over a period of time; the potential at which this charge is stored can be easily calculated. Assuming a plane parallel capacitor configuration and ignoring the fringing fields we have

$$q = CV \qquad \text{and} \qquad C = \frac{A\varepsilon}{d}$$

so that we can write

$$V = \frac{q}{C} = \frac{qd}{A\varepsilon} = \sigma\frac{d}{\varepsilon},$$

where C is the capacitance of the structure, q is the charge, and d is the separation of the plates. For a ceramic substrate 2mm thick, a surface charge of $30\mu C/cm^2$, and a relative dielectric constant near 3000, V = 20,000 volts. The external electric field is about $3x10^8$ V/cm and the internal field is on the order of 10^5 V/cm. It is at once apparent that a capacitor with a potential of 20 kV will not maintain that charge for an indefinite length of time. Secondly, the dielectric breakdown of the insulating overcoating will be on the order of 10^6 V/cm (SiO_2) so that it will not sustain the electric forces and will leak charges into or out of the film. Any instrumentation used to monitor the state of charge of the film, e.g., an electrostatic voltmeter, would further compound the problem rather than help since it would supply an additional leakage path.

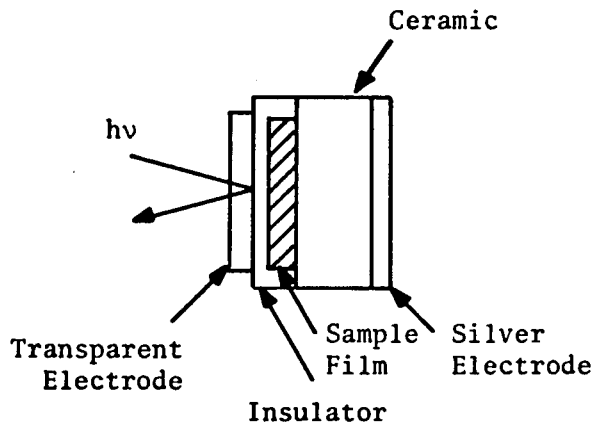


Fig. 6.1. Overcoated Sample Film Deposited On Ceramic

In the case of the sample film deposited on the ceramic, some modifications to the poling procedure will be necessary, since one half of the sample will have to be only partially polarized, while the other half is polarized to saturation. A further complication is that the poling film has to be eliminated because of the free electron screening problem. The solution could be the use of a noncontacting poling electrode such as a wire mesh. If it were placed sufficiently far away from the front surface of the ceramic to prevent shadowing during the evaporation process, it would not have to be removed during the process. The evaporant source size can be made large enough so that neglible shadowing takes place. The limiting factor involved in the substrate-electrode spacing is the maximum value of the poling potential which can be safely introduced into the vacuum chamber. On the other hand, the limitation on the source size is not significant since multiple sources could be used.

A moveable electrode is another possibility; it could be moved out of the way during the evaporation process. In either case, provision should be made to ground the metal film after the evaporation process to ascertain that no residual charge remains on the surface.

In both of these ferroelectric samples, no gap is necessary between sample halves. The elimination of the gap removes the large signal which appears at the second harmonic frequency. The low pass filter is therefore not necessary between the PMT and the lock-in amplifier, and one more possible source of noise is eliminated.

The omission of the gap could prove disadvantageous when either the photovoltaic or the photoconductive detector is used since there is then no signal directly proportional to R. Balancing the reflected intensities will be difficult since the dividing line between the two halves will not be readily apparent.

To prevent ions of the opposite sign from being attracted to the surface, the metal will have to be evaporated soon after the poling process is completed. It takes only a third of a monolayer of appropriately charged ions to completely neutralize all of the surface charge. When the sample film is subsequently evaporated, no field will penetrate through it. It is important to note that it takes a third of a monolayer of charged ions, and that uncharged particles will not affect the result so long as ionization doesn't take place upon collision of the particle with the surface in the presence of the strong electric field. The factor of one third comes from the fact that $30\mu C/cm^2$ is equivalent to the loss of one unit of electronic charge per three surface lattice sites. Ions which arrive after the film has been deposited will not affect the penetration of the field into the film.

The exact method by which the polarization of the substrate will be attained is not yet clear. It might be necessary to use an ultrahigh vacuum system or else to find a conductor which has a very small free electron density and which will form a conductive film for a very small film thickness. The figure of merit to be used in the search for such a material would be $t \cdot n_e$, where t is the minimum thickness

for a conductive film and n_e is the free electron density. The value of this figure should be as small as possible since it represents the number of screening carriers per unit area in the poling film.

As previously mentioned, the question of the discrepancy between the Bi spectrum measured by Cardona and Greenaway and that obtained for thin films will have to be investigated. The Hall coefficient for thin Bi films is not necessarily in agreement with that of bulk samples (Kittel, 1971, p. 289; Jeppesen, Flagg, and Rancourt, 1963). X-ray diffraction may well be necessary to determine the crystalline structure, and Hall effect measurements used to obtain the free carrier concentration.

Instrumentation

As mentioned in Chapter 5, one of the deficiencies of the reflectometer is the low efficiency of the optical system. That chapter also included suggestions for the construction of an improved system to be placed inside a larger light-tight enclosure. This would improve the optical system as well as allow more "elbow room" in the vicinity of the sample mount than there is in the present enclosure.

When the tuning fork executes large amplitude oscillations, the motion cannot be expected to be simple harmonic motion. The excursions to either side of the center position are probably not symmetric, though this has not been verified. A change in the amplitude of the oscillation would then result in a change in the relative amounts of time that each half of the sample is exposed to the light beam. This is equivalent to

a change in the zero point relative to the optical axis. In turn, this leads to an imbalance between the two halves which can be detected by the lock-in amplifier. Due to the sensitivity of the apparatus, a slight change in the balance can be detected. There are many sources which could cause a change in the amplitude; an example is a possible beating of the two tines against each other. Narrowing the gap between the two sample halves will not solve this problem. A narrower gap would mean, however, that a smaller amplitude could be used, but this would be at the expense of the averaging effect which takes place when a large area of each sample half is used.

Another solution to this problem is the use of a forced oscillator instead of a resonant one. In such a configuration, the sample is regidly connected to a loud-speaker coil. This can be driven very conveniently with any common hi-fi amplifier and the amplitude can be easily adjusted. In this way, any frequency can be selected; with the tracking feature of the Ithaco model 391 lock-in amplifier, frequency modulation could be used to further discriminate against noise. This would remove problems associated with resonances in the supporting structure.

Figure 6.2 gives a circuit diagram which will keep the D.C. level of the PMT output constant by varying the high voltage applied to the tube. The A.C. signal from the PMT is then directly proportional to $\Delta R/R$ and this goes to the lock-in amplifier. This eliminates the problems involved in dividing the voltage proportional to ΔR by the one

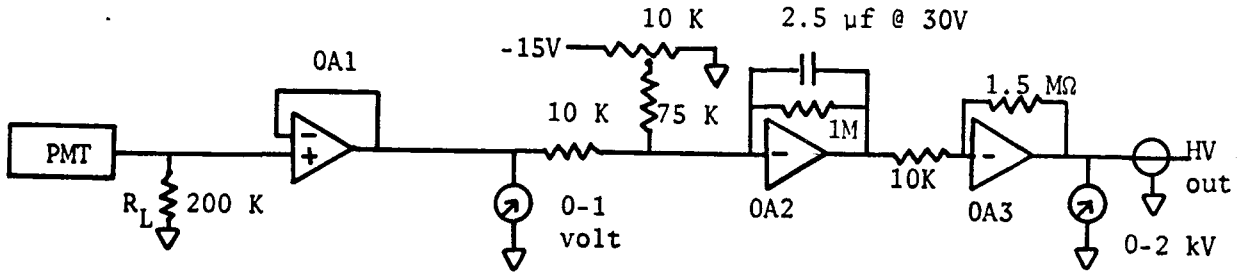


Fig. 6.2. High Voltage Regulating Circuit to Provide a Constant D.C. Output From the PMT

OA1 and OA2 are type 741 operational amplifiers; OA3 can be a Kepco OPS2000 high voltage op amp. OA1 is a voltage follower which is used for impedance matching and as a meter driver. OA2 sets the level, inverts the signal, and drives OA3.

proportional to R after the synchronous detection. Such a circuit was not assembled during the course of this part of the study because of the lack of the high voltage operational amplifier. The alternative approach to the division problem is to obtain a quiet divider that could be placed between the PMT and the lock-in amplifier.

For photovoltaic detectors, a circuit such as given by N. R. Comins and P. D. Grant (1973) is probably the most useful, while for photoconductive detectors, a second modulating frequency is essential to obtain a voltage proportional to R. If the second harmonic of the signal is not available, a suitable signal can be obtained by chopping the light before it enters the reflectometer and then using a tuned or a lock-in amplifier to detect it and produce a D.C. output proportional to R. Only a portion of the light needs to be modulated at this second frequency, so a penalty in the form of light loss need not be paid.

CHAPTER 7

CONCLUSION

The objective of this study was to lay the groundwork for a novel technique for measuring the electroreflectance spectra of metals. This was accomplished by a theoretical investigation of the penetration of an electric field into a metal film and the construction of optical instrumentation to measure the effect.

The theoretical studies showed that it would be very difficult to obtain an electric field in a metal film that has a very high free electron concentration, such as is the case with noble metals. In such a case the film has to be so thin that the observed spectrum would not be representative of bulk material. On the other hand, it was shown that for materials that have a lower free carrier density, such as bismuth and the other semimetals, it is possible to have an electric field penetrate to a significant fraction of the depth that is probed by the light wave. However, as the free electron density increases, the electric field penetration decreases so that it becomes progressively more difficult to observe the effect of the electric field on the reflected light from samples with larger free electron concentrations.

Means of bypassing the difficulties caused by the free electrons were discussed and two basic sample configurations were considered that allowed an overlap between the probing light wave and the electric field

profile. The use of the large surface charge of a polarized ceramic substrate and a grounded thin film sample was a means of eliminating the free electron difficulties. The film was made so thin that the number of charges needed to compensate or shield the externally applied field was a large fraction of the total number available in the film. The alternative approach encapsulated the sample film in a dielectric insulator; unfortunately, the magnitude of the usable electric field had to be reduced significantly from that obtainable with the ceramic sample because of the dielectric breakdown of the insulating film. It was finally decided that the most promising configuration was one where the surface of the ceramic substrate was negatively charged.

A modulation spectrometer was constructed and tested that has a sensitivity two orders of magnitude better than one described in a recently published article in the literature. Ours uses the concept of spatial modulation and is designed to keep the optical path in the instrument constant. It uses only reflecting optics to eliminate chromatic aberrations. Improvements to the instrument are proposed that should improve its signal-to-noise ratio even more.

Preliminary attempts were made to observe the electroreflectance spectrum of bismuth without success. It was theorized that the chromium poling film used in those particular samples completely attenuated the electric field before it was able to penetrate into the bismuth film. This led to the consideration and development of the alternative sample designs.

These accomplishments have advanced us along the path towards the eventual detection of unambiguous electroreflectance spectra. The significance of the work reported here lies in the fact that for the first time a means has been developed to observe the ER spectra of metals without the complicating effects of the free electrons. The way to successful observation of these spectra has been pointed out, and the means by which such goals will be attained have been developed and built. Continuation of these studies is necessary to reach the ultimate goal of the successful interpretation of ER spectral structure as particular transitions between states in the metallic band structure. Certainly interband transitions exist in metallic band structure, and this apparatus and technique have a good chance of being able to measure them.

APPENDIX A

NUMERICAL INTEGRATION AND PLOTTING COMPUTER PROGRAMS

The main program DIFFEQ is a general routine written in Fortram that affects the numerical integration of a general ordinary differential equation of up to eighth order as specified in the subroutine Z. The user can direct that program to use either a Runge-Kutta (RK) method, or one known as the "predictor-corrector" method (pc). The pc method requires a knowledge of the function and its derivatives at two equally spaced points just preceding the point at which these are to be evaluated anew. Since this method is not self-starting given some conditions, the RK method is used to obtain these, and then the pc method can continue the integration. The integration step, Δx, is adjusted to be as large as possible while keeping the error in the integrated functions within a user-specified tolerance.

The overall process consists of calculating the desired function by using the definition of the derivative to go from a continuous curve to one consisting of many small discrete steps. Starting at a given value for x_0, the derivatives are calculated and a new value is found for the function. The value of x is then incremented and the process is repeated until the integration interval is covered when x is equal to the limiting value x_{lim}.

In our particular case, the starting derivative is not known; instead, our second boundary condition is the value of the function at

the second edge of the integration interval. The program therefore searches until it finds a starting slope such that $E(x_{lim})$ has a value close to the desired one. Due to the large range of the numerical values that can occur in this type of equation, a two stage search process was put into the program. In the first stage, large changes are introduced in the starting slope of E until the correct solution is bracketed. The slope at x_0 is adjusted in the subroutine CHECK by multiplying or dividing it by a constant factor G during the coarse adjustment phase. This factor can be specified by the user; a default value of 10 is used if it is not given on an input card. Subsequently, a binary search procedure is used in which the slope error is divided by two at each iteration. The number of iterations used in this search is also user specified.

The accuracy of the final result is difficult to predict due to the nonlinear nature of the equation. The fractional error in the starting slope can be written as

$$\frac{y_c' - y_0'}{y_0'} = G/2^n$$

where y_0' is the true starting slope, y_c' is the starting slope found by the search process, and n is the number of the iterations in the binary search mode.

The logical function subroutine CHECK has two tasks to perform. First, it checks to see if the latest value calculated for the E field

is within the expected bounds as given by the initial and the target values specified by the user. It also checks to make sure that the E field does not change sign. If it determines that the value of E is appropriate, it returns control to the calling program with the value of CHECK as FALSE. If the value of E is out of bounds, or if the end of the interval has been reached, it calculates a new starting slope for the next iterations, as indicated above, and a value of TRUE is returned to the calling program. This routine is called each time a new set of values is calculated in the integration loop.

The main integration program has two modes of operation. When the number of binary iterations to be performed is given as positive, the program iterates the specified number of times to find an approximate solution; after it has finished, it does one more integration and prints out the values as it progresses. Based on the previous number of steps taken, it might skip the printing of some points in an attempt to limit the number printed to between 100 and 200.

In the second mode, which is specified with a negative value for the number of binary iterations, no iterations are performed. Instead, a single integration pass is made across the domain of integration and every |n|th point is printed. This second mode is required when plotting. All points will be plotted regardless of the parameter n.

A word of caution should be added; because of the large range of values that enter into the calculation of this equation, care

should be exercised to ascertain that the finite number of significant figures available in the computer arithmetic is not introducing significant errors in the integration. In some cases it might be necessary to use double precision arithmetic. This was not necessary on the computer used for the calculations in this paper since it has a 60 bit word size (CDC 6400). The program was written with the possibility in mind that double precision might be necessary.

The program is mostly self-explanatory and includes many comment cards. For more information of the integration algorithms, see Ralston and Wilf (1960), Chapters 8 and 9, and McCracken and Dorn (1964).

The input card formats are

Card 1:	(2E10.0)	α, β
Card 2:	(I5,E5.0,7E10.0)	J, G, x_0, Δx_0, x_{lim}, ε_1, T_{pc}, T_{rk}
Card 3:	(8E10.0)	y, y', y", etc. at x_0
Card 4:	(8E10.0)	y, y', y", etc. at x_{lim}
Card 5:	(2I5,L5,I5)	n, m, p, m_{max}

where J is the order of the integration, ε_1 is a dimensionless quantity with values between zero and one that defines the minimum step size allowed through $\Delta x_{min} = (x_{lim} - x_0)\varepsilon_1$, T_{pc} and T_{rk} are the tolerances used in the tests on the accuracy at each integration step, m is the number of gross iteration steps allowed (default value = 10), p is TRUE if plotting is desired, and m_{max} is the number of iterations allowed in the pc method. If m_{max} is negative, the pc method is

bypassed and only the RK method is used. If no value, or zero, or a number greater than 10 is entered in this field on the card, a value of 10 is used.

Figures A.1 to A.3 give the logic flow through the program. The core requirements for the CDC 6400 computer are under $35K_8$ words, and the time required for an average solution is about 10 seconds, if a solution is to be found in a reasonable time. In some cases, the choice of values for α and β makes the solution very sensitive to the choice of $E'(0)$, and the time required to solve such cases becomes quite large unless the starting slope is known accurately.

The program PLTDEQ plots the output produced by program DIFFEQ. Figure A.4 gives the logic flow for this program. Parts of the input and the plotting are overlapped to save time.

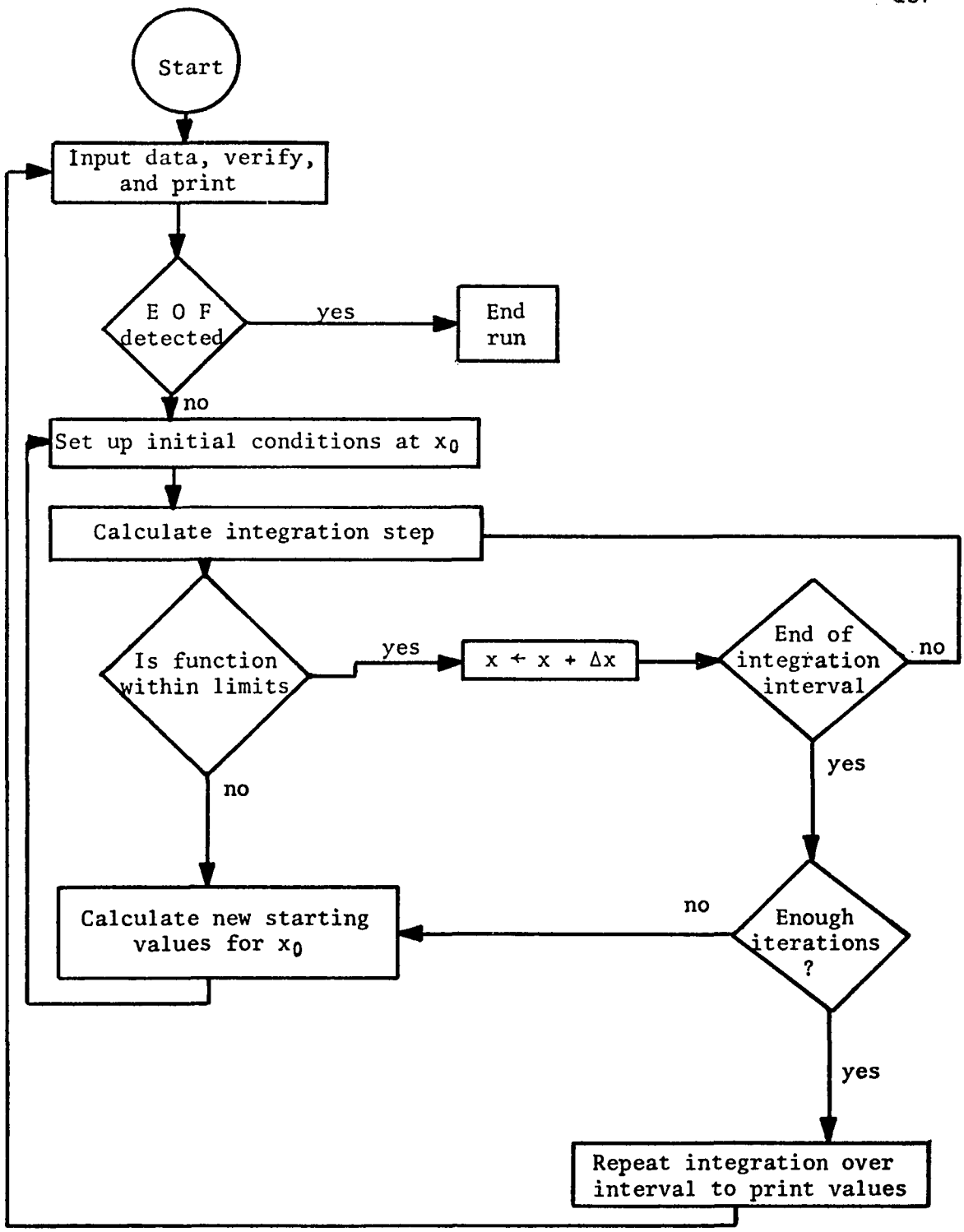


Fig. A.1. Overall Logic Flow of Program DIFFEQ

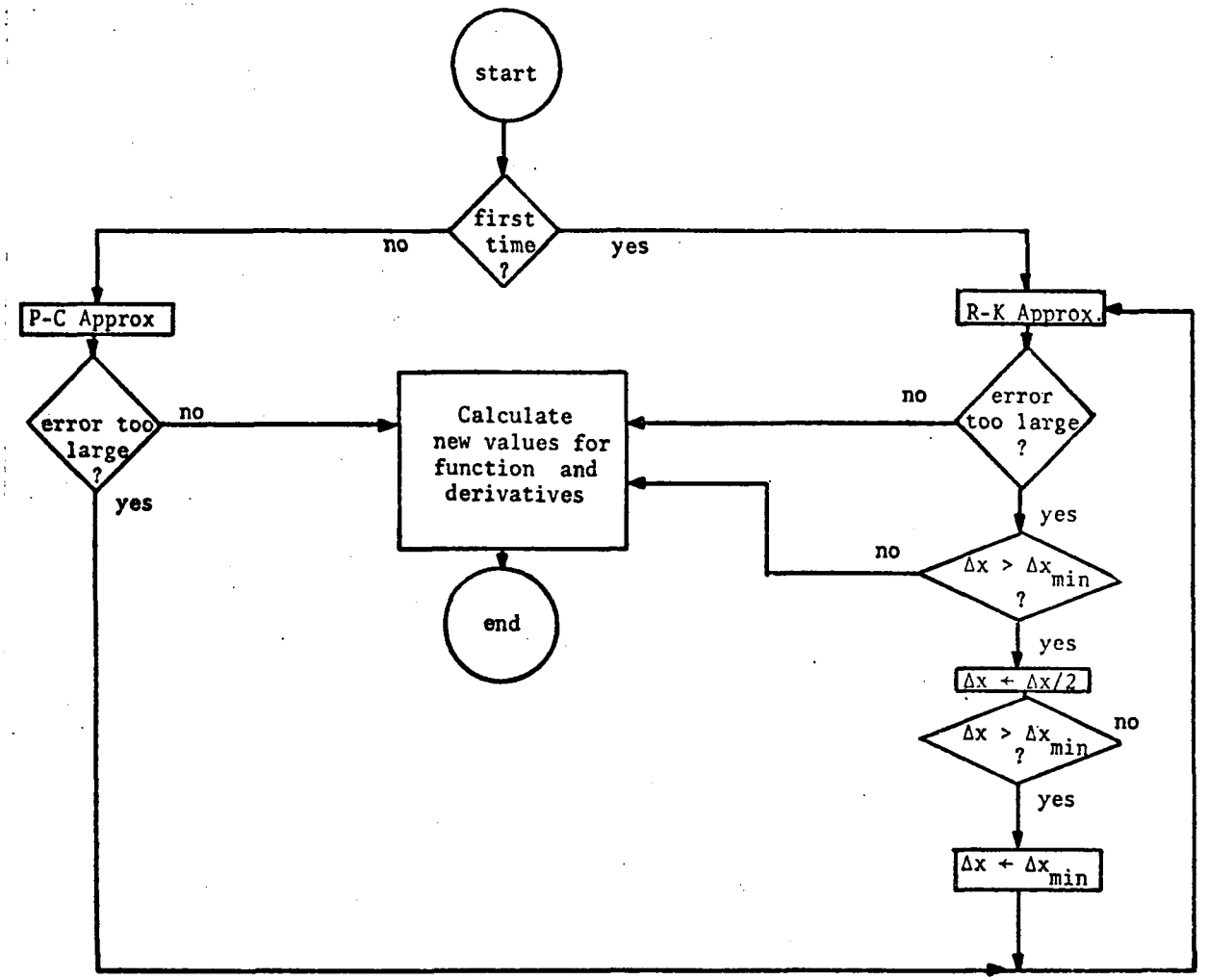


Fig. A.2. Integration Process Logic Flow

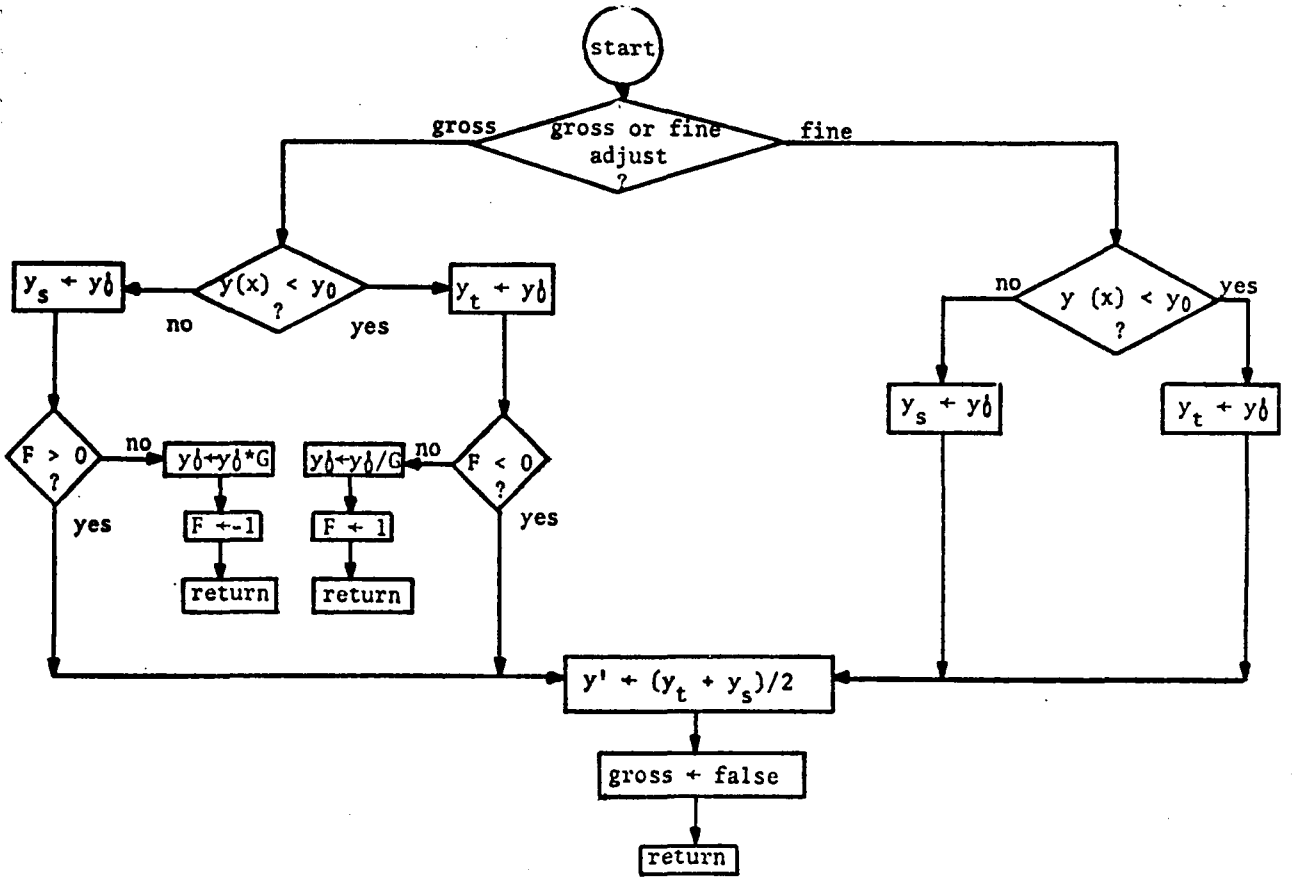


Fig. A.3. Starting Slope Logic (Subroutine CHECK)

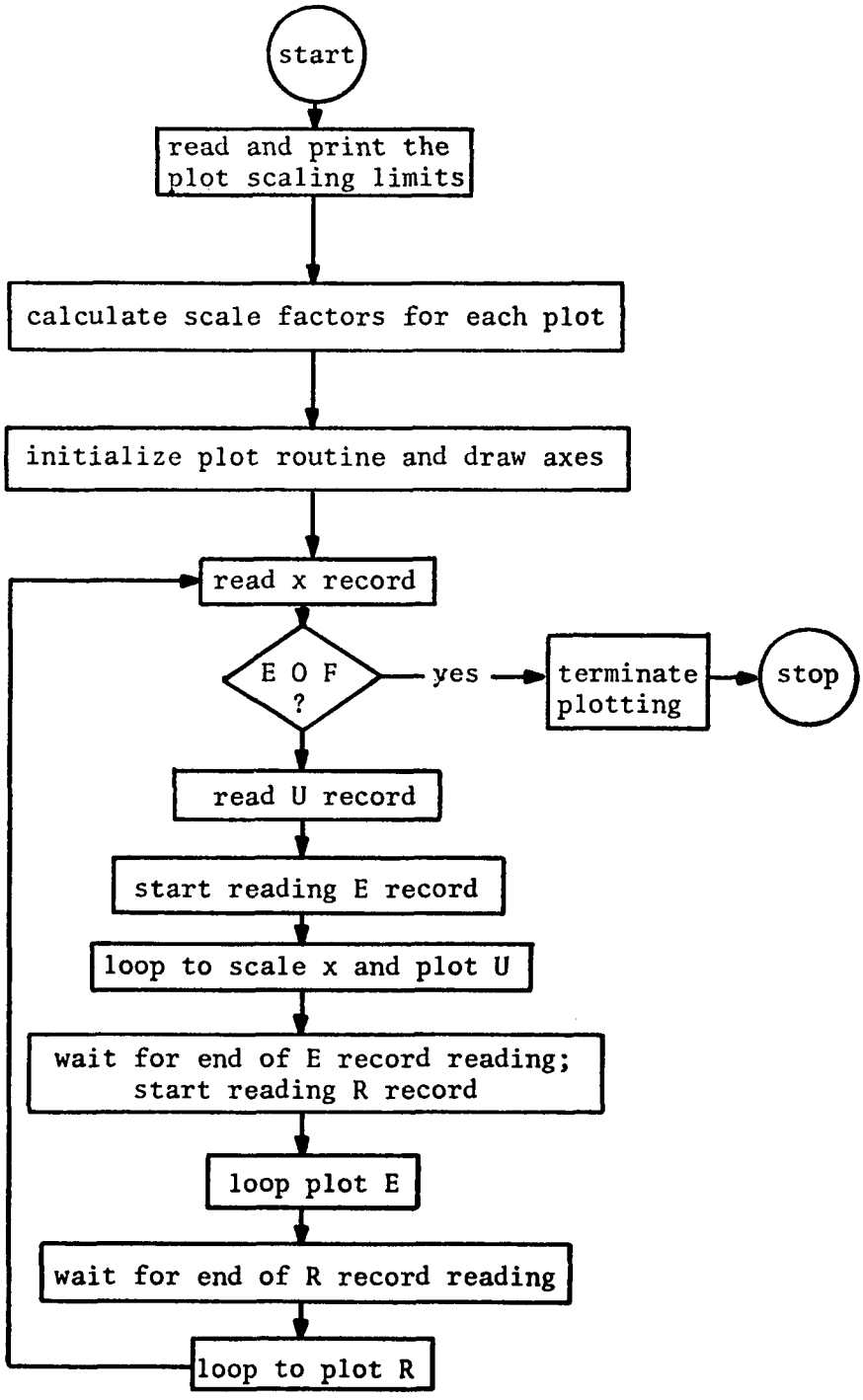


Fig. A.4. Logic Flow of Plotting Program PLTDEQ

```
      PROGRAM DIFFEQ(INPUT=10113,OUTPUT,TAPE10=1018,TAPE1=INPUT)
C
      COMMON Y(8,203),YTARG(8),Y0(8),X(203),YY(8),ZZ(8),RK1(8),RK2(8),
     $   RK3(8),RK4(8),YP(8,10)
C
      COMMON/LOGIC/START,LAST,ILG,YS,YT,IMX,LIMFLG,IBCN,GRADJ,J
      LOGICAL START,LAST,LIMFLG
C
      COMMON/AB/ ALPHA2,BETA
      COMMON/LIMITS/ YLL,YUL,H,IGCNT
C
      LOGICAL FLG,LOG,RKFLG,LAST2,PLTFLG
C
      LOGICAL CHECK
      EXTERNAL CHECK
C
C     INTERNAL STATEMENT FUNCTION TO GET AROUND CHANGE TO DOUBLE EASILY
C
      BBS(X)=ABS(X)      $     ZIGN(X,Y)=SIGN(X,Y)
C
C
C     FLG=.TRUE. IF WE ARE CALCULATING THE LAST POINT IN THE INTERVAL
C        WITH AN H ADJUSTED TO GET US TO THE END EXACTLY
C
C     LIMFLG=.TRUE. IS A FLG WHICH TELLS THE CHECK ROUTINE THAT WE HAVE
C        MADE IT TO THE END OF THE INTERVAL
C
C     LOG IS A DUMMY LOGIC VARIABLE.
C
C     RKFLG IS A FLAG WHICH PREVENTS THE RUNGE-KUTTA SECTION FROM GOING
C        INTO AN INFINITE LOOP WHNE THE MINIMUM STEP SIZE IS USED.
C
C     LAST2 IS SET TO TRUE DURING THE LAST PASS BY FLAG LAST
C
C     START IS TRUE DURING THE COARSE ADJUSTMENT PHASE.
C
C
C     RKTOL IS USED AS A TEST VALUE IN THE R-K SECTION TO CHECK THE STEP
C        SIZE.   VALUE OF 5 PERCENT WAS OBTAINED FROM KAREL REKTORYS,
C        SURVEY OF APPLICABLE MATHEMATICS, M.I.T. PRESS, CAMBRIDGE PRESS
C        , CAMBRIDGE, MASS., 1969, PAGE 1076.
C
 5    CONTINUE
C
C     READ IN EQUATION PARAMETERS
C
      READ 991,ALPHA,BETA
      IF(EOF(1).NE.0) GO TO 900
      PRINT 996,ALPHA,BETA
      ALPHA2=ALPHA*ALPHA
C
C     J IS THE ORDER OF THE EQUATION
C     X4 IS THE BEGINNING OF THE INTEGRATION RANGE
C     XLIM IS THE UPPER LIMIT (OF X) FOR THE INTEGRATION RANGE
C     H IS THE INITIAL INTEGRATION STEP
C     EPS1 IS THE MINIMUM STEP SIZE ALLOWED AS DEFINED BY
C           DX/(XLIM-X(1))
C     TOL1 IS THE TOLERANCE FOR THE PREDICTOR-CORRECTOR TESTS
C     GRADJ IS THE SEARCH FACTOR DURING THE COARSE ADJUSTMENT PHASE.
```

```
C
C
      READ 990,J,GRADJ,X4,H,XLIM,EPS1,TOL1,RKTOL
      XSTPL=XLIM-X4
      H=ZIGN(H,XSTPL)
      IF(RKTOL.LE.0.0) RKTOL=0.05
      IF(TOL1.LE.0.0) TOL1=0.01
      IF(EPS1.LE.0.0) EPS1=0.001
      IF(GRADJ.LE.1.0) GRADJ=10.0
      PRINT 997
      PRINT 999,J,X4,H,XLIM,TOL1,RKTOL,EPS1,GRADJ
C
C     READ IN THE BOUNDARY VALUES FOR THE LEFT SIDE OF INTEGRATION RANGE
C
      READ 991,(Y(I,4),I=1,J)
C
C     READ IN THE TARGET VALUES FOR THE RIGHT SIDE OF INTEGRATION RANGE
C
      READ 991,(YTARG(I),I=1,J)
C
C     IMX IS THE NUMBER OF ITERATIONS IN THE BINARY SEARCH MODE
C     IGRLM IS THE NUMBER OF ITERATIONS ALLOWED IN THE GROSS SEARCH MODE
C     PLTFLG IS TRUE WHEN PLOTTING IS TO BE DONE (ONLY WHEN IMX.LT.0)
C
      READ 983,IMX,IGRLM,PLTFLG,MMAX
      IF(MMAX.GT.10.OR.MMAX.EQ.0) MMAX=10
      IF(IMX.GT.0) PLTFLG=.FALSE.
      PRINT 985
      N=1
      PRINT 993,(Y(I,4),I=1,J)
      PRINT 995
      EP=ABS(EPS1*XSTPL)
      IF(IGRLM.LE.0) IGRLM=10
      PRINT 982,IMX,IGRLM,MMAX
      PRINT 994,EP,(YTARG(I),I=1,J)
      IGCNT=0
      NP=0
      NO=0
      N=4
      X(4)=X4
      YS=0.0
      YT=0.0
      ILG=0
      LAST2=.FALSE.
      START=.TRUE.
C
C     SET UP LIMITS FOR CHECK ROUTINE
C
      TA=YTARG(J-1)
      TB=Y(J-1,4)
      YLL=YUL=0.0
      IF(YLL.GT.TA) YLL=TA
      IF(YLL.GT.TB) YLL=TB
      IF(YUL.LT.TA) YUL=TA
      IF(YUL.LT.TB) YUL=TB
C
C     IF IMX IS NEG, DO NOT LOOP AND PRINT EVERY IABS(IMX) VALUE.
C
      IF(IMX.LT.0) GO TO 120
```

```
C
C      IMX IS POSITIVE OR ZERO
C
       LAST = .FALSE.
       GO TO 125
C
C      IMX IS NEGATIVE
C
 120   LAST=.TRUE.
       NPRNT=IABS(IMX)
C
C      SET UP ORIGINAL CONDITIONS
C
 125   IBCN=0
       IRKCNT=0
       H0=H
       DO 6 I=1,J
       YC(I)=Y(I,4)
 6     CONTINUE
       IF(LAST) GO TO 126
       PRINT 987
       GO TO 127
C
C      RESTART WITH NEW INITIAL CONDITIONS
C
 7     H=H0
       DO 8 I=1,J
       Y(I,4)=Y0(I)
 8     CONTINUE
 127   PRINT 988,NO,X(N),(Y(I,N),I=1,J),YS,YT,(Y0(I),I=1,J)
       IF(IRKCNT.GT.0) PRINT 984,IRKCNT
       IF(.NOT.LAST) GO TO 130
 126   PRINT 989
       LAST2=.TRUE.
       NO=1
       PRINT 998,NO,H,X4,(Y(I,4),I=1,J)
 130   X(4)=X4
       FLG=.FALSE.
       LIMFLG=.FALSE.
       RKFLG=.TRUE.
       N=5
       NO=2
       IRKCNT=0
       GO TO 11
C
C      HALVE THE STEP SIZE FOR THE R-K APPROXIMATION.
C
    14 H=H2
       IF(ABS(H).GT.EP.OR.FLG) GO TO 129
       H=ZIGN(EP,H)
       IRKCNT=IRKCNT+1
       RKFLG=.FALSE.
 129   FLG=.FALSE.
C
C      RUNGE-KUTTA (STARTER).
C
    11 X(N)=X(N-1)+H
       IF(H) 112,930,110
 110   IF(X(N).LE.XLIM) GO TO 117
```

```
      GO TO 270
 112  IF(X(N).LT.XLIM) GO TO 270
 117  H2=H/2.0
      H6=H/6.0
C
C     CALCULATE RUNGE-KUTTA TERMS
C
 119  DO 13 I=1,J
      RK1(I)=Z(I,X(N-1),Y(1,N-1))
      YY(I)=Y(I,N-1)+H2*RK1(I)
 13   CONTINUE
      DO 131 I=1,J
      RK2(I)=Z(I,X(N-1)+H2,YY(1))
      ZZ(I)=Y(I,N-1)+H2*RK2(I)
 131  CONTINUE
      DO 132 I=1,J
      RK3(I)=Z(I,X(N-1)+H2,ZZ(1))
      YY(I)=Y(I,N-1)+H*RK3(I)
 132  CONTINUE
C
C     BEFORE CALCULATING 4TH TERM, CHECK TO SEE IF WE ARE USING A
C        STEP WHICH IS TOO LARGE
C
      DO 133 I=1,J
      IF(.NOT.RKFLG.OR.RK2(I).EQ.0.0) GO TO 16
      IF(ABS((RK2(I)-RK3(I))/RK2(I)).GT.RKTOL) GO TO 14
   16 RK4(I)=Z(I,X(N-1)+H,YY(1))
  133 Y(I,N)=Y(I,N-1)+H6*(RK1(I)+2.0*RK2(I)+2.0*RK3(I)+RK4(I))
      IALP=2HRK
      RKFLG=.TRUE.
C
 140  IF(.NOT.LAST) GO TO 150
      NP=NP+1
      IF(NP.LT.NPRNT) GO TO 150
      NP=0
      PRINT 992,IALP,NO,H,X(N),(Y(I,N),I=1,J)
 150  IF(CHECK(N)) GO TO 810
      IF(MMAX.LE.1) GO TO 250
      N=N+1
      N2=N-2
      NO=NO+1
      IDIR=1
      IF(N-204) 17,31,15
   17 X(N)=X(N-1)+H
      IF(H) 160,800,155
 155  IF(XLIM.GT.X(N)) GO TO 19
      GO TO 270
 160  IF(X(N).GT.XLIM) GO TO 19
      GO TO 270
C
C     PREDICTOR-CORRECTOR METHOD (CONTINUATION).
C
C
C     PREDICT VALUE AT X(N)
C
 19   IF(MMAX.LE.1) GO TO 119
      DO 20 I=1,J
      YY(I)=Z(I,X(N-1),Y(1,N-1))
      YP(I,1)=Y(I,N2)+2.0*H*YY(I)
```

```
 20   CONTINUE
      IALP=1H
      M=1
C
C     ITERATE PREDICTION UNTIL ERROR IS SMALL ENOUGH.  THIS IS LIMITED
C        TO 10 TRIES BEFORE THE STEP IS HALVED.
C
 205  M=M+1
      IF(M.GT.MMAX) GO TO 14
      DO 22 I=1,J
      YP(I,M)=Y(I,N-1)+H2*(YY(I)+Z(I,X(N),YP(1,M-1)))
 22   CONTINUE
      DO 23 I=1,J
      IF(ABS((YP(I,M)-YP(I,M-1))/(YP(I,M)+YP(I,M-1))).GT.TOL1) GO TO 205
 23   CONTINUE
      DO 27 I=1,J
      Y(I,N)=YP(I,M)
 27   CONTINUE
      IF(M-2) 15,103,140
C
 103  IF(.NOT.LAST) GO TO 360
      NP=NP+1
      IF(NP.LT.NPRNT) GO TO 360
      NP=0
      PRINT 998,NO,H,X(N),(Y(I,N),I=1,J)
 360  IF(CHECK(N)) GO TO 810
C
C     X-INTERVAL (H) DOUBLER
C
 250  H=H+H
      H2=2.0*H2
      H6=2.0*H6
      N=N+1
      NO=NO+1
      IDIR=2
      IF(N-204) 30,31,15
   30 X(N)=X(N-1)+H
      IF(H) 260,800,255
 255  IF(X(N).LT.XLIM) GO TO 19
      GO TO 270
 260  IF(X(N).GT.XLIM) GO TO 19
 270  IF(FLG) GO TO 800
C
C     ON THE LAST STEP WE USE THE RK METHOD
C
      H=XLIM-X(N-1)
      FLG=.TRUE.
      GO TO 11
C
C
 31   X(2)=X(N2)
      X(3)=X(203)
      N=N-1
      IF(PLTFLG) BUFFER OUT(10,1)(X(4),X(N))
      DO 320 I=1,J
      IF(.NOT.PLTFLG) GO TO 316
      IF(UNIT(10)) 311,920,920
 311  DO 313 K=4,N
      X(K)=Y(I,K)
```

```
 313  CONTINUE
      BUFFER OUT(10,1)(X(4),X(N))
 316  Y(I,2)=Y(I,N2)
      Y(I,3)=Y(I,203)
 320  CONTINUE
      IF(UNIT(10)) 325,920,920
 325  N=4
      N2=2
      GO TO (17,30),IDIR
C
C     INTERNAL ERROR STOP
C
 15   STOP 6666
C
C     ITERATION DONE - FIND OUT WHAT TO DO NEXT
C
 800  N=N-1
      IF(LAST) GO TO 820
      LIMFLG=.TRUE.
      LOG=CHECK(N)
      IF(IGCNT.GT.IGRLM) GO TO 910
 805  IF(.NOT.LAST) GO TO 7
C
C     CALCULATE PRINT INTERVAL - ABOUT 100 VALUES WILL BE PRINTED
C     AT VARYING SPACINGS WHICH DEPEND ON THE BEHAVIOR OF THE CURVE
C
      NPRNT=NO/100
      GO TO 7
C
C     WAS THIS CHECK ERROR GENERATED ON THE LAST TIME AROUND
C
 810  IF(IGCNT.GT.IGRLM) GO TO 910
      IF(.NOT.LAST2) GO TO 805
C
C     PRINT OUT LAST VALUE
C
 820  IF(IRKCNT.NE.0) PRINT 986,IRKCNT
C
C     DUMP OUT THE LAST BUFFER TO TAPE IF REQUIRED
C
      IF(.NOT.PLTFLG.OR.N.LT.4) GO TO 5
      BUFFER OUT(10,1)(X(4),X(N))
      IF(UNIT(10)) 830,920,920
 830  DO 840 I=1,J
      DO 835 K=4,N
      X(K)=Y(I,K)
 835  CONTINUE
      BUFFER OUT(10,1)(X(4),X(N))
      IF(UNIT(10)) 840,920,920
 840  CONTINUE
      GO TO 5
C
 900  CONTINUE
      STOP 7777
C
C     TOO MANY TRIES IN THE GROSS ITERATION MODE
C
 910  PRINT 981
      GO TO 5
```

```
C
C     ERROR ON DISK WRITE
C
 920  PRINT 980
      STOP 6666
C
C     X INCREMENT EQUAL TO ZERO
C
 930  STOP 5555
C
C
 980  FORMAT(5X19HERROR ON DISK WRITE10(1H*))
 981  FORMAT(5X25HTOO MANY GROSS ITERATIONS 5X10(1H*))
 982  FORMAT(5X20HBINARY ITERATIONS = I3,7X19HGROSS ITERATIONS = I3,
     $   7X17HP-C ITERATIONS = I3)
 983  FORMAT(2I5,L5,I5)
 984  FORMAT(8X9HIRKCNT = I5)
 985  FORMAT(//13X4HY(0)11X5HYP(0))
 986  FORMAT(//5X28HRUNGE-KUTTA MIN. STEP TAKEN I5,7H TIMES.//)
 987  FORMAT(//2X2HN09X      1HX11X6HY(1,N)8X6HY(2,N)10X2HYS12X2HYT
     $   11X5HY0(1) 9X5HY0(2)//)
 989  FORMAT(1H15X1HN9X1HH14X1HX14X2HY112X2HY2)
 995  FORMAT(///)
 997  FORMAT(8X1HJ14X4HX(0)15X5HDX(0)15X4HXLIM15X4HTOL110X5HRKTOL11X
     $   4HEPS1 8X5HGRADJ)
C
C     THE FOLLOWING NEED TO BE CHANGED WHEN GOING FROM SINGL TO
C        DOUBLE PRECISION
C
 988  FORMAT(1XI5,1P9E14.5)
 990  FORMAT(I5,F5.0,7E10.0)
 991  FORMAT(8E10.0)
 992  FORMAT(1XA2,2XI5,1P8E15.6//11X8E15.6)
 993  FORMAT(1H05X1P8E15.5)
 994  FORMAT(//5X17HMIN. STEP SIZE = 1PE15.5//
     $   5X17HTARGET VALUES ARE  //5(1P5E15.5/))
 996  FORMAT(1H15X8HALPHA = 1PE13.6,5X7HBETA = E13.6//)
 998  FORMAT(5XI5,1P6E15.6/11X8E15.6)
 999  FORMAT(1H05XI5,1P3E20.5,0P3F15.5,F12.3)
C
      END
```

```
      LOGICAL FUNCTION CHECK(N)
C
      COMMON/LIMITS/YLL,YUL,DX,IGCNT
      COMMON/LOGIC/ START,LAST,ILG,YPU,YPL,IMX,LIMFLG,IBCN,GRADJ,J
      LOGICAL START,LAST,LIMFLG
      COMMON YY(8,203),YTARG(8),YSTRT(8)
     $   ,X(203),YTEMP(8),ZTEMP(8),RK1(8),RK2(8),RK3(8),RK4(8),YP(8,10)
C
      Y=YY(J-1,N)
      IF(Y.LT.YLL.OR.Y.GT.YUL.OR.LIMFLG) GO TO 30
      CHECK=.FALSE.
      RETURN
C
C     GROSS ADJUSTMENT (START=TRUE) OR BINARY SEARCH (START=FALSE)
C
 30   YP0=YSTRT(J)
      Y1=YTARG(J-1)
      CHECK=.TRUE.
      IF((Y.GE.YUL.OR.Y.GE.Y1).AND.DX.GT.0.0) GO TO 200
      IF((Y.LE.YLL.OR.Y.LE.Y1).AND.DX.LT.0.0) GO TO 200
C
C     LOWER LIMIT EXCEEDED OR END OF INTERVAL AND TOO LOW
C
      YPL=YP0
      IF(.NOT.START) GO TO 300
      IGCNT=IGCNT+1
      IF(ILG.GT.0) GO TO 290
      ILG=-1
      IF(YP0.GT.0.0) GO TO 230
 130  YSTRT(J)=YP0/GRADJ
      RETURN
C
C     UPPER LIMIT EXCEEDED OR END OF INTERVAL AND TOO HIGH
C
 200  YPU=YP0
      IF(.NOT.START) GO TO 300
      IGCNT=IGCNT+1
      IF(ILG.LT.0) GO TO 290
      ILG=+1
      IF(YP0.GT.0.0) GO TO 130
 230  YSTRT(J)=YP0*GRADJ
      RETURN
C
C     END OF THE GROSS ADJUSTMENT PHASE
C
 290  START=.FALSE.
C
C     BINARY SEARCH MODE
C
 300  YSTRT(J)=(YPU+YPL)/2.0
      IBCN=IBCN+1
C
C     HAVE WE TRIED ENOUGH TIMES.
C
      IF(IBCN.GT.IMX) LAST=.TRUE.
      RETURN
      END
```

```
      FUNCTION Z(I,XN,Y)
      COMMON/LOGIC/START,LAST,ILG,YS,YT,IMX,LIMFLG,I9CN,GRADJ,J
      LOGICAL START,LAST,LIMFLG
      DIMENSION Y(8)
      COMMON/AB/ ALPHA2,BETA
C
C     INTERNAL STATEMENT FUNCTION TO GET AROUND CHANGE TO DOUBLE EASILY
C
      BBS(A)=ABS(A)      $    ZIGN(A,B)=SIGN(A,B)
C
      IF(J.EQ.3) GO TO 100
      GO TO (1,2),I
 1    Z=Y(2)
      RETURN
 2    T=1.0-BETA*Y(2)
      Z=ALPHA2*Y(1)*ZIGN(BBS(T)**(1.0/3.0),T)
      RETURN
C
C
 100  GO TO (101,102,103),I
 101  Z=-Y(2)
      RETURN
 102  Z=Y(3)
      RETURN
 103  T=1.0-BETA*Y(3)
      Z=ALPHA2*Y(2)*ZIGN(BBS(T)**(1.0/3.0),T)
      RETURN
      END
```

```
      PROGRAM PLTDEQ(INPUT=1011B,OUTPUT=1011B,TAPE10=100B,TAPE49)
C
C     THIS PROGRAM PLOTS THE OUTPUT PRODUCED BY PROGRAM DIFFEQ
C
      COMMON X(200),E(200),U(200),R(200)
      INTEGER UP,DN
C
      DATA DN,UP/2,3/
      DATA LFN/10/
C
C     SL-     SCALE LENGTHS
C     --SP    SEPARATION BETWEEN EACH PLOT
C     XLM     DISTANCE FROM LEFT LIMIT (LEFT MARGIN)
C
C     READ IN LIMITS FOR THE PLOTS
C
      READ 999,XMIN,XMAX,UMIN,UMAX,EMIN,EMAX,RMIN,RMAX
      READ 999,SLX,SLU,SLE,SLR,UESP,ERSP,XLM
C
C     PRINT PLOTTING PARAMETERS
C
      PRINT 995
      PRINT 996,XMIN,XMAX,UMIN,UMAX,EMIN,EMAX,RMIN,RMAX
      PRINT 993,SLX,SLU,SLE,SLR,UESP,ERSP,XLM
C
C     SET STARTING POINT FOR EACH PLOT IN INCHES
C
      US=0.0
      ES=US+SLU+UESP
      RS=ES+SLE+ERSP
C
C     CALCULATE SCALE FACTORS
C
      XSCL=-SLX/(XMAX-XMIN)
      USCL=SLU/(UMAX-UMIN)
      RSCL=SLR/(RMAX-RMIN)
      ESCL=SLE/(EMAX-EMIN)
C
C     INITIALIZE FILES
C
      REWIND LFN
      CALL INITIAL(8H9701512M,49,0,0,0)
C
C     DRAW AXES - FIRST GET AWAY FROM INITIAL BOX
C
      CALL PLOT(0.0,12.0,-UP)
      CALL PLOT(0.0,-XLM,-UP)
C
C     DRAW X AXIS FOR U
C
      CALL PLOT(0.0,-SLX,DN)
      CALL PLOT(US,0.0,UP)
C
C     DRAW Y AXES - U FIRST, FOLLOWED BY E AND R
C
      CALL PLOT(US+SLU,0.0,DN)
      CALL PLOT(ES,0.0,UP)
      CALL PLOT(ES+SLE,0.0,DN)
      CALL PLOT(RS,0.0,UP)
```

```
      CALL PLOT(RS+SLR,0.0,DN)
C
C     CALCULATE PLACE FOR R ZERO AND PLOT X AXIS
C
      ROF = RS-RSCL*RMIN
      CALL PLOT(ROF,0.0,UP)
      CALL PLOT(ROF,-SLX,DN)
C
C     E X AXIS
C
      CALL PLOT(ES,-SLX,UP)
      CALL PLOT(ES,0.0,DN)
C
C
C     START TO READ DATA - SOME OF THE OPERATIONS ARE OVERLAPPED WITH
C        EXECUTION
C
 20   BUFFERIN(LFN,1)(X(1),X(200))
      IF(UNIT(LFN)) 30,800,900
 30   N=LENGTH(LFN)
      BUFFER IN(LFN,1)(U(1),U(N))
      X(1)=(X(1)-XMIN)*XSCL
      IF(UNIT(LFN)) 40,910,900
 40   BUFFER IN(LFN,1)(E(1),E(N))
      CALL PLOT((U(1)-UMIN)*USCL+US,X(1),UP)
      DO 50 I=2,N
      X(I)=(X(I)-XMIN)*XSCL
      CALL PLOT((U(I)-UMIN)*USCL+US,X(I),DN)
 50   CONTINUE
      IF(UNIT(LFN)) 60,910,900
 60   BUFFER IN (LFN,1)(R(1),R(N))
      CALL PLOT((E(1)-EMIN)*ESCL+ES,X(1),UP)
      DO 70 I=2,N
      CALL PLOT((E(I)-EMIN)*ESCL+ES,X(I),DN)
 70   CONTINUE
      IF(UNIT(LFN)) 80,910,900
 80   CALL PLOT((R(1)-RMIN)*RSCL+RS,X(1),UP)
      DO 90 I=2,N
      CALL PLOT((R(I)-RMIN)*RSCL+RS,X(I),DN)
 90   CONTINUE
      GO TO 20
C
C     DONE - NORMAL E O F
C
 800  CALL PLOT(0.0,0.0,999)
      PRINT 994
      STOP 7777
C
C     PARITY ERROR ON READ.
C
 900  PRINT 997
      CALL PLOT(0.0,0.0,999)
      STOP 6667
C
C     UNEXPECTED E O F
C
 910  PRINT 998
      CALL PLOT(0.0,0.0,999)
      STOP 6666
```

```
C
 993  FORMAT(///42X1HX14X1HU14X1HE14X1HR/
     $   5X22HAXIS LENGTHS IN INCHES 3X4F15.3/
     $   5X16HPLOT SEPARATIONS 40XF6.3,9XF6.3/
     $   5X12HLEFT MARGIN 11XF6.3//)
 994  FORMAT(///45X*P L O T T I N G    D O N E*)
 995  FORMAT(1H15X*PLOT DIFFERENTIAL EQUATIONS*///14X4HXMIN12X4HXMAX12X
     $   4HUMIN12X4HUMAX12X4HEMIN12X4HEMAX12X4HRMIN12X4HRMAX)
 996  FORMAT(1P5X8E16.3)
 997  FORMAT(5X25HPARITY ERROR ON DISK READ  5X10(1H*))
 998  FORMAT(5X27HUNEXPECTED EOF - CHECK PLOT   5X10(1H*))
 999  FORMAT(8E10.0)
C
      E N D
```

REFERENCES

American Institute of Physics, 1972, American Institute of Physics Handbook, 3rd ed., Dwight E. Gray, ed., McGraw-Hill Co., New York.

Aspnes, D. E., 1973, Surface Science 37, 421.

Aspnes, D. E., and A. Frova, 1969, Solid State Commun. 7, 155.

Aspnes, D. E., P. Handler, and D. F. Blossey, 1968, Phys. Rev. 166, 921.

Ballik, E. A., 1971, Appl. Opt. 10, 689.

Beaglehole, D., 1968, Appl. Opt. 7, 2218.

Birth, G. S., and D. P. DeWitt, 1971, Appl. Opt. 10, 687.

Born, Max, and Emil Wolf, 1970, Principles of Optics, 4th ed., Pergamon Press, New York, p. 615.

Cardona, Manuel, 1969, Modulation Spectroscopy, Academic Press, New York.

Cardona, Manuel, and D. L. Greenaway, 1964, Phys. Rev. 133, A1685.

Chemical Rubber Company, 1961, Handbook of Chemistry and Physics, 43rd ed., Charles D. Hodgman, ed., Chemical Rubber Publishing Co., Cleveland, Ohio.

Cheyssac, P., R. Garrigos, R. Kofman, L. Penavaire, J. Richard, and A. Saissy, 1973, Surface Science 37, 683.

Comins, N. R., and P. D. Grant, 1973, Journal of Physics E 6, 751.

Dresselhaus, M. S., 1971, in The Physics of Semimetals and Narrow-Gap Semiconductors, Proceedings of the Conference held at Dallas, Texas, 20-21 March, 1970, eds. D. L. Carter and R. T. Bate, Pergamon Press, New York.

Ehrenreich, H., and H. R. Philipp, 1962, Phys. Rev. 128, 1622.

Falicov, L. M., and P. J. Lin, 1966, Phys. Rev. 141, 562.

Fatuzzo, E., and M. J. Merz, 1967, Ferroelectricity, Wiley-Interscience, John Wiley and Sons, Inc., New York, p. 64.

Feinleib, J., 1966, Phys. Rev. Letters 16, 1200.

Feynman, Richard P., Robert B. Leighton, and Matthew Sands, 1963, The Feynman Lectures on Physics, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, vol. 1.

Garrigos, Robert, Richard Kofman, and Jacques Richard, 1973, Comptes Rendus Acad. Sc. Paris 276, B55.

Golin, Stuart, 1968, Phys. Rev. 166, 643.

Hahn, Robert E., 1973, Research Associate, Optical Sciences Center, personal communication.

Hansen, Wilford N., and Arnold Prostak, 1968, Phys. Rev. 174, 500.

Heavens, O. S., 1965, Optical Properties of Thin Solid Films, Dover Publications, Inc., New York.

Holbrook, J. Alfaro, and R. E. Hummel, 1973, Rev. Sci. Instr. 44, 463.

Huang, Kerson, 1963, Statistical Mechanics, John Wiley and Sons, Inc., New York.

Hummel, R. E., D. B. Dove, and J. Alfaro Holbrook, 1970, Phys. Rev. Lett. 25, 290.

Hummel, R. E., J. Alfaro Holbrook, and J. B. Andrews, 1973, Surface Science 37, 717.

Hunter, W. R., D. W. Angel, and R. Tousey, 1965, Appl. Opt. 4, 891.

Ishibashi, Y., and H. L. Stadler, 1969, J. Phys. Chem. Solids, 30, 2113.

Jackson, John David, 1962, Classical Electrodynamics, John Wiley and Sons, Inc., New York.

Jaffe, B., W. Cook, and H. Jaffe, 1971, Piezoelectric Ceramics, Academic Press, New York.

Jeppesen, M. A., S. R. Flagg, and J. D. Rancourt, 1963, Am. Jour. of Phys. 31, 860.

Kittel, Charles, 1958, Elementary Statistical Physics, John Wiley and Sons, Inc., New York.

Kittel, Charles, 1966, Introduction to Solid State Physics, 3rd ed., John Wiley and Sons, Inc., New York.

Kittel, Charles, 1971, Introduction to Solid State Physics, 4th ed., John Wiley and Sons, Inc., New York.

Kottke, Michael, 1974, Thesis, Univ. of Arizona.

McCracken, Daniel D., and William Dorn, 1964, Numerical Methods and FORTRAN Programming, John Wiley and Sons, Inc., New York, chapter 10.

McIntyre, J. D. E., 1973a, in Advances in Electrochemistry and Electrochemical Engineering, vol. 9, ed. R. H. Muller, Wiley-Interscience, John Wiley and Sons, Inc., pp. 61-166.

McIntyre, J. D. E., 1973b, Surface Science 37, 658.

McIrvine, E. C., 1966a, Phys. Rev. 148, 528.

McIrvine, E. C., 1966b, Surface Science 5, 171.

McNatt, J. L., and P. Handler, 1969, Phys. Rev. 178, 1328.

Munroe, Donald, 1973, "The Heterodyning Lock-In Amplifier," IAN-23, Ithaco, Inc., Ithaca, New York.

Nelson, Alfred L., Karl. W. Folley, and Max Coral, 1960, Differential Equations, 2nd ed., D. C. Heath and Co., Boston.

Parsons, B. J., 1969, Phys. Rev. 182, 975.

Peterson, C. W., and Bruce Knight, 1973, J.O.S.A. 63, 1238.

Prostak, Arnold, and Wilford N. Hansen, 1967, Phys. Rev. 160, 600.

Radio Corporation of America, 1972, Tips on the Use of Photomultipliers, publication PIT-711, Harrison, New Jersey.

Ralston, Anthony, and Herbert S. Wilf, 1960, Mathematical Methods for Digital Computers, Vol. 1, John Wiley and Sons, Inc., New York.

Reitz, John R., and Frederick J. Milford, 1967, Foundations of Electromagnetic Theory, 2nd ed., Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, chap. 5.

Sceats, M. G., and G. C. Morris, 1972, Phys. Stat. Sol. (a) 14, 643.

Seraphin, B. O., 1964, in Physics of Semiconductors (Proc. of 7th Int. Conf.), Dunod, Paris, and Academic Press, New York, p. 165.

Seraphin, B. O., 1967, J. de Phys. 28, C3-73.

Seraphin, Bernhard O., 1972a, in Semiconductors and Semimetals, vol. 9, eds. R. K. Willardson and A. C. Beer, Academic Press, Inc., New York, pp. 1-149.

Seraphin, B. O., 1972b, in Optical Properties of Solids, ed. F. Abeles, North Holland Publishing Co., Amsterdam, chap. 4.

Seraphin, B. O., and R. B. Hess, 1965, Phys. Rev. Lett. 14, 138.

Sommer, A. H., 1973, Appl. Opt. 12, 90.

Stavroudis, Orestes N., 1973, Professor, Optical Sciences Center, personal communication.

Stedman, M., 1968, Chem. Phys. Lett. 2, 457.

Stedman, M., 1970, Symp. Faraday Soc., 4, 64.

Stern, F., 1963, in Solid State Physics, vol. 15, eds. F. Seitz and D. Turnbull, Academic Press, Inc., New York, pp. 327-408.

Stone, John M., 1963, Radiation and Optics, McGraw-Hill Book Co., New York.

Sze, S. M., 1969, Physics of Semiconductor Devices, Wiley-Interscience, John Wiley and Sons, Inc., New York.

Weiss, David E., 1971, Ph.D. Thesis, Temple Univeristy, Philadelphia.

Wolfe, William L., 1965, ed., Handbook of Military Infrared Technology, U. S. Government Printing Office, Washington.

Wooten, Frederick, 1972, Optical Properties of Solids, Academic Press, Inc., New York.