

AD-A098 578

WASHINGTON UNIV ST LOUIS MO DEPT OF ELECTRICAL ENGI--ETC F/G 20/12
SEMICONDUCTOR MILLIMETER WAVELENGTH ELECTRONICS.(U)

OCT 80

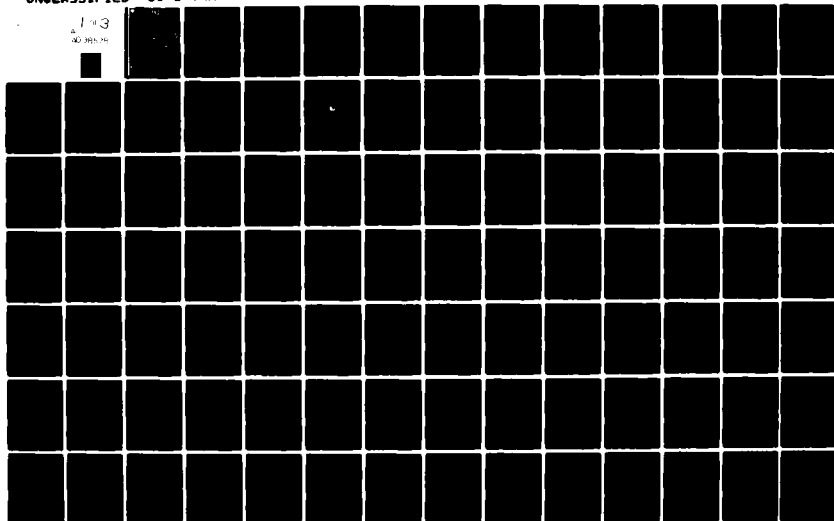
N00014-79-C-0840

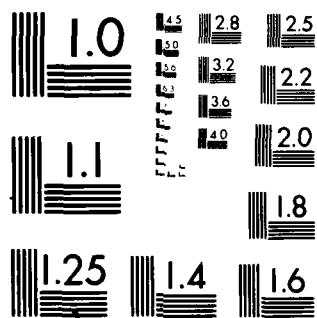
UNCLASSIFIED

80-1-ONR

NL

1-3
AD 80-1-ONR





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A



WASHINGTON
UNIVERSITY
IN ST LOUIS

LEVEL

12

AD A 098 578

SEMICONDUCTOR MILLIMETER WAVELENGTH ELECTRONICS

Department of Electrical Engineering
Washington University
St Louis, MO 63130

DTIC
ELECTE
MAY 5 1981

12) 230

11) Oct 30

13) N00014-79-2-0840

9) ANNUAL REF ONR 80-1
Sep 79 - Aug 80

Office of Naval Research
Arlington, VA 22217

14) 80-2-ONR

DTIC FILE COPY

Reproduction, in whole or in part, is permitted for any purpose of the U.S. Government.

Contract: N00014-79-C-0840 New
Contract Authority: NR SRO-004

Approved for public release; distribution unlimited. 81 5 05 040
403306

TABLE OF CONTENTS

No.	Page
1. Introduction	
1.1 Scope	1-1
1.2 Program Participants	1-2
1.3 Summary	1-2
1.4 Talks and Papers Presented	1-9
2. Materials Growth	
2.1 Epitaxial Material Growth for Device Studies (J. Teng and R. E. Goldwasser)	2-1
2.1.1 Vapor Phase Epitaxial Growth of GaAs	2-1
2.1.2 Substrate Preparation	2-6
2.1.3 Results and Discussion	2-8
2.1.4 Sample Fabrication for Velocity-Field Measurements	2-11
2.1.4.1 Epitaxial Layer	2-14
2.1.4.2 Ohmic Contact Fabrication	2-15
2.1.4.3 Schottky-Barrier Fabrication	2-15
2.2 Impurity Incorporation and Redistribution During Epitaxial Growth (H. Rohdin and M. Muller)	2-19
2.2.1 Introduction	2-19
2.2.2 Assumptions	2-20
2.2.2.1 W and N Treatment: List of Assumptions, Their Consequences and Validity	2-20
2.2.2.2 New Treatment	2-24
2.2.3 The Problem	2-25

TABLE OF CONTENTS
(continued)

No.	Page
2.2.3.1 The Electric Field Problem	2-26
2.2.3.1.1 The Exact Approach	2-28
2.2.3.1.2 Modified W and N Approach	2-32
2.2.3.2 The Impurity Redistribution Problem ..	2-33
2.2.3.2.1 Numerical Approach to the Re- distribution Problem	2-35
2.2.3.3 How to Chose Δx , Δt , $i_{\max}$, and $J_{\max}$...	2-39
2.2.4 Final Remarks	2-41
2.3 Diffusion and Electronic Properties of Cr in GaAs ..	2-43
(H. Rohdin and M. Muller)	
2.3.1 Introduction	2-43
2.3.2 Certain Experimental Results on GaAs:Cr	2-45
2.3.2.1 Redistribution of Cr in GaAs	2-45
2.3.2.1.1 Annealing of Cr Without En- capsulation	2-45
2.3.2.1.2 Annealing of GaAs:Cr With En- capsulation	2-48
2.3.2.1.3 Diffusion of Cr into SI-GaAs: Cr	2-48
2.3.2.1.4 Epitaxial Growth of GaAs on GaAs:Cr	2-48
2.3.2.2 Electronic Properties of Cr in GaAs ..	2-51
2.3.2.2.1 Bulk Grown GaAs:Cr	2-51
2.3.2.2.2 Cr Indiffused in GaAs	2-53
2.3.2.2.3 SI-GaAs:Cr Annealed Under En- capsulation	2-53

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

TABLE OF CONTENTS
(continued)

No.	Page
2.3.2.2.4 Epitaxial Growth of GaAs on SI-GaAs:Cr	2-53
2.3.3 Zn Diffusion in GaAs	2-57
2.3.4 Well Established Electronic Properties of Cr in GaAs	2-61
2.3.5 Cr as an Interstitial Donor in GaAs	2-63
2.3.6 Site Energies and Cr Distribution Over Sites .	2-65
2.3.7 One-Electron Energies	2-68
2.3.7.1 Occupation of the One-Electron Energy Levels	2-68
2.3.8 Explicit Motivation for Autocompensating Cr in GaAs at Room Temperature (300°K)	2-71
2.3.8.1 The Resistivity as a Function of Fermi Level	2-71
2.3.8.2 The Fermi Level as a Function of Cr Concentration	2-72
2.3.9 Vacancies in GaAs	2-74
2.3.9.1 Equilibrium Reaction and Mass Action Laws	2-74
2.3.9.2 Resulting Concentration Profiles, Diffusion	2-77
2.3.9.3 Electric Activity of Vacancies	2-79
2.3.10 Super Fast Interstitial Diffusion	2-79
2.3.11 A Model for the Diffusion and Electronic Properties of Cr in GaAs	2-80
2.3.12 Comments on the Experimental Results	2-83
2.3.12.1 Redistribution of Cr Without En- capsulation	2-83

TABLE OF CONTENTS
(continued)

No.	Page
2.3.12.1.1 Annealing of Cr Without Encapsulation	2-83
2.3.12.1.2 Annealing of GaAs:Cr Without Encapsulation	2-84
2.3.12.1.3 Diffusion of Cr into SI-GaAs: Cr	2-85
2.3.12.1.4 Epitaxial Growth of GaAs on GaAs:Cr	2-85
2.3.12.2 Electronic Properties of GaAs:Cr	2-86
2.3.12.2.1 Bulk Grown GaAs:Cr	2-86
2.3.12.2.2 Cr Indiffused into GaAs	2-86
2.3.12.2.3 SI-GaAs:Cr Anneal With Encapsulation	2-87
2.3.12.2.4 Epitaxial Growth of GaAs on SI-GaAs:Cr	2-88
2.3.13 Conclusion	2-89
 3. Existing Device Studies	
3.1 Harmonic Generation in Transferred Electron Devices (R. E. Goldwasser)	3-1
3.1.1 Introduction	3-1
3.1.2 Harmonic Oscillator Circuits	3-3
3.1.2.1 Coaxial Circuit	3-3
3.1.2.2 Disc Circuit	3-5
3.1.2.3 Waveguide Circuit	3-6
3.1.3 Experimental Results with Flat Doped Diodes .	3-7
3.1.4 Harmonic Power of Flat Profile Wafers	3-9

TABLE OF CONTENTS
(continued)

No.	Page
3.1.5 Harmonic Generation in Cathode Notch Diodes and Future Plans	3-13
3.2 Bulk and Contact Phenomena in Millimeter Wave Diodes	3-20
(P. Chen, F. J. Rosenbaum and R. E. Goldwasser)	
3.2.1 Introduction	3-20
3.2.2 Avalanche Breakdown Analysis	3-20
3.2.2.1 Avalanche Breakdown	3-22
3.2.2.2 I-V Characteristic Relations	3-23
3.2.2.3 Large Signal Analysis	3-24
3.2.3 Mott-Barrier Diffusion Model	3-32
3.2.3.1 I-V Characteristic Relation	3-35
3.2.3.2 Image Force	3-36
3.2.4 One-Dimensional Schottky-Barrier Diode Finite Element Analysis	3-38
3.2.5 Two-Dimensional Schottky-Barrier Diode Analysis	3-46
3.2.5.1 Finite Element Analysis	3-46
3.2.6 Conclusions and Plans for Next Period	3-55
4. New Device Concepts	
4.1 Inertial Transport in Semiconductors - Introduction (B. A. Shrauner and M. Muller)	4-1
4.2 Ballistic Transport and Instabilities	4-3
4.2.1 Ballistic Transport and Multivalued Current- Voltage Relations	4-3
4.2.2 Temporal Instabilities	4-9

TABLE OF CONTENTS
(continued)

No.	Page.
4.2.3 Negative Differential Resistivity and Spatial Instabilities	4-14
4.2.4 Finite Temperature Effects in Ballistic Transport	4-17
4.2.5 The Effect of Non-Parabolic Bands	4-21
4.3 Zener Oscillations	4-28
(M. Muller)	
4.3.1 Introduction	4-28
4.3.2 The Effective Mass Theorem	4-31
4.3.3 A Static Test of the Theorem	4-43
4.3.4 Motion in a Uniform Electric Field	4-46
4.3.5 Tunneling and the Stark Ladder	4-56
4.3.6 Tunneling	4-58
4.3.7 Detailed Calculation of the Tunneling Probability	4-68

Semiconductor Millimeter Wavelength Electronics

1. Introduction

1.1 Scope

→ This report describes progress and results obtained in the first year of the subject research program. The purpose of the program is to investigate new and innovative approaches to the generation and amplification and detection of electromagnetic signals in the millimeter wave spectrum. It incorporates theoretical and applied studies in four areas:

- i) Semiconductor material synthesis and growth
- ii) Electrical characterization of these materials
- iii) Device modeling and fabrication
- iv) New device concepts.

It is our desire to understand more fully the origin and consequences of the physical limitations of semiconductor materials appropriate for use at millimetric wavelengths, and through this knowledge to suggest new materials and/or devices with improved properties.

During the period 1 September 1979 to 30 August 1980, we have initiated the growth of heavily doped GaAs, developed an apparatus for the measurement of high field transport properties, examined some impurity effects in GaAs, and laid the theoretical groundwork for a systematic study of inertial effects in semiconductor charge transport, including the concept of Zener oscillations. We have also carried out Gunn oscillator experiments that have yielded power levels of 5-6 mW at frequencies near 150 GHz. A new millimeter wave mixer diode configuration is also being explored. Progress is summarized in this introductory section. Detailed discussions are presented in the body of the report.

1 - 1

1.2 Program Participants

The following lists those who have contributed to project activities during this reporting period:

<u>FACULTY</u>	<u>GRADUATE ASSISTANTS</u>	<u>UNDERGRADUATE ASSISTANTS</u>
B. Abraham-Shrauner	J. Bornholdt	S. Ewall
R. E. Goldwasser, Co. P.I.	P. Chen	M. Riess
M. W. Muller	G. Homsey	J. Wachsman
F. J. Rosenbaun, P.I.	H. Rohdin	J. Wendt
C. M. Wolfe	J. Teng	
	S. Von Rump	

1.3 Summary

Epitaxial Material Growth for Device Studies

An Arsenic Trichloride-Gallium-Hydrogen epitaxial reactor has been built, calibrated, and operated for the growth of multilayer millimeter wave GaAs structures. The system employs hydrogen sulfide gas doping to obtain the rapid changes in the epitaxial layer concentration that are required in millimeter wave devices. The system is of the rolling furnace design that permits rapid heating and cooling of the wafers. The rapid cooling minimizes dopant redistribution in the layers after growth. Aluminum tubing and fittings have been used in the H_2S lines to minimize doping history effects. The system background doping can be controlled by setting the $AsCl_3$ bubbler control temperature. The doping flow rate has been correlated to the epitaxial layer concentration over three orders of magnitude. The system has been designed for GaAs, however, the system is compatible with the PCL_3 -In Indium Phosphide system requirements. The present system yields highly uniform mirror like layers and requires no significant vapor etching to obtain low dislocation interfaces. Layers as thin as $.15\mu m$ appear uniform.

Impurity Incorporation and Redistribution During Epitaxial Growth

The incorporation of impurities in epitaxially grown layers of semiconductors and their subsequent redistribution often leads to impurity and carrier density profiles that are unforeseen, and that may be unfavorable for certain applications, for example, the formation of high resistance or inverted polarity layers in Gunn devices. An explanation of these effects has been proposed which takes into account the effect on the impurity redistribution of the built-in electric fields. The original work on this phenomenon assumes that during growth the newly incorporated impurities are always in diffusive equilibrium, and that there is no outdiffusion from the substrate.

We have determined that under the usual growth conditions these idealizations cannot be expected to hold, and we have developed a partly analytical and partly numerical model that mirrors the time-dependent behavior more realistically. Programming is nearly complete, and results of the computation should be available during the next reporting period.

Diffusion and Electronic Properties of Cr in GaAs

The most common method for producing semi-insulating GaAs is chromium doping. Remarkable features of this technique are: its insensitivity to Cr concentration during bulk growth, and the variety of redistribution behavior during processing.

We have surveyed the experimental literature on Cr doping and diffusion, and we have developed a model that appears capable of accounting for the observations. The central features of this model are: The existence of at least two electronic states of the Cr ion, a mobile, interstitial donor level, and a set of relatively immobile, substitutional acceptor levels; and the thermodynamic equilibrium of vacancies at the surface and in the bulk.

The initial distribution of Cr among its states is critically dependent on the conditions of growth and impurity incorporation: temperature, other impurities, vacancy concentration, etc. Redistribution during processing depends both on the initial distribution and on processing conditions: temperature, atmosphere, presence or absence of encapsulation, and new impurities.

The behavior of the model is consistent with observed electronic properties and concentration profiles, and is not in conflict with nuclear resonance and optical studies of Cr levels.

System for the Transport Analysis of Semiconductor Materials

The conception, design, and realization of a system for the transport analysis of semiconductor materials is presented. A survey of the relevant literature on transport measurements examines previously developed theory and experimental methods, and some unresolved problems are defined. Based on this survey, a measurement system has been developed using the microwave time-of-flight method. An electron beam is deflected past an externally biased semiconductor sample, producing a time-varying space charge which travels through the sample and induces a current in the external circuit. The detected signal is a function of the transport mechanisms of the material, which can be calculated by a proper interpretation of the measured results. The two properties of interest are velocity and diffusion as a function of applied electric field.

The existing theory of velocity-field measurement is presented, and extended to the general case of nonuniform electric field across the sample caused by the doping profile. An algorithm is described which permits the construction of the velocity-field curve based on microwave phase measurements, taking into account the nonuniform doping density of the sample and hence the non-constant

electric field in it. A model for the effects of diffusion on the measured signal is presented for the case of uniform electric field. The resulting equation is solved and evaluated in terms of the error introduced by assuming that the measured signal is unaffected by diffusion. This simple model predicts errors as high as 50% for currently accepted values of velocity and diffusion for GaAs.

Contributions made to the physical design of the microwave time-of-flight method include the following: Two modulation frequencies, 1 GHz and 12 GHz, are available, chosen according to the theoretical considerations in order to provide the most accurate data possible and to allow both velocity and diffusion measurements. The apparatus permits conventional time-of-flight measurement if desired. The electron accelerating voltage is adjustable, making beam penetration depth into the sample variable. The sample may be heated or cooled. Calibration of the microwave signals is established by direct measurement without a sample. A low noise microwave amplifier provides the signal gain necessary in this case. Measurement accuracy is greatly enhanced by use of an automatic network analyzer system. Finally, the entire experimental procedure is under computer control.

A full report describing the system is presented under separate cover.

Harmonic Generation in Transferred Electron Devices

The harmonic generation in millimeter wave Gunn diodes has been investigated in the 35-150 GHz frequency range. Power outputs at the fundamental, second and third harmonics have been measured on diodes fabricated from flat doped epitaxial wafers and a cathode notch structure. The fundamental power outputs of all the devices tested, peaked at near 40 GHz regardless of their active layer lengths. The harmonic power of the packaged devices

has been measured in three circuit configurations. The second and third harmonic powers have been found to decrease extremely rapidly above 100 GHz for all of the flat doped devices. Significant improvement in the harmonic powers above 100 GHz has been measured on devices that have a cathode doping notch. A power output of 5-6 mW has been observed at 144 GHz on the cathode notch devices. The highest power obtained from the flat doped devices at this frequency is .35 mW.

The improvement which is observed in the harmonic output of the cathode notch structure is likely to be due to a reduction in the formation time of the accumulation layer and a reduction in the cathode dead space that has been predicted by others using Monte Carlo analysis.

Bulk and Contact Phenomena in Millimeter Wave Diodes

In order to reasonably model the behavior of millimeter wave diodes it is necessary to understand processes which occur in the bulk material and at the contacts. We have developed closed form expressions for the current-voltage (I-V) relation describing avalanche breakdown in GaAs in both the small-and-large-signal regimes. Similarly, we have formulated the I-V relation for a particular contact system known as the Mott barrier, which has the property of nearly constant junction capacitance under forward or reverse bias.

The finite element method has been applied to the analysis of Schottky barrier diodes in both one and two dimensional simulations. For GaAs, both the electron velocity and the diffusion coefficient are field dependent. In the Schottky depletion region these quantities also vary in position due to the non-constant electric field. We have included these effects and examined the depletion approximations in a one dimensional Schottky barrier diode.

A new diode configuration suitable for use as a sub-harmonically pumped millimeter wave mixer has been conceived and is being analyzed with the methods described above. Design and performance predictions are under study and an effort has been initiated to fabricate the new structure.

Inertial Transport

Devices of interest in this project are sufficiently small or fast, or operate at high enough frequencies so that the mean free path of the carriers can be comparable with device dimensions, or the mean free time with switching times or oscillation periods. Under such operating conditions, carrier transport is not adequately described by the customary friction-dominated Ohm's law behavior and collision-free or nearly collision-free inertial or ballistic transport may be a more realistic model. Moreover, the achievement of such transport conditions appears desirable, because in this regime transit and response times need not be limited by the carrier saturation velocities of ohmic transport.

Inertial transport in solids is similar in some ways with vacuum electronics, but it differs from that technology in the presence of the fixed charge of immobile impurities, and in the non-Newtonian carrier dynamics imposed by the band structure. Existing accounts of inertial transport have not yet responded adequately to these differences.

We have derived the general solution for the steady-state potential of ballistic transport in a one-dimensional semiconductor as a function of the distance from the cathode for the space-charge limited diode with a constant effective mass. Since the solution implies some strange carrier behavior, we have checked the uniqueness of the current-voltage curve and its temporal and spatial stability. The current-voltage curve is singlevalued for large current

densities. A proof of that has been given and a relation for determining the smallest current for a single-valued relation has been stated.

The steady-state system is found marginally stable by a temporal stability analysis both for fluctuations that vary with the spatial coordinate in the carrier direction and those that vary with spatial coordinates transverse to that direction. The current density-electric field curve exhibits multivalued currents at a higher current than that found for the current-voltage curve. The negative differential resistance is current-controlled. As a result, filament formation which destroys the one-dimensional approximation or unstable device operation may occur when the current is multivalued. Finite temperature effects have been included in a simple fluid model for the electrons.

We have also developed the steady-state theory of a space-charge limited solid-state diode with carriers whose effective mass varies with energy in a manner characteristic of the small-effective-mass direct gap semiconductors most promising for our applications. We find various modifications from the existing, constant m^* theory; for example, an asymptotic first power rather than three-halves power current-voltage relation. The modifications are significant under conditions expected to be realized in normal device operations.

Zener Oscillations

In the extreme high-voltage and/or high frequency submillimeter regime, collision lifetimes can be long compared with an oscillation period. Under these conditions carrier dynamics becomes entirely band-structure dominated, and in the usual semiclassical picture, carriers are expected to carry out real-space Zener oscillations.

The observability of this phenomenon, and of the equivalent Stark quantization of free-carrier orbits (the Stark ladder) has been questioned in the

literature. We have carried out an extensive critical review of this controversy and of its background. The detailed report on this material is intended to serve as a reference for the continuation of this work. We conclude that the phenomena limiting the realization of Zener oscillations are scattering and interband tunneling. It appears that conditions should be achievable under which the oscillations can be observed.

1.4 Talks and Papers Published

M. W. Muller, "Inertial Transport with Non-Parabolic Bands", late news paper given 9-23-80 at The International Symposium on GaAs and Related Compounds, Vienna, Austria.

2. Materials Growth

2.1 Epitaxial Material Growth for Device Studies

J. Teng and R. E. Goldwasser

2.1.1 Vapor Phase Epitaxial Growth of GaAs

The GaAs epitaxial layers are grown in an AsCl_3 -Ga- H_2 flow system of the type first described by Knight et al.¹ This appears to be intrinsically a very reliable system and has the advantage over other systems that all the starting materials are obtainable in a state of high purity. Efforts have been directed, first, to reducing impurities in the system, and second, to controlling physical growth processes which cause variations in electrical properties.

The primary objectives were to (i) set up a reactor of maximum cleanliness and leak-tightness so that residual impurities are minimized, (ii) maintain the purity of the system by careful preparation of substrates and prevention of impurity build-up from outside, (iii) optimize the growth procedure by identifying and controlling the significant parameters.

The reactor and furnace are shown schematically in figure 1. Each zone has an independent temperature controller. The temperature profile of the two-zone furnace is shown in figure 2. The reactor tube, seed holder and melt boat are constructed of quartz. All the gas lines up to the flow valves, except the aluminum H_2S dopant line, are stainless steel. The AsCl_3 bubbler is constructed of Pyrex and kept cooled by a constant temperature refrigerated circulator attached to the water bath.

After construction, the system was dismantled, the quartz and pyrex parts cleaned thoroughly in aqua regia and rinsed in deionized water, the stainless

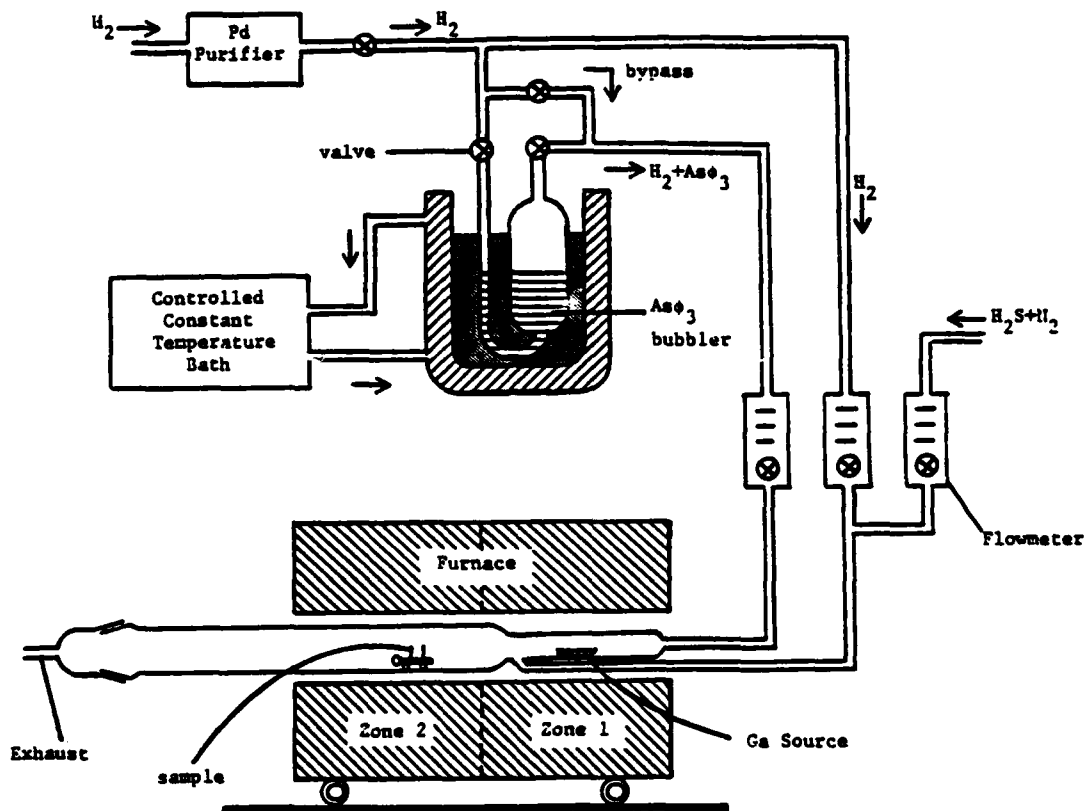


Figure 1. Schematic of AsCl₃-Ga-H₂ vapor phase epitaxial system

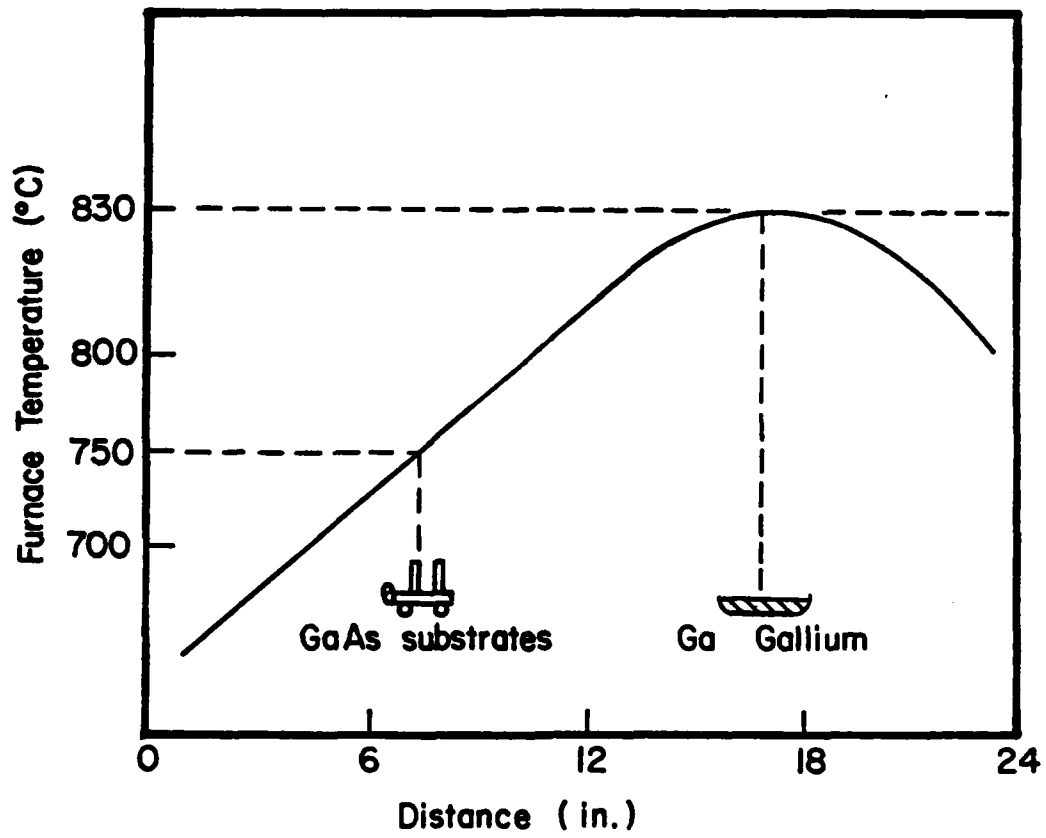


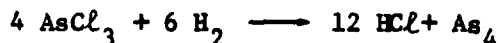
Figure 2. Temperature profile inside reactor

steel tubing cleaned using isopropyl and trichloro ethylene one part each, and dried by nitrogen. This cleaning procedure removes surface impurities which may be introduced by glassblowing. The system has been checked for leak-tightness at each joint.

The AsCl_3 (Mining and Chemical Products, 99.999% pure, in 100 g ampoules) is then loaded and hydrogen passed through it for several hours to distill off the first fraction. If this is not done, the first epitaxial layers have higher carrier concentration than normal. A final clean up of the system to remove impurities is carried out by running the furnace 20°C hotter than a normal deposition run.

The gallium (99.999% pure, 25 g ingots) is then loaded in the source boat. The growth process occurs in two stages, saturation of the source and growth of the layer itself. For the first stage, the Ga source (melting point at 29.78°C), is loaded in zone 1 of the furnace, whose temperature is raised to a level around 840°C . High purity hydrogen from a palladium diffuser bypasses the AsCl_3 bubbler. The H_2 flow rate is set to 120 ml/sec. The H_2 flow rate in the dopant line (10 ppm H_2S in N_2) is set to 200 ml/sec. After one hour, the valve of the bubbler is opened and the hydrogen is bubbled through the AsCl_3 source (kept at 10°C).

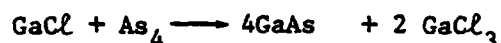
The hydrogen, as a carrier, transports the AsCl_3 into the Ga-source side of the furnace. The initial reaction taking place when the gas mixture heats up is



While the HCl reacts with the Ga source to form GaCl and GaCl_3 , the arsenic vapor is completely absorbed by the liquid Ga source until saturation occurs at -2.25

atomic percent of arsenic. Then a GaAs skin forms around the Ga source and the arsenic vapor passes on down the furnace tube, together with GaCl vapor produced by reaction between the HCl and the gallium.

At this time, the substrate is inserted, with its polished surface toward the stream, and the temperature in zone 2 is lowered to reach a steady temperature of about 750°C. The GaCl and As₄ vapors combine in zone 2, and the epitaxial GaAs begins to grow thereon according to the reaction



However, the temperatures (840°C in zone 1, 830°C in zone 2) and the flow rates are maintained for 6 hours, then the temperature controller in zone 2 is changed to 760°C. After 2 more hours of saturation under these conditions, the AsCl₃ and the furnace are turned off and hydrogen flow rate lowered to 50 ml/min. The Ga source is checked visually for saturation, and the system is ready for GaAs epitaxial layer growth.

Because of differences in reactor design, residual impurities and other factors, the optimum growth conditions vary from one reactor to another. Typical growth conditions are gallium temperatures from 800° to 850°C, growth temperatures from 720° to 760°C, growth temperature gradients from 5°C to 15°C cm⁻¹ hydrogen flow rate from 50 to 300 ml/min.

The growth data and procedure which we use, are as follow:

Ga source temperature: 830°C

Growth temperature: 755° - 760°C

Growth temperature gradient -17°C per inch

The furnace is turned on to the preceeding temperature settings first, while the reactor tubing is not in the furnace. After about one hour, the

temperatures in the furnace are in the steady state. Meanwhile, the prepared GaAs substrate is loaded in the growth region. The hydrogen flow rates in the carrier line (with the AsCl_3 bypassed) and the dopant line are adjusted to 220 ml/min and 200 ml/min respectively. The reactor tubing is then flushed at these two high flow rates of hydrogen for more than 20 minutes, then the hot furnace can be rolled in the growth position. The AsCl_3 is turned on and 10 minutes later, the furnace temperature again approaches its steady state value. After the AsCl_3 is turned on for 5 minutes, we reduce the hydrogen flows in both the carrier and dopant lines to 120 ml/min and 50 ml/min, respectively. At the same time the epitaxial growth is started.

We can increase the carrier concentration of the epitaxial layer by adding H_2S dopants in the gas stream to the growth region. The flow rate of H_2S dopants vs. the carrier concentration of the epitaxial layer measured by capacitance-voltage method is shown in Figure 3. After the specified growth time, the AsCl_3 is turned off (hydrogen bypassed) and the furnace is rolled away from the reactor tubing.

2.1.2 Substrate Preparation

The epitaxial layers were grown on chromium or silicon doped substrates. The crystallographic orientation is 2° off the (100) plane to prevent the formation of hillocks².

After the substrates are cleaved to desired dimensions, they are cleaned with organic solvents in a beaker. Trichloroethylene, acetone, and methanol in sequence are used for this purpose. The substrate is dried with bibulous paper and nitrogen. The sample is then etched in a $5\text{H}_2\text{SO}_4:1\text{H}_2\text{O}_2:1\text{H}_2\text{O}$ solution. This etchant is first cooled for 10 minutes after mixing, while it is being stirred. The sample is then put in the etchant and stirred for 5 minutes.

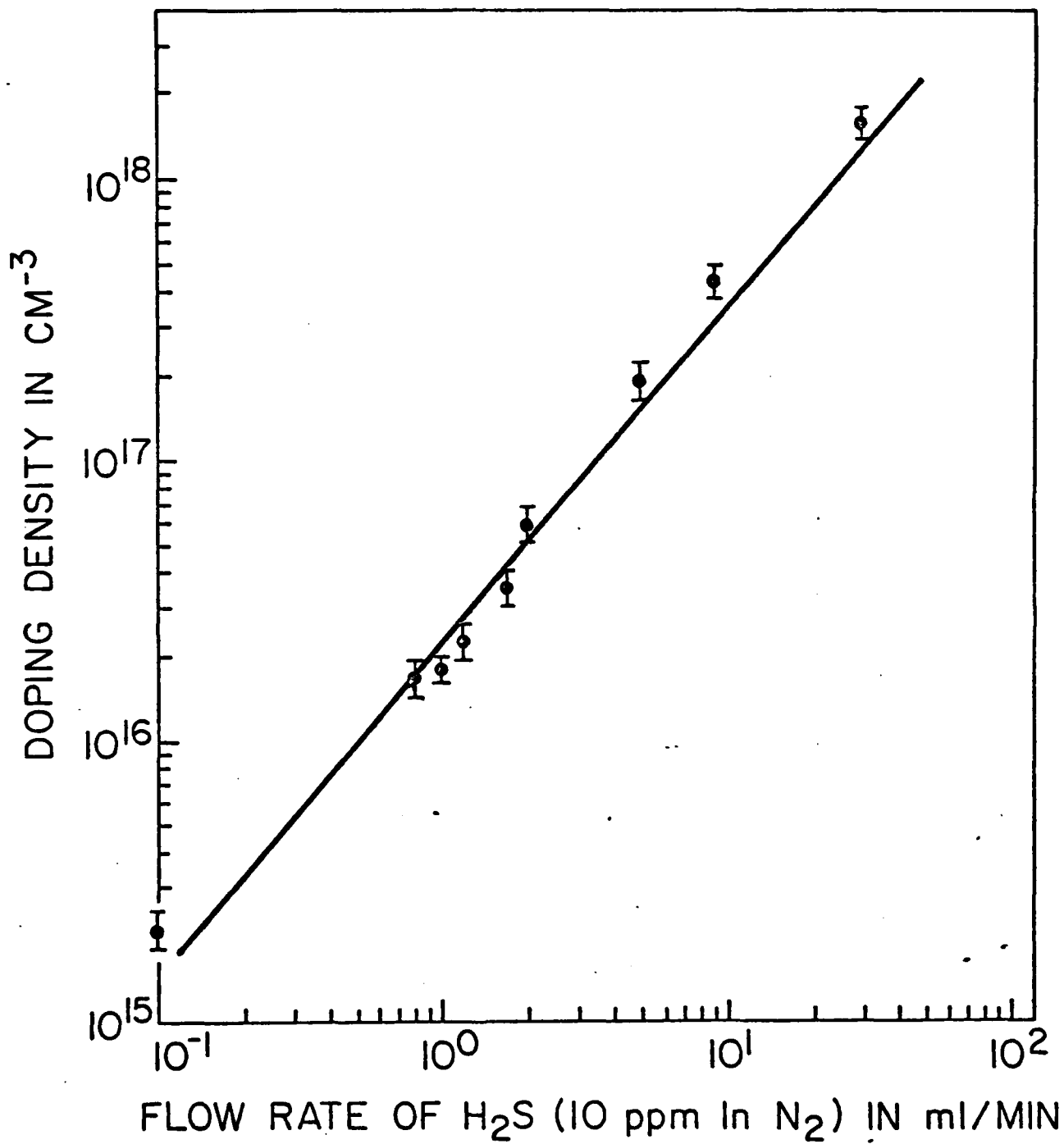


Figure 3. Doping density as a function of H₂S flow rate

With minimal exposure to air, the substrates are then rinsed in deionized water, and dried with bibulous paper and nitrogen.

After etching, the substrate is placed in the sample holder and inserted in the furnace. The reactor is then flushed with hydrogen at high flow rates for 20 to 25 minutes, and the furnace is folled to the growth position.

2.1.3 Results and Discussion

Under the preceeding growth conditions a growth rate of about 10 microns per hour is obtained. The appearance of the grown layers is generally smooth and almost featureless. One phenomenon encountered was that the growth rate tended to drop in successive runs due to the build-up of wall deposits which provided competitive areas of growth.

There is no vapor etching of the sample before growth--the prepared samples appear not to need it--and transfer of impurities from the heavily doped substrate to its surrounding is thus minimized.

The thickness measurement of the epitaxial layer is done by the stain etching method. The sample is sectioned and placed in a solution of $1\text{HF}:3\text{HNO}_3:4\text{H}_2\text{O}$ for about 4 seconds. The stained sample is viewed edgewise in a calibrated microscope. The measurement resolution is limited by how well the microscope could focus on the top rounded surface of the epitaxy and the interface. Depletion layer capacitance-voltage measurements can give a more accuract determination of the epitaxy thickness.

One of the problems which can attend the growth of a high purity epitaxial layer is the diffusion of impurities from the substrate during growth. This problem can be overcome by the growth of a heavily doped, high purity "buffer" layer on to the substrate prior to the growth of the active region. The buffer layer effectively isolates the active layer from the impure substrate. It is

also often desirable to grow a heavily doped GaAs layer after growth of the active region. The reason for doing this is because it offers the best method of obtaining a ohmic, low resistance contact to the active region.

The doping profile can be measured by the capacitance-voltage method. This involves evaporating a number of 20 mil diameter gold Schottky barrier dot diodes on the epitaxial layer. The dot is biased negative with respect to the substrate, the capacitance of the depletion region under the dot is measured as a function of the applied bias. From this data the doping concentration profile can be measured. It is given by ³

$$n = \frac{-C^3}{q\epsilon A^3 \left(\frac{dC}{dV} \right)}$$

where

C = depletion capacitance

V = applied bias to dot

A = dot area.

The measured capacitance may be used to calculate the depletion depth

$$x = \frac{\epsilon A}{C}$$

where x = depletion distance into the epitaxy. This method has been shown to yield quite accurate doping profile measurements ⁴ with only slight error due to spreading capacitance.

The range of x over which the carrier concentration can be measured is limited by the avalanche breakdown of the junction which eventually occurs as the bias voltage is increased. The dependences of the avalanche breakdown voltage of the junction, the maximum penetration, and the depth resolution,

Table 1. Dependence of Profile Properties on Doping Density.

Doping density (cm^{-3})	Maximum Depth (μm)	Maximum d.c. volts	Depth resolution (Debye length) μm
10^{14}	80	500	0.50
10^{15}	12	100	0.50
10^{16}	2	20	0.05
10^{17}	0.3	4	0.015

on the carrier concentration in the layer are approximately as indicated by Table 1. The profiling of a layer thicker than the maximum depth indicated above can be obtained by step etching to build up a piecewise profile plot of the whole layer. Figure 4 shows the profile obtained on one of our samples via the step-etch technique.

A feature of the epitaxial layers is the persistent occurrence of growth 'pyramids' or 'hillocks'. The formation of the 'pyramids' or 'hillocks'⁵ is dependent both on the experimental procedure used and the quality and orientation of the substrate. For the growth system we used, the number of grown pyramids will be increased when the substrate is moved to the lower temperature region (below 745°C) of the reactor tubing.

As the Ga source is used up, the surface area of the gallium decreased and this should slow down the gallium transfer rate and lower the Ga/As ratio. This may explain the systematic decrease in carrier concentration with run sequence and the steep impurity profiles obtained at the end of the useful life of a gallium source, when surface tension forces may suddenly shrink the gallium.

The AsCl_3 -Ga- H_2 reaction has been used to produce low carrier concentration 'background doped' GaAs which has been made into high power Gunn oscillators and other millimeter wave devices. The basis requirement for the improvement of the epitaxial layer is to improve the control and system cleanness of the growth process. With commercially available high purity starting material, carrier concentrations n in the $1\text{--}2 \times 10^{14} \text{ cm}^{-3}$ range can be obtained.

2.1.4 Sample Fabrication for Velocity-Field Measurements

The schematic diagram of the sample used for the velocity-field measurement is shown in Fig. 6. A GaAs epitaxial layer is grown on a high doped GaAs substrate by the vapor-phase AsCl_3 -Ga- H_2 system as indicated in the diagram of

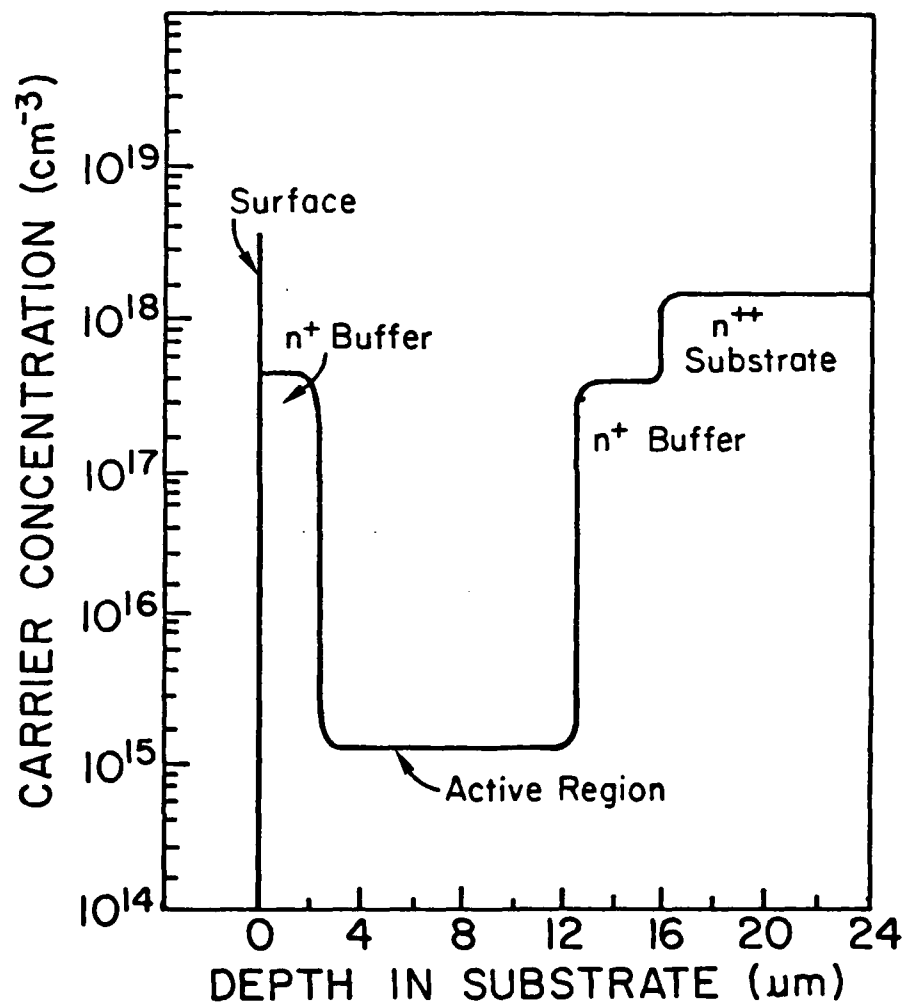


Figure 4. Carrier concentration versus depth for VPE grown layer from Schottky barrier profile measurements

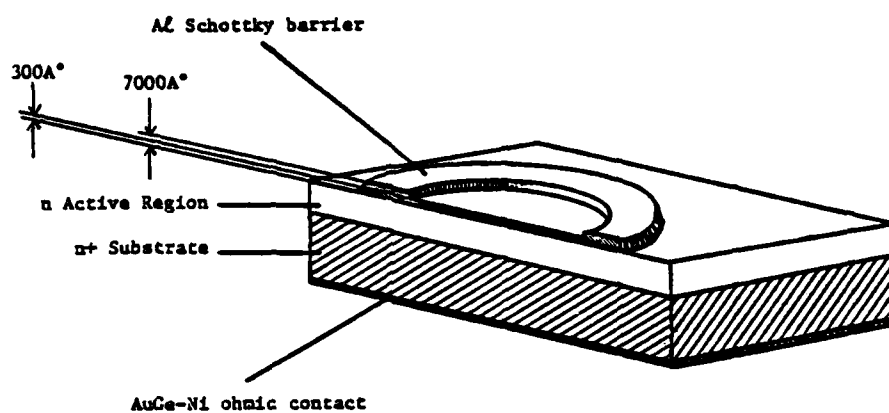


Figure 6. Cross sectional view of sample for velocity-field measurements

Fig. 2. The diode consists of a thin aluminum window and a thick aluminum annulus. In the electron-beam velocity field apparatus, described elsewhere electrons are injected thorough the thin window when their energies are above a critical value that is determined by the contact thickness. The thick annulus provides a good mechanical region that permits ultrasonic wire bonding to the Schottky contact. The substrate contact is made using an alloyed gold-germanium nickel layer. A thick layer of plated gold has been added to permit thermal compression die bonding of the sample to a gold plated copper stud.

2.1.4.1 Epitaxial Layer

The epitaxial reactor used to grow the layers is the vapor phase system described in the preceeding section. The orientation of the GaAs substrate is chosen to be 2° off the (100) planes. The measured carrier concentrations of the substrates are typically $1 \times 10^{18} \text{ cm}^{-3}$. The substrate is degreased by rinsing successively in trichloroethylene, acetone, and methanol for 30 seconds each, and then etched in the $5\text{H}_2\text{SO}_4:1\text{H}_2\text{O}_2:1\text{H}_2\text{O}$ etchant for 5 minutes. It is rinsed in deionized water of resistivity greater than $14 \text{ M}\Omega\text{-cm}$. The substrate is rapidly loaded in the furnace.

The reactor tubing is purged with hydrogen for 20 minutes. The hot furnace is then rolled to the growth position (source 832°C , seed 762°C). The AsCl_3 bubbler is turned on 5 minutes later. The temperature of the furnace is in the steady-state 10 minutes after turning on AsCl_3 . The flow rates of hydrogen are then reduced to 120 ml/min in the AsCl_3 line and 50 ml/min in the dopant line, meanwhile, the $\text{H}_2\text{S-N}_2$ flow rate is adjusted to 45 ml/min . The buffer layer is grown for 10 minutes, the measured thickness and concentration are $2.5 \text{ }\mu\text{m}$ and $4 \times 10^{17} \text{ cm}^{-3}$. The active region is then grown for a predetermined time. The measured thickness and concentration of the active layer are then

measured using either C-V or an MSI doping profile plotter.

2.1.4.2 Ohmic Contact Fabrication

The sample is mounted on a jig by using parafin with the epitaxial layer side down. The back of the sample is slowly lapped with 5 μm grit on a glass plate. The thickness of the sample is measured in a micrometer. The sample is then mounted on a glass slide using parafin to mask the epitaxial layer and etched in 1% bromine-methanol solutions for 5 minutes to obtain a smooth surface. The measured resultant substrate thickness is 135 μm .

The AuGe-Ni-Au ohmic contact system is applied to the back of the substrate. The AuGe used is eutectic composition of 88% Au, 12% Ge by weight. AuGe is evaporated on the sample followed by a nickel evaporation. The AuGe-Ni evaporation ratio must be between 3:1 and 6:1 to obtain thermally stable contacts with low specific contact resistance.

The AuGeNi layer is alloyed at $455^\circ\text{C} \pm 2^\circ\text{C}$ for one minute in a hydrogen atmosphere. The wafer is quenched by removing it quickly. A low thermal mass system is required to obtain a specific contact resistance of less than $5 \times 10^{-6} \Omega/\square$. The wafer is rinsed in hydrofluoric acid for 15 seconds and rinsed in D.I. water. A gold layer of 4 μm is applied to the ohmic contact by plating the wafer in a neutral bath gold solution (Selrex puragold 125).

2.1.4.3 Schottky Barrier Fabrication

The sample is mounted, ohmic contact side down, on a microscope slide with black wax. The sample is again cleaned in the organic solvents, trichloroethylene, acetone, and methanol, and the epitaxial layer is etched lightly just before the sample is mounted in the evaporation system. The etchant is $5\text{H}_2\text{SO}_4 : 1\text{H}_2\text{O}_2 : 1\text{H}_2\text{O}$. After the sample is rinsed in DI water and dried, it is mounted in the evaporation system. The vacuum is then pumped to a pressure of about 9×10^{-7} Torr at

which time evaporation proceeds. The source is slowly heated until the Al melts, the shutter is opened, and evaporation thickness monitored with a transducer mounted close to the samples. The shutter is closed after the desired thickness, 7000A Al, is reached.

Next the sample is removed from the evaporator, the aluminum is covered with photoresist to be used as a etchant mask as shown in Fig. 7(a). A positive photoresist, AZ 1350J, is used for this purpose. The sample is prebaked at 120°C for 15 minutes. The filtered AZ 1350J photoresist is spun on the sample at 4000 RPM, for 60 sec. The photoresist is baked for 30 minutes at 70°C. Then the sample is again coated with photo resist, spun at 5000 RPM for 60 seconds, and then baked at 70°C for 30 minutes.

An optical mask aligner is then used to expose the photoresist except for a pattern of 20 mil diameter 2.5 mil thick circular rings. The pattern is aligned along a cleavage plane. The exposed photoresist is developed and is then dried and postbaked for 20 minutes at 120°C.

The hardened photoresist ring is used as a mask to etch the thick Al layer. The etchant for the aluminum is Aurostrip (Metex Aurostrip #407) diluted in deionized water to a concentration of 80 grams/gallon. The aurostrip solution is heated and stirred, the sample is immersed until all the aluminum is etched except that under the 2.5 mil ring. The resulting structure is shown in Fig. 7(b). The photoresist mask is then removed in hot acetone, cleaned in organic solvents, rinsed in deionized water, and prebaked again at 120°C for 20 minutes.

AZ 1350J photoresist is applied on the sample and baked at 70°C for 30 minutes, and a second layer spun at 5000 RPM for 60 seconds. After another baking at 70°C for 30 minutes, it is exposed on the mask aligner. The purpose of this photoresist layer is to open window at the ring, as shown in Fig. 7(a).

The sample is then postbaked at 120° for 15 minutes after developing.

The sample is again loaded in the evaporator to coat it with a thin Al layer, as shown in Fig. 7(d). The unwanted Al layer can be lifted off by immersing the sample into heated and stirred acetone. It takes about 40 minutes to lift the Al layer. The final structure of the Al Schottky barrier dot is shown in Fig. 7(e).

References

1. Knight, J. R. Effer, D., and Evens, P. R. "The Preparation of High Purity Gallium Arsenide by Vapor Phase Epitaxial Growth", Solid State Electronics, Vol. 8, pp. 178-180, 1965.
2. Blakeslee, A. E., "Effects of Substrate Misorientation in Epitaxial GaAs", Transactions of the Metallurgical Society of AIME, 245, 577 (1969).
3. Thomas, C. O., Kahng, O., Manz, R. C., "Impurity Distributions in Epitaxial Silicon Films", Journal of the Electrochemical Society, 109, (1962), p. 1055.
4. Copeland, J. A., "Diode Edge effect on Doping-Profile Measurements", IEEE Transactions on Electron Devices, Vol. ED-17(5), (May 1970), pp. 404-407.
5. Joyce, B. D. Mullin, J. B., 'Pyramid Formation in Epitaxial Gallium Arsenide Layers', Gallium Arsenide: 1966 Symposium Proceedings, p. 23.

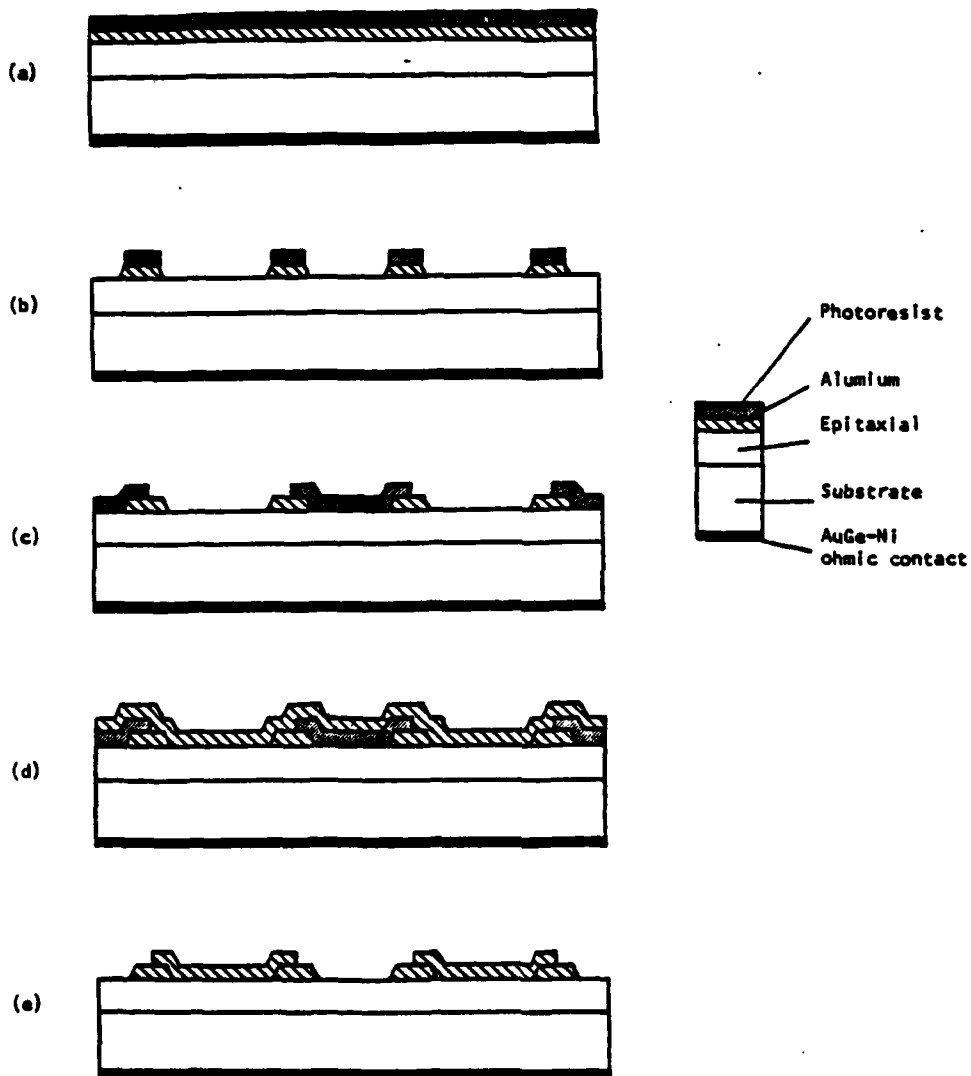


Figure 7. Process used to fabricate aluminum Schottky barrier diodes.

2.2 Impurity Incorporation and Redistribution During Epitaxial Growth

by H. Rohdin and M. Muller

2.2.1 Introduction

The incorporation of impurities in epitaxial layers and their subsequent redistribution can lead to impurity and carrier density profiles that may be unfavorable for certain applications. C. M. Wolfe and K. H. Nichols developed a model, the impurity gradient model (IGM) [1,2], which explains qualitatively some observed features, such as the occurrence of different conductivity regions at the $n-n^+$ layer-substrate interface. It also explains why the observed donor-acceptor compensation is less than expected [2].

In the model the redistribution of impurities is caused by concentration gradient diffusion, and by drift due to a built-in electric field. The sources of the field are charged surface states and impurity concentration gradients. We thus have two coupled problems: The field problem and the redistribution problem. Another feature is the moving boundary i.e., the growth surface is moving with the growth velocity.

The IGM, as developed by Wolfe and Nichols is based on a number of assumptions. Most of these are realistic while others are adopted to make an analytic solution possible. The Wolfe-Nichols treatment (W&N treatment) gives a good qualitative picture, but to get more realistic results we must renounce some of the assumptions made. This will complicate the treatment, and numerical methods must be invoked.

In this report the Wolfe-Nichols assumption and their consequences will be listed and discussed. Then the changes necessary for a more realistic treatment will be presented. The problem will be defined and an approach to the

solution given. The problem will be treated as two semi-distinct ones: the field problem and the redistribution problem, the first being the more challenging.

2.2.2 Assumptions

2.2.2.1 W & N Treatment. List of Assumptions, Their Consequences and Validity

1. The evolution in time of the electric field, which depends on the growth velocity and impurity redistribution is so slow that an electrostatic approach is appropriate. We can thus get the field from Poisson's equation. This assumption is realistic, considering the slow growth rate and low diffusion constant of most impurities.
2. The edges of the relatively thin epitaxial layer do not contribute to the general picture. Thus a one dimensional approach is possible. This is realistic since the lateral dimension of the epitaxial layer is much larger than the thickness.

Assumption 1 and 2 imply that the governing equation for the electric field E is

$$\frac{dE}{dx}(x) = \frac{q}{\epsilon} (p(x) - n(x) + D(x)) \quad (1)$$

where D is the net charge associated with the impurities

$$D = \sum_{i,j} = N_j^n i^{\pm} \quad (2)$$

3. The material is non-degenerate and the electrons and holes are in thermal equilibrium. This results in the following:

1. A simple mass action law for creation and recombination of an electron-hole pair

$$np = n_i^2 \quad (3)$$

- ii. A simple Einstein relation between mobility μ and diffusion constant D of electrons and holes

$$D_x = \mu_x \frac{kT}{q} ; \quad x = n, p \quad (4)$$

- iii. A relation between electric field E and electron concentration to ensure equal and opposite drift and diffusion currents for electrons (and holes).

$$E = - \frac{kT}{q} \frac{d}{dx} \ln n = - \frac{kT}{q} \frac{1}{n} \frac{dn}{dx} \quad (5)$$

Many materials are nondegenerate, especially at the high temperatures during growth. The growth velocity v is in practice far too small to disturb the electron and hole equilibrium. Furthermore the diffusion constants of the impurities are much lower than those of electrons and holes so that the latter easily keep up with the redistribution of the former. The assumption is thus a good one.

4. The epitaxial layer doping is below the intrinsic concentration. The substrate is initially extrinsic (n-type) with constant doping. The redistribution in the epitaxial layer during growth is not large enough to affect the field. There is no drift or diffusion between the epitaxial layer and the substrate. The specific doping condition could well be changed but the essence is that there is no coupling between field and impurity redistribution. This is only realistic if we have very small redistribution of impurities, i.e., if the impurity diffusion constants are sufficiently small.

5. The epitaxial layer is essentially semi-infinite, that is to say, much larger than the Debye length. This enables us to look at the field in the epitaxial layer as a sum of two fields E_s and E_1 , where E_s originates from the surface states on the growing surface and E_1 from the discontinuity

(according to assumption 4) in doping at the layer-substrate interface. This is a good assumption once the growth has been going on for a long enough period, but in the initial stages of growth it is not correct.

6. The substrate is essentially semi-infinite. This gives the necessary value for E_i at the interface. This is a good assumption in most practical cases.

7. The electron concentration at the growth surface is fixed by surface states. This means that the Fermi level is pinned at the surface. This gives a value for E_s at the growth surface. This is a realistic assumption; pinning of the Fermi level has been observed. It would also be reasonable to assume that E_s at the growth surface is fixed by the surface states in accordance with Gauss law in electrostatics, i.e.

$$E(0) = \frac{\sigma_s}{\epsilon} \quad (6)$$

where σ_s is the surface charge density associated with the surface states.

In terms of electron concentration this is according to (5)

$$\frac{dn}{dx}(0) = - \frac{q\sigma_s}{\epsilon kT} n(0) \quad (7)$$

This would contain the assumption that the surface charge density is essentially independent of the doping at the surface and thus independent of the Fermi level at the surface. Depending on how the surface state density depends on energy this may be true or false.

8. No temperature gradients are present, thus no thermal currents. This is a realistic assumption.

9. We study only shallow (hydrogenic) impurities which each only occupy one type of lattice site of constant concentration. This means

i. Only simple diffusion, i.e., no generation of recombination term in the continuity equation for the impurities;

ii. All impurities are singly ionized.

The doping is therefore

$$D = \sum_k \pm N_k \quad (8a)$$

the flow of impurity k , J_k , is

$$J_k = \pm \mu_k N_k E - D_k \frac{\partial N_k}{\partial x} \quad (8b)$$

(upper sign donor; lower acceptor) and the continuity equation for impurity k is

$$\frac{\partial N_k}{\partial t} = - \frac{\partial J_k}{\partial x} \quad (8c)$$

This is, of course, not always the case (e.g. Cr) but it is undoubtedly a technologically interesting case.

10. The impurities are in equilibrium with the electric field. This means that the time derivative of the impurity concentration in the continuity equation are all zero and an analytical expression for the concentration can be found [1,2]. This is a very unrealistic assumption for the slow moving impurities studied.

11. The impurity concentration profiles are calculated during growth. During and after cool-down further redistribution may have occurred. This is a different problem including fixed boundary and cool-down. It is not treated here and we therefore effectively assume that the distributions are frozen in.

12. We study only homoepitaxial growth. We can then easily define an electrostatic potential and we get no surface charge at the layer substrate interface.

13. The numerical factor B in the Einstein relation

$$D = \mu \frac{kT}{Bq} \quad (8d)$$

between impurity diffusion constant D and mobility μ is 1. According to [6] pp. 92-99 this is only true for a simple interstitial mechanism. The simple impurity diffusion that we study is thought to occur through a vacancy mechanism. In this case $B = 1.27$.

2.2.2.2 New Treatment

In the new treatment we want to take a more general viewpoint. This means that we have to do without some of the assumptions discussed above.

The most unrealistic of those is no. 10. We now want to study the case when the impurities are not in equilibrium with the field. This is actually the case for the kind of impurities we are studying which are very slow moving ($D \leq 10^{-14} \text{ cm}^2/\text{s}$ at $T = 750^\circ\text{C}$). We also want to allow redistribution in the substrate and across the layer-substrate interface. We thus would like to relinquish assumption no. 4. We will have coupling between redistribution and field. This will complicate the field calculation considerably. It is not yet clear to us whether the most general field calculation is feasible numerically. Poisson's equation will be nonlinear, with a forcing term, the latter being the net doping, and this is a somewhat difficult problem, even numerically. We will try to overcome the difficulties, but if this is not possible, we can, in view of the low mobilities of the impurities, use a modified W & N approach, i.e. we could calculate the field once and for all with the non-redistributed doping profile. We would, however, not treat the epitaxial layer as infinite in the initial stage of growth. (Assumption 5 in W & N treatment). In either case we will attempt to use the assumption that the field rather than the po-

tential at the growth surface is fixed by surface states (assumption 7 above). Assumption 6 above will make it possible to use the boundary condition

$$E(\infty) = 0 \text{ and/or } n(\infty) = n_s \quad (9)$$

where n_s is the equilibrium electron concentration fulfilling charge neutrality deep in the substrate, i.e.

$$n_s - \frac{n_i^2}{n_s} - D_s = 0 \quad (10)$$

where D_s is the net doping deep in the substrate. It will also enable us to avoid the problem of the moving boundary by putting the origin at the growth surface. Now the other boundary will be moving, but as this is at infinity it does not complicate things. We will keep the temperature constant in time (as well as in space).

In summary: We will use assumption 1-3, 6-9, 11 and 12. 7 will be modified in the sense that we keep $E(0)$ fixed. 8 will also include no time variations in the temperature. We will allow B in assumption 13 to take an appropriate value. The other approximations of W & N will be relinquished in the new treatment.

2.2.3 The Problem

The problem as mentioned, is a coupled one. The redistribution of impurities depends on the field and the field depends on the distribution of impurities. Only under certain restrictions can the field be calculated separately from the redistribution. The general approach will be to calculate the field and with this field to calculate the redistribution during the next time interval, then calculate the new field etc. First the field problem will be treated and then the redistribution.

2.2.3.1 The Electric Field Problem

If possible, an exact solution will be attempted. In adopting this approach we have encountered several numerical problems and therefore a simpler approach may be necessary. Both will be presented here.

Let us introduce the following quantities to get equations containing only dimensionless quantities:

$$v = \frac{n}{n_i} ; \quad n_j = \frac{N_j}{n_i} ; \quad \delta = \frac{D}{n_i} \quad (11)$$

where n_i is the intrinsic electron concentration. Furthermore

$$\xi = x/L_{D_i} ; \quad L_{D_i} = \sqrt{\frac{\epsilon kT}{q^2 n_i}} \quad (12)$$

L_{D_i} is the intrinsic Debye length.

$$e = E/E_0 ; \quad E_0 = \frac{kT}{qL_{D_i}} \quad (13)$$

(5), (11), (12), (13) give

$$e = - \frac{d}{d\xi} \ln v = - \frac{1}{v} \frac{dv}{d\xi} \quad (14)$$

(1), (3), (11), (12), (13) give

$$\frac{de}{d\xi} = \frac{1}{v} - v + \delta \quad (15)$$

(14), (15) give

$$\frac{d^2}{d\xi^2} \ln v = v - \frac{1}{v} - \delta \quad (16)$$

or

$$\frac{d^2 v}{d\xi^2} = \frac{1}{v} \left(\frac{dv}{d\xi} \right)^2 + v^2 - v\delta - 1 \quad (17)$$

(9), (11), (12) give

$$e(\infty) = 0; \quad v(\infty) = v_s \quad (18)$$

(10), (11) give

$$v(\infty) = v_s = \frac{\delta_s}{2} + \sqrt{\frac{\delta_s^2}{4} + 1} \quad (19)$$

(6), (13) give

$$e(0) = \frac{\sigma_s}{\epsilon E_0} = e_0 \quad (20)$$

or with (14)

$$\frac{dv}{d\xi}(0) = -e_0 v(0) \quad (21)$$

If we deal with homoepitaxy we can define a continuous electrostatic potential ψ which we can set to zero at $x = \infty$, i.e. deep in the substrate:

$$E_1(x) - (E_1)_s = -q\psi(x) \quad (22)$$

E_1 is the intrinsic Fermi level. Define a normalized potential ϕ

$$\phi = \frac{q\psi}{kT} \quad (23)$$

The normalized electron concentration can be written

$$v(\xi) = v_s e^{\phi(\xi)} \quad (24)$$

since we have non-degenerate material. Define

$$u(\xi) = \ln v(\xi) = u_s + \phi(\xi) \quad (25)$$

(14), (25) give

$$e = - \frac{d\phi}{d\xi} \quad (26)$$

(15), (25) give

$$\frac{de}{d\xi} = - 2 \sinh(u_s + \phi(\xi)) + \delta(\xi) \quad (27)$$

(26) and (27) are a set of two coupled first order ordinary nonlinear differential equations. They can, of course, also be written as a second order ordinary nonlinear differential equation

$$\frac{d^2\phi}{d\xi^2} = 2 \sinh(u_s + \phi(\xi)) - \delta(\xi) \quad (28)$$

with boundary condition

$$\phi(\infty) = 0 \quad (29)$$

$$\frac{d\phi}{d\xi}(0) = -e_0 \quad (30)$$

2.2.3.1.1 The Exact Approach

The electric field can be calculated either from the electron concentration through (14) or from the potential through (26). If the first method is chosen, the pertinent differential equation is (17) and the boundary conditions are (19) and (21). If one uses the potential the governing equation is (28) and the boundary conditions are (29) and (30). (17) and (28) are connected by (24).

In both cases we have a non-homogeneous non-linear differential equation with two boundary conditions, one being at the growth surface and the other at infinity or rather deep in the substrate. Both equations have a rather

arbitrary forcing term which contains the doping. Only when the doping is uniform is it relatively easy to solve the problem. One then chooses to solve (28). We have tried to solve both (17) and (28) numerically, so far without complete success. We have tried the following methods:

1. Treat the problem as an initial value problem, i.e., make two coupled ordinary differential equations of (17) or (28) respectively. In the latter case one possibility is the system (26), (27). To be able to solve this we need to know the initial value of both the function itself and its derivative. But we only know one of these, or the ratio between them. The other boundary condition that we know is at infinity. We thus have to adopt some kind of "aiming" method, i.e. pick a pair of initial values in such a way as to fulfill the boundary condition at infinity (or in practice at a point sufficiently far in the substrate). This involves iteration and is time consuming. Furthermore both (17) and (28) are ill-conditioned for prescribed initial value and initial derivative. To illustrate the difficulty, consider the equation

$$\frac{d^2 y}{dx^2} = y; \quad y(0) = y_0, \quad y(\infty) = 0 \quad (31)$$

In this example an analytical solution is easily found, namely

$$y = A_1 e^x + A_2 e^{-x} \quad (32)$$

With this potential, one would for the physical reasons $y(\infty) = 0$ have picked $A_1 = 0$. Unfortunately the computer has no capability for physical reasoning, so the term $A_1 e^x$ would appear in the numerical solution if we had chosen to solve (31) as an initial value problem

$$\frac{dy_1}{dx} = y_2$$

$$\frac{dy_2}{dx} = y_1 \quad (33)$$

with initial value for $y_1 = y$ known to be

$$y_1(0) = y_0$$

but the initial value for $y_2 = \frac{dy}{dx}$ unknown. A solution based on picking $\frac{dy}{dx}(0)$ to satisfy $y(\infty) = 0$ seems impossible in practice. For our problem this is even more true since the right hand sides of (17) and (28) are much more sensitive to the unknown variable than is the right hand side of (31).

2. Use a combination of finite difference approximation and iteration [3], i.e. we have the nonlinear ordinary second order differential equation

$$y'' = f(x, y, y'); \quad a \leq x \leq b; \quad (34)$$

and we intend to make better and better approximation $y^{(n)}$ of y . The finite difference approximation to (34) is

$$\frac{y_{i-1}^{(n+1)} - 2y_i^{(n+1)} + y_{i+1}^{(n+1)}}{h^2} = f(x_i, y_i^{(n+1)}, (y')_i^{(n+1)}); \quad i=1, 2, \dots, i_{\max} \quad (35)$$

where subscript i indicates the value at mesh point i

$$x_1 = a$$

$$x_{i_{\max}} = b \quad (36)$$

and h is the distance between subsequent mesh points

$$x_{i+1} - x_i = h \quad \text{all } i$$

$$x_{i_{\max}} - x_1 = h(i_{\max} - 1)$$

Furthermore

$$(y')_i = \frac{1}{2h} (y_{i+1} - y_{i-1}) \quad (37a)$$

is the finite difference approximation of y' . We cannot use the $(n+1)$ th approximation on the right hand side of (35) since that is the one we are seeking. We thus have to approximate. Define

$$f_i^n \equiv f(x_i, y_i^n, (y')_i^n) \quad (38)$$

and $\left(\frac{\partial f}{\partial y}\right)_i^n$, $\left(\frac{\partial f}{\partial y'}\right)_i^n$ analogously. Then an approximation to (35) is

$$\begin{aligned} & y_{i-1}^{(n+1)} \left\{ 1 + \frac{h}{2} \left(\frac{\partial f}{\partial y'}\right)_i^{(n)} \right\} + y_i^{(n+1)} \left\{ -2h^2 \left(\frac{\partial f}{\partial y}\right)_i^{(n)} \right\} + \\ & + y_{i+1}^{(n+1)} \left\{ 1 - \frac{h}{2} \left(\frac{\partial f}{\partial y'}\right)_i^{(n)} \right\} = \\ & = h^2 \left\{ f_i^{(n)} - y_i^{(n)} \left(\frac{\partial f}{\partial y}\right)_i^{(n)} - (y')_i^n \left(\frac{\partial f}{\partial y'}\right)_i^{(n)} \right\}; \quad i=1, 2, \dots, i_{\max} \quad (39) \end{aligned}$$

It is easy to incorporate boundary values into this scheme. If one boundary is at infinity one picks $x_{i_{\max}}$ large enough so that the boundary condition is effectively fulfilled there. Unfortunately this does not seem to converge to a steady value for y_i for either of (17) and (28). For (28), not only is there no convergence, but the result is unstable.

We are still trying to find a way to solve (17) and/or (28), without simplification. If this is unsuccessful, we have to adopt a simpler approach which is a modification of the W & N treatment.

2.2.3.1.2 Modified W & N Approach

We will here assume that we can calculate the field as if there were no redistribution. This is probably a fairly good assumption with the low diffusion constants of the impurities under study. The epitaxial layer will then approximately have the intended doping concentration which we assume is a constant D_e . We also assume that the substrate doping is a constant, D_s . In regions where δ is constant and $\frac{d\phi}{d\xi} \neq 0$ (28) can, after multiplying both sides with $\frac{d\phi}{d\xi}$, be written

$$\left(\frac{d\phi}{d\xi}\right)^2 = 4\cosh(u_s + \phi) - 2\delta\phi + c \quad (40)$$

where c is an integration constant specific for each region. The potential ϕ is still set to zero deep in the material i.e. in the substrate. (40) yields:

$$\xi = \pm \int_{\phi_0}^{\phi} \frac{d\phi'}{\sqrt{4\cosh(u_s + \phi') - 2\delta\phi' + c}} \quad (41)$$

The plus or minus sign is chosen from physical reasoning. It may reverse within a region at a point where $\frac{d\phi}{d\xi} = 0$. The boundary condition (30) gives

$$e_0^2 = 4\cosh(u_s + \phi_0) - 2\delta_e\phi_0 + c_e \quad (42)$$

where ϕ_0 is unknown in our treatment while e_0 is known. (Subscript e for epitaxial layer). The boundary condition (29) yields

$$c_s = -4\cosh u_s \quad (43)$$

Continuity of the field and potential at the interface (assumption 12) gives with the aid of (40)

$$c_e = c_s + 2\phi_{es}(\delta_e - \delta_s) \quad (44)$$

where ϕ_{es} is the potential at the interface. (42) - (44) provide a relation for ϕ_{es}

$$\phi_{es} = \frac{e_0^2 - 4\cosh(u_s + \phi_0) + 4\cosh u_s + 2\delta_e \phi_0}{2(\delta_e - \delta_s)} \quad (45)$$

(41) gives for the width ℓ_e of the epitaxial layer (if there is no sign reversal in (41))

$$\ell_e = \pm \int_{\phi_0}^{\phi_{es}(\phi_0)} \frac{d\phi'}{\sqrt{4\cosh(u_s + \phi') - 4\cosh u_s - 2\delta_e \phi' + 2(\delta_e - \delta_s)\phi_{es}(\phi_0)}} \quad (46)$$

We can indirectly determine ϕ_0 as a function of ℓ_e from (46). Then we can determine c_e as a function of ℓ_e by (45), (43), (44). With (41) we can then determine ϕ indirectly as a function of ξ for a specific ℓ_e in both the epitaxial layer and in the substrate

2.2.3.2 The Impurity Redistribution Problem

This is an easier problem than the field calculation since it is linear once the field is known. The flow J_k of impurity type k in a field and concentration gradient is (assumption 9)

$$J_k = \pm \mu_k N_k E - D_k \frac{\partial N_k}{\partial x} \quad (8b)$$

with (8d), (11), (12), (13)

$$J_k = \frac{D_k n_i}{L_{D_i}} \left[\pm B_k e \eta_k - \frac{\partial \eta_k}{\partial \xi} \right] \quad (47)$$

Define, $j_k = \frac{D_k n_1}{L_{D_1}}$ (48)

then

$$J_k = j_k \left[\pm B_k e \eta_k - \frac{\partial \eta_k}{\partial \xi} \right] \quad (49)$$

The continuity equation for impurity k is (assumption 9)

$$\frac{\partial N_k}{\partial t} = - \frac{\partial J_k}{\partial x} \quad (8c)$$

As mentioned in section 2.2.2.2 we can with assumption 6 avoid the moving boundary by setting the origin of the x-axis at the growth surface. Now the other boundary will be moving but this boundary is at infinity and not a specified distance from the growing layer or the interface. If we pick a region of a given length before growth, such that one boundary is the growth surface (initially the interface) and the other at a distance deep in the substrate (so deep that it is effectively at infinity), we can grow until the "equivalent infinity boundary" no longer can be said to be deep in the substrate. We thus study a fixed boundary problem for the allowed period of growth. When we change the coordinate, J_k must be replaced by $J_k + N_k v$ where v is the growth velocity, With this and (8c), (47) we get

$$\frac{\partial \eta_k}{\partial t} = \frac{D_k}{L_{D_1}^2} \left[\frac{\partial^2 \eta_k}{\partial \xi^2} \mp B_k \frac{\partial}{\partial \xi} (e \eta_k) - \frac{v L_{D_1}}{D_k} \frac{\partial \eta_k}{\partial \xi} \right] \quad (50)$$

We normalize the diffusion constants D_k to a typical diffusion constant D_0 , typical for the studied temperatures and class of impurities:

$$D_k = d_k D_0 \quad (51)$$

With our class of impurities D_0 could be picked to be $10^{-15} \text{ cm}^2 \text{ s}^{-1}$ for $T = 750^\circ \text{C}$ (see [4] p. 418). We define a normalized time τ , normalized to $L_{D_1}^2/D_0$

$$t = \tau \frac{L_{D_1}^2}{D_0} \quad (52)$$

We also introduce a normalized growth velocity u , normalized to D_0/L_{D_1}

$$v = u \cdot D_0 / L_{D_1} \quad (53)$$

(50)-(53) yield the normalized transport equation for impurity species k .

$$\frac{\partial \eta_k}{\partial \tau} = d_k \frac{\partial^2 \eta_k}{\partial \xi^2} + B_k d_k \frac{\partial}{\partial \xi} (e \eta_k) - u \frac{\partial \eta_k}{\partial \xi} \quad (54)$$

If $e(\xi, \tau)$ is known, this is a linear partial differential equation for

$\eta_k = \eta_k(\xi, \tau)$. It is a boundary value problem in ξ :

$$\eta_k(0, \tau) = \eta_{ek}(\tau) \quad \text{all } \tau \quad (55)$$

$$\eta_k(\infty, \tau) = \eta_{sk} \quad \text{all } \tau \quad (56)$$

and an initial value problem in τ :

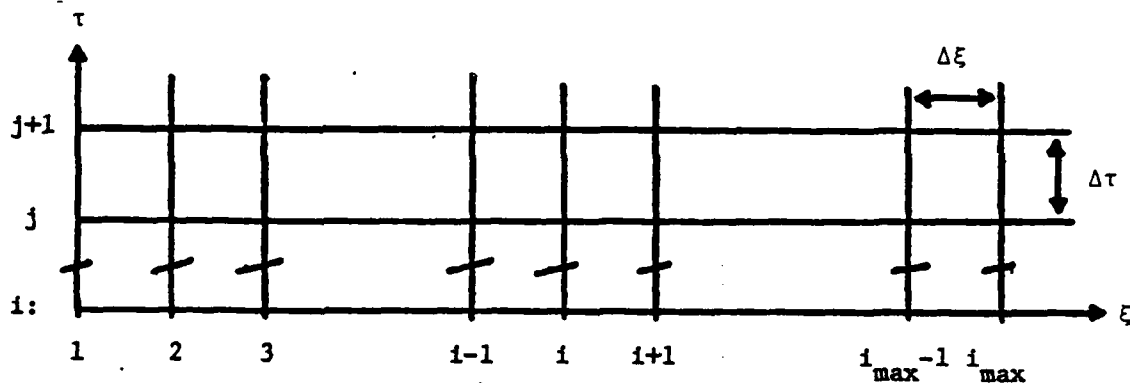
$$\eta_k(\xi, 0) = \eta_{sk} \quad \xi > 0 \quad (57)$$

2.2.3.2.1 Numerical Approach to the Redistribution Problem

In fact, (54) is not a linear equation since $e(\xi, \tau)$ depends on all η_k 's. Our approach will be to calculate e with the present distribution, i.e. at time τ , of η_k 's according to the previous section. Then we will calculate η_k at time $\tau + \Delta\tau$ where $\Delta\tau$ is a small time interval, assuming that e remains con-

stant during this time interval. The smaller $\Delta\tau$, the better this approximation is. We then calculate e with the new impurity distribution, etc. We cannot numerically take the entire ξ -interval $(0, \infty)$ into account, but we have to cut it off at some distance $\xi = L$ such that there is virtually no field or redistribution for $\xi \geq L$.

We have already divided the time scale into discrete times $(j-1)\Delta\tau$, $j = 1, 2, 3, \dots, j_{\max}$. We will also divide the studied ξ -region $(0, L)$ into discrete intervals $(i-1)\Delta\xi$, $i = 1, 2, 3, \dots, i_{\max}$ where $\Delta\xi$ is a small distance. These discrete values in time and position will generate a net, and we will seek the value of n_k at each mesh point. This involves approximating (54) by its finite difference approximation (FDA) in terms of the values of n_k at the mesh points.



This can either be done explicitly or implicitly [5]. The latter technique is the most stable, since it takes into account the values of η_k at the next time interval which is what one wishes to calculate. To calculate the η_k 's at the $(j+1)$ th time we now write the following implicit FDA for (54) at $i = 2, 3, \dots, i_{\max} - 1$ (where $\eta_k(\xi_i, \tau_j) = \eta_k(i, j)$ etc.):

$$\begin{aligned} \frac{\eta_k(i, j+1) - \eta_k(i, j)}{\Delta \tau} &= d_k \frac{\eta_k(i-1, j + \frac{1}{2}) - 2\eta_k(i, j + \frac{1}{2}) + \eta_k(i+1, j + \frac{1}{2})}{(\Delta \xi)^2} \\ &\mp B_k d_k \frac{e(i+1, j) \eta_k(i+1, j + \frac{1}{2}) - e(i-1, j) \eta_k(i-1, j + \frac{1}{2})}{2\Delta \xi} \\ &- u(j + \frac{1}{2}) \frac{\eta_k(i+1, j + \frac{1}{2}) - \eta_k(i-1, j + \frac{1}{2})}{2\Delta \xi} \end{aligned} \quad (54a)$$

with

$$\eta_k(1, j) = \eta_{ek}(j) \quad \text{all } j \quad (55a)$$

$$\eta_k(i_{\max}, j) = \eta_{sk} \quad \text{all } j \quad (56a)$$

and

$$\eta_k(i, 1) = \eta_{sk} \quad i > 1 \quad (57a)$$

The left hand side of (54a) is exactly equal to $\frac{\partial \eta_k}{\partial \tau}(\xi_i, \tau_j + \delta \tau_j)$ for some (unknown) value of $\delta \tau_j$ ranging from 0 to $\Delta \tau$. We have therefore placed the FDA of the right hand side of (54) at the midpoint between the j th and the $(j+1)$ th time. This is what makes the scheme implicit.

We have to solve the matrix equation

$$\begin{bmatrix}
 b & -c_2 & \dots & 0 & \dots \\
 -a_3 & b & -c_3 & \dots & \dots \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 \vdots & -a_1 & b & -c_1 & \dots \\
 0 & \dots & \dots & -a_{i_{\max}-1} & b
 \end{bmatrix}
 \begin{bmatrix}
 \eta_k(2, j+1) \\
 \eta_k(3, j+1) \\
 \vdots \\
 \eta_k(i, j+1) \\
 \vdots \\
 \eta_k(i_{\max}-1, j+1)
 \end{bmatrix}
 =
 \begin{bmatrix}
 d_2 \\
 d_3 \\
 \vdots \\
 d_1 \\
 \vdots \\
 d_{i_{\max}-1}
 \end{bmatrix}$$

where

$$b = \frac{1}{d_k \Delta \tau} + \frac{1}{(\Delta \xi)^2}$$

$$a_i = \frac{1}{2(\Delta \xi)^2} \pm B_k \frac{e(i-1, j)}{4\Delta \xi} + \frac{u(j + \frac{1}{2})}{4d_k \Delta \xi} ; \quad i = 3, \dots, i_{\max}-1$$

$$c_i = \frac{1}{2(\Delta \xi)^2} \mp B_k \frac{e(i+1, j)}{4\Delta \xi} - \frac{u(j + \frac{1}{2})}{4d_k \Delta \xi} ; \quad i = 2, \dots, i_{\max}-2$$

$$d_1 = a_1 \eta_k(i-1, j) + \left(\frac{1}{d_k \Delta \tau} - \frac{1}{(\Delta \xi)^2} \right) \eta_k(i, j) + c_1 \eta_k(i+1, j); \quad i = 3, 4, \dots, i_{\max}-2 \quad (59)$$

$$d_2 = 2a_2 \eta_k(1, j) + \left(\frac{1}{d_k \Delta \tau} - \frac{1}{(\Delta \xi)^2} \right) \eta_k(2, j) + c_2 \eta_k(3, j)$$

$$d_{i_{\max}-1} = a_{i_{\max}-1} \eta_k(i_{\max}-2, j) + \left(\frac{1}{d_k \Delta \tau} - \frac{1}{(\Delta \xi)^2} \right) \eta_k(i_{\max}-1, j) + 2c_{i_{\max}-1} \eta_k(i_{\max}, j)$$

This is easily done by Gauss's elimination ([5] pp. 20). By making $\Delta \xi$ sufficiently small we can make the first term dominant in a_i and c_i . The problem then reduces to the simple Crank-Nicholson scheme considered by [5] pp. 60-65. This means that our scheme is stable for all $\frac{d_k \Delta \tau}{(\Delta \xi)^2}$ for sufficiently small $\Delta \xi$.

2.2.3.3 How to Chose Δx , Δt , $i_{\max}$ and $j_{\max}$

A field falls off to zero essentially within a few appropriate Debye lengths. This means that we have to pick Δx to be a rather small fraction of the smallest appropriate Debye length. The time interval Δt should fulfill

$$|J| \Delta t \ll N \Delta x \quad \text{everywhere} \quad (60)$$

so as to make the relative change of impurity concentration everywhere small during each time interval. The flow J is either drift, diffusion, or translation flow. The last arises from the motion of coordinates with the growth surface.

$$|J \text{ (drift)}| = \mu N E$$

$$|J \text{ (diff)}| = D \frac{\Delta N}{\Delta x} \quad (61)$$

$$|J \text{ (transl)}| = N v$$

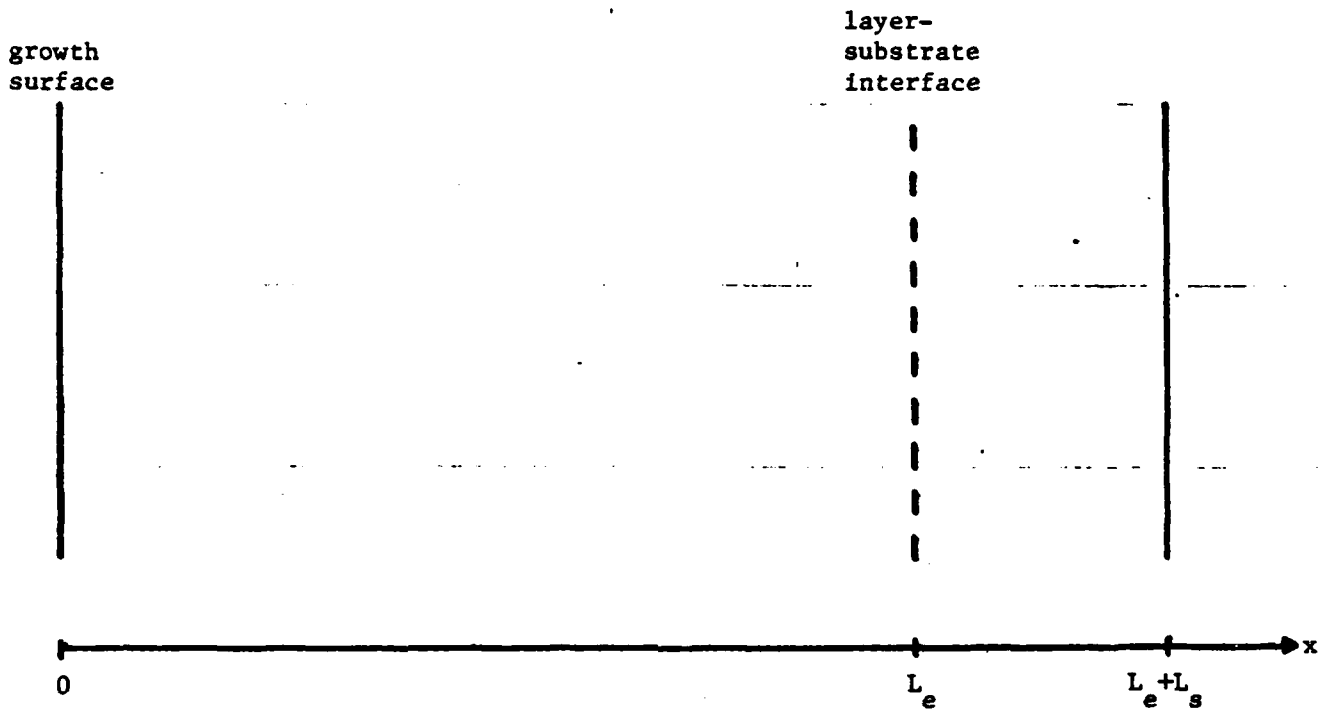
$j_{\max}$ is determined by Δt and the total time we want to grow, t_g .

$$t_g = (j_{\max} - 1) \Delta t \quad (62)$$

The length of the total studied region is

$$L = (i_{\max} - 1) \Delta x = L_e + L_s \quad (63)$$

where L_e is the thickness of the epitaxial layer determined by t and $v(t)$ and L_s is the depth of substrate that we consider at time t .



At $t = t_g$, L_s must be large enough so that at $x = L$ the properties of the substrate are essentially those at infinity. This means that the field has to be zero there and that there is no redistribution due to diffusion either. The fields fall off to zero within a few substrate Debye lengths L_{Ds} , so L_s must fulfill, say,

$$L_s \geq 4 \cdot L_{Ds}$$

If so at $x = L$ we have pure diffusion. This diffusion is approximately equal to what the diffusion would have been if the substrate and epitaxial layer

had been semi-infinite. In this case for the change in impurity concentration to be negligible, say 0.1 - 1% of the maximum value of the impurity concentration, L_s has to fulfill

$$L_s > 4\sqrt{Dt_g} \quad (64)$$

So finally L_s must fulfill

$$L_s \geq \max \{4L_{Ds}, 4\sqrt{Dt_g}\} \quad (65)$$

this puts a lower limit on $i_{\max}$. Finally we may want to choose D_0 so that we get τ expressed in some convenient unit like minutes or hours.

2.2.4 Final Remarks

We have discussed a number of aspects of the problem of field assisted redistribution of impurities during epitaxial growth. The redistribution part is completely outlined while the field problem needs more study. The amount of research effort and computer time that should be devoted to this problem depends in part on the impurities to be considered. For relatively immobile shallow dopants the simpler modified W & N approach should be adequate. If we also wish to study such problems as Cr redistribution in GaAs, a more accurate approach is needed. Since Cr is the most widely used dopant in semi-insulating GaAs, it is of particular interest to us.

Computing is in progress and we hope that before long we will have some interesting results.

References

1. C. M. Wolfe and K. H. Nichols, Appl. Phys. Lett., Vol. 31, No. 5, p. 356.
2. Final Scientific Report No. 59356-6, Washington University, St. Louis.
3. L. Fox and D. F. Mayers, Computing Methods for Scientists and Engineers, Clarendon Press, Oxford, 1968.
4. D. Shaw (editor), Atomic Diffusion in Semiconductors, Plenum Press, London and New York, 1973.
5. G. D. Smith, Numerical Solution of Partial Differential Equations, Oxford University Press, New York and London, 1965.
6. L. A. Girifalco, Atomic Migration in Crystals, Blaisdell Publishing Company, 1964.

2.3 Diffusion and Electronic Properties of Cr in GaAs

by H. Rohdin and M. Muller

2.3.1. Introduction

In the course of the work on field assisted redistribution of impurities during epitaxial growth, treated elsewhere in this report, we encountered several interesting redistribution properties of Cr in GaAs.

Cr in GaAs is a very interesting system technologically. Cr has the ability to produce semi-insulating (SI) GaAs and thus to compensate the residual donors in bulk grown GaAs. The semi-insulating properties seem remarkably independent of the amount of Cr added during the growth process. During high temperature treatment, such as post-implantation anneal, pre-epitaxial growth anneal and the epitaxial growth itself, redistribution of Cr occurs, which may degrade the wished-for electronic properties. It is therefore important to understand the causes of the redistribution of Cr in GaAs.

In the hope of being able to incorporate Cr transport into our model we started to investigate the properties of this system. So far the investigation has essentially been based on a rather extensive literature search in various fields, the most interesting being redistribution properties of Cr ions in GaAs, electronic properties of GaAs:Cr, optical properties of GaAs:Cr, and magnetic properties of the different Cr ions in GaAs. We soon realized that the problem was not an easy one. A lot of seemingly contradictory features are puzzling.

We have tried to develop a model that is consistent with most data in the different fields. So far it is tentative and experiments have to be performed to test it. We plan to do this in the near future. Here we will present the work done to date.

We will start by presenting some results from experiments performed on GaAs doped with Cr. We will point out the interesting features and the often encountered feeling of variability and contradictions in the results. We will take a brief look at Zn transport in GaAs. This is probably the most extensively studied case of acceptor diffusion in GaAs, and it can serve as a basis for discussion of Cr diffusion in GaAs. Then follows a survey of the well-established electronic behavior of Cr in GaAs inferred from electric measurements, EPR data and photoluminescence and photoconductivity data. Then comes a crucial section where we try to motivate the existence of interstitial Cr acting as a donor. This is not a new idea, but it never seems to have gained credibility since its suggestion by Broom [1] in 1967. A discussion in general terms about energies associated with different sites and charge states of Cr in GaAs, and on the corresponding one-electron energies will follow and serve as an introduction to an explicit motivation for auto-compensation of Cr in GaAs. Vacancies play an important role in the model and a short treatment of the pertinent vacancy chemistry and resulting concentration profiles will be given. A short sketch of the possibility of superfast diffusion will precede the presentation of our tentative model for Cr in GaAs. Finally we will apply the model to explain some of the experimental results in section 2.3.2 together with the explanations given by the different authors.

2.3.2. Certain Experimental Results on GaAs:Cr

2.3.2.1 Redistribution of Cr in GaAs

2.3.2.1.1 Annealing of Cr without encapsulation

Kasahara and Watanabe [2] and Tuck et al. [3] reported on Cr distribution in GaAs after annealing of Cr doped GaAs with encapsulation. Graphs taken from the two papers are shown in fig. 1 and 2 respectively.

[2] observed that for zero As pressure and an annealing time of one hour there was no outdiffusion for temperatures below 700°C. For temperatures above 800°C there was outdiffusion and the Cr concentration decreased gradually approaching the surface where it tended to zero.

For As-pressure larger than zero, when [2] annealed at a temperature of 850°C for one hour, the surface concentration no longer tended to zero. The amount of outdiffused Cr decreased.

[3] reported that one-hour annealing at 750°C resulted in a uniform lowering of the bulk Cr concentration. A high peak developed at the surface. For higher As pressure there was less lowering of the bulk concentration, a higher surface peak and all together less total outdiffusion.

There are three interesting things to note when comparing the results in [2] and [3]:

- 1) For higher As pressure the total outdiffusion was lowered. In [3] the Cr concentration in the entire crystal was affected. In [2] just the surface region was affected.
- 2) For essentially the same annealing temperature and time [2] and [3] got completely different profiles.
- 3) In the GaAs:Cr that [2] used, the Cr was incorporated during bulk growth, while in [3] the Cr had been indiffused in the form of ^{51}Cr .

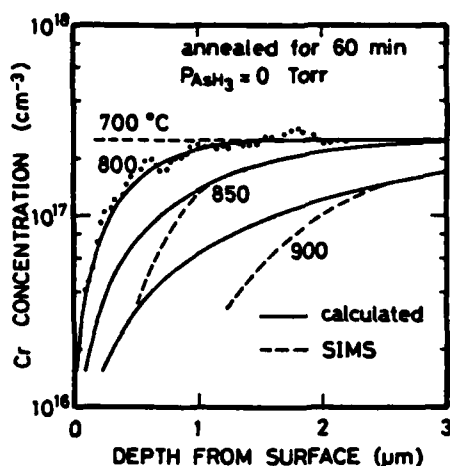


Fig. 1. Redistributed Cr profiles after annealing at temperatures ranging from 700-900°C for 60 min under $P_{\text{AsH}_3} = 0$ Torr. Broken lines show the experimental results except the case of 800°C where the experimental data are shown by the dots. Calculated profiles from eq. (1) in the text are shown by the solid lines.

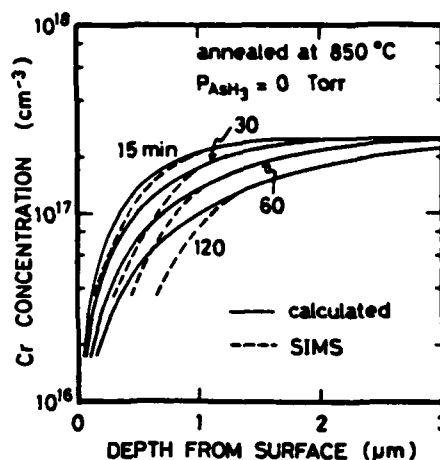


Fig. 2. Chromium profiles after annealing at 850°C for 15-120 min under $P_{\text{AsH}_3} = 0$ Torr. Broken lines and solid lines are experimental and calculated profiles, respectively.

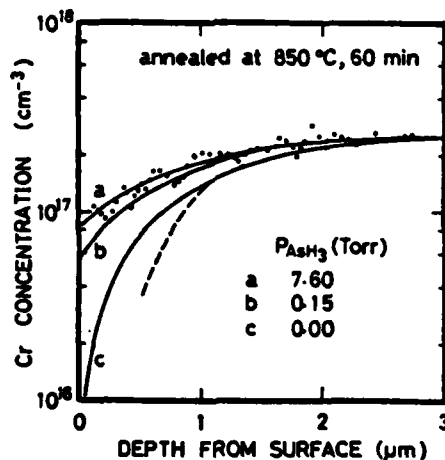


Fig. 3. Chromium profiles after annealing at 850°C for 60 min under $P_{\text{AsH}_3} =$ (a) 7.6 Torr, (b) 0.15 Torr, and (c) 0 Torr. Experimental SIMS data are represented by the dots for the case of $P_{\text{AsH}_3} = 7.60$ Torr. A broken line (c) is the smoothed-out experimental profile. Solid lines were calculated by eq. (1).

Fig. 1. Cr profiles in GaAs after annealing without encapsulation taken from [2]

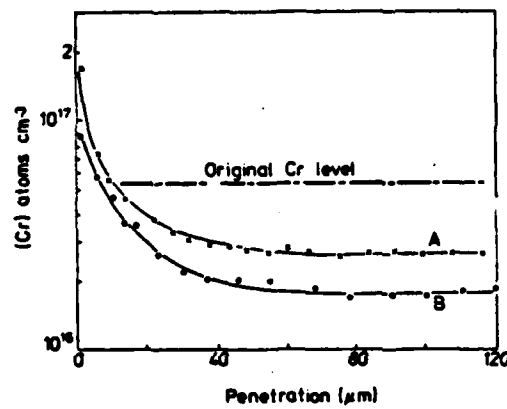


Figure 4. 'Drive-out' diffusion profiles for one-hour 750 °C reanneals A, in an epitaxial reactor and B, in an evacuated ampoule (see text).

Fig. 2. Cr profile in GaAs after annealing without encapsulation taken from [3]

2.3.2.1.2 Annealing of GaAs:Cr with encapsulation

Favennec and L'Haridon [4], Huber et al. [5] and Asbeck et al. [6] all studied annealing of GaAs:Cr with a Si_3N_4 encapsulation. Graphs from these papers of Cr distribution after anneal are shown below in fig. 3-5.

They all observed a very thin Cr peak at the substrate-encapsulant interface and a thicker Cr depletion zone below the interface. Note that the depletion and the extrapolation of this, i.e. the Cr concentration minus the surface build-up, is considered to be the "real" diffusion profile, to which an erfc-function can be fitted. When the redistribution is so large that electrical conversion occurs the substrate is considered "unqualified". In fig. 5 the upper curve is probably from an unqualified substrate while the other two probably are from qualified ones.

2.3.2.1.3 Diffusion of Cr into SI-GaAs:Cr

This was also studied by Tuck et al. [3]. Fig. 6 is taken from this paper.

There was an exceedingly rapid penetration of Cr through GaAs at temperatures between 800 and 1100°C. The profiles did not show a simple erfc-form. Instead there was a rapid fall-off at the surface while in the bulk the Cr concentration was essentially constant. Note that this was the way the substrates used in the outdiffusion experiments treated in 2.3.2.1.1 were prepared. The surface region was removed before outdiffusion.

2.3.2.1.4 Epitaxial growth of GaAs on GaAs:Cr

Tuck et al. [3] and Wilson et al. [7] both performed epitaxial growth of GaAs on Cr-doped GaAs. In [3] VPE growth at 750°C of undoped and S-doped epitaxial layers were performed on substrates doped with Cr during bulk growth

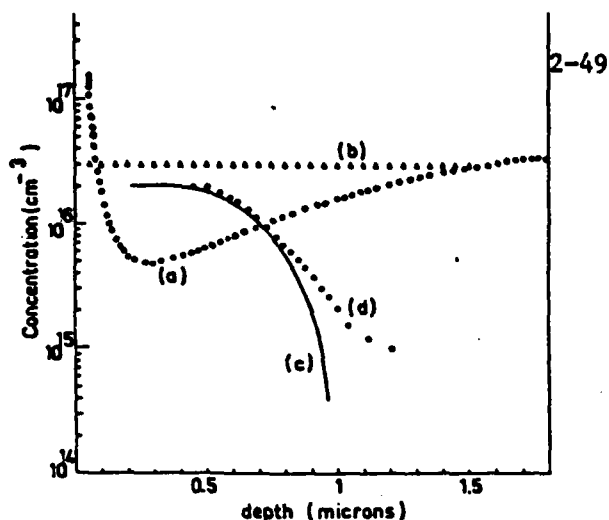


Fig. 3. Taken from [4]

FIG. 3. For substrates B, chromium (curve a) and silicon (curve b) distributions; calculated (curve c) and measured (curve d) electron carrier distributions.

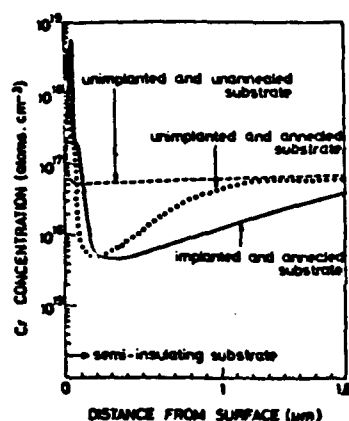


Fig. 4. Taken from [5]

FIG. 1. Cr profiles in GaAs samples.

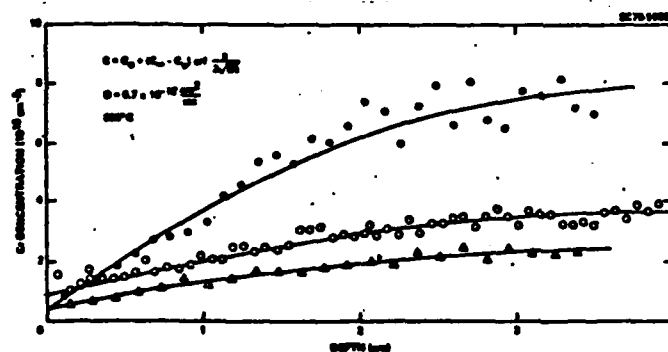


Fig. 5. Taken from [6]

Fig. 1. Cr distribution measured by SIMS in substrates annealed with a Si_3N_4 cap. The experimental data have been fit with error function curves corresponding to a diffusion coefficient $D = 8.7 \times 10^{-13} \text{ cm}^2/\text{sec}$. Data points within $0.1 \mu\text{m}$ of the GaAs surface have been omitted (see text).

Fig. 3-5. Cr profiles in GaAs after annealing with encapsulation

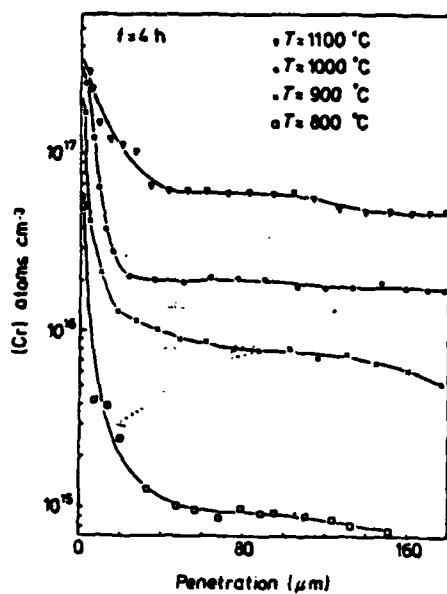


Figure 2. Profiles for four-hour diffusions of Cr into semi-insulating GaAs at 800, 900, 1000 and 1100 °C.

a

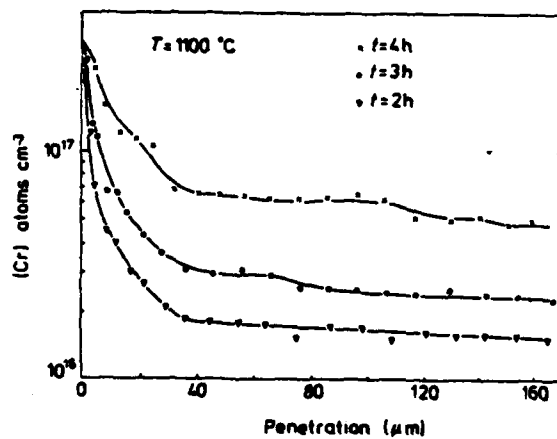


Figure 3. 1100 °C diffusion profiles of Cr in GaAs for $t = 2, 3$ and 4 h.

b

Fig. 6. Cr profiles in GaAs after indiffusion of Cr taken from [3]

and by indiffusion of Cr radiotracer as described in 2.3.2.1.3. The result is shown in figure 7.

In [7] undoped layers of GaAs were grown by LPE at 800°C on SI-GaAs:Cr where the Cr was incorporated during bulk growth. The result is shown in the figure 8.

[3] observed large outdiffusion and noted that S-doping inhibited outdiffusion. [7] observed less outdiffusion and an erfc-function could be fitted to the Cr profile.

Note that in [3] the outdiffusion was larger than in [7]. Note also the difference in incorporation of the observed Cr ions.

2.3.2.2 Electronic Properties of Cr in GaAs

2.3.2.2.1 Bulk grown GaAs:Cr

Cronin and Haisty [8] were the first to report on the electronic properties of GaAs doped by Cr during bulk growth. They observed that if one adds a sufficient amount of Cr to the melt the resulting GaAs is semi-insulating. This is so even when the added Cr amount is much larger than the residual donor concentration. The table in fig. 9, taken from [8], illustrates this.

Determination of the conductivity type of the crystals by Hall measurements and thermal probe did not always yield the same answer. According to Dr. C.M. Wolfe the thermal probe technique probably was the most reliable. If so, the type can be either p or n. This agrees with the observations of Zucca [9] who also noted that excess Cr still produced semi-insulating GaAs.

Brozel et al. [10] observed in their experiments that a larger Si concentration during bulk growth resulted in a larger Cr concentration. They

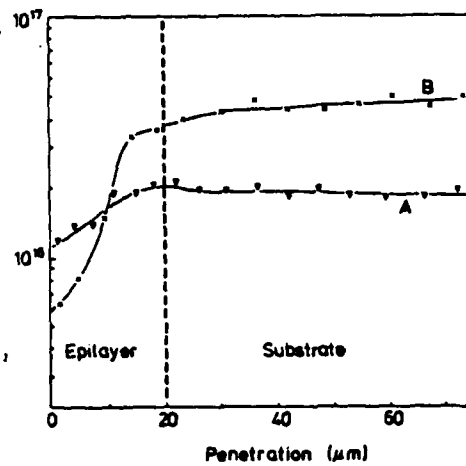


Figure 5. Out-diffusion during epitaxy of A, undoped and B, S-doped layers.

Fig. 7. Taken from [3]

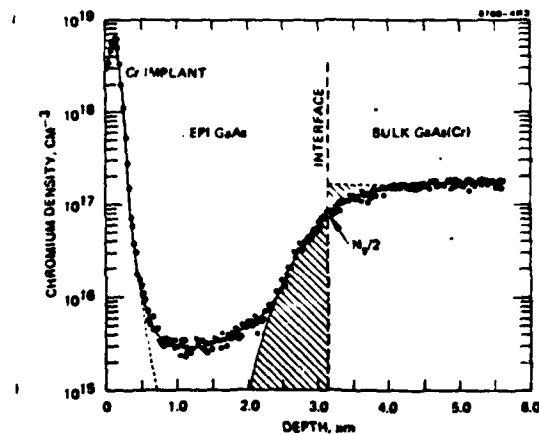


FIG. 3. SIMS depth distribution of ^{52}Cr in an undoped epitaxial layer of GaAs grown on a GaAs(Cr) substrate and implanted with ^{52}Cr . Profiled through the epitaxial layer and into the bulk GaAs(Cr).

Fig. 8. Taken from [7]

Fig. 7,8. Cr profiles in GaAs substrate and epilayer after outdiffusion into the latter

also concluded that Cr is a double acceptor in GaAs. This is the commonly accepted picture of the electronic action of Cr in GaAs. Note that the common observation is that the concentration of compensating Cr is adjusted to the concentration of the residual donor to be compensated. [10] also observed a geometrical adjustment.

2.3.2.2.2 Cr indiffused into GaAs

Tuck et al. [3] performed Hall measurements on GaAs samples into which radioactive Cr had been indiffused. Their result: If the sample initially was SI-GaAs:Cr it became p-type ($p=6 \times 10^{14} \text{ cm}^{-3}$) after indiffusion. If the sample initially was n-GaAs:Te it became high resistive (n or $p < 10^{14} \text{ cm}^{-3}$) in a region below the surface while beyond this the sample remained n-type.

2.3.2.2.3 SI-GaAs:Cr annealed under encapsulation

Asbeck et al. [6] observe a spurious n-type layer just below the GaAs-Si₃N₄ interface as illustrated by figure 10 taken from [6].

In addition, Favennec and L'Haridon [4], Huber et al. [5] and Asbeck et al. [6] all reported a thicker n-conducting layer below the interface for non-qualified SI-GaAs:Cr substrates corresponding to the Cr depletion region discussed earlier. For some substrates [4] got p-conversion. The three cases observed by [4] are illustrated in fig. 11 below.

2.2.4 Epitaxial growth of GaAs on SI-GaAs:Cr

In K.H. Nichols' doctoral dissertation [11] fig. 12, based on the results of Yamasaki et al. [12], appears. It shows an interfacial region with higher electron concentration than expected. Essentially the same phenomenon was reported by Khokhlov et al. [13]. They grew nondoped VPE layers on SI-GaAs:Cr. Fig. 13 is taken from their paper.

Table I. Properties of chromium-doped GaAs at 300°K

Crystal	Cr analysis (ppm by wt)	R_H ($\text{cm}^2/\text{coulomb}$)	R_H/ρ ($\text{cm}^2/\text{v-sec}$)	ρ (ohm-cm)	Type By Hall coefficient	By thermal probe
3-37	0.2-0.3	1.2×10^{11}	1130	1.06×10^8	n	p
3-41	0.4	2.8×10^8	2120	1.34×10^{-1}	n	n
3-42	0.2-0.5	8.8×10^{10}	256	3.43×10^8	n	p
3-44	0.4	3.5×10^{11}	630	5.61×10^8	n	p
3-45*	0.4-0.5	3.0×10^8 4.0×10^{11}	—	3.30×10^8	n	n
555-215	1.8	3.2×10^{11}	1120	2.84×10^8	n	p
555-216	3.5	2.9×10^{11}	693	4.17×10^8	n	p
3-50	360-400	1.3×10^{10}	614	2.15×10^8	n	p

* Crystal 3-45 inhomogeneous. Two values for Hall coefficient are measurements at each end of sample.

Fig. 9. Electrical data of GaAs doped with Cr during bulk growth taken from [8]

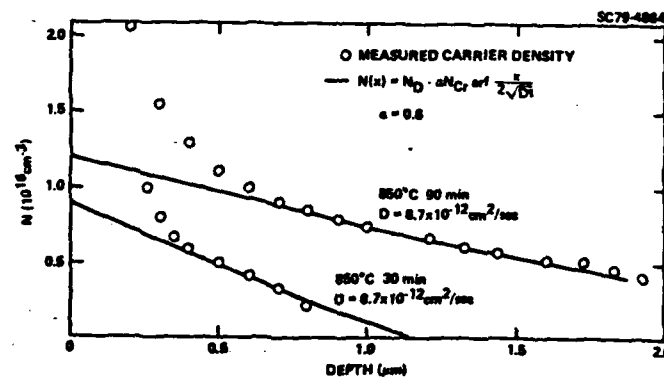


Fig. 3. Comparison of carrier density profiles in unimplanted, annealed samples with predicted profiles based on Cr redistribution.

Fig. 10. Electron concentration profiles in GaAs:Cr after annealing with encapsulation taken from [6]

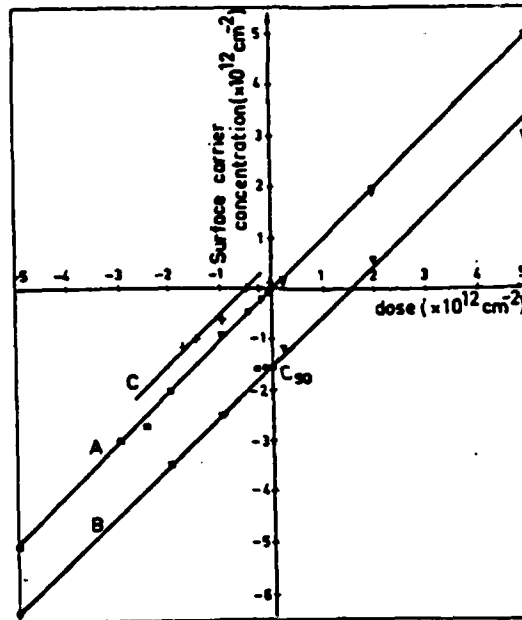


FIG. 2. Surface carrier concentration (hole > 0 , electron < 0) versus the dose ($\phi > 0$ if Zn, $\phi < 0$ if Se) for A, B, and C substrates.

Fig. 11. Surface carrier concentration in GaAs:Cr substrates after implantation and annealing

- A - qualified substrates (no conversion)
- B - unqualified substrates (n-conversion)
- C - unqualified substrates (p-conversion)

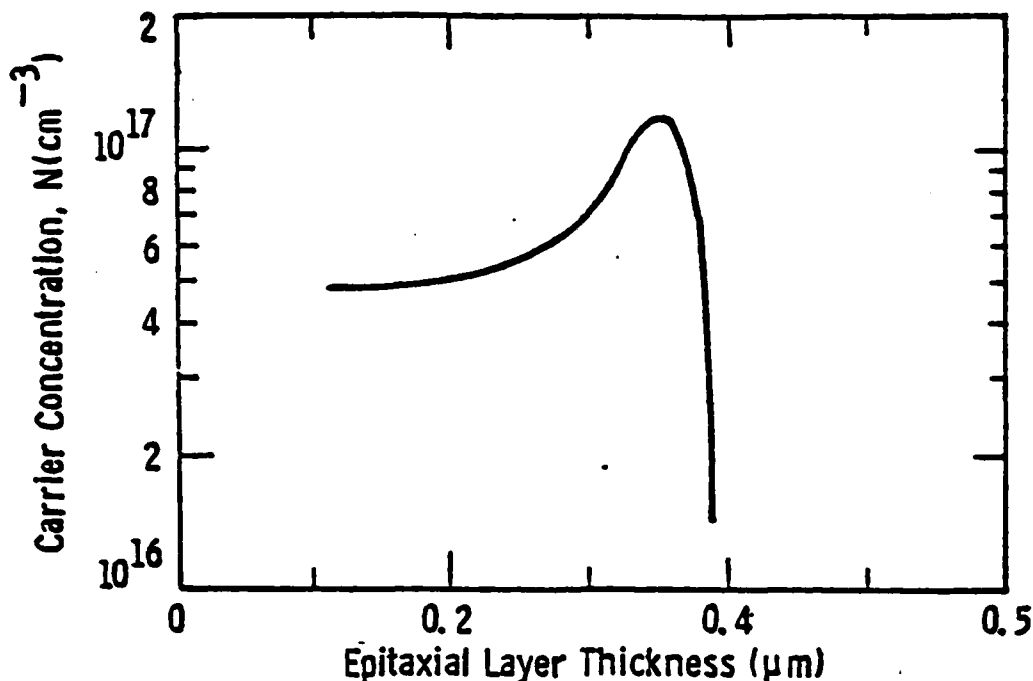


Fig. 12. Carrier concentration versus epitaxial layer thickness for an n-type epitaxial layer (LPE) grown on a Cr-doped substrate with an n^+ interfacial region. Taken from [11]

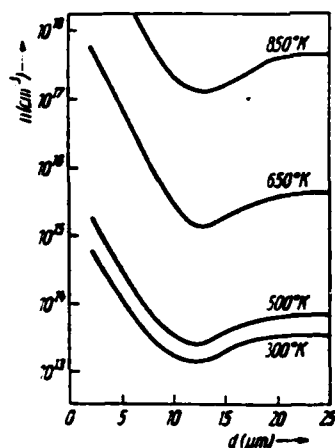


Fig. 3. Temperature change in the electron density profile along the film thickness. d is the film thickness from the film-substrate interface

Fig. 13. Taken from [13]. Non doped VPE layers on SI-GaAs:Cr substrates

Fig. 12, 13. Electron concentration profiles in epitaxial layers grown on SI-GaAs:Cr substrates

Both [12] and [13] got an increase of the electron concentration in the layer close to the layer-substrate interface. Fig. 13 shows a region of lower electron concentration as well.

Schlachetzki and Salow [14] grew non doped LPE-layers at different temperatures and from different source materials. For each source material (I and II) they got n-type, p-type or semi-insulating layers depending on the growth temperature. For temperatures close to a temperature T_0 , specific for each source material, they got SI-GaAs layers. For growth temperatures larger than T_0 they got p-type layers and for growth temperatures below T_0 they got n-type layers. Fig. 14 below is taken from [14].

Khokhlov et al. [13] also reported that for high temperatures the carrier density in their VPE layers increased faster with temperature than the intrinsic concentration. This is illustrated below in figure 15 taken from [13].

2.3.3. Zn Diffusion in GaAs

Zn diffusion in GaAs is well explained by a substitutional-interstitial diffusion model in which the Zn ions are distributed over both substitutional and interstitial sites. More specifically, the assumptions associated with this model are:

- 1) Zn can exist substitutionally, substituting for Ga (Zn_{Ga}), where it acts as a shallow single acceptor.
- 2) Zn can also exist interstitially (Zn_i), acting as a shallow single donor.
- 3) Ga vacancies (V_{Ga}) are neutral and electrically inactive.
- 4) There is an equilibrium reaction in which Zn changes site:

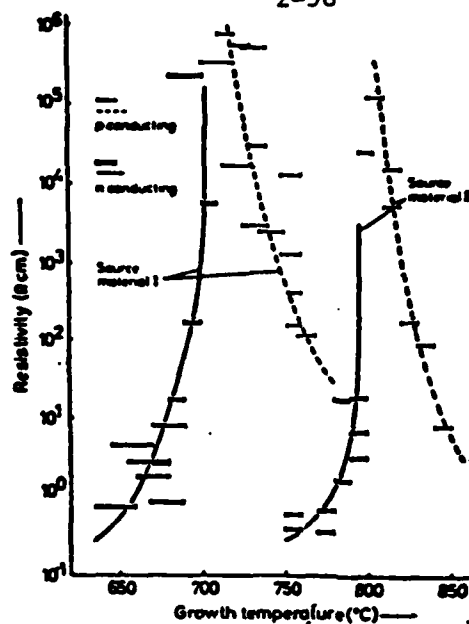


Fig. 2. Resistivity of epitaxial layers grown from different source material in dependence on growth temperature. Bars indicate temperature interval of growth. (Low temperature side: *n*-conducting material; high temperature side: *p*-conducting material)

Fig. 14. Taken from [14]. Non-doped LPE layers on SI-GaAs:Cr substrates

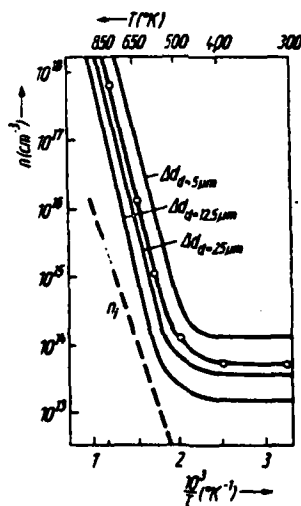
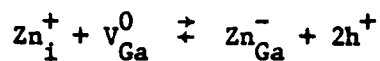


Fig. 4. Dependence of the electron density on the reciprocal temperature. $\Delta d \rightarrow 0$; $\circ$ experiment for $d = 25 \mu\text{m}$; n_i intrinsic electron density in GaAs

Fig. 15. Taken from [13]. Non-doped VPE layers grown on SI-GaAs:Cr substrates



with the mass action law

$$[\text{Zn}_i^+] = [\text{Zn}_{\text{Ga}}^-] (\gamma p)^2 / K(T) P_{\text{As}_2}^{1/2}$$

where γ is the hole activity, $K(T)$ the equilibrium constant and P_{As_2} the As-pressure.

- 5) The concentration of Zn_i is much smaller than that of Zn_{Ga} i.e.

$$[\text{Zn}_i^+] \ll [\text{Zn}_{\text{Ga}}^-] \approx [\text{Zn}]$$

These assumptions lead to an over-all diffusion constant (including drift in the built in field) which is proportional to the square of the Zn_{Ga} -concentration:

$$D(\text{Zn}) \sim [\text{Zn}_{\text{Ga}}^-]^2$$

Non-equilibrium effects do contribute but the treatment gives excellent agreement with experiment as long as the Zn concentration is not too high. Casey wrote a review article [15] which emphasizes Zn diffusion in GaAs. Figure 16 is taken from this and shows the good agreement between theory and experiment. Note the difference between these profiles and those obtained by Tuck et al. [3] shown in section 2.3.2.1.3: Zn is absorbed by Ga-vacancies

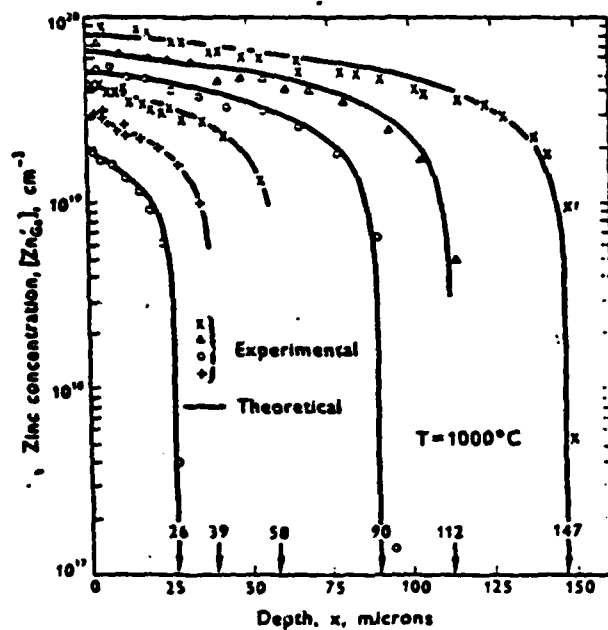


Figure 6.16. Comparison of Cunnell and Gooch's [12] experimental Zn diffusion profiles in GaAs with Weisberg and Blanc's [73] calculated diffusion profiles with $D(Zn) \propto [Zn_{Ga}]^2$. The arrows indicate the 'effective zero' for each theoretical curve.

Fig. 16. Taken from [15]

in an equilibrium process which leads to a sharp cut off in the diffusion profile while Cr show extremely high penetration leading essentially to a uniform bulk concentration as well as a surface build up. This indicates a completely different transport mechanism for Cr. In our model, however, the Cr diffusion mechanism does have some similarities with Zn diffusion as we will see.

2.3.4. Well Established Electronic Properties of Cr in GaAs

The free Cr-atom has the electronic configuration $3d^5 4s$. In GaAs, Cr is considered to occupy Ga vacancies. This basically leads to a tetrahedral symmetry for the Cr atoms. Cr contributes 3 electrons to form bonds with the 4 surrounding As atoms. This leaves a Cr^{3+} core which generally is supposed to be able to accept one or two electrons.

We thus have at least 3 charge states of Cr on Ga-sites:

- Cr^{3+} - neutral
- Cr^{2+} - one electron accepted
- Cr^{+} - two electrons accepted

In reality the first two states do not show tetrahedral symmetry. Instead Cr^{3+} and Cr^{2+} are unstable to Jahn-Teller distortion as reported by Krebs and Stauss [16], [17].

Figure 17, taken from [18], shows a phenomenological ionization diagram including lattice relaxation for Cr in GaAs. Both acceptor levels are deep. Figure 18 shows a more detailed scheme taken from [19]. This is based on EPR data. We can see how the Jahn-Teller distortion of the Cr^{2+} center has lifted the orbital degeneracy of the d-levels. From optical data Stocker and Schmidt [20] deduced the level scheme shown in fig. 19.

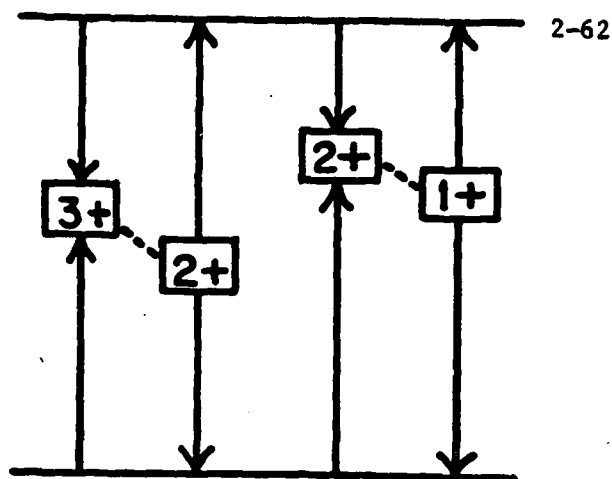


Fig. 17. Taken from [18]

FIG. 1. A schematic phenomenological ionization diagram for Cr in GaAs, showing possible electron transitions (vertical arrows) and subsequent relaxation relative to the bands due to lattice coupling (broken lines). Upward transitions represent optical excitation involving exchange of carriers with the bands; downward transitions represent carrier capture. Cr^{3+} is regarded as a double acceptor in GaAs, with Cr^{2+} and Cr^{1+} representing first and second ionization states, respectively.

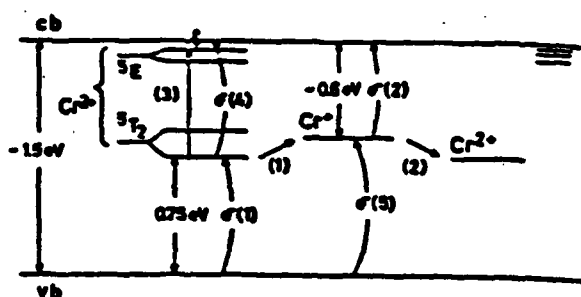


Fig. 18. Taken from [19]

Fig. 3. Cr levels in GaAs as inferred from the ESR data. The splitting of the 5E and 3T_2 states of Cr^{3+} is expected because of a static, tetragonal Jahn-Teller distortion^{7, 8}.

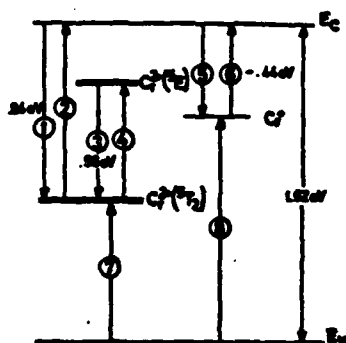


Fig. 19. Taken from [20]

Fig. 5. Energy levels of Cr in GaAs. Arrows indicated experimentally observed optical transition

Details in these schemes may differ due to the different underlying experiments but the overall picture of two deep acceptor levels emerges clearly.

2.3.5. Cr as an Interstitial Donor in GaAs

As mentioned in the introduction, Broom proposed in 1967 [1], to explain why he had

$$\left| \frac{N_a}{N_d} - 1 \right| \ll 1$$

in his Cr doped GaAs samples, "an auto compensation mechanism in which part of the total number of Cr atoms is either neutral, or possibly at interstitial sites, acting as donors". The idea came from a paper by Allen [21] concerning autocompensation by oxygen in GaAs. In the case of Cr the idea has never seemed to get wide acceptance. We believe that Cr actually is autocompensating. This seems, together with other ideas to be presented in the following sections, to be able to account for many of the experimental results assembled in section 2.3.2.

In addition to Broom's reason there are other indications of an interstitial Cr-donor.

Deveaud and Favennec [22] observed a new PL-line in substrates that had been Cr-implanted and annealed, and also in contaminated epitaxial layers, i.e. layers into which Cr had outdiffused from the substrate. A very interesting thing about the latter is that, although the region into which Cr had outdiffused contained almost as much Cr as the substrate, it was not semi-insulating. Furthermore the new PL line was much stronger than those usually assigned to Cr. They suggested that Cr may be located

elsewhere. Overall their result is consistent with the existence of an interstitial Cr-donor.

Schlachetzki and Salow [14] and André and LeDuc [23] got n-conductivity in GaAs crystals grown from the liquid phase. [23] concluded that Cr forms a shallow donor contrary to Cr-doped crystals grown from the melt.

A reason why Cr may find it favorable to exist in GaAs at a place different from the substitutional site can be found in [19] where the authors, Kaufmann and Schneider, say that to their knowledge Cr^{3+} has never been observed in tetrahedral ligand coordination in any semiconductor. Furthermore, they say, there exist no stable tetrahedral Cr^{3+} complexes.

Once a Cr^{3+} occupies a Ga-vacancy the instability of the tetrahedral configuration seems, in the light of what is said in [16] and [17], to result in Jahn-Teller distortion.

For the sake of comparison we list some of the pertinent properties of Cr together with those of two impurities, Zn and Cu, which do occupy both substitutional and interstitial sites.

	Cr	Zn	Cu
Electronic activity in GaAs	deep double acceptor on Ga-site	shallow single acceptor on Ga-site; shallow single donor interstitially	double acceptor sub- stitutionally; single donor interstitially
Electronic configuration	$3d^5 4s$	$3d^{10} 4s^2$	$3d^{10} 4s$
Ionic radius for +1-ion (Å)	0.81	0.88	0.96
First ionization energy (eV)	6.764	9.391	7.724
Atomic number	24	30	29

Apart from the fact that Cr is a transition metal, the properties of Cr do not differ much from those of Zn and Cu. Most interesting is that the ionic radius of Cr is the smallest of the three, which intuitively would make an interstitial site favorable.

2.3.6. Site Energies and Cr Distribution Over Sites

If Cr can exist both substitutionally and interstitially, it is of course of interest for both the electronic properties of GaAs:Cr and the transport properties of Cr in GaAs to determine the Cr distribution among sites.

In view of what has been said about the stability of the different Cr ions we will try to motivate the principal and phenomenological site energy scheme in figure 20.

Both Cr^{3+} and Cr^{2+} are unstable to Jahn-Teller distortion, the former being the least stable. The minimum energy position should therefore be displaced from the Ga-site. Krebs and Stauss [16] suggest that Cr^{3+} belongs to a class of centers which exhibits Jahn-Teller distortion with large stabilization energies. If so, the barrier for Cr^{3+} against going interstitial should be rather large. As, according to [19], Cr^{2+} is less unstable than Cr^{3+} , the corresponding barrier is expected to be even larger for Cr^{2+} . This should be even more true for Cr^+ which is stable to Jahn-Teller distortion.

For interstitial Cr the barrier is probably large as well. This is indicated by the observation in [22] that Cr outdiffused into a growing layer did not produce semi-insulating material i.e. did not in general go into a substitutional site where it may act as an acceptor.

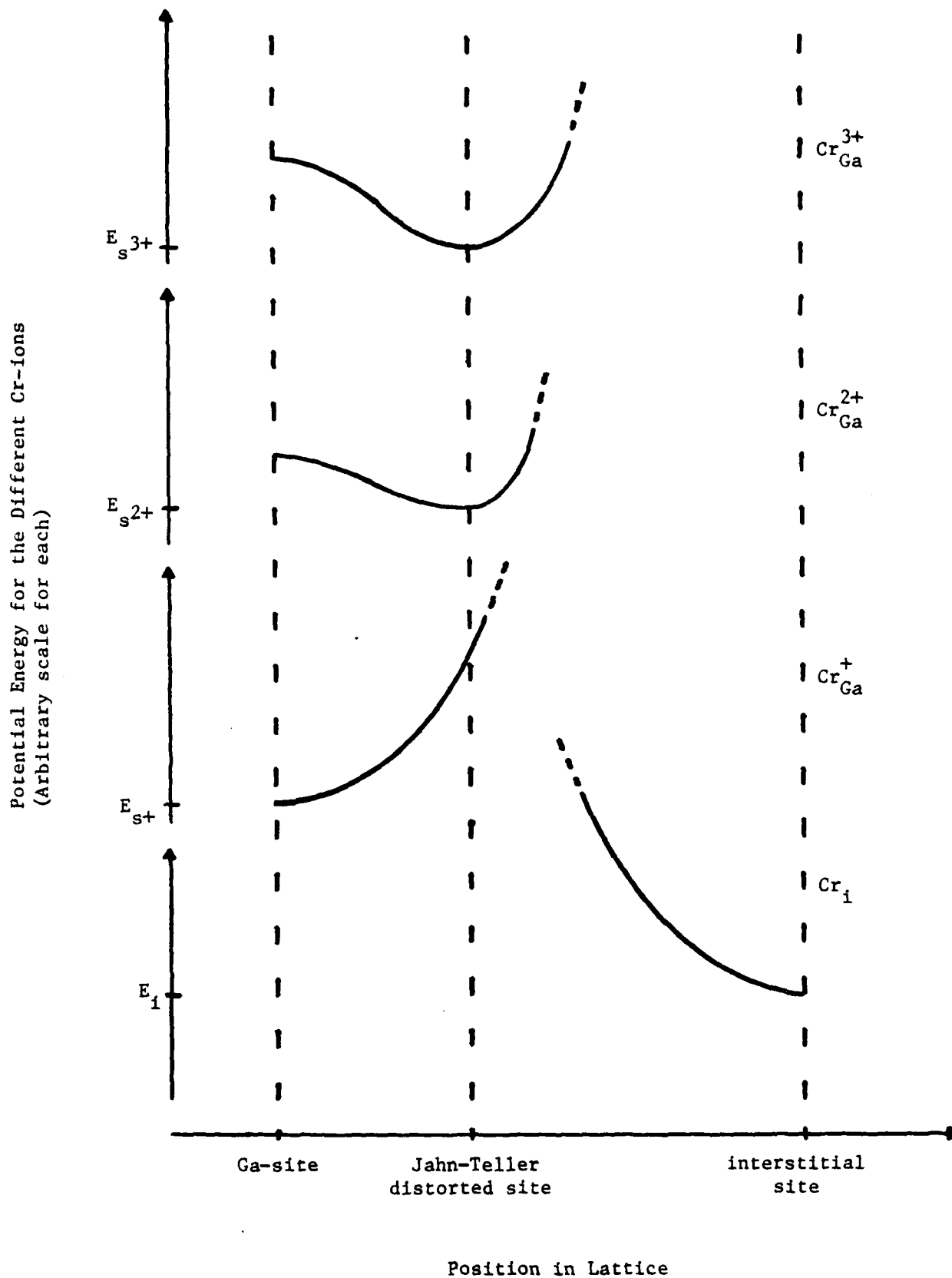


Fig. 20. Phenomenological site energy scheme for the different Cr ions in GaAs

The discussion of the site energy scheme is not intended to contain any detailed information. The essence of it is to convey our belief that the barriers involved are substantial, leading, in general, to a nonequilibrium distribution of Cr ions over the sites.

The redistribution among sites should depend on:

- 1) The temperature: the higher the temperature, the more easily the Cr ions may overcome the barriers.
- 2) The doping: in more n-type material fewer Cr ions leave their substitutional sites, since most of the Cr ions are in the more stable states Cr^{2+} or Cr^+ .
- 3) The vacancy concentration: the larger the concentration of Ga vacancies, the larger is the tendency of Cr ions to be substitutional.

Only for very high temperatures should it be easy for Cr ions to come into equilibrium. At bulk growth from the melt it seems reasonable that the Cr ions could be incorporated in their equilibrium distribution. If so, the latter should coincide with the distribution necessary for a high degree of compensation, since nearly perfect compensation is almost always observed for sufficiently high concentration of Cr when incorporated during growth from the melt ([1],[8],[9],[10]).

During epitaxial growth the situation is different. First of all, the redistribution after incorporation into the lattice is inhibited by the lower temperatures. Furthermore the incorporation into the lattice may differ with growth temperature and vacancy concentration (which depends on the temperature and the type of epitaxial growth).

In the case of indiffusion of Cr the situation again is different. As when enters an already grown lattice it seems likely that the Cr to a

larger extent should remain interstitial compared to incorporation of Cr during epitaxial growth.

2.3.7. One-electron Energies

Approximate values for the one-electron energies associated with $\text{Cr}_{\text{Ga}}^{2+}$ and $\text{Cr}_{\text{Ga}}^{+}$ can be inferred from the level schemes in section 2.3.4. The level associated with $\text{Cr}_{\text{Ga}}^{+}$ should lie somewhat above the intrinsic Fermi level while that associated with $\text{Cr}_{\text{Ga}}^{2+}$ should lie close to it but below, so that when the material is compensated (essentially intrinsic), most Cr_{Ga} are $\text{Cr}_{\text{Ga}}^{2+}$. The one-electron energy of Cr_i should be close to the conduction band since the donor is thought to be shallow [14].

A phenomenological one-electron scheme is shown in figure 21. A level associated with a shallow residual donor is included.

2.3.7.1 Occupation of the One-electron Energy Levels

If we proceed in a fashion similar to that in Seeger [25] p. 35-40 we can get the occupation of the different one-electron levels.

Notation:

g -statistical weighting factor

$[x]$ - concentration of species x

subscript s on substitutional Cr

subscript i on interstitial Cr

$$[\text{Cr}_s] = [\text{Cr}_s^{+}] + [\text{Cr}_s^{2+}] + [\text{Cr}_s^{3+}] \quad (1)$$

$$[\text{Cr}_i] = [\text{Cr}_i^0] + [\text{Cr}_i^{+}] \quad (2)$$

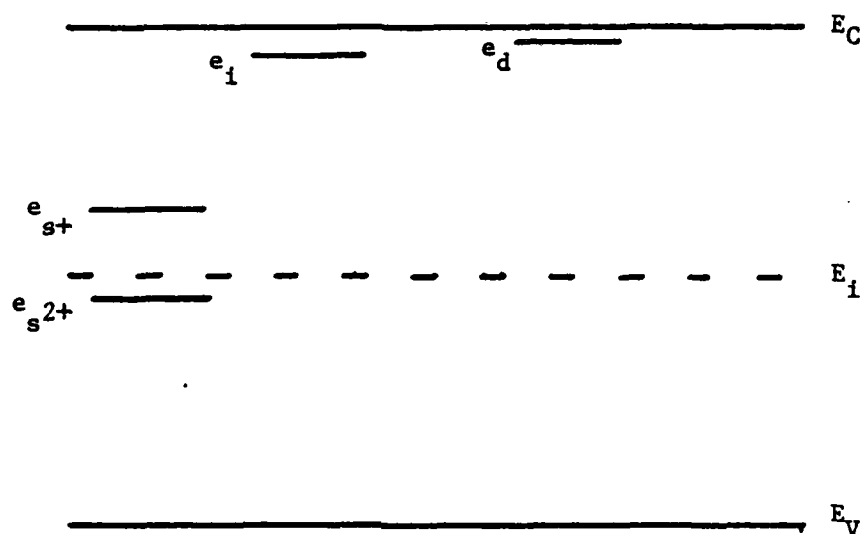


Fig. 21. Phenomenological one-electron scheme for GaAs containing Cr and a shallow donor (e.g. Si)

$$\frac{[Cr_s^{2+}]}{[Cr_s]} = \left\{ \frac{1}{g_{s^{2+}}} e^{(e_{s^{2+}} - E_F)/kT} + 1 + g_{s^+} e^{-(e_{s^+} - E_F)/kT} \right\}^{-1} \equiv f_2(E_F) \quad (3)$$

$$\frac{[Cr_s^+]}{[Cr_s]} = \left\{ \frac{1}{g_{s^+}} e^{(e_{s^+} - E_F)/kT} + 1 + \frac{1}{g_{s^{2+}}} e^{(e_{s^{2+}} - E_F)/kT} \frac{1}{g_{s^+}} e^{(e_{s^+} - E_F)/kT} \right\}^{-1}$$

$$\equiv f_1(E_F) \quad (4)$$

$$\frac{[Cr_i^0]}{[Cr_i]} = \left\{ \frac{1}{g_{i^0}} e^{(e_{i^0} - E_F)/kT} + 1 \right\}^{-1} \quad (5)$$

$$\frac{N_d^0}{N_d} = \left\{ \frac{1}{g_d} e^{(e_d - E_F)/kT} + 1 \right\}^{-1} \quad (6)$$

For non-degenerate material:

$$n = n_i e^{(E_F - E_i)/kT} \quad (7)$$

$$p = n_i^2/n \quad (8)$$

where E_i is the intrinsic Fermi level:

$$E_i = \frac{1}{2} (E_V + E_C) + \frac{3}{4} kT \ln \frac{m_h}{m_e} \quad (9)$$

The Fermi level is as usual determined in the bulk by the condition of charge neutrality

$$[\text{Cr}_i^+] - [\text{Cr}_s^{2+}] - 2[\text{Cr}_s^+] + N_d^+ + p - n = 0 \quad (10)$$

2.3.8. Explicit Motivation for Autocompensating Cr in GaAs at Room Temperature (300K)

Suppose now that there is no interstitial Cr donor level and that all the Cr is substitutional on Ga sites and acts as double deep acceptors. Assume room temperature ($T=300\text{K}$) and consider the valence band as one single band.

2.3.8.1 The Resistivity as a Function of Fermi Level

Impurity scattering is not dominant at room temperature, so the mobility is essentially independent of doping. We use Kittel's [24] values for the mobilities:

$$\mu_n = 8800 \text{ V cm}^2/\text{s}$$

$$\mu_p = 400 \text{ V cm}^2/\text{s}$$

The resistivity is then given by (two bands)

$$\rho = \rho_i \frac{1 + 1/b}{e^x + (1/b)e^{-x}} \quad (1)$$

where ρ_i is the intrinsic resistivity

$$\rho_i = [q \mu_n n_i (1+1/b)]^{-1}, \quad (2)$$

$$b = \mu_n / \mu_p \quad (3)$$

and

$$x = \frac{E_F - E_i}{kT} \quad (4)$$

The logarithm of $\frac{\rho}{\rho_i}$ is plotted below in figure 22a as a function of $\frac{E_F - E_i}{kT}$. The maximum value of ρ occurs for $\frac{E_F - E_i}{kT} \approx -1.5$. Outside the region $-5 < \frac{E_F - E_i}{kT} < 2$ the resistivity is less than 10% of this and falls exponentially the further away from $\frac{E_F - E_i}{kT} = -1.5$ one gets.

2.3.8.2 The Fermi Level as a Function of Cr Concentration

At room temperature, $T = 300K$, the band gap is

$$E_g = 1.42 \text{ eV}$$

The electron mass and mean value of the hole masses are

$$m_e = 0.07 m_0$$

$$m_h = 0.33 m_0$$

The intrinsic carrier concentration is

$$n_i = 1.75 \cdot 10^6 \text{ cm}^{-3}$$

The intrinsic conductivity is

$$\rho_i = 3.9 \cdot 10^8 \Omega \text{cm}$$

The intrinsic Fermi level lies at

$$E_i = \frac{1}{2} (E_v + E_c) + 0.1 \text{ meV}$$

which is essentially in the middle of the gap. We chose the following values which are inferred from the level schemes in section 4

$$e_{s2+} = E_i$$

$$e_{s+} = E_i + 0.3 \text{ eV}$$

For the shallow residual donor we pick

$$e_d = E_c - 0.01 \text{ eV}$$

We do not know the statistical factors but let us assume that these are all 2, as in the case of shallow impurities.

$$g_{s2+} = g_{s+} = g_d = 2$$

For the residual donor concentration we pick a typical value

$$N_d = 10^{10} n_i$$

We use formulas (3), (4), (6), (7), (8), (10) of section 2.3.7.7 to numerically determine the Fermi level as a function of the Cr concentration. The result is shown in figure 22b and c. 22c is just a more detailed plot of the region in 22b where the Fermi level drops steeply. The interval in Cr concentration over which the resistivity is larger than 90% of its maximum value is less than two orders of magnitude. This is not consistent with the results obtained by Cronin and Haisty [8] who did not observe any significant difference in resistivity when they changed the Cr content 3 orders of magnitude. This indicates that some kind of autocompensation mechanism is likely.

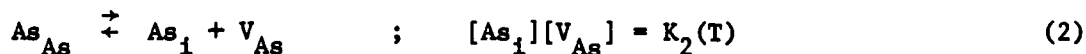
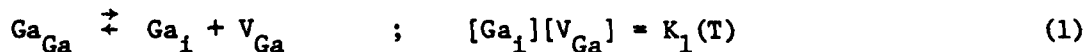
2.3.9. Vacancies in GaAs

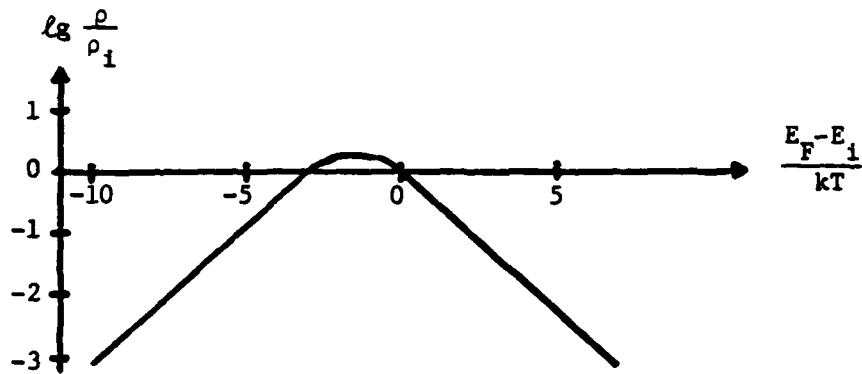
Since the vacancy concentration is intimately connected to the diffusion of substitutional species it is worth taking a look at vacancy formation and distribution in GaAs.

2.3.9.1 Equilibrium Reaction and Mass Action Laws

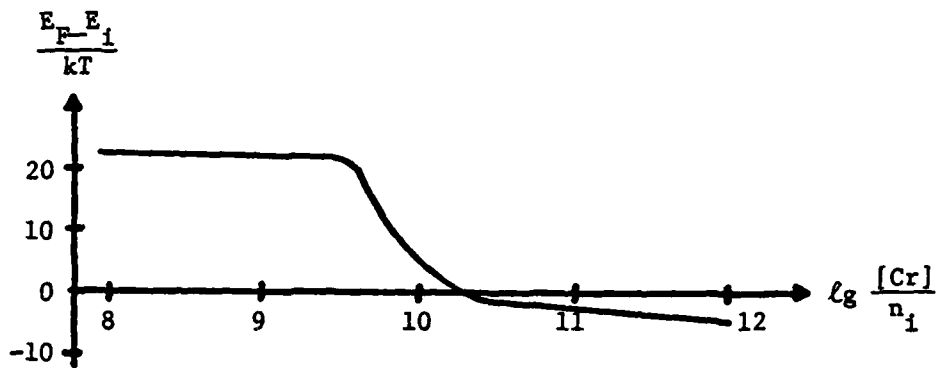
Let us assume first that the vacancies are in equilibrium with the rest of the crystal. There are a number of reactions in which the vacancies are produced. The presence of a surface will change the situation there, so we distinguish between the surface (S) and the bulk (B) far from the surface. Subscript g stands for gas phase, l for liquid phase and s for solid phase.

Bulk and Surface:

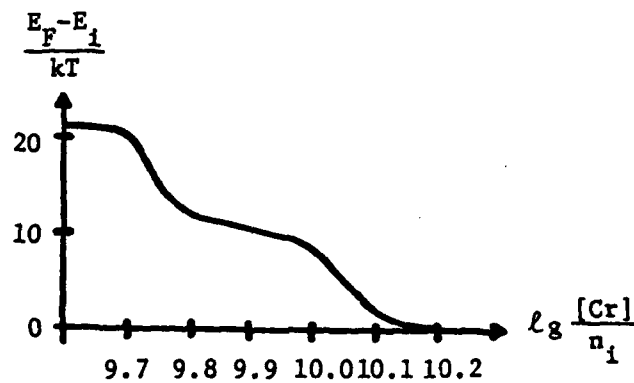




(a)



(b)



(c)

Fig. 22. (a) The resistivity as a function of the Fermi level in GaAs
 (b) The Fermi level as a function of the Cr concentration assuming no interstitial Cr and $N_d = 10^{10} n_i$ at $T = 300K$
 (c) Same as (b) for a smaller interval

Surface

$$\text{Ga}_i \rightleftharpoons \text{Ga}_1 \quad ; \quad \frac{[\text{Ga}_i]_S}{[\text{Ga}_1]_S} = K_4(T) \quad (4)$$

$$\text{Ga}_{\text{Ga}} \rightleftharpoons \text{Ga}_1 + V_{\text{Ga}} \quad ; \quad \frac{[\text{Ga}_{\text{Ga}}]_S}{[\text{Ga}_1]_S [V_{\text{Ga}}]_S} = K_5(T) \quad (5)$$

$$\text{As}_i \rightleftharpoons \frac{1}{n} \text{As}_{n_g} \quad ; \quad \frac{[\text{As}_i]_S}{P_{\text{As}_n}^{1/n}} = K_6(T) \quad (6)$$

$$\text{As}_{\text{As}} \rightleftharpoons \frac{1}{n} \text{As}_{n_g} + V_{\text{As}} \quad ; \quad \frac{[\text{As}_{\text{As}}]_S}{P_{\text{As}_n}^{1/n} [V_{\text{As}}]_S} = K_7(T) \quad (7)$$

$$\text{Ga}_1 + \frac{1}{n} \text{As}_{n_g} \rightleftharpoons \text{GaAs}_s \quad ; \quad [\text{Ga}_1]_S P_{\text{As}_n}^{1/n} = K_8(T) \quad (8)$$

From (1) - (3) for the bulk case we get

$$[V_{\text{Ga}}]_B [V_{\text{As}}]_B = \frac{K_{1B}(T) K_{2B}(T)}{K_{3B}(T)} = K_9(T) \quad (9)$$

From (1) - (3) for the surface case and from (4) - (8) we get

$$[V_{\text{As}}]_S = \frac{K_{2S}(T)}{K_6(T)} P_{\text{As}_n}^{-1/n} = K_{10}(T) P_{\text{As}_n}^{-1/n} \quad (10)$$

$$[V_{\text{Ga}}]_S = \frac{K_{1S}(T)}{K_4(T) K_8(T)} P_{\text{As}_n}^{1/n} = K_{11}(T) P_{\text{As}_n}^{1/n} \quad (11)$$

$$[V_{\text{As}}]_S [V_{\text{Ga}}]_S = \frac{K_{1S}(T) K_{2S}(T)}{K_4(T) K_6(T) K_8(T)} = K_{12}(T) \quad (12)$$

$$[\text{As}_{\text{As}}]_S = K_7(T) K_{10}(T) = K_{13}(T) \quad (13)$$

$$[\text{Ga}_{\text{Ga}}]_S = K_5(T) K_8(T) K_{11}(T) = K_{14}(T) \quad (14)$$

The most interesting expressions are (9) - (12)

2.3.9.2 Resulting Concentration Profiles. Diffusion

Note that the vacancy concentration at the surface and infinitely far away from the surface are different, as one might expect. It is also reasonable to assume for 'normal' As-pressure that the vacancy concentration at the surface is much larger than in the bulk. Between the surface and bulk we have a transition region where the vacancy concentration falls off to the bulk concentration. This fall-off is smoothed by diffusion of vacancies from the surface. Note that the surface concentrations of vacancies are constant (for constant P_{As}) if the diffusion is slow enough so as not to drain the surface region and displace the equilibrium point. We thus expect a erfc-form of the vacancy concentration sufficiently close to the surface. This is observed by Chiang and Pearson [26] who also find the diffusion coefficients

$$D(V_{As}) = 7.9 \times 10^3 \exp(-4.0 \text{ eV/kT}) \text{ cm}^2 \text{ s}^{-1}$$

$$D(V_{Ga}) = 2.1 \times 10^{-3} \exp(-2.1 \text{ eV/kT}) \text{ cm}^2 \text{ s}^{-1}$$

In the bulk one expects the same concentration of V_{Ga} and V_{As} since the enthalpy change for the formation of neutral V_{Ga} and V_{As} is the same ([26] p. 2989). At the surface the concentration depend on P_{As} but one expects normally that $[V_{As}]_S > [V_{Ga}]_S$. Figure 23 is taken from [26] and shows the vacancy profiles close to the surface. A total picture of $[V_{As}]$, $[V_{Ga}]$ in accordance with figure 23 is shown in figure 24.

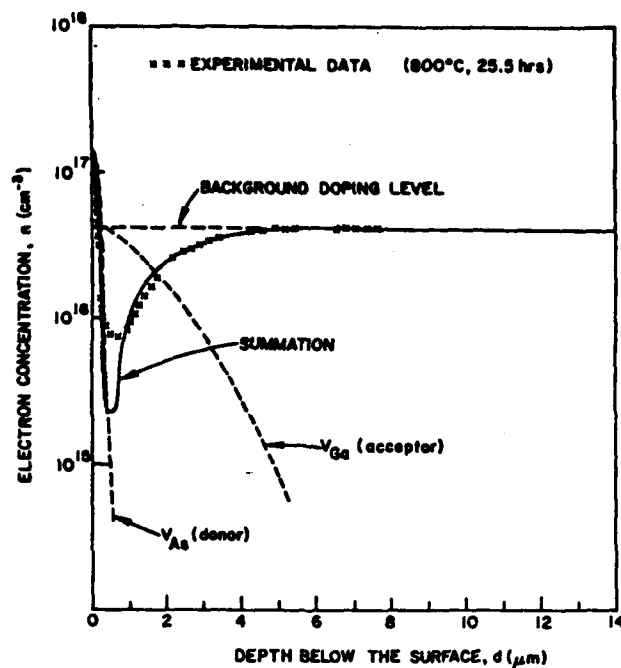


FIG. 1. Comparison of experimental results with theoretical calculations assuming complementary error function diffusion profiles for Ga and As vacancies. Sample was originally undoped, $n = 3 \times 10^{15} \text{ cm}^{-3}$.

Fig. 23. Near-surface vacancy distribution in GaAs taken from Chiang and Pearson [26]

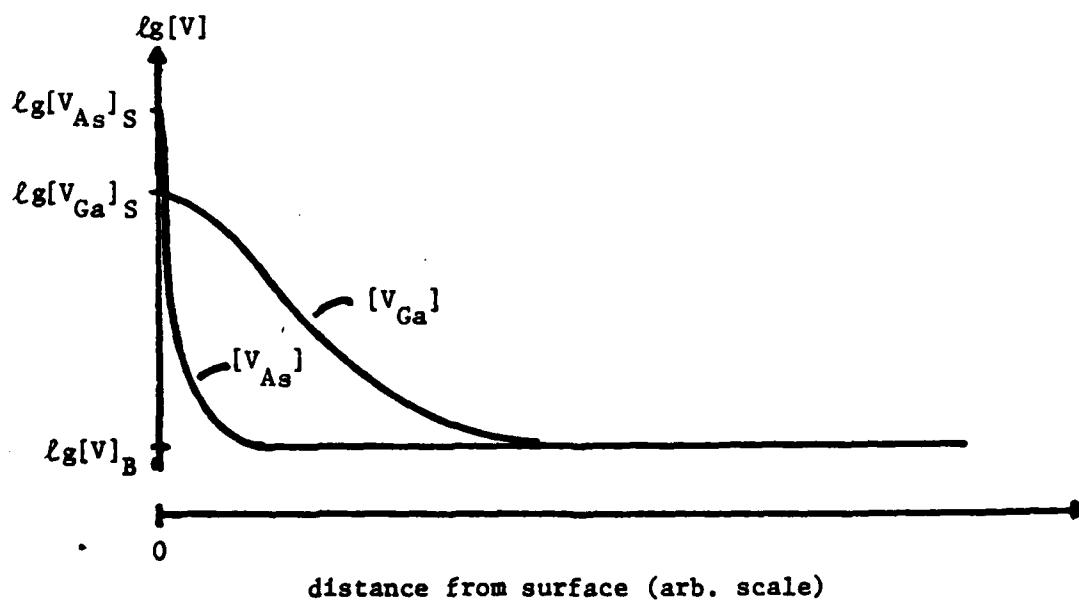


Fig. 24. Phenomenological overall picture of the vacancy distribution in GaAs

The impression one generally gets is that As is very volatile and that GaAs, when heated, loses much more As than Ga. We do not think this is true in general. The difference between As and Ga is that, outside the sample in the ampoule, As, due to its high vapor pressure, is the species that is noticed, while Ga is just present as a liquid film on the sample.

For long annealing when the bulk concentrations start to get much larger than the equilibrium value we expect to get dislocation climb which will eat vacancies.

2.3.9.3 Electric Activity of Vacancies

[26] argues that V_{As} is a donor and V_{Ga} an acceptor. This will fit well into our model. But there are other opinions. Dr. C.M. Wolfe thinks vacancies are not electrically active, and this could also be accommodated in our model to be presented below.

2.3.10. Super Fast Interstitial Diffusion

Cu is known to diffuse extremely fast in GaAs. It is believed that it diffuses interstitially. Weiser [27] developed a theory for interstitial diffusion in the diamond lattice. He found that one would expect some ions of intermediate size to diffuse interstitially, virtually without any potential barrier between equilibrium sites. He proposes that this may be the case for Cu in GaAs which would explain its diffusion behavior. It is not possible to make any direct comparisons with Cr which has an unfilled d-shell and has 16% smaller ionic radii than Cu, but it is conceivable that interstitial Cr may also diffuse rapidly in GaAs.

2.3.11. A Model for the Diffusion and Electronic Properties of Cr in GaAs

We are now ready to summarize what has been said so far in a few points that will form the basis of a model that seems capable of explaining many of the experimental results in section 2.3.2 in a unifying way.

1) Cr can, in addition to being a double deep acceptor on Ga-sites, also occur interstitially where it is a shallow donor.

2) The initial distribution of Cr between interstitial and substitutional sites is in general not an equilibrium one, due to the large barriers.

It depends on:

- a. under what circumstances Cr was incorporated, i.e. temperature, vacancy concentration, and doping
 - b. how the Cr was incorporated, i.e. whether it was during bulk growth, in diffusion, epitaxial growth, etc.
- 3) The redistribution of Cr depends on:
- a. the initial distribution of Cr between interstitial and substitutional sites, i.e. on the distribution dealt with in 2) and on the redistribution among sites. The latter depends on temperature, doping, and the local vacancy concentration.
 - b. the spatial distribution of V_{Ga} , i.e. on the initial distribution, net generation and diffusion. The generation depends on temperature and As-pressure at the surface.
 - c. the response to the driving forces which depends on temperature. The driving forces are concentration gradients, electric fields, strain fields.

AD-A098 578

WASHINGTON UNIV ST LOUIS MO DEPT OF ELECTRICAL ENGI--ETC F/G 20/12
SEMICONDUCTOR MILLIMETER WAVELENGTH ELECTRONICS. (U)

OCT 80

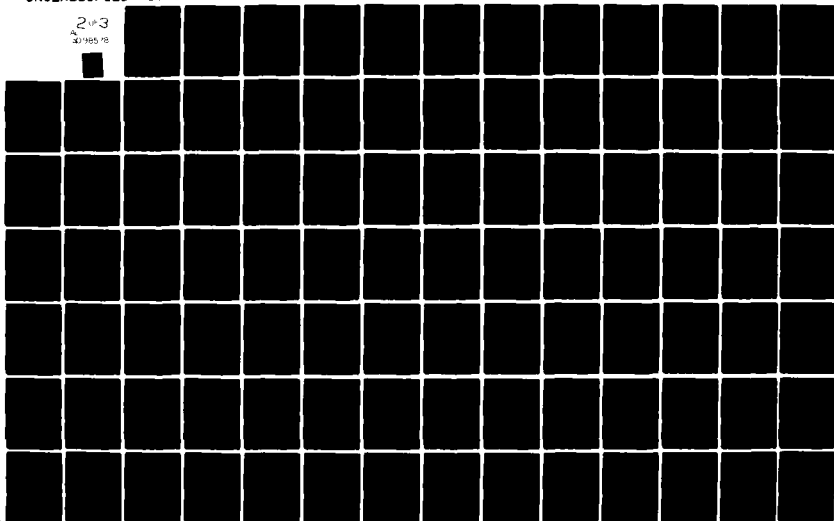
N00014-79-C-0840

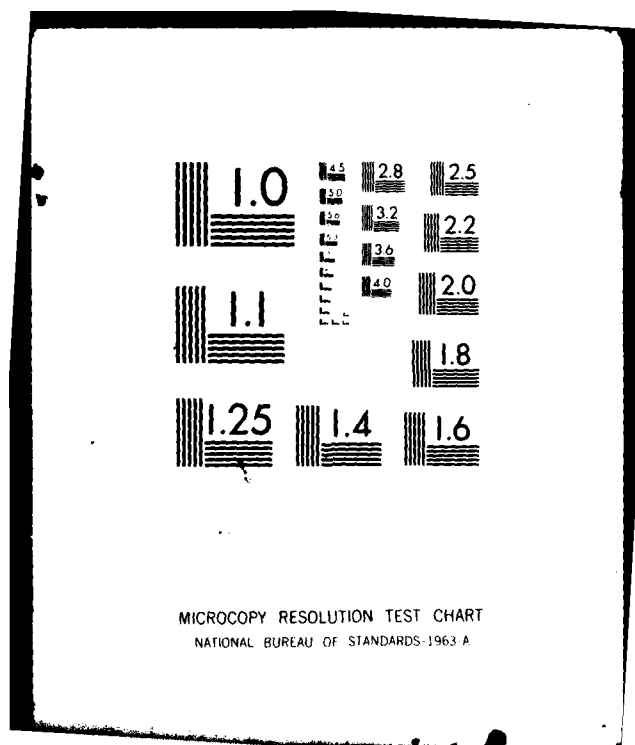
UNCLASSIFIED

80-1-0NR

NL

243
30 1985 78





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963 A

- 4) The interstitial Cr is very mobile. It moves interstitially, virtually without any potential barrier. The substitutional Cr moves slowly by a substitutional mechanism.

If Cr_i is a shallow donor we have approximately

$$[Cr_i] = [Cr_i^+] \quad (1)$$

For Cr_s we have to take all the charge states into account

$$[Cr_s] = [Cr_s^{3+}] + [Cr_s^{2+}] + [Cr_s^+] \quad (2)$$

$[Cr_s^{2+}]$ and $[Cr_s^+]$ are determined by the Fermi level E_F according to equation (3) and (4) in section 2.3.7

$$[Cr_s^{2+}] = [Cr_s] f_2(E_F) \quad (3)$$

$$[Cr_s^+] = [Cr_s] f_1(E_F) \quad (4)$$

The presence of different species which diffuse and react with each other necessitates a set of coupled differential equations to describe the redistribution. Suppose for simplicity that a one dimensional approach is appropriate.

$$\frac{\partial [Cr_i]}{\partial t} = D(Cr_i) \frac{\partial^2 [Cr_i]}{\partial x^2} - m(Cr_i) F(Cr_i) [Cr_i] - \frac{\partial [Cr_s]}{\partial t} \quad (5)$$

$$\frac{\partial [\text{Cr}_s^{n+}]}{\partial t} = D(\text{Cr}_s^{n+}) \frac{\partial^2 [\text{Cr}_s^{n+}]}{\partial x^2} - m(\text{Cr}_s^{n+}) F(\text{Cr}_s^{n+}) [\text{Cr}_s^{n+}] + r_n [\text{V}_{\text{Ga}}] [\text{Cr}_i] - p_n [\text{Cr}_s^{n+}], \quad n = 1, 2, 3 \quad (6)$$

$$\frac{\partial [\text{V}_{\text{Ga}}]}{\partial t} = D(\text{V}_{\text{Ga}}) \frac{\partial^2 [\text{V}_{\text{Ga}}]}{\partial x^2} - s([\text{V}_{\text{Ga}}] - [\text{V}_{\text{Ga}}]_0) - \frac{\partial [\text{Cr}_s]}{\partial t} \quad (7)$$

The D's are the diffusion coefficients, the m's are mobilities and the F's are forces acting on the different species. The terms

$$r_n [\text{V}_{\text{Ga}}] [\text{Cr}_i] - p_n [\text{Cr}_s^{n+}] \quad n = 1, 2, 3$$

model the net capture of interstitial Cr. The term

$$-s([\text{V}_{\text{Ga}}] - [\text{V}_{\text{Ga}}]_0)$$

models the net production of Ga-vacancies. $[\text{V}_{\text{Ga}}]_0$ is the equilibrium Ga-vacancy concentration. According to the stability discussion in section 2.3.6, p_n should be largest for $n = 3$. Depending on the Fermi level any of the r_n can be the largest.

2.3.12. Comments on the Experimental Results

We will now go through the experimental results that we reviewed in section 2.3.2. We will give our interpretation and when possible compare it with that of the authors.

2.3.12.1 Redistribution of Cr in GaAs

2.3.12.1.1 Annealing of Cr without encapsulation

Kasahara and Watanabe [2] observed that for $P_{As} > 0$ the Cr outdiffusion decreases and the surface concentration does not tend to zero. They explain this by a smaller thermal conversion at the surface, resulting in less surface conductance.

According to our model the substitutional Cr diffusion should be enhanced by a larger P_{As} since this results in a higher V_{Ga} concentration. We suggest that the Cr_i that is present diffuses rapidly toward the surface, where it partly leaves the sample and partly is captured by Ga vacancies. For larger P_{As} we will get a larger capture of Cr_i and less outdiffusion. Thus the surface concentration will not tend to zero.

[2] also report that for large temperatures and diffusion times they got more outdiffusion than expected, especially for $P_{As} = 0$. They propose that an increase in V_{Ga} -concentration should increase Cr diffusion. As mentioned we think this unlikely. Instead we propose that for $P_{As} = 0$ we have few V_{Ga} . There will be little capture of Cr_i which outdiffuses rapidly without accumulating at the surface.

The results suggest that P_{As} controls the surface condition but has negligible effect on the bulk. We think that, due to the effect of P_{As} on $[V_{Ga}]$ in the surface region, P_{As} will affect both Cr diffusion modes. It will affect the interstitial diffusion much further into the bulk, since this

is the fastest mode. But in this case $[Cr_1]$ is probably much smaller than $[Cr_g]$ which means that the bulk will be essentially unaffected, while the surface region will be affected in the way just described.

Tuck et al. [3] got much larger outdiffusion than [2]. We think that this is because the detectable Cr in this case is indiffused into the material, resulting in a much larger proportion of Cr_1 . The bulk concentration is lowered uniformly. This would be the case if Cr diffuses superfast, which is what we suggest about Cr_1 . The authors suggest that the migration occurs interstitially which is consistent with our picture.

The high surface peak that develops could be explained by the vacancy distributions that we discussed. In regions of high V_{Ga} concentration there would be a large capture of Cr_1 causing a build up of Cr. The authors too suggest indiffusion of Ga vacancies as an explanation.

For larger P_{As} they got a higher surface peak and less total outdiffusion. They explain this by the more stable conditions, due to higher P_{As} . We suggest a more specific explanation: At the surface there will be larger capture of Cr_1 for higher P_{As} . This leads to an effectively lower mobility there, and to less total out diffusion. In the bulk the mobility is still the same which leads to the observed constant bulk concentration. This is higher due to the lower effective surface mobility.

2.3.12.1.2 Annealing of GaAs:Cr with encapsulation

[4], [5] and [6] all observed a Cr depletion zone below the interface and a much thinner Cr pile-up close to the interface. The depletion region is considered to be due to normal outdiffusion of Cr while the thin pile-up is an anomaly proposed by [5] to be due either to interface contamination or non-equilibrium incorporation of Cr. We agree about the depletion region.

It is probably due to substitutional outdiffusion of Cr. The pile-up at the interface may also be due to forces such as strain from interface and electric fields due to interfacial states. Since the interface concentration according to [5] is higher than the solubility it is likely to be Cr_1 .

2.3.12.1.3 Diffusion of Cr into SI-GaAs:Cr

As mentioned in section 2.3.2, Tuck et al. [3] observed rapid fall-off in Cr concentration near the surface and exceedingly high penetration of Cr leading to uniform bulk concentration. The authors suggest that these features are due to indiffusion of Ga-vacancies and to interstitial diffusion respectively. This agrees with our vacancy picture and our suggestion that incorporation of Cr by indiffusion occurs interstitially, and is essentially what was said in section 2.3.12.1.1.

2.3.12.1.4 Epitaxial growth of GaAs on GaAs:Cr

Tuck et al. [3] get large outdiffusion and note that S-doping of the epilayer inhibits the outdiffusion. They explain the latter by (1) the presence of a substitutional impurity (S) on one sublattice is sufficient to affect the solubility of another (Cr) on the other sublattice and (2) the vacancy concentration is a function of the Fermi level. We explain the large outdiffusion as a result of the large amount of interstitial Cr in the substrate. In our picture n-doping enhances capture of Cr_1 since it induces the Cr to go into a compensating acceptor state. This lowers the effective diffusion constant. Wilson et al. [7] get less outdiffusion than [3]. Their substrates are prepared in the usual way to get SI-GaAs:Cr, which for not too large Cr content gives little interstitial Cr that can outdiffuse.

2.3.12.2 Electronic Properties of GaAs:Cr

2.3.12.2.1 Bulk grown GaAs:Cr

Cronin and Haisty [8] were first to report that Cr doping of GaAs gives SI material for sufficiently large Cr concentration, even for a large excess of Cr. They explained this by Cr being a deep acceptor and found unverified indications that the excess Cr precipitated in a second phase. In our picture the almost perfect compensation is due to autocompensation of Cr.

Zucca [9] reported that SI-GaAs:Cr can be either n- or p-type. In our picture, if the incorporation of Cr is not ideal, too little or too much Cr, can be incorporated, causing n- or p-type material, respectively. Zucca also observed that for excess Cr, N_a/N_d was still close to 1. He explained this by invoking a deep donor which he suggested was oxygen. He could not verify this, and in the most Cr-rich sample, which also was compensated, he could not detect enough oxygen to explain the observation. We instead invoke our autocompensation mechanism to account for the phenomenon.

Brozel et al. [10] observed that increasing the Si concentration during growth also increased the Cr concentration and that parts of the crystal which had high Si concentration also had high Cr concentration. This indicates that Cr also distributes geometrically so as to compensate shallow donors, as suggested by the coupled transport equations.

2.3.12.2.3 Cr indiffused into GaAs

Tuck et al. [3] got p-type ($p=6.4 \times 10^{14} \text{ cm}^{-3}$) material after indiffusion of Cr into SI-GaAs:Cr. This is surprising to us since we would have expected the indiffused Cr to be, mostly shallow interstitial donors. When they indiffused Cr into n-GaAs:Te the material became high resistive

(n or $p < 10^{14} \text{ cm}^{-3}$) in a $75 \mu\text{m}$ thick region below the surface, while beyond this the material remained n-type. This is less surprising. We expect the n-doping to enhance the capture of Cr_1 , especially in this case when the material was not pre-doped with Cr.

2.3.12.2.4 SI-GaAs:Cr annealed with encapsulation

Asbeck et al. [6] observed a spurious n-type layer just below the interface. This fits well with our picture. We have argued before from the point of view of solubility that the thin Cr peak at the interface probably is Cr_1 , in which case the material should be n-type here.

Favennec and L'Haridon [4] observed three kinds of substrates: qualified (A), those that showed n-conversion (B) and a few that showed p-conversion (C). The n-conversion was explained by outdiffusion of Cr, leaving a Cr depleted region. The p-conversion was more complicated, but could occur according to the authors if (1) four levels are present at the same time: a shallow acceptor, a shallow donor, a deep acceptor and a deep donor, and if (2) the deep donor diffuses towards the surface.

We think that the behavior can be explained in terms of the initial distribution of Cr among sites and the number of residual donors.

A. There is just enough Cr to compensate not too many residual donors. In absolute numbers then, little Cr will outdiffuse substitutionally. The material is compensated even after anneal. The substrate is qualified.

B. There is just enough Cr to compensate many residual donors. In absolute numbers, much Cr will outdiffuse substitutionally and leave unneutralized donors. The material will be n-type in a region below the interface. The substrate is unqualified (B).

C. There is more Cr than necessary to compensate. There will then be a large amount of Cr_i which will outdiffuse faster than Cr_s , leaving more Cr_s than necessary to compensate. The material is p-type in a rather thick region below the surface. The substrate is unqualified (C).

2.3.12.2.5 Epitaxial growth of GaAs on Si GaAs:Cr

Yamasaki et al. [12] observed an n^+ -layer just above the interface between the substrate and the n-type layer. We suggest that Cr in the form of Cr_i outdiffused into the epilayer.

Khokhlov et al. [13] observed higher electron concentration in the epitaxial layer close to the interface between the substrate and undoped layer. This is essentially the same observation.

Schlachetzki and Salow [14] grew undoped epitaxial layers on Cr doped substrates. The layers were n-type for growth temperatures T_g below some critical temperature T_o , p-type for $T_g > T_o$ and high resistive for $T_g \approx T_o$. T_o depended on the source material. They explained this by suggesting different incorporation of donors and acceptors in the source material at different T_g even though there was no intentional doping. We think that the observation can be explained by Cr that outdiffuses into the epilayer. At different temperatures and source materials the circumstances for incorporation are different. For instance, at high temperatures the vacancy concentration is higher, which enhances the capture of Cr_i . The Cr that outdiffuses is probably mostly Cr_i . For low temperatures these will probably remain interstitial, resulting in an n-type layer.

Khokhlov et al. [13] also reported that for higher T the electron concentration in their epilayers increased faster with T than the intrinsic

electron concentration. They explained this by an increased density of electron suppliers due to thermal rearrangements of defects. This fits well into our picture with the rearrangement being Cr_s to Cr_i .

2.3.13. Conclusion

Cr-doped semi-insulating GaAs is one of the most important substrate materials for a variety of devices, and it offers enough advantages so that it will undoubtedly continue to be used in this capacity in spite of its frequently problematic technology.

The semiquantitative model presented here searches for the causes of the sometimes erratic behavior of the material. It appears sufficiently promising to warrant further study.

We shall try to design experiments that can pin down the more tentative features of the model. Such an approach can take advantage of working with material that must be fabricated in any event in connection with the device work under this contract. The work, if it is successful, would contribute to bringing the technology under better control.

References

- [1] R.F. Broom, J. Appl. Phys., Vol. 38, No. 9, p. 3483 (1967).
- [2] J. Kasahara and N. Watanabe, Jpn. J. Appl. Phys., Vol. 19, No. 3, p. L151 (1980).
- [3] B. Tuck, G.A. Adegboyega, P.R. Jay and M.J. Cardwell, Inst. Phys. Conf. Ser. No. 45, Ch. 2, p. 114 (1979).
- [4] P.N. Favennec and H. L'Haridon, Appl. Phys. Lett., Vol. 35, No. 9, p. 699 (1979).
- [5] A.M. Huber, G. Morillot, N.T. Linh, P.N. Favennec, B. Deveaud and B. Toulouse, Appl. Phys. Lett., Vol. 34, No. 12, p. 858 (1979).
- [6] P.M. Asbeck, J. Tandon, B.M. Welch, C.A. Evans, Jr. and V.R. Deline, IEEE Electron Device Letters, Vol. EDL-1, No. 3, p. 35 (1980).
- [7] R.G. Wilson, P.K. Vasudev, D.M. Jamba, C.A. Evans, Jr. and V.R. Deline, Appl. Phys. Lett., Vol. 36, No. 3, p. 215 (1980).
- [8] G.R. Cronin and R.W. Haisty, J. Electrochem. Soc., Vol. 111, No. 7, p. 874 (1964).
- [9] R. Zucca, J. Appl. Phys., Vol. 48, No. 5, p. 1987 (1977).
- [10] M.R. Brozel, J. Butler, R.C. Newman, A. Ritson, D.J. Stirland and C. Whitehead, J. Phys. C: Solid St. Phys., Vol. 11, p. 1857 (1978).
- [11] Final Scientific Report No. 59356-6, Washington University, St. Louis (1979).
- [12] H. Yamasaki, J.G. Oakes, D.L. Barrett and T.M.S. Heng, Inst. Phys. Conf. Ser. No. 33b, p. 281 (1976).
- [13] V.I. Khokhlov, Yu. G. Sidorov and S.A. Dvoretiskii, Phys. Stat. Sol. (a), Vol. 25, p. 311 (1974).
- [14] A. Schlachetzki and H. Salow, Appl. Phys. 7, p. 195 (1975).

- [15] D. Shaw (editor), Atomic Diffusion in Semiconductors, Plenum Press, London and New York 1973.
- [16] J.J. Krebs and G.H. Stauss, Phys. Rev. B, Vol. 15, No. 1, p. 17 (1977).
- [17] J.J. Krebs and G.H. Stauss, Phys. Rev. B, Vol. 16, No. 3, p. 971 (1977).
- [18] A.M. White, J.J. Krebs and G.H. Stauss, J. Appl. Phys., Vol. 51, No. 1, p. 419 (1980).
- [19] U. Kaufmann and J. Schneider, Solid St. Commun., Vol. 20, p. 143 (1976).
- [20] H.J. Stocker and Martin Schmidt, Proceedings of the 13th International Conference on the Physics of Semiconductors, North-Holland, p. 611 (1976).
- [21] J.W. Allen, Nature, Vol. 187, p. 403 (1960).
- [22] B. Deveaud and P.H. Favennec, Solid St. Commun., Vol. 24, p. 473.
- [23] E. André and J.M. LeDuc, Mater. Res. Bull., Vol. 4, p. 149 (1969).
- [24] C. Kittel, Introduction to Solid State Physics, 5th ed., John Wiley & Sons, Inc. 1976.
- [25] K. Seeger, Semiconductor Physics, Springer-Verlag 1973.
- [26] S.Y. Chiang and G.L. Pearsen, J. Appl. Phys., Vol. 46, No. 7, p. 2986 (1975).
- [27] K. Weiser, Phys. Rev., Vol. 126, No. 4, p. 1427 (1962).

3. Existing Device Studies

3.1 Harmonic Generation in Transferred Electron Devices

3.1.1 Introduction

Short Gunn diodes are dominated by an accumulation layer instability that is launched from a region near the cathode. The charge layer grows in time and moves toward the anode where it is collected. As a consequence of the noninstantaneous transfer of low energy electrons from the central valley to the satellite valley, the electrons are not accurately characterized by the static velocity field curve for the material. The electron velocity increases above the static peak and after a characteristic time slows to its equilibrium value. Rees⁽¹⁾, using Monte Carlo simulations, examined the behavior of low $n \cdot L$ product devices at frequencies whose periods are comparable to the response time of the velocity-field curve. Assuming a deformation potential of 10^9 eV cm⁻¹ for intervalley scattering, a maximum operating frequency of 100 Ghz has been predicted for space charge free devices. In a similar Monte Carlo calculation by Jones and Rees⁽²⁾, the growth and decay of accumulation layers were simulated for a few special cases in GaAs.

The results of that simulation indicate that a non-zero time is required to establish the negative differential conductivity of the electrons and hence a delay exists in the nucleation of the space charge after the field has risen above threshold. This delay results in dead time which is not reduced by shortening the sample. In addition to the time delay of the charge nucleation, a dead zone between the cathode contact and the anode exists in which no charge growth occurs. This region results from the non-zero time that is required to bring a low energy injected electron above the valley energy separation and the time

required to transfer to the satellite valley. Simulations show that this region may be as much as $3 \mu\text{m}$ for average electric fields of 6 kv cm^{-1} . The simulations also indicate that this region is significantly reduced for large electric fields in the cathode region. If these simulations are correct, one would anticipate a significant change in the maximum operating frequency if electrons of high energy are injected into the region or if the electric field in the cathode region is increased above its usual value.

The decay of the satellite valley electrons from the high to low field region has been observed in simulations by Curtice and Purcell (1970) and Jones⁽²⁾ to invariably fall rapidly. This fall in some cases can be as short as .7 ps. The primary reason for this fast fall is predicted from the cooling of the distribution that occurs when electrons are scattered in a direction against the field. These electrons are rapidly brought to low energies by the electric field. This rapid decay of the accumulation layer will lead to a large harmonic content in the output current. It is therefore interesting to observe the harmonic content and use this information to estimate the rate at which the accumulation layer is quenched. This rate is likely to be indicative of the scattering rate from the satellite valley and the cooling rate of the field. Furthermore one should be able to obtain higher power operation in this mode than the fundamental at frequencies for which the accumulation layer formation time is a significant portion of an rf cycle.

To obtain an idea of the merit of this technique of generation of millimeter wave power, several devices have been evaluated in circuits designed to permit harmonic operation of the devices. The circuits, devices and results are reported in the following sections.

3.1.2 Harmonic Oscillator Circuits

3.1.2.1 Coaxial circuit

The basic elements of the coaxial circuit are shown in Figure 1. The fundamental resonator is formed by the TEM line section between the packaged Gunn diode and the low pass filter. A thin dielectric rod is placed near the diode to permit slight tuning of the resonant frequency. The harmonic resonator is formed by a waveguide line section behind the resonator post. The line section is cutoff at the fundamental and hence does not significantly affect the frequency of operation. Several dielectric rods are inserted into the line section to adjust its resonant frequency and the coupling to the main resonator. In actual operation very small changes (~ 200 MHz) of the fundamental frequency are observed through variations of the harmonic resonator. The output of the harmonic power is coupled to the load via the front iris which opens to a waveguide that is cutoff at the fundamental. This waveguide section is removeable to allow monitoring of the fundamental power and frequency. Initial measurements are made on diodes using the coaxial circuit without the fundamental filter.

The circuit is adjusted in frequency by setting the length of the coaxial line section and its diameter. At lower frequencies where the package parasitics are small, the line section is nearly half wavelength. As the frequency is increased the package parasitics force the resonator length to drop faster than $1/f$ and at 60 Ghz the resonator is nearly quarter-wavelength indicating a parallel resonance of the unmodified package near 60 Ghz. The frequency of operation can be further increased up to 70 Ghz by additional shortening of the resonator.

To permit operation above 70 Ghz modifications to the package were found necessary. Using a diamond file the 0.030" diameter ceramic was

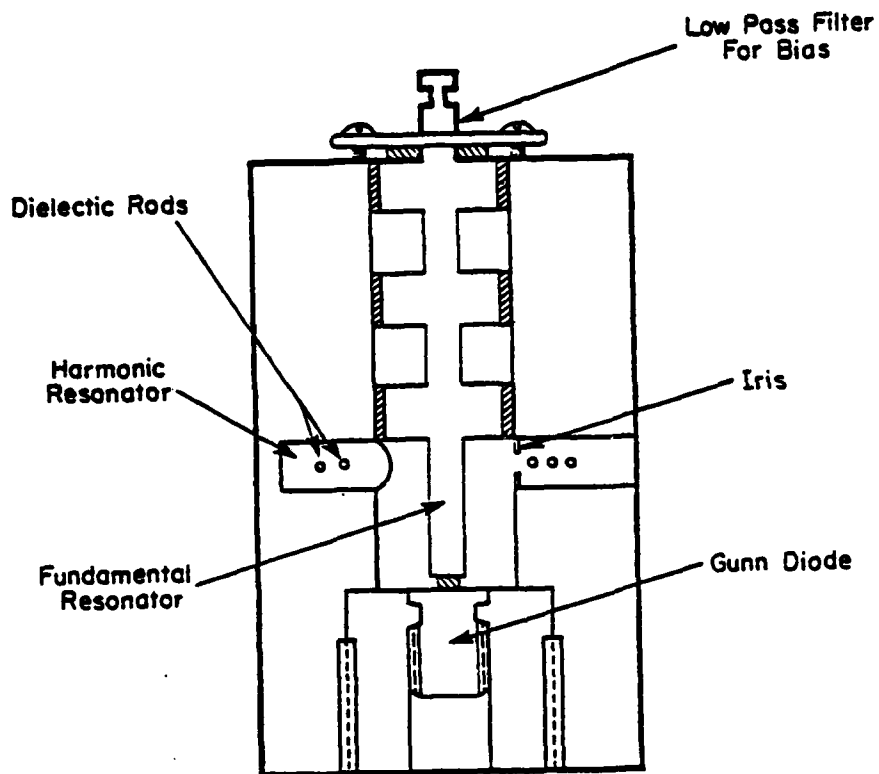


Figure 1

Cross sectional view of coaxial circuit

reduced to 0.023" with a corresponding increase in operating frequency to 78 Ghz. Further reductions in the package wall thickness showed little change in operating frequency and were often fatal to the mechanical integrity of the package. Additional wires and ribbons connecting the chip and the top of the ceramic package provided only slight changes in the operating frequency. For this reason designs are now under way to use unpackaged diodes in this circuit.

The harmonic power is nearly zero until the rear resonator is adjusted to the harmonic frequency. The iris width is then increased until the power is maximized with respect to iris size. Further matching is then done using the 0.020" diameter quartz rods. Tuning is iterated to obtain the maximum harmonic power. The power is measured using a series of high pass filters. In most cases the only significant power is at the harmonic at which the unit has been peaked.

3.1.2.2 Disc circuit

The disc or hat circuit is constructed using a radial resonator at the fundamental frequency. The diode is mounted in the broad wall of a waveguide that is cutoff at the fundamental frequency and propagating at the desired harmonic. The location of the diode is chosen to form a $\lambda_g/2$ resonator between the back wall and the diode post. The fundamental radial disc resonator is fabricated at the end of a line section on the low pass feedthrough and the length between the resonator and the low pass filter is chosen to be slightly less than $\lambda_o/4$. Dielectric rods are inserted in the side walls of the waveguide and used to tune the coaxial section between the feedthrough and the disc. The harmonic coupling is optimized by choosing the thickness of the disc and the impedance of the coaxial section that joins the feedthrough and the disc. Additional

dielectric rods are inserted in the harmonic resonator region and in the reduced height waveguide section and are used to couple the circuit to the load. The circuit described here is the one used by most of the labs reporting 94 Ghz Gunn sources.

Several prominent modes are found to be present in this circuit which often lead to the existence of two fundamental signals. Under these conditions significant power has been observed at the sum of these two frequencies. This output can often be as large as the harmonic and in two cases it can be optimized more easily than the harmonic case.

The maximum fundamental frequency of the disc resonator is determined by the diameter of the disc, its thickness, and the diode package. The disc diameter for 50 Ghz is ~ 0.053 " for a disc thickness of 0.015". The frequency increases as the disc is reduced in diameter. To maintain a constant coupling the diameter of the coaxial section must also be reduced. In addition, the dielectric loading of the package above 60 Ghz significantly reduces the required disc diameter and it is therefore necessary to use a different circuit design unless the package is reduced in volume. Using a package with a wall thickness of 0.005" instead of 0.010" has resulted in a frequency of operation in excess of 70 Ghz.

3.1.2.3 Waveguide circuit

The waveguide circuit consists of a reduced height TE_{101} cavity at the fundamental frequency. The cavity is significantly foreshortened to compensate for the diode package volume and for the dielectric rods inserted in the narrow wall which tune the harmonic resonance. The diode is located near the back wall of the cavity with the best location being centered in the broad wall with the diode's dielectric touching the back

of the cavity. A low pass filter centered at the harmonic is used to provide dc bias. The post diameter is chosen to be the same diameter as the diode in order to maximize the post resonant frequency and minimize fringing capacitance in the diode region. The iris width is chosen to be $\lambda_0/2$ at the desired harmonic. A section of waveguide which is cutoff at the fundamental is used to separate the fundamental from the load. The waveguide section is attached with several screws and can be easily removed to determine the fundamental power and frequency.

The dielectric rods in the side wall lower the higher order modes (TE_{mon}) faster than the fundamental and it is therefore possible to resonate both the fundamental and the harmonic with some rod manipulation. The circuit losses appear to be somewhat lower than the disc or coaxial circuits and the output powers in the region of 80-100 Ghz (fundamental 40-50 Ghz) appear to be 1-2 db higher than with the alternate circuits. The complex mode possibilities allow wide changes in the harmonic coupling circuit and provide an interesting higher Q source. Several tries have been made to scale this circuit up to the 75 Ghz region have lead to the conclusion that the package parasitics are too large to permit fundamental operation above 60 Ghz. Designs are now in progress which eliminate most of the package in this circuit.

3.1.3 Experimental Results with Flat Doped Diodes

The power versus frequency performance of constant doped GaAs Gunn diodes has been examined in the coaxial, disc and waveguide circuits. The maximum observed output powers at the fundamental, second, and in some cases, third and fourth harmonics have been measured and their bias voltage dependences have been noted. An extremely sharp drop in output power has been observed for frequencies above 100 Ghz. The nature

of the drop is in part circuit connected but it is likely that much of the performance degradation is v-E response time related.

The devices used in this section are fabricated from multilayer epitaxial material having both buffer layers and epitaxial cathode contacts. The carrier concentration of the cathode contact layer and the buffer layer is $3-4 \times 10^{17} \text{ cm}^{-3}$. The doping profiles are nearly constant in the active region and in all cases decrease slightly toward the cathode (surface). The layers are all grown on 2° off (100) toward (110) silicon doped $1-3 \times 10^{18} \text{ cm}^{-3}$ substrate material. The substrate thickness is kept less than $6.5 \times 10^{-3} \text{ cm}$. The units are of the flip chip mesa construction. All units tested are mounted in a threaded stud ceramic package having a ceramic height of 0.015" typical and a starting wall thickness of 0.010". The inside diameter of the ceramic is 0.013" and the pedestal extends into the ceramic 0.005". The top leads of the diode (anode leads) are 0.002" diameter fully annealed gold wire that is thermal compression bonded in an "X" configuration. For most of this work a lid was not placed over the metalized ceramic in an attempt to minimize extra inductances.

Diodes were fabricated on epitaxial material with active layer thicknesses in the range $1.5 \text{ } \mu\text{m}$ to $3.5 \text{ } \mu\text{m}$. Active layers with thicknesses less than $1.5 \text{ } \mu\text{m}$ showed no signs of oscillations either at dc or at millimeter waves. Doping densities as high as $5 \times 10^{15} \text{ cm}^{-3}$ were tried in an attempt to obtain a high N-L product. However even these high doped samples did not show Gunn effect. Table 3.1.3A lists wafers which were found to yield microwave oscillations. Those which yielded the highest fundamental powers are listed in Table 3.1.3.B. The final diode current was adjusted using in package etching and in most cases the diode current was between 700 and 800 mA.

The output power recorded in Table B are the best values obtained from devices made from the wafer and operated in the coaxial configuration. This corresponds to the highest fundamental power measured. Several trends should be noted from Table B. The highest fundamental powers are all obtained at the lowest frequency for which the diode was measured. In all cases the power declined as the frequency was increased. This would tend to indicate that the effective lengths of the samples are all too long for the highest frequency of operation. This however has not been borne out by thinner samples. There is likely a minimum length for which no space charge growth occurs.

Several samples were operated in a pulsed mode at the low duty. A 300 nsec pulse was used and the operating voltage for maximum power output under pulsed operation at 70 and 75 Ghz did not change significantly from the cw values of the best diodes. It is therefore likely that the bias voltage decline is not significantly a thermal problem but rather a result of maximizing the space charge growth.

3.1.4 Harmonic Power of Flat Profile Wafers

The harmonic power of the evaluation diodes has been measured in the coaxial, the disc, and the waveguide circuits. The power outputs at the harmonics have been measured at frequencies up to 148 Ghz. The output powers for all of the flat doped diodes decline very sharply above 100 Ghz and the output powers at 148 Ghz were only a few tenths of a milliwatt. The decline in output power is so steep that it is difficult to attribute the decrease to a single cause.

The maximum power output at the second harmonic of wafer 3C33-1 is displayed in Figure 2. This wafer showed large low frequency output power at 35 Ghz. The output power at the second harmonic was in excess of 25%

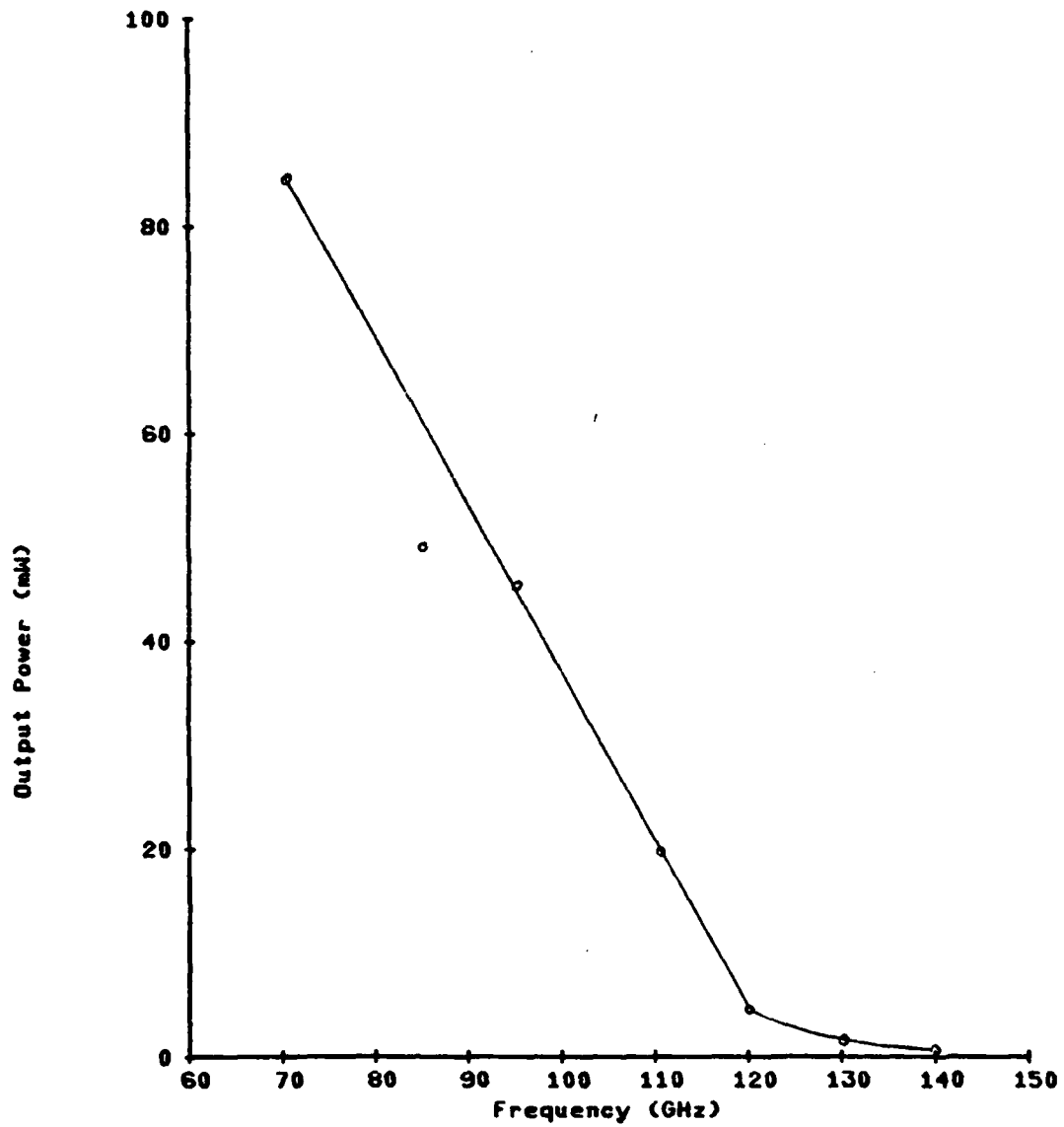


Figure 2

Second harmonic output of Gunn diode fabricated from wafer 3C33-1

Table 3.1.3.A Properties of Wafers Exhibiting
Millimeter Wave Oscillations

Wafer	Nd (cm^{-3}) (cathode side)	Active Length (10^{-4} cm)	Drop Back (I_p/I_v) (pulsed)
3C33-1	7×10^{15}	3.0-3.5	1.6
3C32-1	7.8×10^{15}	3.0	1.4
3C32-2	7.5×10^{15}	2.7	1.4
3D6-1	1.5×10^{16}	2.3	1.6
3D8-1	2.5×10^{16}	2.0-2.2	1.5
3D10-1	2.5×10^{16}	1.8-2.0	1.4
3D11	3.5×10^{16}	1.8-2.0	1.3
3E11	2.0×10^{16}	1.8-2.0	1.4
3E12	4×10^{16}	1.8	1.3
3D13-1	4×10^{16}	1.5	1 (No OSC)
3E13-2	4×10^{16}	1.5	1 (No OSC)
JAl-5-1	2×10^{16}	2.3	1.5
JAl-5-2	2×10^{16}	1.5	1 (No OSC)

Table 3.3.1.B Conditions for Maximum Power Generated
at the Fundamental Frequency

Wafer	Fundamental Frequency (GHz)	Power (mW)	V _{Bias} (volts)
3C33-1	35	320	6.1
	50	105	4.3
	60	30	3.0
	70	.1	2.3
3D8-1	45	200	6.5
	50	160	5.0
	60	110	4.7
	70	70	4.3
	75	67	3.9
3D10-1	50	170	5.5
	60	103	5.3
	70	57	4.6
	75	50	4.0
JA1-5-2	40	180	5.7
	50	165	4.9
	60	100	4.6
	70	70	3.9
	75	60	3.7

of the output power obtainable at 35 Ghz. At frequencies up to 94 Ghz the highest powers were obtained from the waveguide harmonic circuit. At frequencies above 94 Ghz the disc resonator and coaxial circuits yielded nearly equal powers. The sharp decline in the harmonic power might be expected from the fundamental power roll-off observed.

The third harmonic power was obtained in similar circuits. The output power at the third harmonic for wafer 3C33-1 is shown in Figure 3. The surprising observation is the extremely severe decline in output above the fundamental frequency of 35 Ghz. This would not have been anticipated from the fundamental measurements.

The output power at the second harmonic has been graphed in Figure 4 for devices from wafer JAl-5-2. These diodes yielded up to 45 mW at 90 Ghz however the output powers declined very rapidly above 100 Ghz. The output power at 148 Ghz is 0.3 mW. Several other wafers were evaluated and in each case the output power was found to decline sharply above 100 Ghz for any harmonic.

Three probable causes for the rapid decline in output power are the package parasitics, the substrate skin depth and the v-e response time. Further work will be needed to sort the three causes. The new circuits will allow unpackaged plated heatsink devices to be used with much lower parasitic capacitances and inductances. A technique to deposit a thick layer of gold over most of the exposed substrate is being developed. In addition further work with the circuits may allow a wider range of matching.

3.1.5 Harmonic Generation In Cathode Notch Diodes and Future Plans

A significant portion of the flat doped diode in the region near the cathode remains below threshold in short samples. Jones⁽²⁾ has

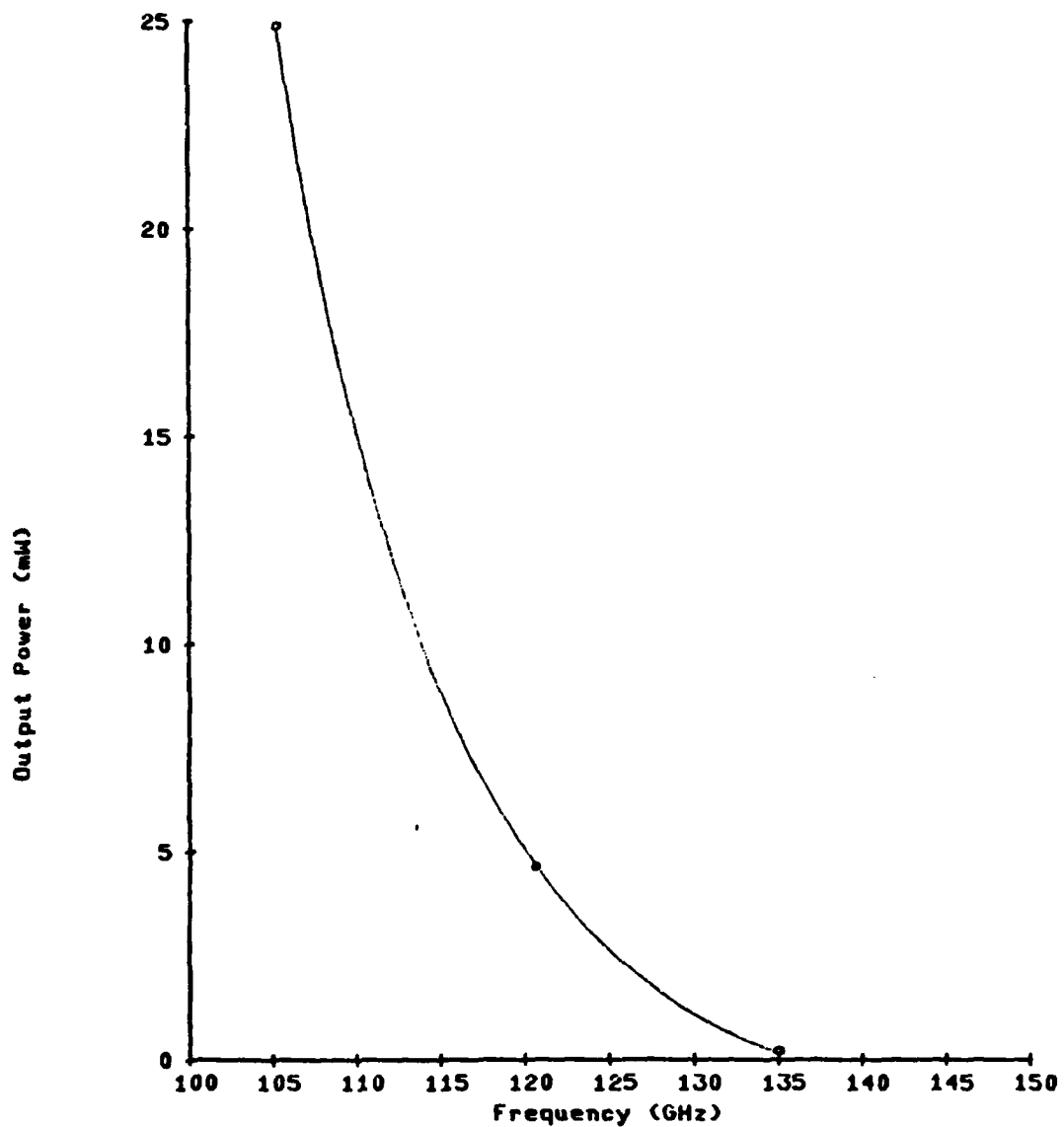


Figure 3

Third harmonic output of Gunn diode fabricated from wafer 3C33-1

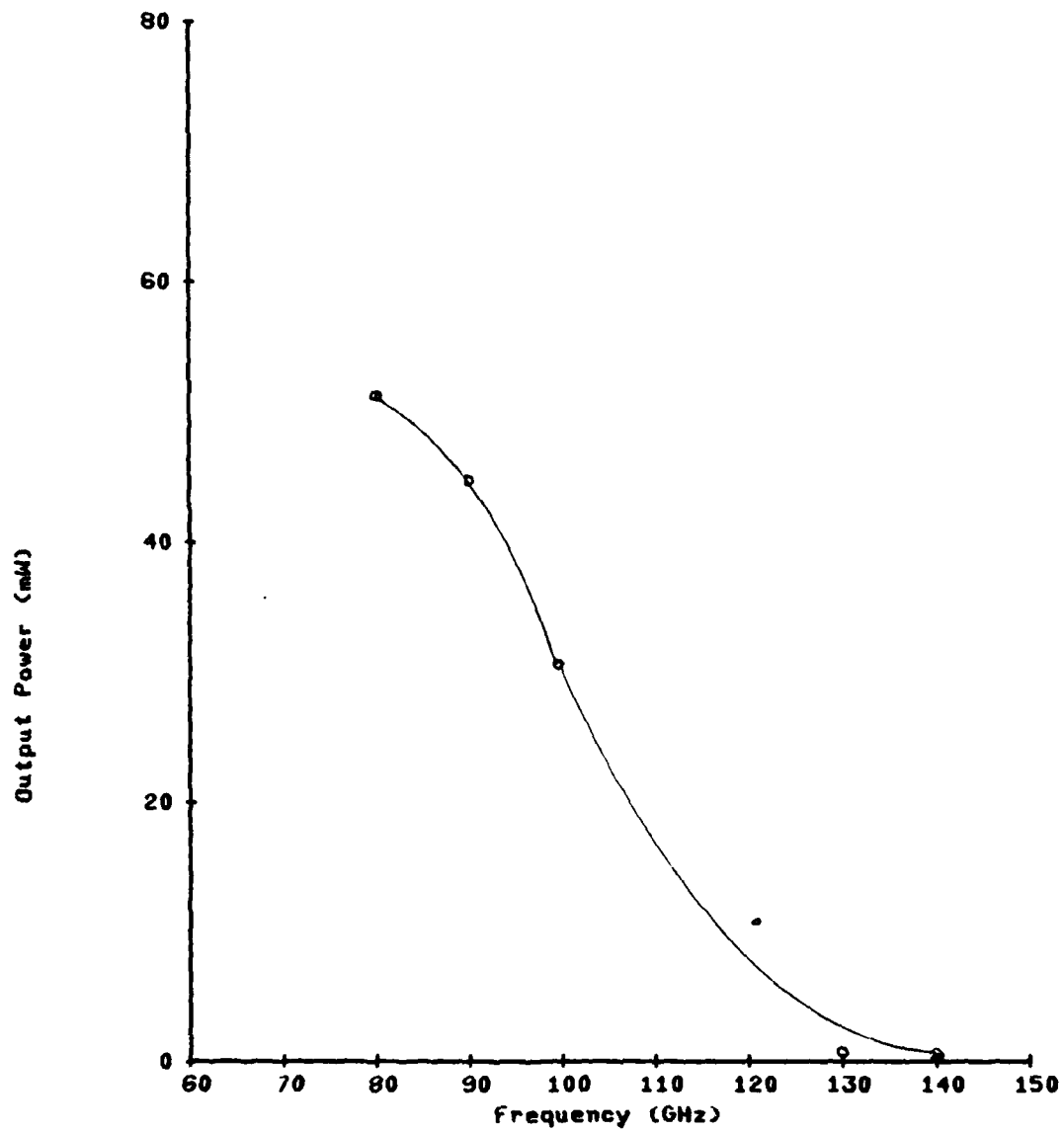


Figure 4

Second harmonic output of Gunn diode fabricated from wafer JA1-5-2

simulated accumulation layers and his study indicates that the layer formation time is extremely sensitive to the cathode region. In addition once the layer begins to form its growth rate is highly sensitive to small changes in the electric field. Furthermore, the likely velocity overshoot is only large if the electric field in the region of accumulation layer formation is large. It is therefore interesting to examine high frequency harmonic generation of diodes that have a thin low doped region next to the cathode contact. The effect of this doping notch is to increase the electric field in the cathode region to obtain a more rapid growth of the accumulation layer and obtain an overall reduction in the response time.

Two wafers were grown that have a low doped region at the cathode contact. One of them has a 0.3-0.4 μm region between the cathode contact and the active layer with a doping density as measured by C-V of $2-2.5 \times 10^{15} \text{ cm}^{-3}$. The other wafer employs a so called "metallic" contact without the aid of a heavily doped region for the cathode contact. Diodes were fabricated out of both wafers and the wafer employing the "metallic" contact showed significant power at 50 Ghz. These diodes have been measured in the fundamental and harmonic circuits. The active layer excluding the lightly doped region is 2.2 μm thick and the active region carrier concentration is $1.8 \times 10^{16} \text{ cm}^{-3}$.

The fundamental output power is shown in Figure 5. The output at 40 Ghz is 150 mW and the bias voltage is 6.8 volts. The power declines slowly to 45 mW at 74 Ghz in the coaxial circuit. Large changes in the operating current have been observed with changes in temperature. The harmonic powers of these devices show significant variations between units. However the better devices, as can be seen in Figure 6, produce

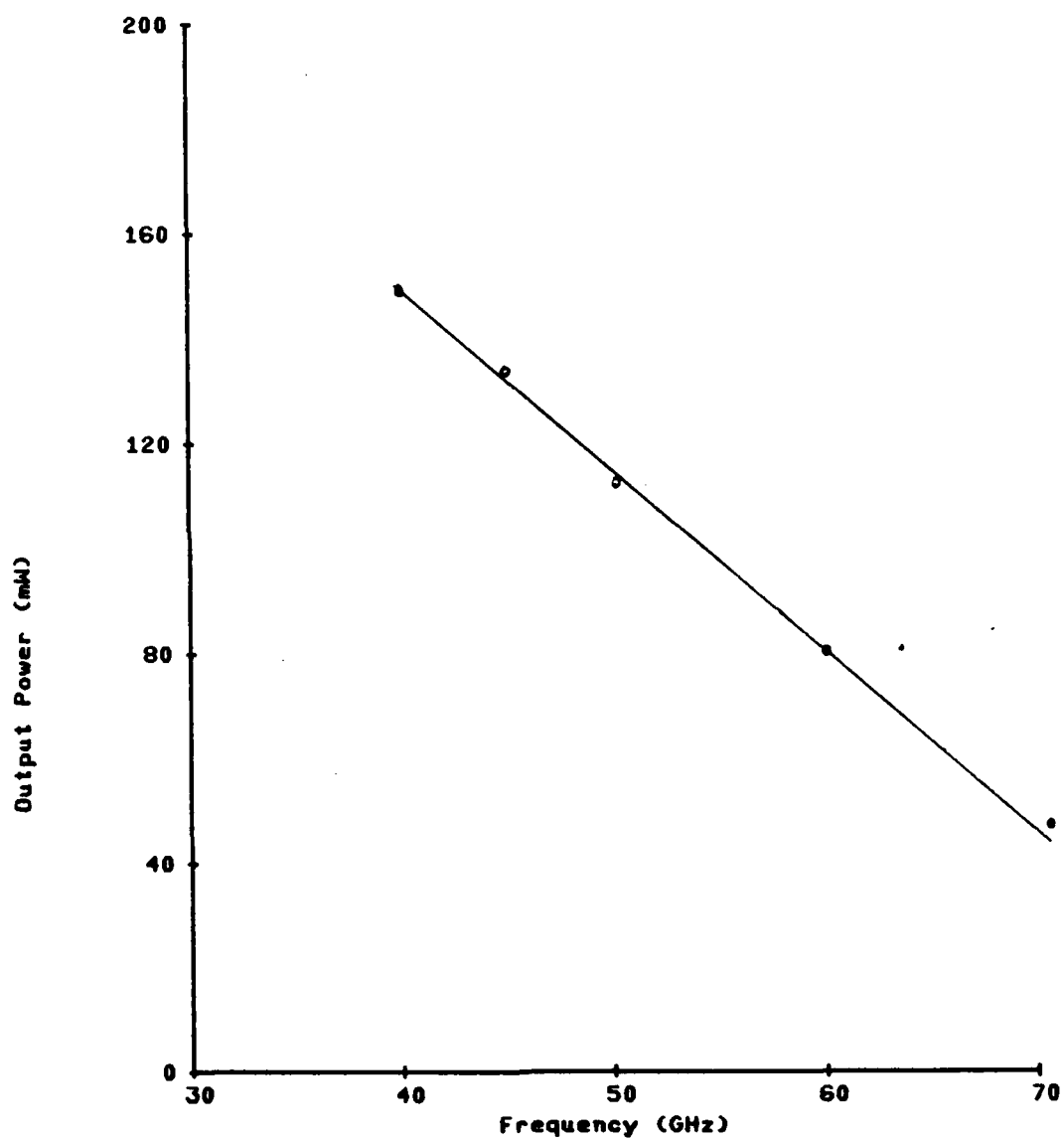


Figure 5

Fundamental output of cathode notch diodes

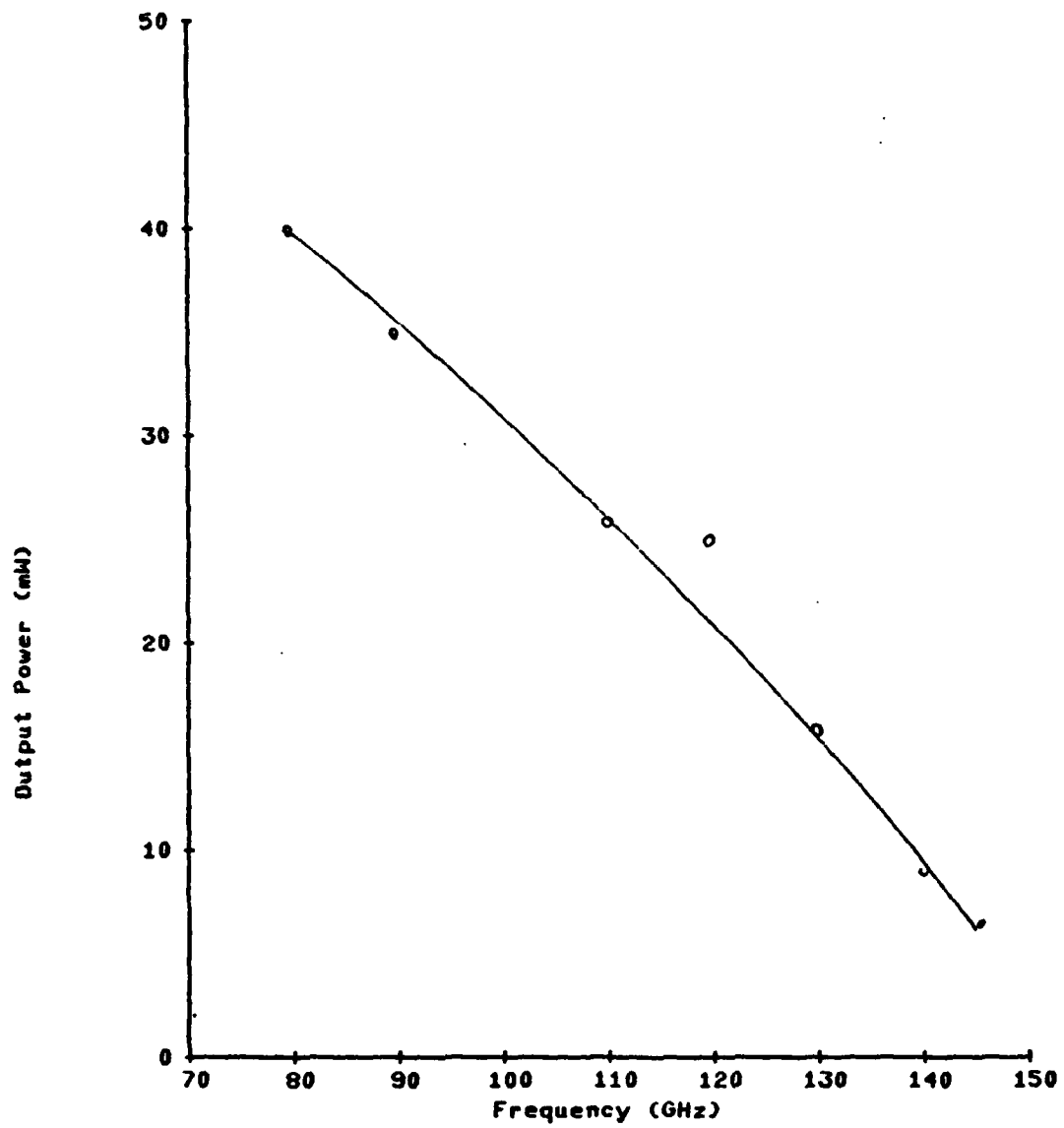


Figure 6

Second harmonic output of cathode notch diodes

substantial power at frequencies above 100 Ghz. These devices have yielded up to 5-6 mW at 144 Ghz. A slightly higher device capacitance prevents the packaged devices from operation above 144 Ghz and measured values of the third harmonic appears low. This may, in fact, be a circuit limitation enhanced by the package parasitics.

The cathode appears to be a possible area for improvements in generation above 100 Ghz. In the next period hot electron injection techniques will be evaluated. The injection can be done with either a heterojunction or a barrier- n^+ region. The techniques of harmonic evaluation will be applied to In P diodes late in the period.

References

1. H.D. Rees, "Time Responses of the High-field Electron Distribution Function in GaAs", IBM J. Res. Develop., pp. 537-542, Sept. (1969).
2. D. Jones and H.D. Rees, "A reappraisal of instabilities due to transferred electron effect", J. Phys. C. Solid State Physics, Vol. 6 (1973).

3.2

BULK AND CONTACT PHENOMENA IN
MILLIMETER WAVE DIODES

P. Chen, F.J. Rosenbaum, R.E. Goldwasser

3.2.1 Introduction

In the study of semiconductor devices for use at millimeter wavelengths it is important to know the behavior of carriers in the bulk material and at the device contacts. Since device dimensions are necessarily small, electric fields sufficiently large to produce avalanche breakdown may be encountered, even for modest applied voltages. Furthermore, many existing devices such as mixer diodes and Impatts employ depleted regions where high electric fields are present. In order to better understand the operation and limitations of mm-wave devices we have been studying the current-voltage relations for diodes in the presence of avalanche breakdown, and for diodes with nearly constant reverse depletion capacitance (Mott barrier or punch-through barrier junctions).^{1,2} One and two-dimensional finite element analyses of the carrier concentrations and field distributions have been developed in order to examine the details of the charge depletion approximation commonly made in the analysis of depleted regions. We have included the effects of field dependent mobility and diffusion in GaAs. We have also conceived of and are analyzing a new device geometry which may prove useful for sub-harmonically pumped mixers at millimeter wavelengths.³⁻⁵ A simulation for the new device provides a means of quickly and economically evaluating its performance.

3.2.2 Avalanche Breakdown Analysis^{1,6}

Assume a one dimensional Schottky Barrier diode as shown in Figure 1. The bulk region is doped n-type, at a level N_d . A reverse-bias voltage V is applied to the device between the S.B. and ohmic contacts. For simplicity, assume the doping and the space charge in the depletion region are uniform.

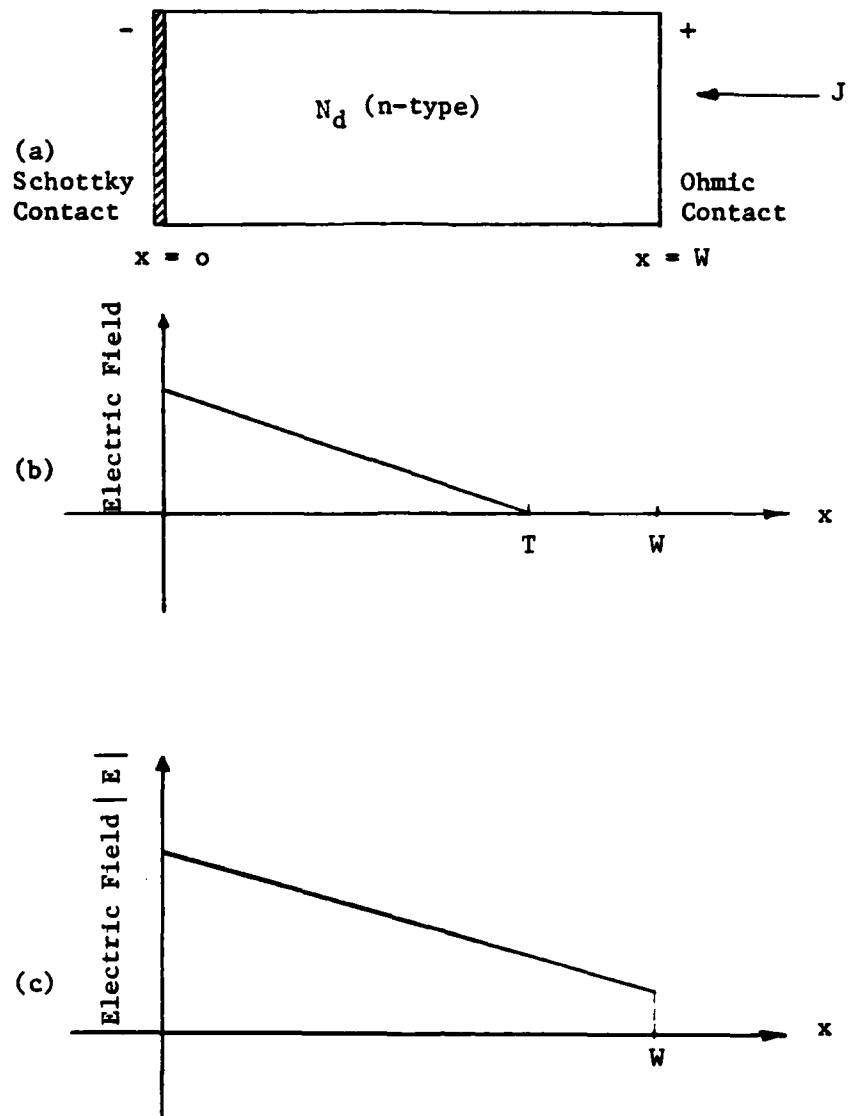


Figure 1 (a) One-dimensional Schottky Barrier diode.

(b) field distribution when the depletion length $T < W$; W is the diode length.

(c) $T > W$.

The electric field distribution in the depletion region, $T < W$, is given by

$$E(x) = \frac{q}{\epsilon} N_d x - \frac{2V}{T} \quad (1)$$

where $T = \frac{2V}{qN_d}$.

If the depletion region extends across the device, $T \geq W$, then the field is given by

$$E(x) = \frac{q}{\epsilon} N_d x - \frac{V + V_p}{W} \quad (2)$$

where $V_p = \frac{q N_d W^2}{2\epsilon}$

is punch-through voltage. For convenience, we use positive values for the avalanche breakdown field and the voltage.

3.2.2.1 Avalanche Breakdown

If sufficiently large reverse bias voltage is applied, avalanche breakdown will occur. The electron and hole currents can be determined from the continuity equations. In the dc case, assume

$$\frac{d}{dx} J_n(x) = \alpha(E) J_n(x) + \beta(E) J_p(x) \quad (3)$$

$$\frac{d}{dx} J_p(x) = \alpha(E) J_n(x) - \beta(E) J_p(x) \quad (4)$$

where $E = E(x)$, has been defined before and α , β are the electron and hole ionization coefficients, respectively. The total current density, J , is a constant.

$$J = J_n(x) + J_p(x). \quad (5)$$

Solving the two differential equations above, we obtain

$$J = \frac{J_n(o) + J_p(W) \exp \left[-\int_0^W (\alpha - \beta) dx \right]}{1 - \int_0^W \alpha \exp \left[-\int_0^x (\alpha - \beta) dx' \right] dx} \quad (6)$$

and the avalanche breakdown voltage can be found from the denominator when

$$1 - \int_0^W \alpha \exp \left[-\int_0^x (\alpha - \beta) dx' \right] dx = 0 \quad (7)$$

Stillman and Wolfe⁶ suggest that α and β have the following form:

$$\begin{aligned} \alpha &= 2 \times 10^6 \exp [-2 \times 10^6/E] \\ \beta &= 10^5 \exp [-5 \times 10^5/E] \end{aligned} \quad (8)$$

where E is the magnitude of the field and is a function of x .

3.2.2.2 I-V Characteristic Relations

In n-type material the electron current density is large, so $J_n(o) \gg J_p(w)$, $\int_0^W (\alpha - \beta) dx < 1$, and the total current density becomes

$$J \approx \frac{J_n(o)}{1 - \int_0^W \alpha \exp \left[-\int_0^x (\alpha - \beta) dx' \right] dx} \quad (9)$$

where

$$J_n(o) = J_{sat} \approx A^* T^2 \exp \left(-\frac{q\phi_{Bn}}{kT} \right). \quad (10)$$

We may define an avalanche multiplication factor, M , such that

$$M = \frac{1}{1 - \int_0^W \alpha \exp \left[-\int_0^x (\alpha - \beta) dx' \right] dx} \quad (11)$$

and so

$$J \approx MJ(o) \quad (12)$$

The behavior of M can be examined simply by taking $\alpha = \beta$.

Then

$$M = \frac{1}{1 - \int_0^W \alpha dx} \quad (13)$$

Now, $\int_0^W \alpha dx$ is an exponential integral, and using polynomial and rational approximations, one can calculate the value of M corresponding to the applied voltage V . For the case of $\alpha \neq \beta$, polynomial and rational approximations can be used inside the denominator integral and the value of M found numerically. Some representative multiplication voltage curves are shown in Figure 2.

The doping density N_d affects the breakdown voltage as shown in Figure 3. Also plotted there are several experimental results taken from the literature.

After some manipulation it is found that the M curves can be approximately represented by the following expression⁷:

$$M = \frac{1}{[1 - (\frac{V}{V_{BD}})^n]^m} \quad (14)$$

When N_d is between 10^{16} and 10^{18} cm^{-3} , the n, m values which fit the data are given in Table 1.

3.2.2.3 Large Signal Analysis

The preceeding section dealt with the dc breakdown characteristics of GaAs. Breakdown can also occur if a sufficiently large time-varying reverse bias voltage is applied to the diode. To analyze the time dependent case, assume a one-dimensional structure with equal electron and hole ionization coefficients, $\alpha = \beta$. As shown in Figure 4, $0 \leq x \leq x_a$ is the avalanche region. The current in this region satisfies the

Figure 2 Two representative current multiplication curves
respect to applied reverse bias voltages.
(α, β see Equation (8)).

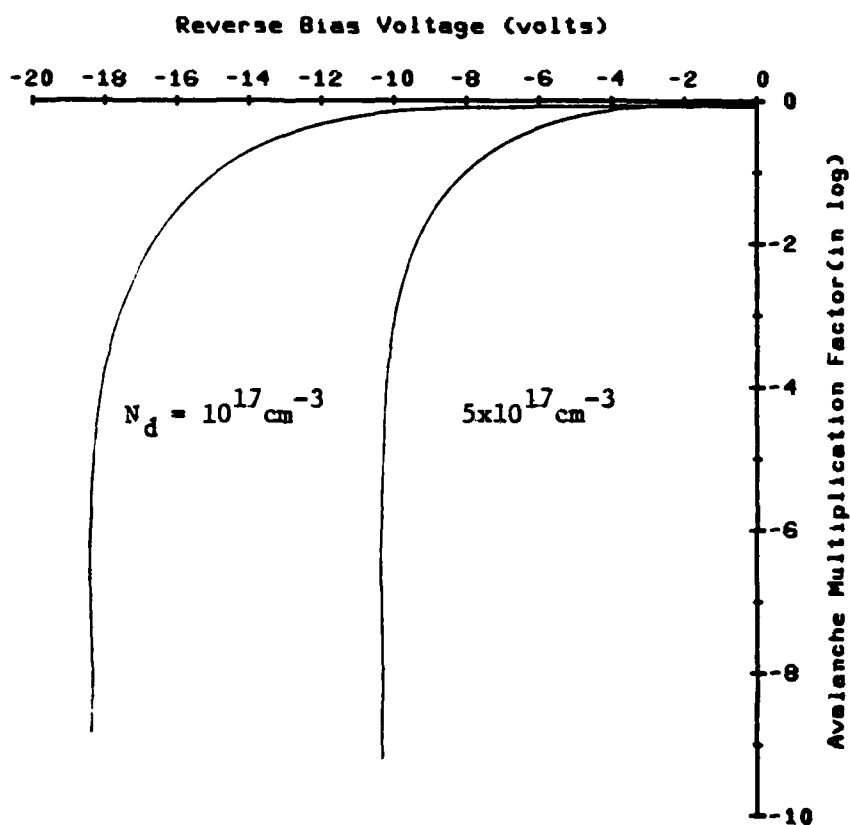
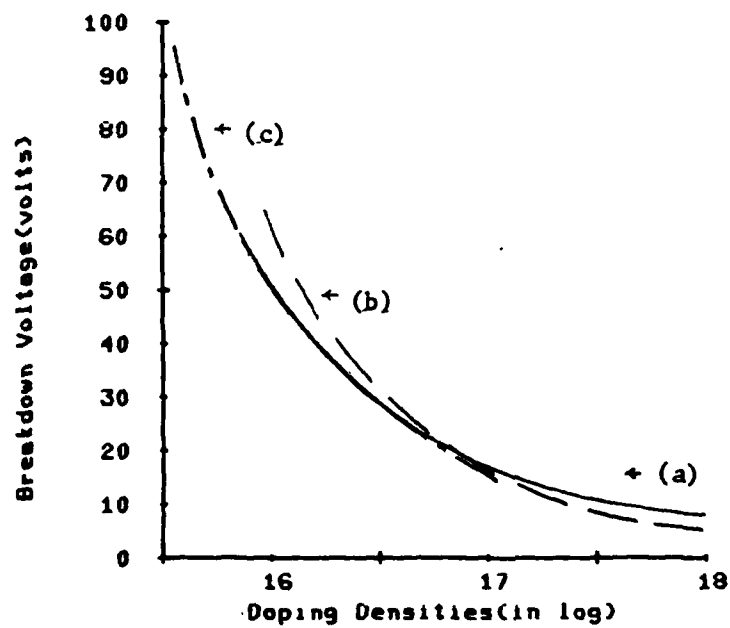


Figure 3 (a) Calculated breakdown voltage for GaAs as a function of doping density.
(b) Experimental results taken from Sze¹, pp 115, which assume $\alpha = \beta$.
(c) Taken from Stillman and Wolfe⁶.



N_d	0.1	0.25	0.50	0.75	1.0	2.5	5.0	7.5	10
n	3.6	3.4	3.2	3.0	2.8	2.4	2.0	1.6	1.7
m	0.9	1.2	1.6	1.5	1.4	1.6	1.8	1.6	1.7

Table 1 GaAs Avalanche multiplication exponents as a function of doping density.
(The unit used for N_d is 10^{17} cm^{-3}).

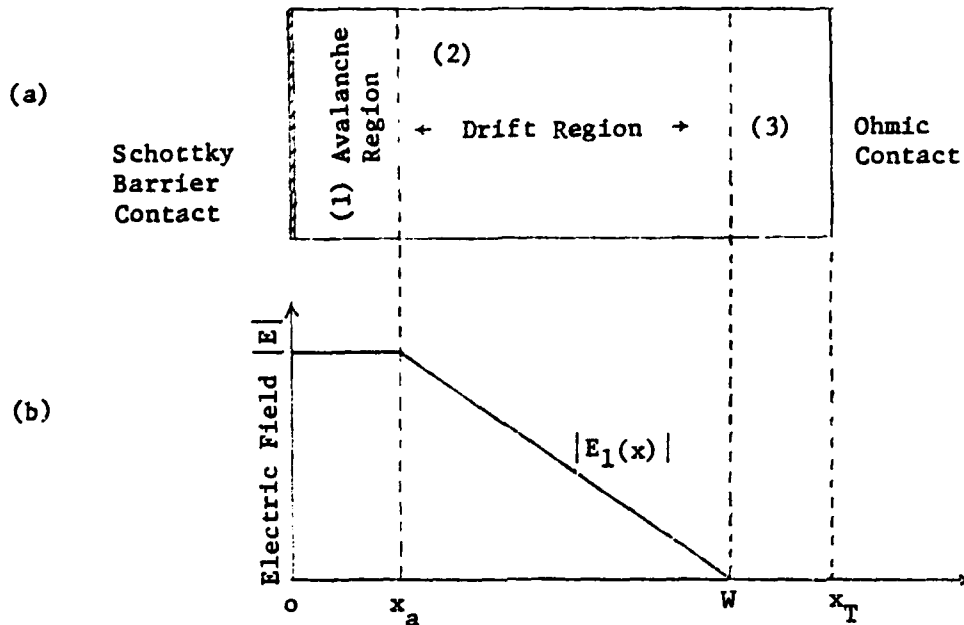


Figure 4 (a) Model of one-dimensional Schottky Barrier diode with (1) avalanche region, (2) drift region, and (3) inactive region.

(b) Electric field distribution in each region.

following continuity equation⁸:

$$\frac{d}{dt} J_{inj}(t) = \frac{2}{\tau_a} J_{inj}(t) \left[\int_0^{x_a} \alpha(E) dx - 1 \right] + \frac{2}{\tau_a} J_{sat} \quad (15)$$

Here, the transit time through the avalanche region is

$$\tau_a \triangleq \frac{x_a}{v_s}, \quad J_{sat} \triangleq J_{ns} + J_{ps} = J_{ns} \quad (\text{for n-type}),$$

and v_s is the saturation velocity of electrons.

The field in the avalanche region is given by

$$E(x,t) = E_1(x) + E_2(t) \quad (16)$$

where $E_1(x)$ is determined by the background doping. At any instant of time, the field throughout the avalanche region is determined by the value of the field at any reference plane. For convenience, we take $x = x_a$ and the avalanche field is defined as $E_a(t) \triangleq E(x_a, t)$. The ionization integral may be expressed as

$$\int_0^{x_a} \alpha(E) dx = F(E_a(t)) \quad (17)$$

After some rearrangement, one obtains

$$E_a(t) = E_p - \frac{V_p}{x_T} + \frac{V_T(t)}{x_T} - \frac{V_s}{\epsilon x_T} \left[\int_{t-\tau_d}^t (\tau_d + t' - t) J_{inj}(t') dt' - \tau_d^2 J_{ps} \right] \quad (18)$$

where $E_p \triangleq E_1(x_a)$, the punch-through field,

$$V_T(t) \triangleq \int_0^{x_T} E(x,t) dx; \quad x_T \text{ is total length of the active region.}$$

$$V_p \triangleq \int_0^{x_T} E_1(x) dx$$

$$E_1(x) \triangleq -\frac{q}{\epsilon} \left[\int_0^x N_d(x') dx - \int_0^{x_T} N_d(x') dx' \right] \quad (19)$$

$$t' \triangleq t - \frac{x - x_a}{v_s}; \quad \tau_d \triangleq \frac{x_T - x_a}{v_s}$$

$$E(x, t) \triangleq \frac{q}{\epsilon} \int_0^x (N_d + p-n) dx.$$

Also, we assume a relatively narrow avalanche region of width x_a , with ionization coefficient α , and a wider drift region where $\alpha \approx 0$. The displacement current in the avalanche region is assumed constant.

Now, if we know the applied voltage $V_T(t)$, and ionization coefficient α , the current under avalanche operation can be found from the following expressions, assuming a reasonable value of x_a :

$$\frac{d}{dt} J_{inj}(t) = \frac{2}{\tau_a} J_{inj}(t) \left[\int_0^{x_a} \alpha(E) dx - 1 \right] + \frac{2}{\tau_a} J_{sat} \quad (20)$$

$$E_a(t) = E_p - \frac{V_p}{x_T} + \frac{V_T(t)}{x_T} - \frac{v_s}{\epsilon x_T} \left[\int_{t-\tau_d}^t (\tau_d + t' - t) J_{inj}(t') dt' - \tau_d^2 J_{ps} \right] \quad (21)$$

Using direct iteration or the Newton-Raphson method, it is possible to obtain an approximate solution for $J_{inj}(t)$.

Once the injected current density is known the large signal I-V relation can be calculated as follows:

$$J_{inj}(t) = J_n(x, t) + J_p(x, t) + \epsilon \frac{\partial E(x, t)}{\partial t}. \quad (22)$$

Integration over the entire active region gives:

$$\begin{aligned}
 J_{inj}(t) &= \frac{\epsilon}{x_T} \frac{d}{dt} \int_0^{x_T} E(x,t) dx + \frac{1}{x_T} \int_0^{x_T} (J_n(x,t) + J_p(x,t)) dx \\
 &= \frac{\epsilon}{x_T} \frac{d}{dt} V_T(t) + J_{ind}(t)
 \end{aligned} \quad (23)$$

where $J_{ind}(t)$ is the space average of the conduction current:

$$J_{ind}(t) \triangleq \frac{1}{x_T} \left[\int_0^{x_a} J_{inj}(t) dx + \int_{x_a}^{x_T} J_n(x,t) dx + x_d J_{ps} \right] \quad (24)$$

where $x_d \triangleq x_T - x_a$ and $J_p(x,t) \approx J_{ps}$ in the drift region.

Thus,

$$J_{ind}(t) \triangleq \frac{x_a}{x_T} J_{inj}(t) + \bar{x}_d \frac{1}{\tau_d} \int_{t-\tau_d}^t J_{inj}(t') dt' \quad (25)$$

where

$$\bar{x}_d \triangleq \frac{x_d}{x_T} = \frac{x_T - x_a}{x_T}$$

$$\tau_d \triangleq \frac{x_d}{v_s} = \frac{x_T - x_a}{v_s}$$

$$t' = t - \frac{x - x_a}{v_s}$$

The term $\frac{\epsilon}{x_T} \frac{d}{dt} V_T(t)$ or $J_{inj}(t) - J_{ind}(t)$ is the induced current, which comes from variation of the terminal applied voltage $V_T(t)$. Equations (20) and (21) can be used to examine, numerically, the large signal behavior of an avalanching region. This formulation may prove useful in the simulation of breakdown behavior in millimeter wave mixer diodes and in field effect transistors.

3.2.3 Mott Barrier Diffusion Model^{1,13,14}

A Mott barrier, also called punch-through barrier, is defined as one in which the epitaxial layer is much narrower than the required depletion width, under the depletion approximation, such that the layer is depleted even under forward bias. For simplicity, we assume an ideal contact and no surface states. The structure is shown in Figure 5. The band diagrams at various biasing conditions are also shown in Figure 5. Assuming that the required depletion width is W_0 ($W_0 > W_2$) and the built-in voltage is V_0 , the electric field is given by

$$E(x) = \frac{q}{\epsilon} [N^-(x-W_1) - N(W_2 - W_1) - N^+(W_0 - W_2)] \quad 0 \leq x \leq W_1 \quad (26)$$

$$= \frac{q}{\epsilon} [N(x - W_2) - N^+(W_0 - W_2)] \quad W_1 \leq x \leq W_2$$

$$= -\frac{q}{\epsilon} N^+(W_0 - x) \quad W_2 \leq x \leq W_0$$

$$\begin{aligned} \text{and } V_0 &= \frac{1}{2} \frac{qN^+}{\epsilon} [W_0^2 - (1 - \frac{N}{N^+}) W_2^2 - \frac{N}{N^+} (1 - \frac{N^-}{N}) W_1^2] \\ &= \frac{1}{2} \frac{q}{\epsilon} [N^+ W_0^2 - (N^+ - N) W_2^2 - (N - N^-) W_1^2]. \end{aligned} \quad (27)$$

Therefore,

$$W_0 = \left[\frac{2 \epsilon V_0}{qN^+} + (1 - \frac{N}{N^+}) W_2^2 + \frac{N}{N^+} (1 - \frac{N^-}{N}) W_1^2 \right]^{1/2} \quad (28)$$

When we apply a voltage V , the depletion width W will be:

$$W = \left[\frac{2 \epsilon (V_0 - V)}{qN^+} + (1 - \frac{N}{N^+}) W_2^2 + \frac{N}{N^+} (1 - \frac{N^-}{N}) W_1^2 \right]^{1/2} \quad (29)$$

and $W \geq W_2$.

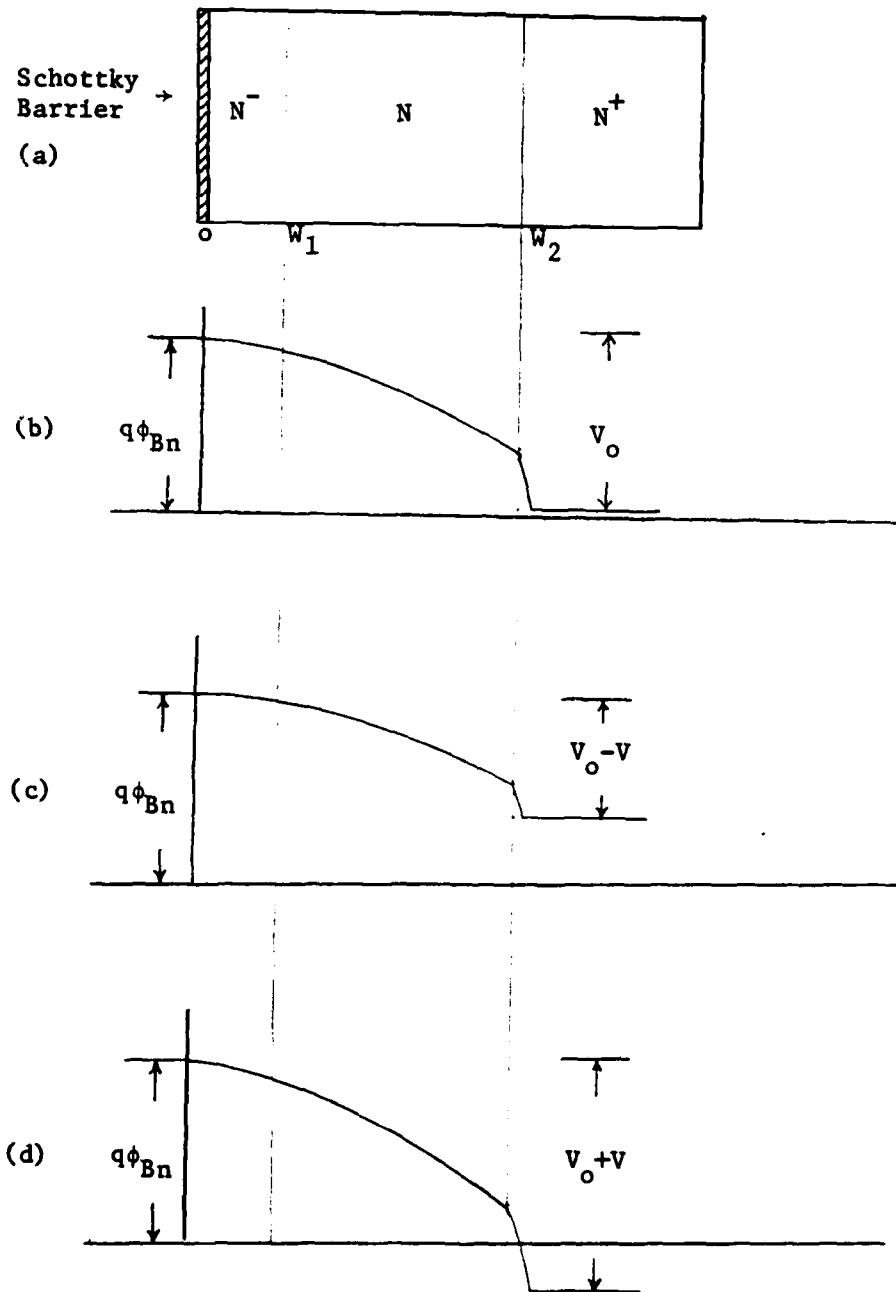


Figure 5 (a) Mott Barrier model and the band diagrams at
 (b) zero-bias, (c) forward-bias and (d) reverse-bias.

Figure 6 I-V curve for Mott Barrier diode as in Figure 5.

$$N^- = 2 \times 10^{15} \text{ cm}^{-3}$$

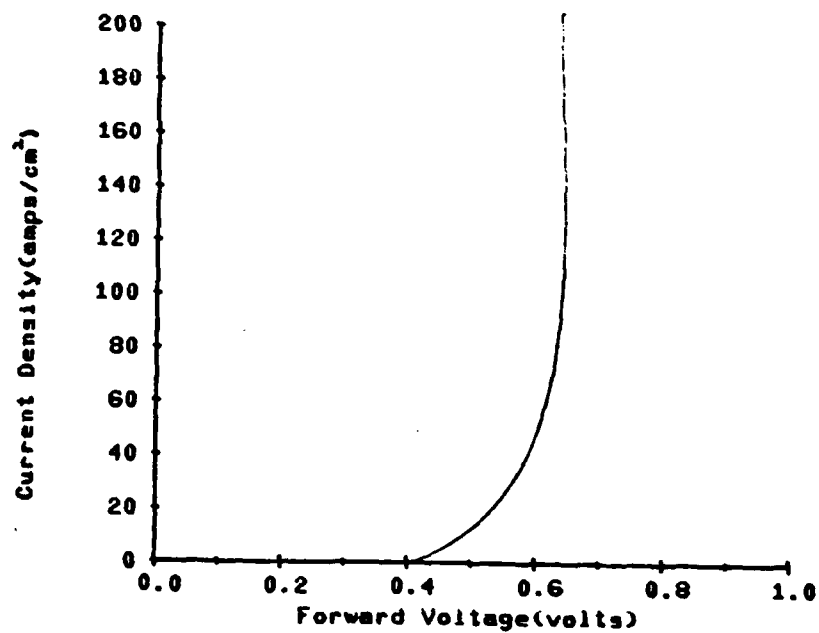
$$N = 1 \times 10^{17} \text{ cm}^{-3}$$

$$N^+ = 2 \times 10^{18} \text{ cm}^{-3}$$

$$W_1 = 0.046 \times 10^{-5} \text{ cm}$$

$$W_2 - W_1 = 1.0 \times 10^{-5} \text{ cm}$$

$$\text{Built-in Voltage } V_0 = 0.9 \text{ volt}$$



3.2.3.1 I-V Characteristic Relation

Charge transport is governed by the current equation and Poisson's Equation. We use the field dependent mobility for GaAs^{9,10,11} and write the required equation as follows:

$$J = J_n = qn\mu_n(E) + qD_n(E) \frac{\partial n}{\partial x} \quad (30)$$

where

$$\mu_n(E) = \frac{\mu_0 + \frac{v_s}{E_c} \left(\frac{E}{E_c} \right)^3}{1 + \left(\frac{E}{E_c} \right)^4} \quad (31)$$

$$D_n(E) = \frac{kT}{q} \frac{v_n(E)}{E} + \frac{3}{2} \tau v_n^2(E) \quad (32)$$

$$\nabla^2 \psi = -\frac{q}{\epsilon} (N_d - n) \quad (33)$$

$$E = -\frac{d\psi}{dx} \quad (34)$$

$$\frac{dJ}{dx} = 0 \quad (35)$$

where

$$\begin{aligned} N_d &= N^- & 0 \leq x < W_1 \\ &= N & W_1 \leq x < W_2 \\ &= N^+ & W_2 \leq x \end{aligned}$$

$$n(0) = N^- \exp\left(-\frac{q}{kT} V_0\right)$$

$$n(W) = N^+$$

It is not easy to solve the above nonlinear system in closed form, although it is possible to obtain an approximation numerically. (See Figure 6). In order to proceed with the numerical analysis, it is necessary to know the barrier height. As a matter of fact, besides the properties of the metal used for the Schottky Barrier contact, the image force between metal and bulk material also affects the barrier height. This is called the Schottky Effect.

3.2.3.2 Image Force

Figure 7 shows the energy diagram for electrons near the Schottky metal contact. The height of the potential barrier is reduced because of the attractive force between an electron and its image in the contact metal. If we take this image force into account, since the location of the potential maximum x_m is very close to the metal surface; let $x_m \ll W_1$. The electric field at that point is zero,

$$E(x_m) = \frac{q}{\epsilon} [N^-(x_m - W_1) - N(W_2 - W_1) - N^+(W - W_2)] + \frac{q}{\epsilon} \frac{1}{16\pi x_m^2} = 0 \quad (36a)$$

or

$$\frac{1}{16\pi x_m^2} = N^-(W_1) + N(W_2 - W_1) + N^+(W - W_2) \quad (36b)$$

so

$$x_m = \frac{1}{\left[16\pi [N^-(W_1) + N(W_2 - W_1) + N^+(W - W_2)] \right]^{1/2}} \quad (36c)$$

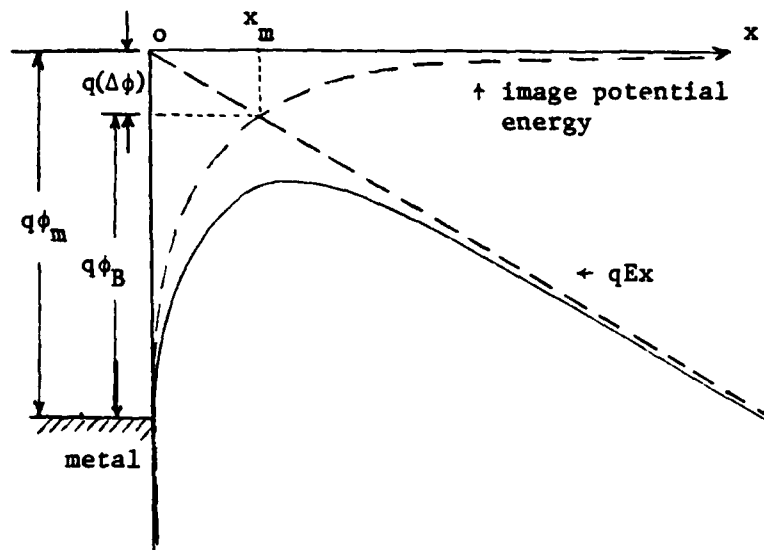


Figure 7 Potential energy of electrons in a Schottky Barrier diode, taking barrier lowering due to the image force into account¹. (x_m is in the order of 10^{-9} cm).

Thus, the potential is

$$\begin{aligned}
 V(x) &= - \int_0^x E(x) dx \\
 &= \frac{q}{\epsilon} \left\{ [N^- (W_1 x - \frac{1}{2} x^2) + N(W_2 - W_1) x + N^+ (W - W_2) x] \right. \\
 &\quad \left. - \left[\frac{x}{16\pi x_m^2} \right] \right\} - \phi_{B_n} \quad \begin{array}{l} \text{-----} \quad 0 \leq x < x_m \\ \text{-----} \quad x_m \leq x \leq W_1 \end{array} \quad (37)
 \end{aligned}$$

$$V(x) = \frac{q}{\epsilon} \left\{ [N(W_2 x - \frac{1}{2} x^2) + N^+ (W - W_2) x] + \frac{1}{16\pi x} \right\} - \phi_{B_n} \quad \text{-----} \quad W_1 \leq x < W_2 \quad (38)$$

$$V(x) = \frac{q}{\epsilon} \left\{ [N^+ (W x - \frac{1}{2} x^2)] + \frac{1}{16\pi x} \right\} - \phi_{B_n} \quad \text{-----} \quad W_2 \leq x \leq W \quad (39)$$

Thus, the image force lowers the barrier energy by an amount

$$\Delta\phi = \frac{q}{\epsilon 16\pi x_m}$$

3.2.4 One-Dimensional Schottky-Barrier Diode Finite Element Analysis^{10,11,12}

In order to obtain the electron densities and potentials everywhere inside the diode, one solves Poisson's equation and the continuity equations. However, closed form solutions are apparently impossible when nonlinear velocity-field relation is taken into account. In this section we consider only one-dimensional cases and examine the details of the static solution in a Schottky-barrier diode.

The relevant equations are ^{9,10,11}.

$$\nabla^2 \psi = \frac{q}{\epsilon} (n - N_d) \quad (40)$$

$$0 = q \frac{\partial n}{\partial t} = \nabla \cdot \vec{J}_n \quad (\text{we assume } \frac{\partial n}{\partial t} = 0) \quad (41)$$

$$\vec{E} = - \nabla \psi \quad (42)$$

$$\vec{J}_n = q n \vec{v}_n + q D_n \nabla n \quad (43)$$

where

$$\vec{v}_n = \frac{\mu_o + \left(\frac{v_s}{E_o} \right) \left(\frac{|E|}{E_o} \right)^3}{1 + \left(\frac{|E|}{E_o} \right)^4} \vec{E} \quad (44)$$

Assume the one-dimensional structure shown in Figure 8 with the Schottky-Barrier on the left hand side, a region with doping density, N_d , and an applied voltage, V . We have elected to use the finite element method to solve this problem. We divide the entire length into $L+1$ elements with equal lengths. There are L nodes inside, and each element has length $G = \frac{W}{L+1}$; V_i , n_i represent the electron potential and density respectively at node i . If L is large enough, one can assume the potential between node i and node $i+1$ is

$$\psi_i = a_i + b_i x \quad i \leq i \leq L \quad (45)$$

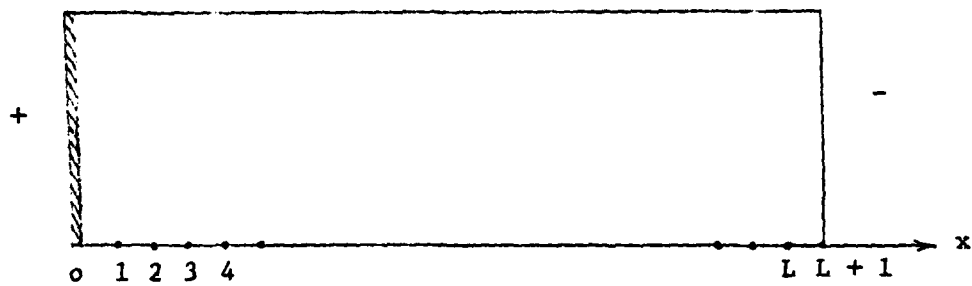


Figure 8 One-Dimensional Schottky Barrier diode with L equally separated nodes inside.

It can be shown that

$$a_i = \frac{1}{G} [x_{i+1} v_i - x_i v_{i+1}] \quad (46a)$$

$$b_i = \frac{1}{G} [v_{i+1} - v_i] \quad (46b)$$

$$\frac{\partial \psi_i}{\partial x} = b_i \quad (46c)$$

$$\frac{\partial^2 \psi_i}{\partial x^2} = \frac{1}{G} [b_{i+1} - b_i] \quad (46d)$$

Similarly, if V is replaced by n , we obtain the density between each node as

$$m_i = C_i + d_i x \quad 1 \leq i \leq L \quad (47a)$$

$$C_i = \frac{1}{G} [x_{i+1} n_i - x_i n_{i+1}] \quad (47b)$$

$$d_i = \frac{1}{G} [n_{i+1} - n_i] \quad (47c)$$

$$\frac{\partial m_i}{\partial x} = C_i \quad (47d)$$

$$\frac{\partial^2 m_i}{\partial x^2} = \frac{1}{G} [C_{i+1} - C_i] \quad (47e)$$

Substitution into Poisson's equation and the continuity equations yields

$$v_{i+2} + v_i - 2v_{i+1} = -\frac{q}{\epsilon} G^2 \left[N_d - \frac{1}{2} (N_i + N_{i+1}) \right] \quad (48a)$$

$$\frac{1}{2} (N_i + N_{i+1}) \mu_n(E_i) \left[-\frac{1}{G^2} (v_{i+2} + v_i - 2v_{i+1}) \right] \quad (48b)$$

$$+ \frac{1}{G} (n_{i+1} - n_i) \mu_n(E_i) \left[-\frac{1}{G} (v_{i+1} - v_i) \right] \quad 1 \leq i \leq L$$

$$+ D_n(E) \frac{1}{G^2} (n_{i+2} + n_i - 2n_{i+1}) = 0$$

where

$$E_i = -\frac{\partial \psi_i}{\partial x}$$

We take the following parameters for GaAs.

$$\mu_0 = 6875 \text{ cm}^2/\text{v sec}$$

$$v_s = 1.1 \times 10^7 \text{ cm/sec}$$

$$E_0 = 3.2 \text{ KV/cm}$$

and apply the boundary conditions:

$$n(0) = N_d \exp \left[- \frac{q}{kT} (V_0 - V) \right]$$

$$n(L+1) = N_d$$

$$V(0) = -V_0$$

$$V(L+1) = -V$$

where V_0 = built-in voltage

V = applied voltage

Computer print out shows that the potential increases gradually from left to right. However, the electron density increases very rapidly in the segment nearest to the Schottky contact. The values obtained are plotted in Figures 9(a) and 9(b). It is noted that the electron density decreases from 10^{17} to 10^{14} , but that the usual depletion approximation $n = 0$ is strictly valid only immediately under the contact. However, the potential distribution is close to that under the depletion approximation; the difference being no more than

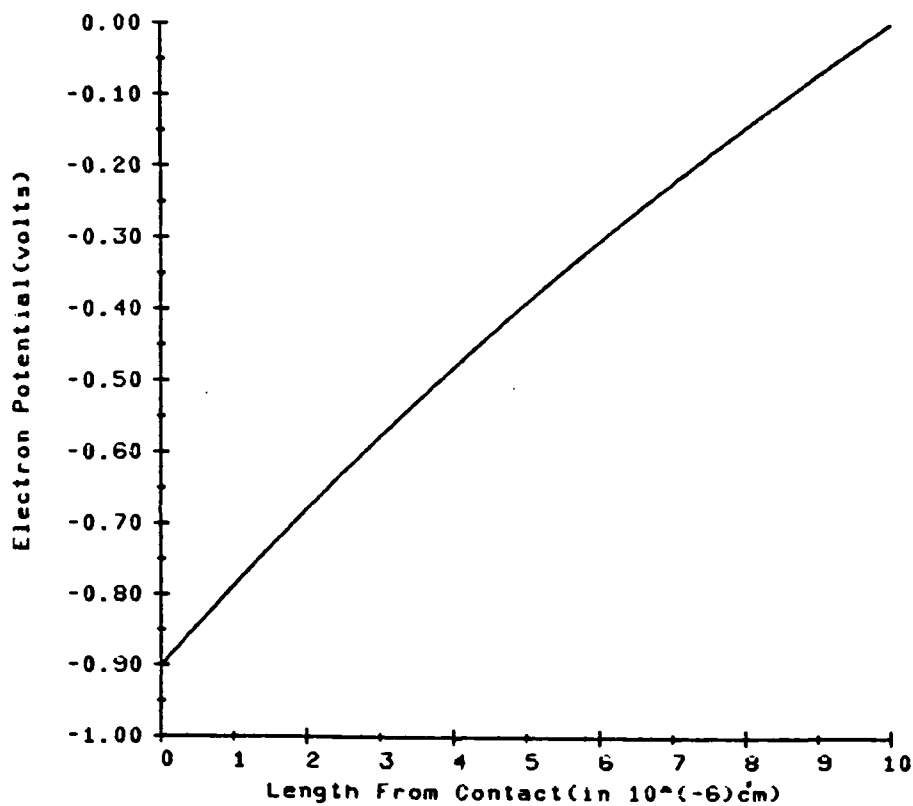


Figure 9: Computer print out for one-dimensional Schottky Barrier diode (punched through at zero bias):
 Barrier height = 0.9 volt, $N_d = 10^{17} \text{ cm}^{-3}$
 (a) Potential in entire bulk region

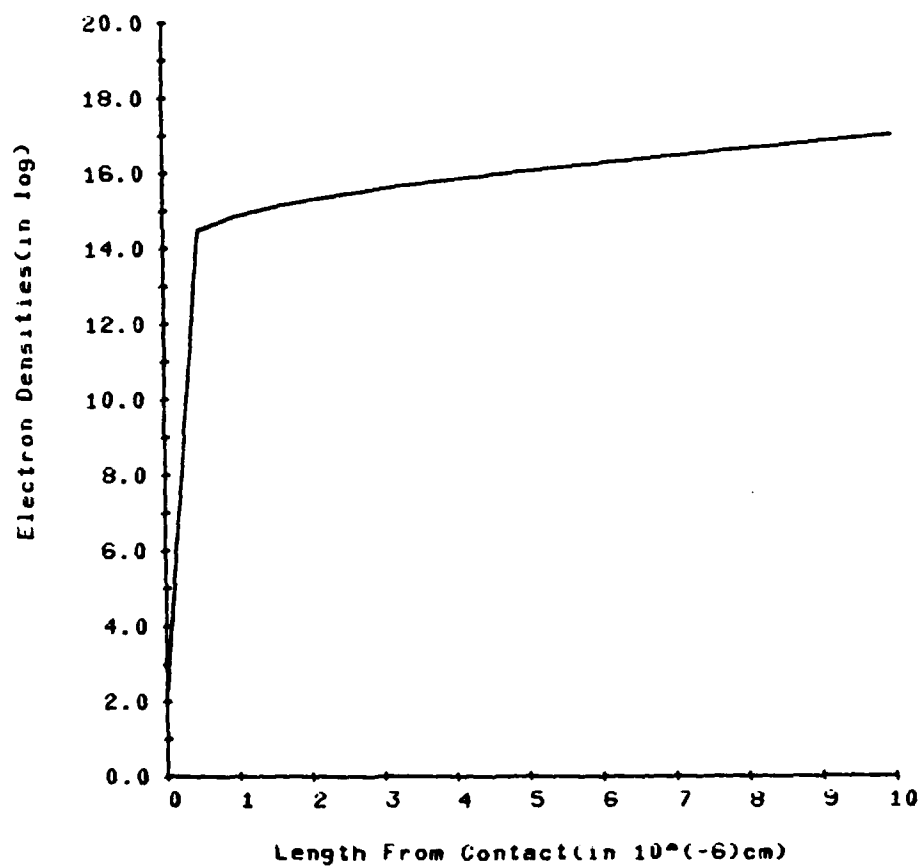


Figure 9: (b) Electron density in entire bulk region

5%. The electric field is nearly constant ($\frac{\partial \psi}{\partial x} \approx \text{constant}$). Note that the Debye length for this example is 1.89×10^{-6} cm.

3.2.5 Two-Dimensional Schottky Barrier Diode Analysis

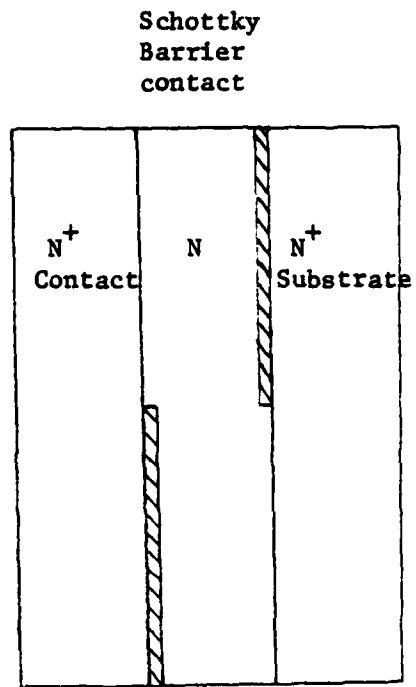
Subharmonically pumped (antiparallel) diode pairs are proving useful as millimeter wave mixers. Their primary advantage is that the local oscillator frequency is half that of the desired signal, thus easing the burden on 100 GHz (and higher) solid state sources. However, there are a variety of problems associated with fabricating and mounting matched diode pairs in millimeter wave circuits. We have conceived of the following fabrication scheme which may lead to improved matching and packaging of subharmonically pumped diodes.

Consider the geometry of Figure 10 which shows an antiparallel GaAs diode pair fabricated on a single N^+ substrate. A Schottky metal, such as Al is deposited over a portion of the substrate, and an N doped GaAs layer is grown around and over it. Techniques for overgrowing metal layers are currently under investigation at MIT Lincoln Laboratories and appear to be feasible. After the desired diode thickness is obtained a second Schottky barrier is deposited and an ohmic contact is applied to the entire upper surface completing the device. Such a structure should be highly symmetric and uniform so that well balanced I-V curves should be obtained. Furthermore, the millimeter wave parasitic elements are all external to the diode pair, so that circuit balance should be improved and impedance matching simplified.

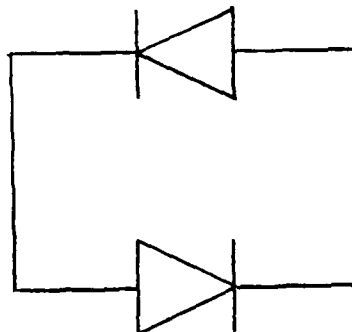
In order to exploit these apparent advantages we have begun to analyze the carrier distribution in such a structure to determine whether low leakage resistance pairs can be defined and, if so, to obtain design data.

3.2.5.1 Finite Element Analysis

The equations needed for our analysis are^{9, 10, 11}.



(a) Antiparallel diode pair configuration



(b) Equivalent circuit

Figure 10. Two-dimensional Schottky Barrier diode and the equivalent circuit

$$\nabla^2 \psi = \frac{q}{\epsilon} (n - N_d)$$

$$\nabla \cdot \vec{J}_n = 0 \quad (\text{assume } \frac{\partial n}{\partial t} = 0)$$

(49)

$$\vec{E} = -\nabla \psi$$

$$\vec{J}_n = q_n \vec{v}_n + q D_n \nabla n$$

where

$$\vec{v}_n = \frac{\mu_0 + \left(\frac{v_s}{E_0} \right) \left(\frac{|E|}{E_0} \right)^3}{1 + \left(\frac{|E|}{E_0} \right)^4} \vec{E}, \quad D_n(E) = \frac{kT}{q} v_n(E) + \frac{3}{2} \tau v_n^2(E)$$

Instead of solving second order differential equations it is possible to reduce them to first order relations with the aid of a vector formula:

$$\iint f \left(\nabla \cdot \frac{\vec{J}_n}{q} \right) d\Omega = \iint \nabla \cdot \left(f \frac{\vec{J}_n}{q} \right) d\Omega - \iint \nabla f \cdot \frac{\vec{J}_n}{q} d\Omega \quad (50)$$

We approach the numerical solution of these equations by means of a finite element method. The first term on the right hand side will vanish if the elements with which we are dealing are inside the diode, and also vanish if the boundary values are constant or the component of the current along the normal of the boundary is zero. Similar arguments apply to the first term of the right hand side of the following equation:

$$\iint \nabla f \cdot \nabla \psi d\Omega = \iint \nabla \cdot (f \nabla \psi) d\Omega - \iint f \nabla^2 \psi d\Omega \quad (51)$$

So we have

$$\iint \nabla f \cdot \frac{\vec{J}_n}{q} d\Omega = 0$$

and

$$\iint \nabla f \cdot \nabla \psi d\Omega = - \iint f \frac{q}{\epsilon} (n - N_d) d\Omega \quad (52)$$

Now the two-dimensional device of Figure 10, when divided into a suitable number of elements for analysis, appears as shown in Figure 11. Summing over all elements for which j is a node, using

$$\begin{aligned} f &= \phi_j \\ \psi &= \sum_{i=1}^3 \phi_i \psi_i \\ n &= \sum_{i=1}^3 \phi_i n_i \end{aligned} \quad (53)$$

yields

$$\sum_{\substack{\text{element} \\ \text{with } j}} \iint_{\Delta} \left\{ \sum_{i=1}^3 \left[\nabla \phi_j \cdot \nabla \phi_i \right] \psi_i + \frac{q}{\epsilon} \left[\sum_{i=1}^3 (\phi_j \phi_i) n_i - \phi_i n \right] \right\} dx dy = 0 \quad (54)$$

and

$$\sum_{\substack{\text{element} \\ \text{with } j}} \iint_{\Delta} \sum_{i=1}^3 \left\{ (\phi_j \phi_i) \frac{\partial n_i}{\partial t} + (\nabla \phi_j) \cdot \left[\phi_i n_i \vec{v}_n + D_n (\nabla \phi_i) n_i \right] \right\} dx dy = 0 \quad (55)$$

where $\frac{\partial n_i}{\partial t} \stackrel{\Delta}{=} 0$ by our assumption.

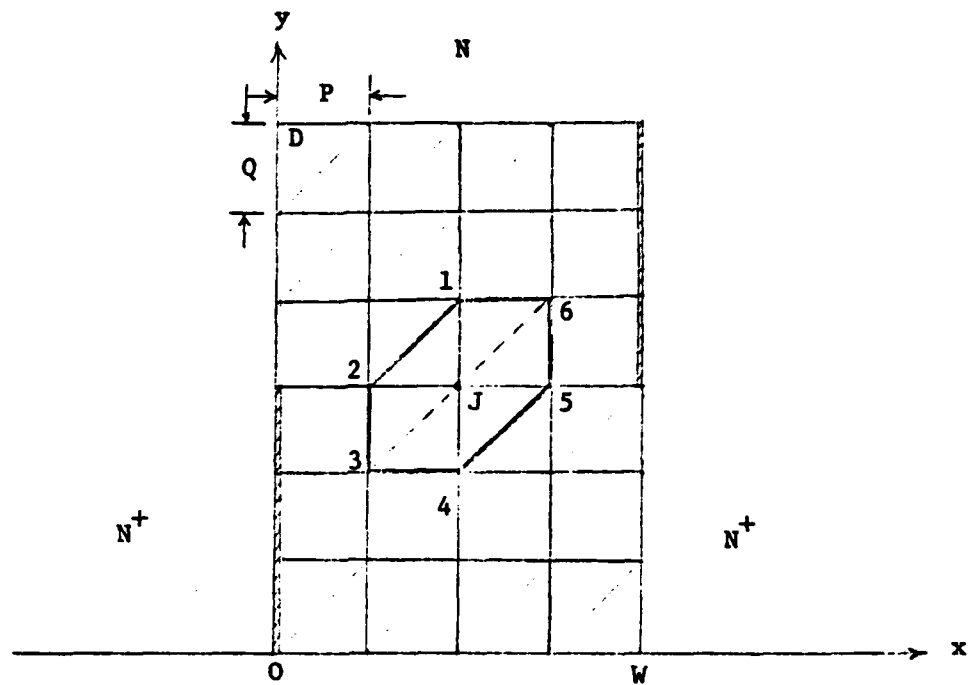


Figure 11. Application of linear function method to the entire bulk region, using triangular elements.

Since the elements are triangles, some properties¹² about them are needed in our analysis. See Figure 12. Take the origin at the centroid of a triangle. Then

$$\frac{1}{3} (x_1 + x_2 + x_3) = 0$$

$$\frac{1}{3} (y_1 + y_2 + y_3) = 0$$

$$\int x dx dy = 0$$

$$\int y dx dy = 0 \quad (56)$$

$$\int x^2 dx dy = \frac{\Delta}{12} (x_1^2 + x_2^2 + x_3^2)$$

$$\int y^2 dx dy = \frac{\Delta}{12} (y_1^2 + y_2^2 + y_3^2)$$

$$\int xy dx dy = \frac{\Delta}{12} (x_1 y_1 + x_2 y_2 + x_3 y_3)$$

$$\text{where } \Delta = \text{area of triangle} = \frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}$$

$$\text{Define } \phi_j = (a_j + b_j x + c_j y) / 2\Delta \quad (57)$$

where Δ = area of one element

$$= \frac{1}{2} PQ$$

To find the a_j, b_j, c_j values, assume $\psi = \alpha_1 + \alpha_2 x + \alpha_3 y$

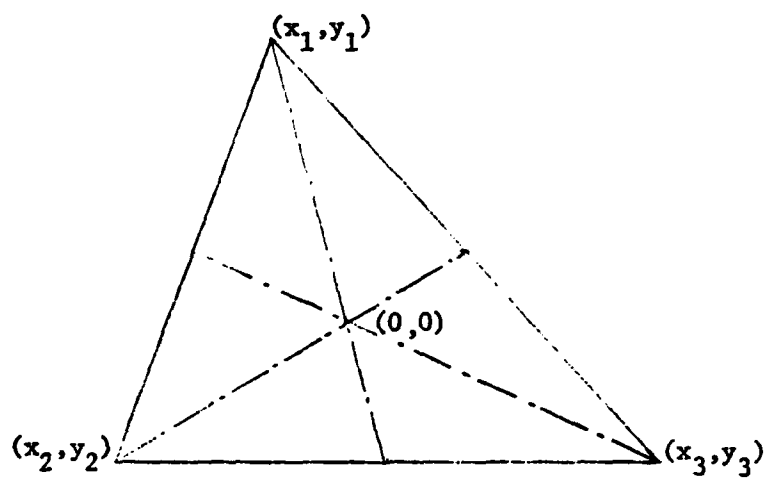


Figure 12. Triangle with origin at the centroid

Then,

$$\begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \quad (58)$$

$$\begin{aligned} \psi = [1 \ x \ y] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} &= [1 \ x \ y] \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix}^{-1} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \\ &= [1 \ x \ y] \frac{1}{2\Delta} \begin{bmatrix} \frac{2}{3}\Delta & \frac{2}{3}\Delta & \frac{2}{3}\Delta \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} \quad (59) \end{aligned}$$

We can easily obtain b_j, c_j values from the relation above and it is understood that in the centroid coordinate $a_j = \frac{2}{3}\Delta$, $j = 1, 2, 3$.

Two kinds of triangles used in analysis are:

- i) Triangle 12j, j34, 6j5;

It can be shown, that for this type

$$b_1 = Q; \quad b_2 = 0; \quad b_3 = -Q$$

$$c_1 = -P; \quad c_2 = P; \quad c_3 = 0$$

- ii) Triangle j23, 5j4, 6lj

It can be shown that for this type

$$b_1 = -Q; \quad b_2 = 0; \quad b_3 = 0$$

$$c_1 = P; \quad c_2 = -P; \quad c_3 = 0$$

where P, Q is the length of the elements in x and y direction respectively.

After some manipulation, Poisson's equation and the Continuity equation can be written:

$$\sum_{\substack{\text{element} \\ \text{with } j}} \left[\sum_i \psi_i \left(\frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j}{\partial x} + \frac{\partial \phi_i}{\partial y} \frac{\partial \phi_j}{\partial y} \right) + \frac{1}{12} \frac{q}{\epsilon} (n_j + n_1 + n_2 + n_3) - \frac{1}{3} \frac{q}{\epsilon} N_d \right] = 0 \quad (50)$$

$$\Delta \sum_{\substack{\text{element} \\ \text{with } j}} \left[\frac{1}{3} (n_1 + n_2 + n_3) \left(v_{nx} \frac{\partial \phi_j}{\partial x} + v_{ny} \frac{\partial \phi_j}{\partial y} \right) + \sum_i D_n(E) n_i \cdot \left(\frac{\partial \phi_i}{\partial x} \frac{\partial \phi_j}{\partial x} + \frac{\partial \phi_i}{\partial y} \frac{\partial \phi_j}{\partial y} \right) \right] = 0 \quad (61)$$

where

$$\frac{\partial \phi_i}{\partial x} = \frac{b_i}{2\Delta}; \quad \frac{\partial \phi_j}{\partial x} = \frac{b_j}{2\Delta}$$

$$\frac{\partial \phi_i}{\partial y} = \frac{c_i}{2\Delta}; \quad \frac{\partial \phi_j}{\partial y} = \frac{c_j}{2\Delta}$$

v_{nx}, v_{ny} the velocity in x, y directions, respectively, and

$$\vec{E} = -\nabla\psi = -\nabla \left(\sum_{i=1}^3 \phi_i \psi_i \right)$$

$$E_x = - \sum_{i=1}^3 \psi_i \left(\frac{\partial \phi_i}{\partial x} \right); \quad E_y = - \sum_{i=1}^3 \psi_i \left(\frac{\partial \phi_i}{\partial y} \right)$$

$$|E| = \sqrt{E_x^2 + E_y^2}$$

$$v_{nx} = \mu_n(E)E_x; \quad v_{ny} = \mu_n(E)E_y$$

Figure 13 shows a flow chart of the calculation.

With the aid of a HP 9825A desk top calculator, we obtained the results shown. The potential distributions on the two boundaries $y = 0$ and D are close to the one dimensional depletion approximation (for example: See Figure 14, (a), (b) node number 1, 8, 15, 22, 29 and 7, 14, 21, 28, 35), if we take $D \geq 2W$. Some leakage is observed on a strip area around $y = \frac{1}{2} D$, in our results. This phenomena might be improved by increasing the lengths of both Schottky Barrier contacts a little bit such that two contacts overlap in y -direction.

3.2.6 Conclusion and plans for next period

We intend to pursue the analysis of the diode pair just described. Static I-V curves will be predicted and a parametric study made on the effect of the device geometric and material parameters. For example the overlap of the Schottky contacts should increase the leakage resistance by changing the position of the depletion regions. However, the capacitance between the diodes will also increase. Small signal scattering parameters will be predicted to examine the trade-off encountered in the design. Preliminary growth work to uncover the problems in epitaxial growth over metals will be initiated.

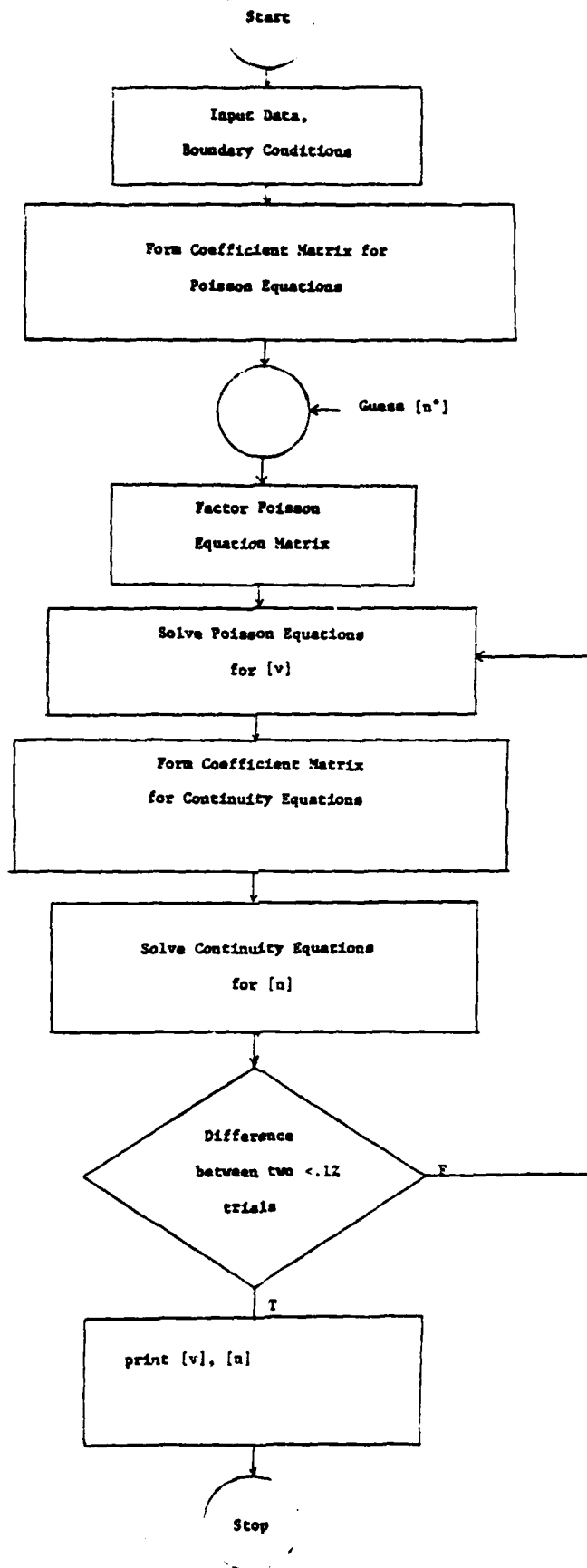


Figure 13 Flow chart for two-dimensional S.B. diode analysis using finite element method

Figure 14. An example of two-dimensional Schottky Barrier diode analysis using finite element method

Built-in voltage	0.90 volt
Applied voltage	0 volt
Energy Relaxation time	1.0×10^{-13} sec
Doping density	$1 \times 10^{17} \text{ cm}^{-3}$

Total number of nodes inside the bulk region 35

a.) Node Designation

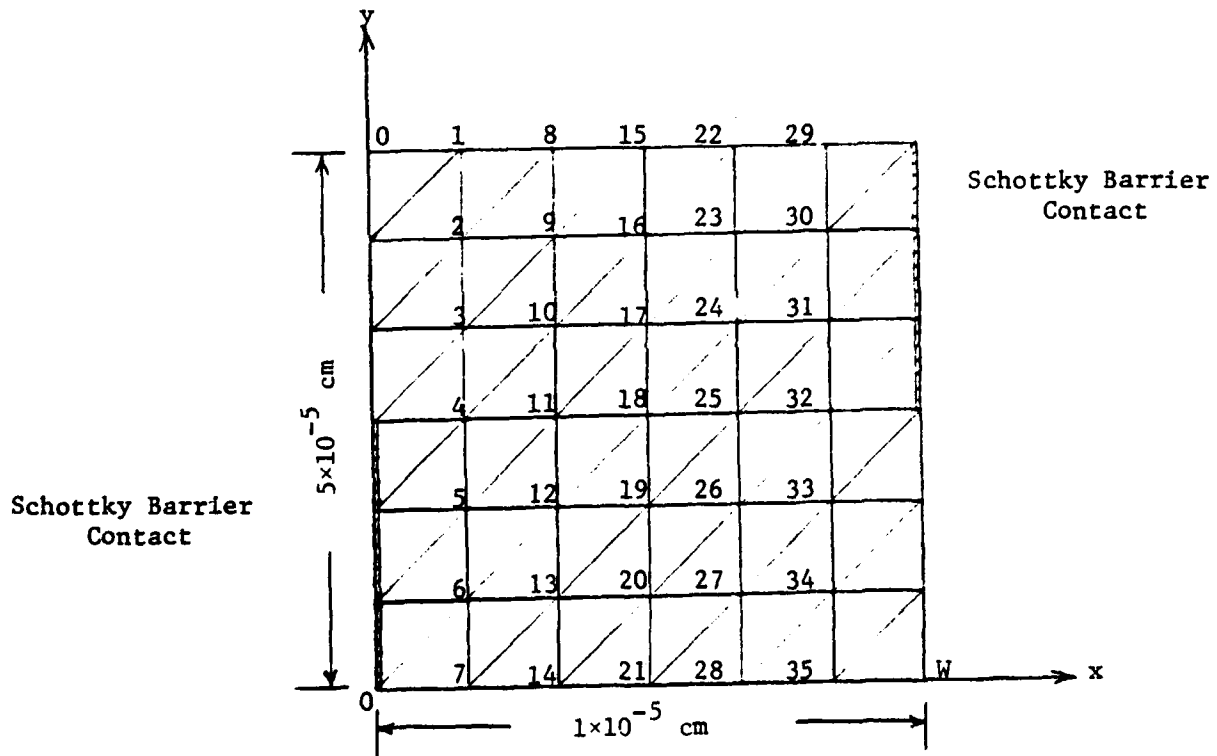


Figure 14(b). Numerical results (the first and last columns are boundary nodes, the number from 1 to 35 is the node number, respectively)

(1) potential distribution at each node

-0.9	1 -0.74	8 -0.58	15 -0.43	22 -0.28	29 -0.13	0
-0.9	2 -0.72	9 -0.56	16 -0.40	23 -0.26	30 -0.13	0
-0.45	3 -0.36	10 -0.28	17 -0.21	24 -0.13	31 -0.06	0
0	4 -0.01	11 -0.03	18 -0.04	25 -0.03	32 -0.01	0
0	5 -0.05	12 -0.13	19 -0.21	26 -0.28	33 -0.36	-0.45
0	6 -0.12	13 -0.26	20 -0.40	27 -0.56	34 -0.72	-0.90
0	7 -0.13	14 -0.28	21 -0.43	28 -0.58	35 -0.74	-0.90

(2) Density distribution at each node (unit: 10^{16} cm^{-3} : the zeros on the first and the last column are actually $8.1 \times 10^{-15} \times \text{unit}$ i.e. $8.1 \times 10^{-15} \times 10^{16} = 81$)

10	1 9.9	8 9.5	15 8.8	22 7.6	29 5.1	0
10	2 9.0	9 8.2	16 7.4	23 6.3	30 4.3	0
5.0	3 5.0	10 4.8	17 4.9	24 5.3	31 6.3	0
0	4 1.9	11 2.4	18 2.7	25 3.4	32 4.6	0
0	5 2.8	12 3.9	19 4.6	26 5.0	33 5.4	5.0
0	6 4.7	13 6.7	20 7.8	27 8.6	34 9.5	10
0	7 5.2	14 7.6	21 8.9	28 9.5	35 9.9	10

References

1. Sze, S. M., Physics of Semiconductor Devices, Wiley, N.Y. 1969, pp. 363-424, 111-115.
2. Keen, N., Haas, R. and Perchtold, E., "A Very Low Noise Mixer at 115 GHz, Using a Mott Diode Cooled to 20 K", Electron. Lett. Vol. 14, No. 25, pp. 825-825, Dec. 1978.
3. Cohn, M., Degenford, J. E., and Newman, B. A., "Harmonic Mixing with an Antiparallel Diode Pair", Vol. MTT-23, pp. 667-673, August 1975.
4. Carlson, E. R., Schneider, M. V., and McMaster, T. F., "Subharmonically Pumped Millimeter-Wave Mixers", IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-26, pp. 706-715, Oct. 1978.
5. Kerr, A. R., "Noise and Loss in Balanced and Subharmonically Pumped Mixers: Parts I and II", IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-27, pp. 938-950, Dec. 1979.
6. Stillman, G. E., and Wolfe, C. M., "GaAs Electroabsorption Avalanche Photodiode Detectors", Inst. Phys. Conf. Ser. No. 24, 1975, Chapter 4.
7. Ziel, V. D., "Solid State Physical Electronics", third edition, Prentice Hall, 1976, pp. 283-309.
8. Schroeder, W. E., "Nonlinear Properties of IMPATT Devices", Technical Report, RADC-TR-72-267, University of Michigan, Oct. 1972, pp. 1-92.
9. Barnes, J. J., "A Two-Dimensional Simulation of MESFETS", Technical Report, RADC-TR-76-153, University of Michigan, May, 1976.
10. Yamaguchi, K., Asai, S. and Kadera, H., "Two-Dimensional Numerical Analysis of Stability Criteria of GaAs FET's", IEEE Trans. on Electron Devices, Vol. ED-23, No. 12, Dec., 1976.
11. Yamaguchi, K., and Kadera, H., "Drain Conductance of Junction Gate FET's in the Hot Electron Range", IEEE Trans. on Electron Devices, Vol. ED-23, No. 6, June 1976.
12. Zienkiewicz, O. C., "The Finite Element Method", 3rd Edition, McGraw-Hill, 1977.
13. Mott, N. F., "Note on the Contact Between a Metal and an Insulator or Semiconductor", Proc. Camb. Phil. Soc., 34, 568., 1938.
14. Hemisch, H. K., "Rectifying Semiconductor Contacts", Oxford at the Clarendon Press., Oxford 1957.

15. Thim, H. W., "Computer Study of Bulk GaAs Devices with Random One-Dimensional Doping Fluctuations", J. Appl. Phys., Vol. 39, pp. 3897-3904, 1968.
16. Yamaguchi, K., Toyabe, T. and Koderu, H., "Effect of Field-Dependent Carrier Diffusion on the Two-Dimensional Analysis of a Junction Gate FET", Japan, J. Appl. Phys., Vol. 14, pp. 1069-1070, 1975.

4. New Device Concepts

B. Abraham-Shrauner and M. Muller

4.1 Inertial Transport in Semiconductors - Introduction

In devices that use high-mobility, low effective mass semiconductors, the mean free time between collisions of a carrier can become equal to, or longer than an oscillation period, or the mean free path between collisions equal to, or longer than a typical device dimension.

Under such conditions the carrier transport in semiconductors is not governed by the traditional friction-dominated Ohm's law behavior. The motion of the carriers for the short times or distances of primary interest for device operation may be more accurately described by the inertial or ballistic equation of motion.

An extreme version of such ballistic motion, assuming that the time between collisions is long compared with the time that a carrier needs to traverse a Brillouin zone, is discussed at length in Section 4.3 under the rubric of Zener Oscillations. The questions addressed there deal with the limits of validity of the effective mass theorem, tunneling, and similar concerns that arise in the ultra high field regime of semiconductor transport.

Under somewhat less extreme operating conditions that are beginning to be imposed on conventional but very small devices using extrinsic material, response time and high frequency performance is also thought to be capable of improvement through achievement of inertial transport conditions.

The paradigmatic device for studying inertial transport is the short space-charge limited semiconductor diode. Calculations of the behavior of such diodes, without collisions, and with few collisions¹⁾ have been reported. The work published to date has not yet taken into account several considerations that may be important, on which we have begun work:

- (i) an algebraic study of the uniqueness of the current voltage relation
- (ii) a linear stability analysis of fluctuations propagating in the longitudinal and transverse direction
- (iii) filament formation for current-controlled negative differential resistance and its possible instabilities
- (iv) finite carrier temperature and cathode electric fields
- (v) the effect of non-parabolic band structure.

4.2 Ballistic Transport and Instabilities

by B. Abraham-Shrauner

4.2.1 Ballistic Transport and multivalued current-voltage relations

In this section the derivation of the basic relation for the electrostatic potential as a function of distance from which can be found the d.c. current-voltage characteristics is given. The derivation is included because our results and those reported in the literature^{1,2} show some strange features that may under certain conditions cause instabilities rather than a stable operation regime for a device.

Our model is a low temperature semiconductor, for example GaAs at 77°K, with a mean free path for collisions greater than the short length of the device (.1 → 1 μm). Because the lateral dimensions are large, we use a one-dimensional model with a d.c. voltage of .1 to 1 volt across the semiconductor length. The energy of the electrons, the mobile carriers, at the cathode is much less than their energy at the anode and is assumed zero. Since there are a large number of electrons at the cathode in this space-charge limited regime, the electric field is assumed zero at the cathode. Diffusion and scattering of electrons are ignored; a steady-state solution is assumed.

The appropriate equations are

Poisson's equation

$$\frac{d^2U}{dx^2} = \frac{-e(n_0 - n)}{\epsilon} = \frac{-\rho}{\epsilon} \quad (1)$$

where U is the electrostatic potential, ρ is the volume charge density, e is the electronic charge of a proton, n is the electron density, n₀ is the doping density, ε is the electrical permittivity in SI units and x is the distance from the cathode, the equation of continuity

$$\nabla \cdot \vec{J} = \frac{dJ}{dx} = 0 \quad (2)$$

for the current density and the conservation of energy

$$\frac{1}{2} m^* v^2 - eU = 0 \quad (3)$$

where m^* is the electron effective mass $\approx .068 m_e$ for m_e the free electron mass, v is the bulk carrier speed of the electrons in the x-direction and the current relation

$$J = -env. \quad (4)$$

The solution of the above set of coupled equations follows the solution by Child³ and Langmuir⁴ for the space-charge limited vacuum diode that gave a $\frac{3}{2}$ power law for the current density. The electron density in Poisson's equation is eliminated by means of eqs (2) to (4) in terms of the potential which gives

$$\frac{d^2 U}{dx^2} = \frac{-e}{\epsilon} \left[n_0 + \frac{J}{e} \left(\frac{m^*}{2eU} \right)^{1/2} \right]. \quad (5)$$

For mathematical simplicity and ease of comparison we employ dimensionless variables.

$$u = \frac{U}{U_0}, \quad U_0 = \frac{m^* j^2}{2e^3 n_0^2}$$

$$w = \frac{x}{x_0}, \quad x_0 = j \left(\frac{\epsilon m^*}{2e^4 n_0^3} \right)^{1/2}$$

$j = -J > 0$, $\epsilon = \epsilon_0 \epsilon_R$. Then eq (5) becomes

$$\frac{d^2 u}{dw^2} = u^{-1/2} - 1. \quad (6)$$

The first integral is found by multiplying eq (6) by $\frac{du}{dw} dw$ and integrating,

$$\frac{1}{2} \left(\frac{du}{dw} \right)^2 = 2 \sqrt{u} - u + C. \quad (7)$$

The boundary conditions are $u(0) = \frac{du}{dw}(0) = 0$ at the cathode, $x = 0$. Thus $C = 0$.

The square root of both sides of eq (7) gives a separable equation that can be reduced to quadratures.

$$\int_0^u \frac{du'}{\sqrt{2} \sqrt{2u'^{\frac{1}{2}} - u'}} = w, \quad 0 \leq w \leq \pi\sqrt{2}. \quad (8)$$

The positive sign for $\frac{du}{dw}$ is determined by the fact that the electric field points from the positive anode to the cathode and $E = -\frac{dU}{dx}$. Then the integral can be found by letting $u' = t^4$. The result is valid only up to $w = \pi\sqrt{2}$. At this point $\frac{du}{dw} = 0$ but $\frac{d^2u}{dw^2} < 0$ there or for $w > \pi\sqrt{2}$, $\frac{du}{dw} < 0$ until $w = 2\pi\sqrt{2}$. The solution for $u(w)$ given here as an implicit function is periodic.

$$w = \left[(n+1)\pi\sqrt{2} - w_0 \right] \left[\frac{1 - (-1)^n}{2} \right] + \left[n\pi\sqrt{2} + w_0 \right] \left[\frac{1 + (-1)^n}{2} \right] \quad (9)$$

$$n\pi\sqrt{2} < w < (n+1)\pi\sqrt{2}$$

where $w_0 = -2^{\frac{1}{2}} u^{\frac{1}{4}} \left(2 - u^{\frac{1}{2}} \right)^{\frac{1}{2}} + 2^{\frac{3}{2}} \sin^{-1} \frac{u^{\frac{1}{4}}}{2^{\frac{1}{2}}} \quad (10)$

is valid for $0 \leq w \leq \pi\sqrt{2}$ as solution and appears in the general solution.

This implicit function $u(w)$ is not invertible analytically although one may solve for $u^{\frac{1}{4}}/2^{\frac{1}{2}}$ in terms of a sine function and demonstrate the periodic dependence of u on w . In Figure 1 the dimensionless potential u versus the dimensionless length w is plotted for two periods. The cathode is located at $w = 0$; the anode is at any other value of w that is appropriate. The curve is somewhat misleading since u and w are larger for smaller currents. If the anode is located at $w = 2\pi\sqrt{2}$, this predicts that electrons speed up and then slow down to zero at the anode with no potential across the semiconductor. Such behavior

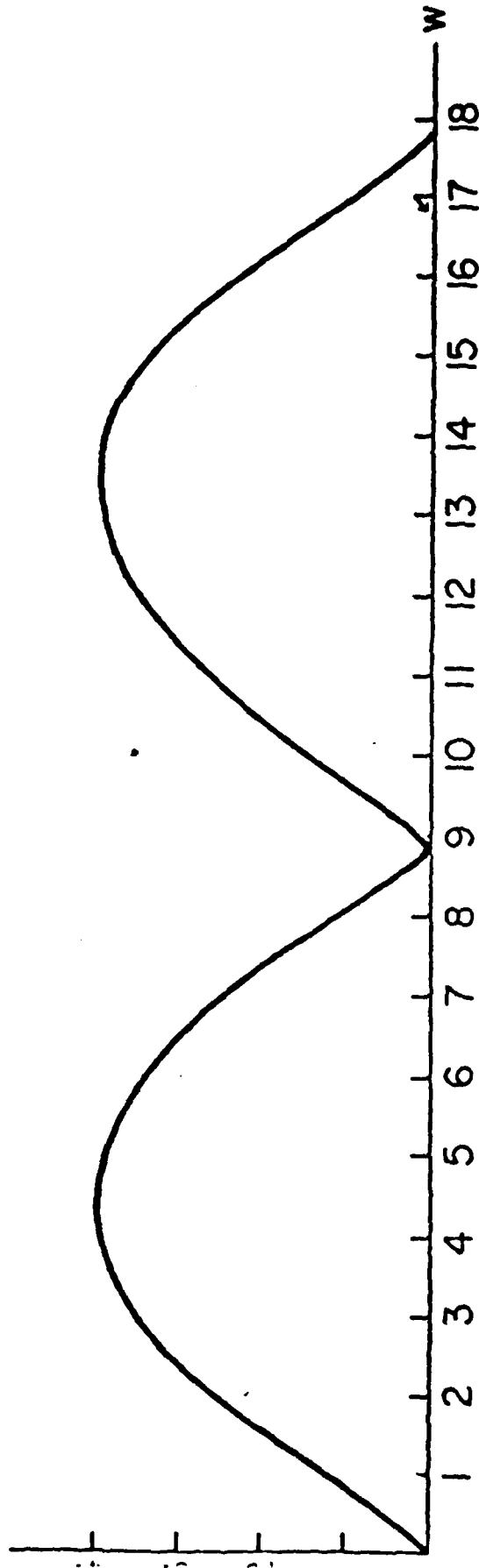


Figure 1. Dimensionless potential u versus dimensionless length w .

is bizarre but no more than the cases where electrons slow down, then speed up etc. In order to investigate this strange predicted behavior we first check the uniqueness of the current-voltage relation. Both M. Muller (private communication) and M. Shur² have found multivalued current densities for some potentials. Since multivalued currents are related to negative resistance which in turn can lead to instabilities, we first check the range of the unique values of current density for a given potential.

The range for which the current density is unique for a given potential is checked here. For $0 < w < \pi\sqrt{2}$ (we exclude the end points) we show that if (u_1, w_1) are chosen so that w_1 is in the range just given, then there is no other point (u_2, w_2) for which the potential U is the same but for which the current density J ($J = -j$) is different. Now (w_1, u_1) obey eq (10) and (w_2, u_2) obey

$$w_2 = 2\pi\sqrt{2} + 2^{\frac{1}{2}} u_2^{\frac{1}{2}} \left(2 - u_2^{\frac{1}{2}} \right)^{\frac{1}{2}} - 2^{\frac{3}{2}} \sin^{-1} \frac{u_2^{\frac{1}{2}}}{2^{\frac{1}{2}}} \quad (11)$$

for the range $\pi\sqrt{2} < w_2 < 2\pi\sqrt{2}$. We shall prove the uniqueness for this range of w_2 and then extend the proof. Now from the current density dependence of U_0 , x_0 we have

$$w_1 = \frac{j_2}{j_1} w_2 = \alpha w_2 \quad (12a)$$

$$u_1 = \frac{j_2^2}{j_1^2} u_2 = \alpha^2 u_2. \quad (12b)$$

Eqs. (10)-(12) must be satisfied if the current density is multivalued where $x = L$ is the device length in w_1 and w_2 and U is fixed in u_1 and u_2 . We shall prove that an inequality exists so that this set of equations is not satisfied. We show that

$$\frac{w_1}{\alpha} < 2\pi\sqrt{2} - \frac{2^{\frac{1}{2}} u_1^{\frac{1}{2}}}{\alpha^{\frac{1}{2}}} \left(2 - \frac{u_1^{\frac{1}{2}}}{\alpha} \right) - 2^{\frac{3}{2}} \sin^{-1} \frac{u_1^{\frac{1}{2}}}{(2\alpha)^{\frac{1}{2}}} \quad (13)$$

where eq (12a) was used for w_2 on the left hand side with w_1 given by eq (10) and eq (11) with eq (12b) for u_2 was used for the right hand side of inequality (13). We note that $w_1 = w_0(u_1)$, and find

$$\frac{w_0(u_1)}{\alpha} + w_0\left(\frac{u_1}{2}\right) < 2\pi\sqrt{2} \quad (14)$$

must hold. Now any real u must be less than 4. For w_0 with argument less than 4, $w_0 < \pi\sqrt{2}$. Hence, $w_0\left(\frac{u_1}{2}\right) < \pi\sqrt{2}$. Next we show

$$w_1(u_1) < \alpha w_1\left(\frac{u_1}{2}\right) < \pi\sqrt{2}. \quad (15)$$

If this is true, then the inequality (14) follows. To show (15), we use a series expansion for $-2^{\frac{1}{2}} u^{\frac{1}{2}} \left(2 - u^{\frac{1}{2}} \right)^{\frac{1}{2}}$ and $2^{\frac{3}{2}} \sin^{-1} \frac{u^{\frac{1}{2}}}{2^{\frac{1}{2}}}$ from tables. We find for (5) if $T_1 = \frac{u_1}{2^{\frac{1}{2}}}$

$$\left[\frac{2 T_1^3}{3} + \frac{1.3.8 T_1^5}{2.4.5} + \frac{1.3.5.12 T_1^7}{2.4.6.7} + \dots \right] < \left[\frac{\alpha T_1^3}{3\alpha^{\frac{1}{2}}} + \frac{1.3.8 T_1^5}{2.4.5\alpha^{\frac{3}{2}}} + \frac{1.3.5.12 T_1^7}{2.4.6.7\alpha^{\frac{5}{2}}} + \dots \right] \quad (16)$$

holds since $\alpha < 1$ is required for w_2 larger than w_1 . Since the first two terms of the inequality in (15) are shown in (16) and the third is known, we have shown (14). The proof can be repeated for $w_2 > 2\pi\sqrt{2}$ and one finds that an inequality holds also since the left hand side is at most that found in (14) and the right hand side is equal to or larger than that found in (14).

On the other hand, if the point (w_1, u_1) is chosen so that $\pi\sqrt{2} < w_1 < 2\pi\sqrt{2}$, then multivalued current densities may arise. The condition for multiple values

of current density is

$$2\pi\sqrt{2} + w_0 \left(\frac{u_1}{\alpha^2} \right) = \frac{2\pi\sqrt{2} - w_0 \left(\frac{u_1}{\alpha} \right)}{\alpha} \quad (17)$$

if (u_2, w_2) lie in $2\pi\sqrt{2} < w_2 < 3\pi\sqrt{2}$. The smallest w_1 for which multiplicity exists occurs when $u_2 = \frac{u_1}{2} = 4$. The ranges of w_1 and w_2 restrict α and then eq (17) becomes for the smallest w_1

$$\frac{3}{2}\alpha = 1 - \frac{1}{2\pi\sqrt{2}} w_0 (4\alpha^2) \quad (18)$$

which is a transcendental equation. One can also plot the quantities $\frac{2u}{w^2}$ and $\frac{1}{w}$. The potential is proportional to the first quantity and the current density to the second quantity where the new dimensionless potential no longer has the current density j present. In Figure 2 we have plotted $\frac{2u}{w^2}$ and $\frac{1}{w}$ where the current density is multivalued. This occurs graphically for smallest $w_1 \approx 7.75$, $u_1 \approx 1.5$.

The multivalued current density is associated with negative resistance as can be seen from the negative slopes in Figure 2. The negative resistance may give rise to spatial or temporal instabilities. We first analyze possible temporal instabilities.

4.2.2 Temporal Instabilities

The steady-state solution discussed for the electrostatic potential in ballistic transport may fail to exist because the initial state does not evolve into a steady-state solution or because various instabilities occur. We consider the latter possibility here. The approach for temporal instabilities used was first applied by Rowlands⁵ for cold plasmas and extended by Infeld⁶. The mobile carriers considered in the derivation of the ballistic transport current-voltage

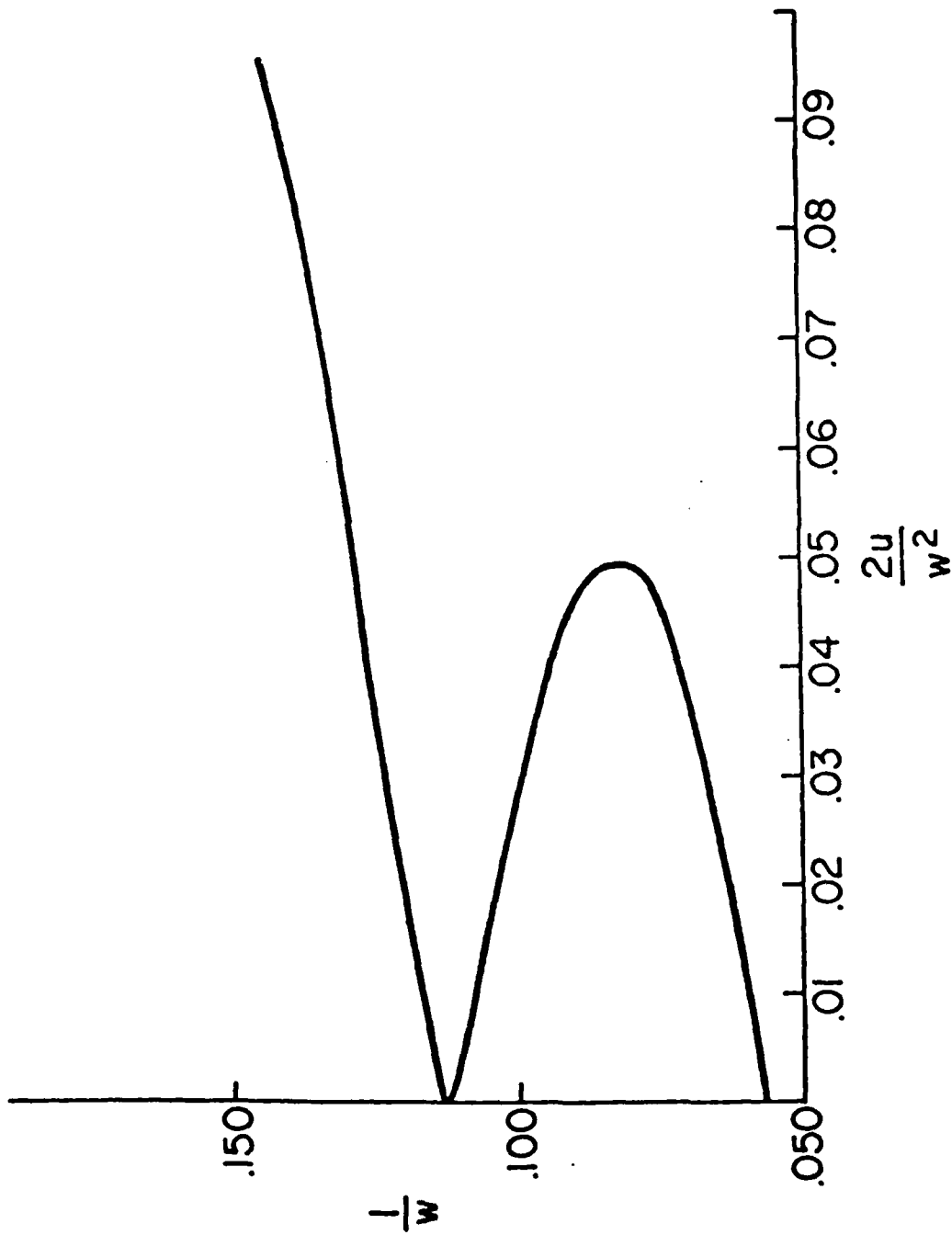


Figure 2. Normalized current density $\frac{1}{w}$ versus normalized potential $\frac{2u}{w^2}$

characteristic are treated there in the cold plasma or beam approximation. The stability analysis for finite temperature carriers is considerably more complicated.

First, the one-dimensional stability analysis where all dependent variables vary with x is considered. The electron carriers are modeled as a beam or they are equivalent together with the doping background to a cold plasma.

The fluid equations for the electrons are

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x} (nv) = 0, \quad (19a)$$

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \frac{e}{m} \frac{\partial u}{\partial x} = \frac{-eE}{m}, \quad (19b)$$

$$\frac{\partial^2 U}{\partial x^2} = \frac{-\partial E}{\partial x} = - \frac{(n_0 - n)}{\epsilon} e. \quad (19c)$$

We assume a steady-state solution exists with fluctuations about the steady state. The fluctuating variables are functions of position x and time t . Each dependent variable is the sum of a steady-state term plus a fluctuating term where the latter is assumed to be a small perturbation initially about the steady-state term.

The perturbation expansion is

$$\begin{aligned} n(x,t) &= \bar{n}(x) + \delta n(x,t) \\ v(x,t) &= \bar{v}(x) + \delta v(x,t) \\ U(x,t) &= \bar{U}(x) + \delta U(x,t) \\ E(x,t) &= \bar{E}(x) + \delta E(x,t). \end{aligned} \quad (20)$$

The steady state equations are

$$\frac{\partial}{\partial x} (\bar{n} \bar{v}) = 0 \quad (21a)$$

$$\bar{v} \frac{\partial \bar{v}}{\partial x} = \frac{e}{m} \frac{\partial U}{\partial x} = -\frac{e\bar{E}}{m}, \quad (21b)$$

$$\frac{\partial^2 U}{\partial x^2} = -\frac{\partial \bar{E}}{\partial x} = -\frac{(n_0 - \bar{n})e}{\epsilon}. \quad (21c)$$

Eq (21a) for density is equivalent to the continuity of current relation, eq (2). The momentum flux equation (21b) integrates to the conservation of energy equation (3) and eq (21c) is just eq (1).

The equations for the perturbed variables are

$$\frac{\partial \delta n}{\partial t} + \frac{\partial}{\partial x} (\bar{n} \delta v) + \frac{\partial}{\partial x} (\bar{v} \delta n) = 0, \quad (22a)$$

$$\frac{\partial \delta v}{\partial t} + \bar{v} \frac{\partial \delta v}{\partial x} + \delta v \frac{\partial \bar{v}}{\partial x} = \frac{e}{m} \frac{\partial \delta U}{\partial x} \quad (22b)$$

$$\frac{\partial^2 \delta U}{\partial x^2} = -\frac{\delta n e}{\epsilon}. \quad (22c)$$

The stability of this system is tested by assuming a harmonic dependence in time $\exp(-j\omega t)$ and by solving for ω . If the imaginary part of ω is positive, the system is unstable. As the steady state is a function of x , the quantities $\bar{n}$, $\bar{v}$ vary with x . Therefore, the usual expansion in a Fourier series in x of the perturbed variables will not do for this problem. A variable change is introduced that renders the equations more tractable. New perturbed quantities are defined that enable us to pass to a new variable τ , the trajectory time for a fluid particle, that replaces x and removes any explicit dependence on x in the equations. Then one can expand the equations in a Fourier series in τ .

The perturbed variables are

$$\begin{aligned} \delta \beta &\equiv \bar{n} \delta v + \bar{v} \delta n, \\ \delta \eta &\equiv \bar{v} \delta v. \end{aligned} \quad (23)$$

To pass to these variables multiply eq (22a) by $\bar{v}$ and eq (22b) by $\bar{n}$ and add the results together. Next change to the variables in eq (23); multiply eq (22b) by $\bar{v}$ and change variables. The two resultant equations are

$$\frac{\partial}{\partial t} \delta\beta + \bar{v} \frac{\partial \delta\beta}{\partial x} + \bar{n} \frac{\partial \eta}{\partial x} = \frac{-e}{m} \bar{n} \delta E, \quad (24a)$$

and

$$\frac{\partial \delta\eta}{\partial t} + \bar{v} \frac{\partial}{\partial x} \delta\eta = \frac{-e}{m} \bar{v} \delta E. \quad (24b)$$

Then multiply eq (24a) by $\bar{v}/\bar{n}$ and subtract the result from eq (24b). Take the x-derivative of this equation, multiply by $\bar{n}$ and add this to the time derivative of eq (24a). Finally multiply by $\bar{v}$ and we find

$$\bar{v} \frac{\partial^2 \delta\beta}{\partial t^2} + \bar{v}^2 \frac{\partial^2 \delta\beta}{\partial x \partial t} + \frac{\partial^2}{\partial t} \left(\frac{\bar{v}^2}{2} \delta\beta \right) + \frac{\partial}{\partial x} \left(\frac{\bar{v}^3}{2} \frac{\partial \delta\beta}{\partial x} \right) = \frac{-e}{m} n_0 v_0 \frac{\partial \delta E}{\partial t}. \quad (25)$$

We next alter eq (25) by first noting the result from Maxwell's equations

$$\frac{\partial \delta E}{\partial t} = \frac{\delta j}{\epsilon} = \frac{e \delta\beta}{\epsilon}. \quad (26)$$

Secondly, one term is replaced in eq (25) where

$$\frac{\partial}{\partial x} \frac{\bar{v}^3}{2} \frac{\partial}{\partial x} \delta\beta = \bar{v} \frac{\partial \bar{v}}{\partial x} \frac{\partial}{\partial x} (\bar{v} \delta\beta) - (\bar{v} \delta\beta) \frac{\partial^2}{\partial x^2} \left(\frac{\bar{v}^2}{2} \right) \quad (27)$$

$$\text{where } \frac{\partial^2}{\partial x^2} \frac{\bar{v}^2}{2} = \frac{\partial^2}{\partial x^2} \left(\frac{e\bar{U}}{m} \right) = \frac{e^2}{m \epsilon} (n_0 - \bar{n}) \quad (28)$$

by eqs (21b)-(21c). Then eq (25) becomes

$$\frac{\partial^2}{\partial \tau^2} (\bar{v} \delta\beta) + 2 \frac{\partial^2}{\partial t \partial \tau} (\bar{v} \delta\beta) + \frac{\partial^2}{\partial t^2} (\bar{v} \delta\beta) + \omega_{p0}^2 (\bar{v} \delta\beta) = 0 \quad (29)$$

where $\bar{v} \frac{\partial}{\partial x} = \frac{\partial}{\partial \tau}$ and

$$\tau = \int_0^x \frac{dx'}{\bar{v}(x')}, \quad T_L = \tau(L) \text{ and}$$

$$\omega_{p0}^2 = \frac{e^2 n_0}{m \epsilon_0} \text{ is the plasma frequency squared.}$$

Also

$$\bar{v}^2 \frac{\partial}{\partial x \partial t} \delta\beta = 2 \bar{v} \frac{\partial}{\partial x \partial t} (\bar{v} \delta\beta).$$

The eq (29) is a linear partial differential equation if $\bar{v} \delta\beta$ is the dependent variable. A Fourier expansion in τ is possible of $\bar{v} \delta\beta$.

$$\bar{v} \delta\beta = \sum_n B_n \exp \left[\frac{j 2\pi n \tau}{T_L} \right]. \quad (30)$$

If we employ the harmonic dependence in time, eq (29) simplifies and the condition for nontrivial solutions is

$$\left(\omega - \frac{2\pi n}{T_L} \right)^2 = \omega_{p0}^2 \quad (31)$$

From this relation for the frequency ω we see that it is real or the steady-state solution is marginally stable if fluctuations that vary in x are allowed. This result was also found by Infeld for a traveling wave in an electron-ion plasma if terms of order of the electron mass to the ion mass are neglected.

Infeld and Rowlands⁷ considered transverse spatial variations in the fluctuating quantities and found, in the absence of particle trapping, that the system is again marginally stable. As a result we exclude any simple temporal instabilities.

4.2.3 Negative Differential Resistivity and Spatial Instabilities

The current-voltage curve for ballistic transport shows multivalued current densities for some values of potential (d.c. voltage). The slope of the current-voltage curve then may be negative for low values of potential and current density. See Figure 2 for the range of negative resistance.

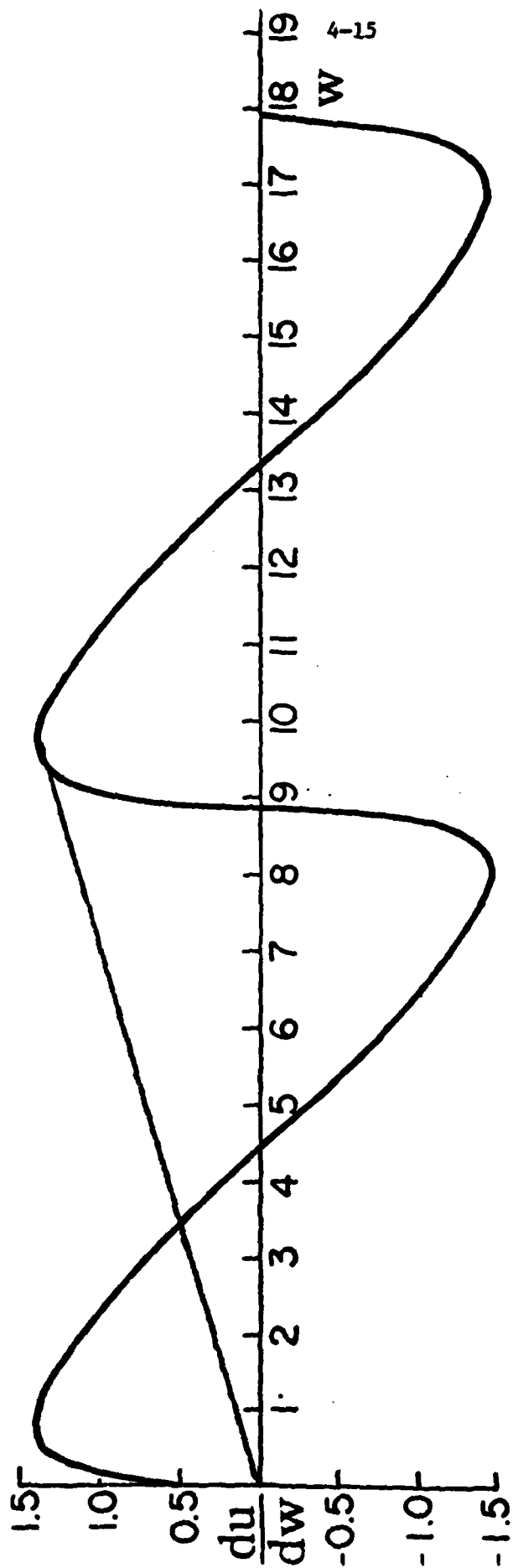


Figure 3. Dimensionless electric field $\frac{du}{dw}$ versus dimensionless length w .

Ridley⁸ has discussed the effect of differential negative resistance and the instabilities associated with it. The current density-electric field relation is fundamental in his discussion. If different values of the electric field occur for a fixed current density, the differential negative resistance is called voltage-controlled; if different values of the current density occur for a fixed electric field, the differential negative resistance is called current-controlled. For the ballistic transport in semiconductors the curve for the current density versus the electric field is plotted in Figure 4. We see that the differential negative resistance is current-controlled.

The range of multivalued current density as a function of the electric field need not be the same as the range of multivalued current density as a function of the electrostatic potential. If for the ballistic transport case the potential is fixed and two values of the current taken, the values of the electric field will, in general, be different. We cannot rely on the current-voltage curves for the current density-electric field multivalued cases.

The electric field is $E = -\frac{dU}{dx}$ for the magnitude of the one-dimensional electric field intensity. If we consider the dimensionless electric field

$$\frac{du}{dw} = \frac{x_0}{U_0} \frac{dU}{dx} \quad (32)$$

where x_0 and U_0 were defined above eq (6). A plot of $\frac{du}{dw}$ versus w is given in Figure 3. The value of $\frac{dU}{dx}$ is next assumed fixed and we vary j in x_0 and U_0 . Then we have

$$\frac{du}{dw} (w_1) / \frac{du}{dw} (w_2) = \frac{1}{\alpha} = \frac{w_1}{w_2} = \frac{E_1}{E_2} \quad (33)$$

where we have introduced E_1 and E_2 as the dimensionless electric field. Next we keep $\frac{dU}{dx}$, $x = L$ and the other semiconductor parameters fixed except for the

current density. Then a straight line is drawn through the points w_1, E_1 and w_2, E_2 . This gives

$$\frac{E_2}{E_1} = \frac{aw_2 + b}{aw_1 + b} = \frac{w_2}{w_1} \quad (34)$$

or $b = 0$. Hence, a straight line drawn through the origin and one point of the $w - E$ curve intersects any other points that have the same electric field. In Figure 3 such a line is drawn such that the smallest w for multiple values of current occur. An analytic condition is possible also. The smallest value of w , called w_1 here, is less than $\pi\sqrt{2}$. Hence, the multivalued current density occurs for smaller w or larger current density for the electric field than for the potential. This result can also be seen in the plot in Figure 4. Here a dimensionless electric field for fixed $x = L$ and independent of current density is plotted as the abscissa $\frac{du}{dw}/w$; the dimensionless current density is plotted as the ordinate $1/w$.

For current-controlled negative resistance filament formation is predicted. These can form in some cases a new stable state⁹ but the spatial instability that leads to filament formation destroys the one-dimensional approximation. Spatial inhomogeneity in the transverse direction exists. In addition, if the negative differential resistance is placed in an external circuit, stable filament formation is difficult to obtain in practice, as Ridley pointed out.

In conclusion, if a cold plasma or beam approximation for the electrons is employed, the regime of likely device operation for ballistic transport seems to be for $w < \pi\sqrt{2}$, $u < 4$ and the smaller voltages and current densities (larger u and w) are probably unobtainable.

4.2.4 Finite Temperature Effects in Ballistic Transport

The effects of finite temperature on ballistic transport could be important

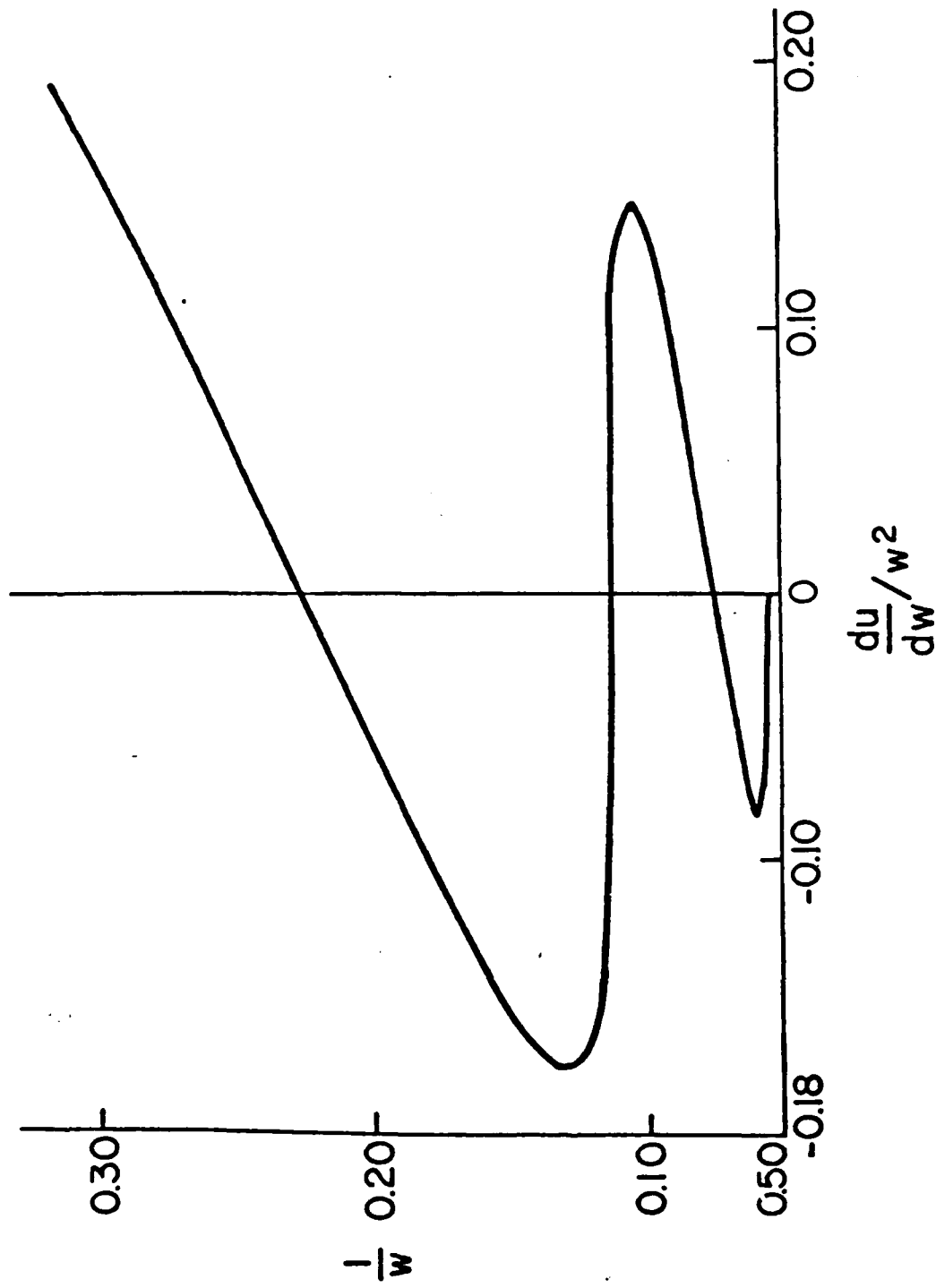


Figure 4. Normalized current density $\frac{1}{w}$ versus normalized electric field $\frac{du}{dw^2}$.

as device voltages are reduced. The model just discussed can be corrected and the infinite carrier density at the cathode removed. Temporal instabilities are affected by temperature but here since the cold system is stable, one expects the warm system to be more stable. The work on the finite temperature corrections has just begun. A preliminary calculation has been made where the usual ballistic approximations are modified by assuming that the electrons leave the cathode with finite energy. A calculation with a one-particle distribution function would give a more fundamental picture, but for low temperature probably does not affect the results much.

The derivation of the finite temperature correction for initial beam energy proceeds as before. We replace eq (3) by

$$\frac{1}{2} m^* v^2 - eU = w = \frac{1}{2} m^* v_0^2 \quad (35)$$

Poisson's equation becomes

$$\frac{d^2 U}{dx^2} = \frac{-e}{\epsilon} \left[n_0 - \frac{j}{e \left(\frac{2eU}{m^*} + v_0^2 \right)^{1/2}} \right] \quad (36)$$

In dimensionless coordinates we have

$$\frac{d^2 u}{dw^2} = \left(u + \frac{v_0^2}{2} \right)^{1/2} - 1, \quad j = e n_0 v_0 \quad (37)$$

where v_0 is the initial carrier speed at the cathode and v_0 the effective carrier speed if the current density were to have the doping density n_0 . The only difference between eq (37) and eq (6) is the replacement of $u^{1/2}$ by $\left(u + \frac{v_0^2}{2} \right)^{1/2}$.

We therefore can integrate by a translation of u and find for $0 \leq w \leq \pi\sqrt{2}$ that

$$w = - \left[4 \left(u + \frac{v_0^2}{v_0^2} \right)^{\frac{1}{2}} - 2u - \frac{4v_0}{v_0} \right]^{\frac{1}{2}} + \frac{\sqrt{2}\pi}{2} - \sqrt{2} \sin^{-1} \left[\frac{1 - \left(u + \frac{v_0^2}{v_0^2} \right)^{\frac{1}{2}}}{1 - \frac{v_0}{v_0}} \right], \quad \frac{v_0}{v_0} < 1. \quad (38)$$

For small $\frac{v_0}{v_0}$, the likely case, the value of w is not altered much. The solution does not exist for $\frac{v_0}{v_0} = 1$.

4.2.5 The Effect of Non-Parabolic Bands (M. Muller)

In a device that operates in a collision-free, or nearly collision-free regime, the carriers that contribute to the device current reach energies corresponding to the voltages applied to the device. In small and fast devices, these are of the order of one volt. In most semiconductors, especially those with small effective mass carriers, the bands depart from parabolicity at energies well below 1eV. The effective mass generally increases with energy. Thus the current is smaller than one would predict for a parabolic band structure.

To assess the magnitude of this effect, we use the conduction band structure of a model narrow-gap semiconductor as computed by two-band $\bar{k} \cdot \bar{p}$ perturbation theory¹⁰⁾. Such a model can provide a fair approximation to the real band structure in the vicinity of the conduction band extremum.

Assuming a direct-gap semiconductor, and placing the zero of energy at the conduction band minimum, the theory gives the hyperbolic band structure

$$\epsilon = \frac{1}{2} \epsilon_g \left[\sqrt{1 + \frac{2\hbar^2 k^2}{m^* \epsilon_g}} - 1 \right] \quad (39)$$

The theory also provides a relation between the effective mass m^* and the bandgap ϵ_g , but for the present purpose these two parameters may be chosen independently, for instance to fit a measured band structure.

To obtain the carrier velocity as a function of energy, we use

$$\frac{\partial \epsilon}{\partial k} = \frac{\hbar^2 k^2}{m^*} \left(1 + \frac{2\hbar^2 k^2}{m^* \epsilon_g} \right)^{-1/2} \quad (40)$$

and then solve eq. (1) for k^2

$$k^2 = \frac{2m^*}{\hbar^2} \left(\epsilon + \frac{\epsilon_g^2}{4\epsilon_g} \right) \quad (41)$$

Substituting this in eq. (40) gives

$$\begin{aligned}
 \frac{\partial \epsilon}{\partial k} &= \sqrt{\frac{2}{m^*}} \frac{h(\epsilon + \epsilon_g^2/\epsilon_g)^{1/2}}{1 + 2\epsilon/\epsilon_g} \\
 &= 2h v_s \frac{\frac{\epsilon}{\epsilon_g} + \frac{\epsilon_g^2}{\epsilon_g^2}}{1 + 2\epsilon/\epsilon_g} \\
 &= h v_s \frac{(2m^* v_s^2 \epsilon + \epsilon_g^2)}{m^* v_s^2 \epsilon + \epsilon_g^2}
 \end{aligned} \tag{42}$$

where we have defined a band-structure limited velocity

$$v_s \equiv (\epsilon_g/2m^*)^{1/2} \tag{43}$$

Figure 5 shows the current best plot of the GaAs band structure¹¹⁾ with both eq. (39) using the experimental ϵ_g and m^* , and a $\hbar^2 k^2/2m^*$ parabola superposed on the plot. The hyperbola of eq. (39) is evidently a much better fit.

The Poisson equation for a semiconductor uniformly doped with n_0 donors/cm³ carrying an electron current density J is

$$\frac{d^2 V}{dx^2} = \frac{\rho}{\epsilon} = -\frac{qn_0}{\epsilon} + \frac{J}{\epsilon v} \tag{44}$$

where ϵ is the dielectric constant and v is the electron velocity.

Electrons injected at $x = 0$ acquire a velocity

$$v = \frac{1}{\hbar} \frac{\partial \epsilon}{\partial k} \tag{45}$$

when their energy $\epsilon = qV \equiv \phi$.

Using the band structure gradient from eq. (42) brings the Poisson equation into the form

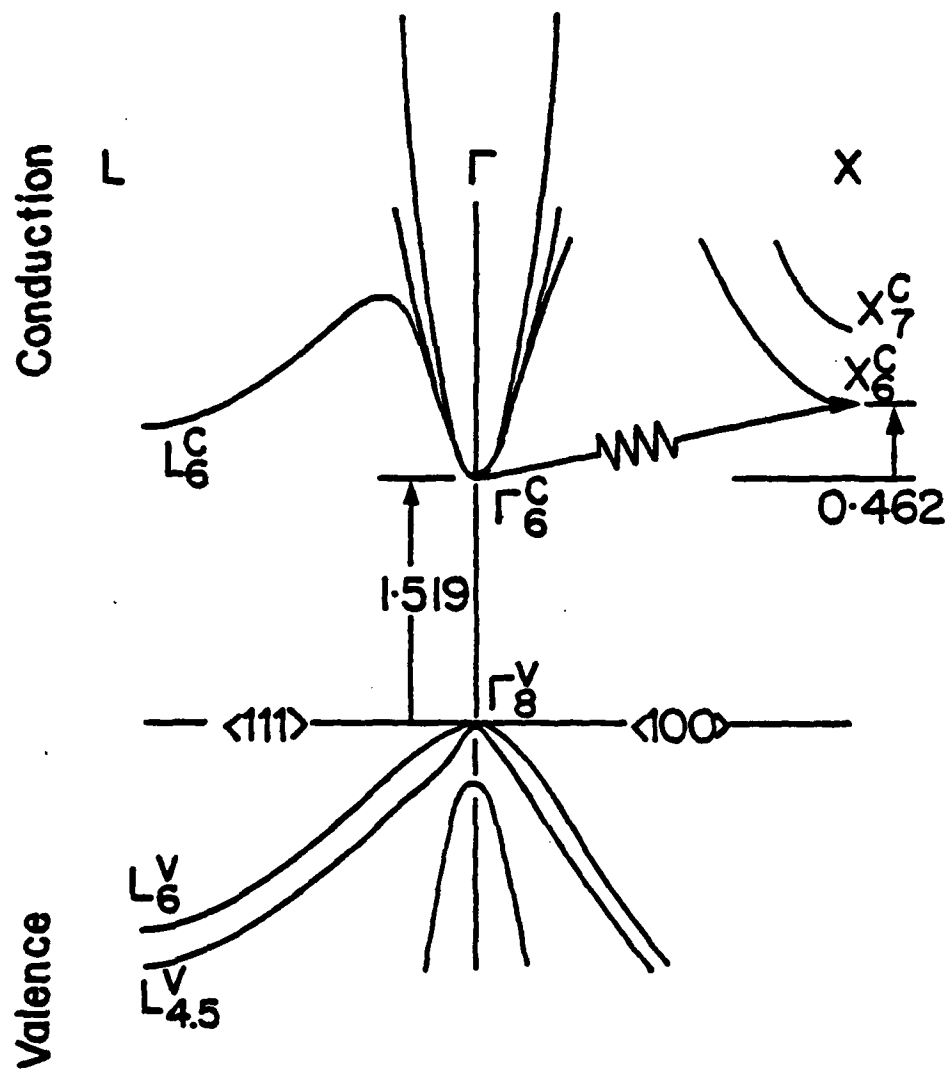


Figure 5. GaAs band structure with parabolic and hyperbolic conduction band approximation

$$\phi'' = -\frac{q^2 n_0}{\epsilon} + \frac{Jq}{\epsilon v_s} \frac{(m^* v_s^2 + \phi)}{(2m^* v_s^2 + \phi)^{1/2}} \quad (46)$$

where a prime denotes d/dx . Eq. (46) can be integrated once to give

$$(\phi')^2 - (\phi'(0))^2 = \frac{2Jq}{\epsilon v_s} (\phi^2 + 2m^* v_s^2 \phi)^{1/2} - \frac{2q^2 n_0}{\epsilon} \phi \quad (47)$$

Assuming space charge limitation $\phi'(0) = 0$, and changing to reduced variables

$$u \equiv \frac{\phi}{2m^* v_s^2}, \quad j = \frac{J}{qn_0 v_s}, \quad \lambda = \frac{x}{(\epsilon m^* / n_0 q^2)^{1/2} v_s} \quad (48)$$

we can write the diode equation, the integral of eq. (47), in the form

$$\int_0^u du [j(u^2 + u)^{1/2} - u]^{-1/2} = \lambda \quad (49)$$

where $u = v$ ($V = V_0$, the applied voltage) at $\lambda = \lambda$ ($x = L$, the diode length).

The normalized variables are:

Diode voltage in units of $2m^* v_s^2 / q$ —four times the kinetic energy the carriers would have if their effective mass were constant at the band edge value, and they were moving with the limiting velocity v_s ; if the $\bar{k} \cdot \bar{p}$ perturbation theory is taken literally, this unit is equal to ϵ_g / q ;
Current density in units of $qn_0 v_s$ —the current that would flow if the mobile carriers from the doping were moving with velocity v_s ;

Diode length in units of $v_s \tau_p$, with τ_p the plasma oscillation period of the semiconductor.

Curves of current density versus voltage are shown in Figure 6 as full curves. For comparison the current density-voltage curves one would compute assuming constant effective mass are shown dashed.

We have not extended the diode characteristics into the low-current region at the lower ends of the curves. The problems encountered by modeling this

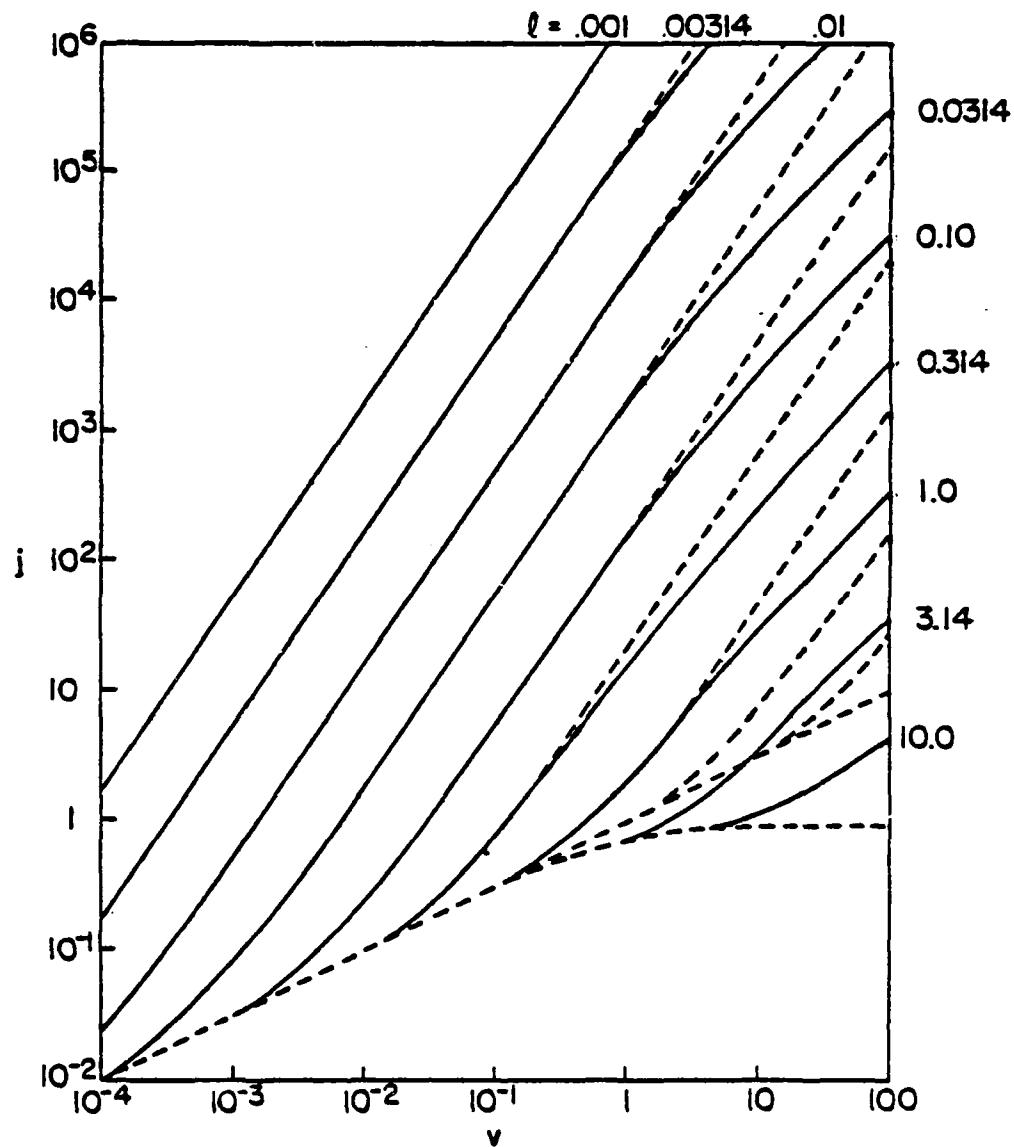


Figure 6. Space-charge limited current-voltage relation of a collisionless semiconductor

AD-A098 578

WASHINGTON UNIV ST LOUIS MO DEPT OF ELECTRICAL ENGI--ETC F/G 20/12
SEMICONDUCTOR MILLIMETER WAVELENGTH ELECTRONICS. (U)

OCT 80

N00014-79-C-0840

UNCLASSIFIED

80-1-ONR

NL

3-3

AL-ONR '80

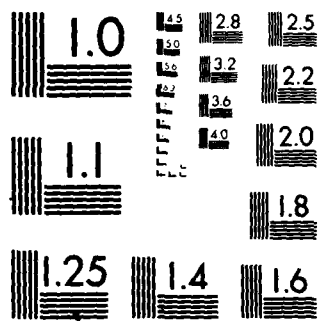
END

DATE

FILED

8-88

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

operating regime are discussed in section 4.2.

The vacuum diode analog $3/2$ power law holds only for $V_0 \leq \epsilon_g$, and then only for $l \leq 1$. That means very short diodes--for GaAs with $n_0 = 10^{16}$, shorter than $0.25\mu\text{m}$. The limiting higher voltage behavior is $J \propto V$ rather than $J \propto V^{3/2}$; the J-V curves bend over.

In the typical size and operating voltage range of GaAs devices the corrections indicated here appear significant but not dramatic. They are likely to be rather more important in devices made from smaller bandgap, lower m^* material.

References

1. M. S. Shur and L. F. Eastman, IEEE Trans. El. Dev., ED-26, 1677 (1979); IEEE El. Dev. Lett. EDL-1, 147 (1980).
2. M. S. Shur, "Ballistic Transport in a Semiconductor with Collisions", (preprint)
3. C. D. Child, "Discharge from Hot CAO", Phys. Rev 32, 492 (1911).
4. I. Langmuir, "The Effect of Space Charge and Residual Gases on Thermionic Currents in High Vacuum", Phys. Rev 2, 450 (1913).
5. G. Rowlands, "Stability of non-linear plasma waves", J. Plasma Physics 3, 567 (1969).
6. E. Infeld, "On the stability of nonlinear cold plasma waves", J. Plasma Physics 8, 105 (1972).
7. E. Infeld and G. Rowlands, "On the stability of nonlinear cold plasma waves", J. Plasma Phys. 13, 173 (1975).
8. B. K. Ridley, "Specific Negative Resistance in Solids", Proc, Phys. Soc. 82. 954 (1963).
9. M. W. Muller and H. Guckel, "Negative Resistance and Filamentary Currents in Avalanching Silicon $P^+ - I - N^+$ Junctions", IEEE Transactions on Electron Devices, 15, 560 (1968).
10. E. O. Kane, in Semiconductor and Semimetals Vol. I, R. K. Willardson and A. C. Beer, eds, Academic Press, New York 1966.
11. D. E. Aspnes, in Proc. 6th Int's Symp. GaAs and Rel-Comp., L. F. Eastman, ed., Inst. Phys. London 1977.

4.3 Zener Oscillations

M. Muller

4.3.1 Introduction

C. Zener,¹⁾ more than forty years ago, pointed out a curious feature of the dynamics of band electrons which could, in principle, lead to a high-frequency oscillation.

In the absence of any other interaction, the motion of a crystal electron in a steady electric field follows the law

$$\hbar \dot{\bar{k}} = q \bar{E}, \quad \bar{k} = \frac{q \bar{E}}{\hbar} t,$$

That is to say, the electron describes a rectilinear trajectory in the extended Brillouin zone of $\bar{k}$ - space. Since the energy ϵ is a periodic function of $\bar{k}$, and the velocity $\bar{v}$ is

$$\bar{v} = \frac{1}{\hbar} \nabla_{\bar{k}} \epsilon,$$

the trajectory of an electron wave packet is oscillatory. The period of the oscillation is the time required to traverse the reduced Brillouin zone

$$\dot{\bar{k}} \tau = \frac{2\pi}{a}, \quad \tau = \frac{2\pi\hbar}{aqE} = \frac{4.1 \times 10^{-7}}{aE}$$

where a is the lattice parameter in Å, E the field in $v \text{ cm}^{-1}$ and the amplitude of the oscillation is

$$\ell_{\text{osc}} = \int_{\text{half cycle}} v dt = \frac{1}{\hbar} \int \nabla_{\bar{k}} \epsilon dt = \frac{1}{\hbar \dot{\bar{k}}} \int_{\text{band}} \nabla_{\bar{k}} \epsilon d\bar{k} = \frac{\Delta}{qE}$$

*We assume for simplicity that the field is applied along a principal lattice vector.

where Δ is the width of the energy band.

This elementary description of the phenomenon is quoted from the proposal for the present project. We also commented that Zener recognized at the time that competing phenomena would rule out the occurrence of such oscillations, and that there had not been any report of an unambiguous experimental verification. It appeared possible to us that progress in materials technology had reached a point that would justify a more detailed study of the phenomenon.

The eventual objective of studying Zener oscillations is to utilize them in a device for the generation of submillimeter radiation. The manifest advantage of such a device would be its ability to use a uniform steady electric field as a source of power, and its direct voltage tunability. The form that a device would take is not yet clear; our effort at this point is focused on the feasibility of observing the central phenomenon.

Zener oscillations are a special form of charge transport in crystals. The reason why they have not been observed is generally believed to be carrier scattering. (It has been argued that Zener oscillations might not be observable even in the absence of scattering; we will comment on this controversy at length below.) Very recently collisionless carrier transport, called "ballistic" or "inertial" transport, has attracted a good deal of attention²⁾. It is believed that "conventional" devices, such as short-channel FET's, may be capable of superior performance if the carrier transit times are made short compared with the mean time between collisions. In consequence of taking a viewpoint suggested by the study of Zener oscillations we have been able to contribute to the theory of ballistic transport in thin layers of semiconductor. The

contribution appears elsewhere in this report. Here we shall discuss those topics bearing directly on the oscillations:

The range of validity of the effective mass approximation

Tunneling, interband transitions, and the Stark ladder

Scattering

This last topic is at the heart of transport theory. In principle the theoretical apparatus for a description of high-field transport is in existence, but its application to device electronics is not yet a reality^{2a)}. We should like to devote some serious work to this area, but we are not in a position to begin such an effort within the scope of this project. We will look at scattering phenomenologically; such a viewpoint is probably adequate for our purpose. It does appear appropriate, however, to discuss the first two topics in some detail. Much of this discussion is tutorial, but surprisingly some aspects of the dynamics of crystal electrons are still controversial.

4.3.2 The Effective Mass Theorem

We can start by looking more carefully at the equations on p. 4-28. They are a curious mixture of classical and quantum dynamics. The classical view is implied in the integration $x = \int v dt$; the quantum mechanics appears in the wave vector of the electron and in the band structure.

The approximation on which the equations are based is to regard the crystal electron (or hole) as a particle with a (tensor) mass

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 \epsilon}{\partial k_i \partial k_j}.$$

This implies that the motion of the particle is described by a Schrödinger equation giving, for example, hydrogenic states for a coulombic defect, or as another example, scattering that can be calculated by perturbation theory; and, through the correspondence principle, classical behavior for suitably constructed wave packets, with a velocity

$$v_i = \frac{1}{\hbar} \frac{\partial \epsilon}{\partial k_i}$$

and dynamics

$$\hbar \dot{\vec{k}} = q(\vec{E} + \frac{1}{c} \vec{v} \times \vec{H}).$$

One expects this classical dynamics to describe, for example, cyclotron resonance, and acceleration and mobility in slowly varying or static electric fields. Thus the usual theory of mobility starts with

$$\begin{aligned} \vec{r} &= \int_0^t \vec{v} dt = \frac{1}{\hbar} \int_0^t \frac{\partial \epsilon}{\partial \vec{k}} dt = \frac{1}{\hbar \vec{k}} \int_0^t \frac{\partial \epsilon}{\partial \vec{k}} \cdot \frac{d\vec{k}}{dt} dt \\ &= \frac{1}{qE} \int_{\vec{k}_0}^{\vec{k}} \frac{\partial \epsilon}{\partial \vec{k}} \cdot d\vec{k} = \frac{1}{qE} \left[\epsilon(\vec{k}_0 + \frac{q}{\hbar} \vec{E} t) - \epsilon(\vec{k}_0) \right]. \end{aligned}$$

If the electron starts at a band extremum, where

$$\epsilon(\bar{k}-\bar{k}_0) = \frac{\hbar^2(\bar{k}-\bar{k}_0)^2}{2m^*}, \quad \text{then}$$

$$\bar{r}(t) = \frac{\hbar^2 \left(\frac{q}{\hbar} Et\right)^2}{2m^*qE} = \frac{qE}{2m^*} t^2.$$

This is the motion until the carrier is scattered--the inertial or ballistic motion. The usual

$$\bar{v} = \mu E = \frac{q}{m^*} \langle \tau \rangle$$

results from averaging over scattering lifetimes when $\langle \tau \rangle$ is much shorter than any other time interval of interest.

When the scattering lifetime is long, the inertial motion continues, but the carrier reaches energies where the energy must be described by the full band structure rather than the parabolic approximation valid near an extremum; the resulting motion is the Zener oscillation described on p. 4-28.

To assess the validity of this description, it is necessary to examine the theory to see where approximations are made, and how they might be expected to fail.

The starting point is the one-electron approximation. This discards questions of correlation, plasma oscillations and other effects important at high carrier densities.

The one-electron wave functions of the infinite perfect crystal are generated by the periodic potential Hamiltonian

$$H_0 \psi_{nk} = \epsilon_n(\bar{k}) \psi_{nk}$$

with $H_0 = -\frac{\hbar^2}{2m} \nabla^2 + V(\vec{r})$

$$V(\vec{r}+\vec{\ell}) = V(\vec{r})$$

where ℓ is any lattice vector, and they are of the Bloch form

$$\psi_{n\vec{k}} = u_{n\vec{k}} e^{i\vec{k} \cdot \vec{r}}.$$

In these equations, n is the band index.

When there is a field present, the Hamiltonian is $H = H_0 + U$, the Schrödinger equation is

$$(H_0 + U)\psi(\vec{r}, t) = i\hbar \frac{\partial \psi(\vec{r}, t)}{\partial t}$$

and one can attempt to write the wave function as a wave packet of Bloch functions

$$\psi(\vec{r}, t) = \sum_{n, k} a_n(\vec{k}, t) \psi_{n\vec{k}}(\vec{r})$$

Rather than substituting this expansion directly in the Schrödinger equation and solving by perturbation theory, one can take advantage of the periodicity of $\epsilon_n(\vec{k})$ to transform the equation; since we can write

$$\epsilon_n(\vec{k}) = \sum_{\vec{\ell}} \epsilon_{n\vec{\ell}} e^{i\vec{k} \cdot \vec{\ell}}$$

where

$$\epsilon_{n\vec{\ell}} = \frac{1}{N} \sum_{\vec{k}} \epsilon_n(\vec{k}) e^{-i\vec{k} \cdot \vec{\ell}},$$

the energy can be formally represented as a function of the momentum operator by replacing $\bar{k}$ in the Fourier expansion by the operator $-i \text{ grad}$:

$$\epsilon_n(-i\nabla) = \sum_{\bar{l}} \epsilon_{n\bar{l}} e^{\bar{l} \cdot \nabla} = \sum_{\bar{l}} \epsilon_{n\bar{l}} (1 + \bar{l} \cdot \nabla + \frac{1}{2}(\bar{l} \cdot \nabla)^2 + \dots).$$

Note that the operator $\exp(\bar{l} \cdot \nabla)$ translates any function of position by the lattice vector $\bar{l}$:

$$e^{\bar{l} \cdot \nabla} f(\bar{r}) = f(\bar{r} + \bar{l}), \text{ so that operating on a Bloch function}$$

$$\begin{aligned} \epsilon_n(-i\nabla) \psi_{n\bar{k}}(\bar{r}) &= \sum_{\bar{l}} \epsilon_{n\bar{l}} \psi_{n\bar{k}}(\bar{r} + \bar{l}) \\ &= \sum_{\bar{l}} \epsilon_{n\bar{l}} e^{i\bar{k} \cdot \bar{l}} \psi_{n\bar{k}}(\bar{r}) = \epsilon_n(\bar{k}) \psi_{n\bar{k}}(\bar{r}). \end{aligned}$$

We can use this to evaluate part of the Hamiltonian acting on the wave packet.

$$H\psi = \sum_{n\bar{k}} (H_0 + U) a_n(\bar{k}, t) \psi_{n\bar{k}}(\bar{r})$$

$$\sum_{n\bar{k}} [\epsilon_n(\bar{k}) + U] a_n(\bar{k}, t) \psi_{n\bar{k}}(\bar{r})$$

$$\sum_n [\epsilon_n(-i\nabla) + U] \sum_{\bar{k}} a_n(\bar{k}, t) \psi_{n\bar{k}}(\bar{r}) = i \hbar \frac{\partial \psi}{\partial t}$$

This looks almost like a Schrödinger equation for the wave packet not involving the periodic lattice potential, except the kinetic energy operator $\epsilon_n(-i\nabla)$ is not outside the summation over bands.

One obvious way to deal with this difficulty is to assume immediately that the motion of the electron is adequately described by using Bloch functions from one band only, thus

$$\psi(\bar{r}, t) = \sum_{\bar{k}} a(\bar{k}, t) \psi_{n\bar{k}}(\bar{r})$$

giving immediately

$$H\psi = [\epsilon_n(-i\nabla) + U]\psi = i\hbar\dot{\psi}.$$

The identification of $\epsilon_n(-i\nabla) = \epsilon_n(\frac{\bar{p}}{\hbar})$ as the kinetic energy operator then completes the effective mass acceleration theorem, but the approximation discards any possibility of band mixing or band-to-band transitions. Since these effects are of primary interest, we must use better approximations. Actually one can do a little better without the one-band approximation by using Wannier rather than Bloch functions to construct the wave packet. The Wannier functions are (maximally) localized functions defined by

$$\omega_n(\bar{r}-\bar{\ell}) = N^{-1/2} \sum_{\bar{k}} e^{-i\bar{k}\cdot\bar{\ell}} \psi_{n\bar{k}}(\bar{r})$$

so that

$$\psi_{n\bar{k}}(\bar{r}) = N^{-1/2} \sum_{\bar{\ell}} e^{i\bar{k}\cdot\bar{\ell}} \omega_n(\bar{r}-\bar{\ell})$$

Each Wannier function is centered on a lattice site; unlike atomic orbitals, Wannier functions on different sites are orthogonal

$$\begin{aligned} \int \omega_n^*(\bar{r}-\bar{\ell}) \omega_n(\bar{r}-\bar{\ell}') d\bar{r} &= \frac{1}{N} \sum_{\bar{k}, \bar{k}'} e^{i\bar{k}\cdot\bar{\ell}} e^{-i\bar{k}'\cdot\bar{\ell}'} \int \psi_{kn}^* \psi_{k'n} d\bar{r} \\ &= \frac{\delta_{nn'}}{N} \sum_{\bar{k}} e^{i\bar{k}\cdot(\bar{\ell}-\bar{\ell}')} = \delta_{nn'} \delta_{\bar{\ell}\bar{\ell}'} \end{aligned}$$

The wave packet is constructed as

$$\psi(\bar{r}, t) = \sum_{n\bar{\ell}} f_n(\bar{\ell}, t) \omega_n(\bar{r}-\bar{\ell});$$

that is to say, a separate envelope function is defined for each band. From here, the standard method of perturbation theory is followed. The wave packet is inserted in the Schrödinger equation $(H_0 + U)\psi = i\hbar\dot{\psi}$, the equation is multiplied by $\omega_n^*(\bar{r}-\bar{\ell}')$, and integrated over the crystal

$$\begin{aligned} \sum_{n\bar{\ell}} \int d\bar{r} \omega_n^*(\bar{r}-\bar{\ell}') (H_0 + U) \omega_n(\bar{r}-\bar{\ell}) f_n(\bar{\ell}, t) \\ = i\hbar \frac{\partial}{\partial t} f_n(\bar{\ell}', t); \end{aligned}$$

but

$$\begin{aligned} H_0 \omega_n(\bar{r}-\bar{\ell}) &= N^{-1/2} \sum_{\bar{k}} e^{-i\bar{k} \cdot \bar{\ell}} H_0 \psi_{n\bar{k}} \\ &= N^{-1/2} \sum_{\bar{k}} e^{-i\bar{k} \cdot \bar{\ell}} \epsilon_n(\bar{k}) \psi_{n\bar{k}} \\ &= N^{-1} \sum_{\bar{k}} e^{-i\bar{k} \cdot \bar{\ell}} \epsilon_n(\bar{k}) \sum_{\bar{\ell}'} e^{i\bar{k} \cdot \bar{\ell}'} \omega_n(\bar{r}-\bar{\ell}') \\ &= \sum_{\bar{\ell}'} \epsilon_{n, \bar{\ell}-\bar{\ell}'} \omega_n(\bar{r}-\bar{\ell}') \end{aligned}$$

giving

$$\sum_{n\bar{\ell}} [\delta_{nn'} \epsilon_{n, \bar{\ell}-\bar{\ell}'} + U_{nn'}(\bar{\ell}, \bar{\ell}')] f_n(\bar{\ell}, t) = i\hbar \frac{\partial}{\partial t} f_n(\bar{\ell}', t)$$

with

$$U_{nn'}(\bar{\ell}, \bar{\ell}') \equiv \int \omega_n^*(\bar{r}-\bar{\ell}') U(\bar{r}) \omega_n(\bar{r}-\bar{\ell}) d\bar{r}.$$

The envelope function $f_n(\bar{\ell})$ up to now has consisted of a discrete set of numbers, one for each lattice point, but it can also be viewed as a continuous function which, on the lattice points, takes on these numerical values:

$$f_n(\bar{\ell}) = [f_n(\bar{r})]_{\bar{r}=\bar{\ell}}.$$

With this viewpoint we can use

$$\sum_{\bar{\ell}} \epsilon_{n\bar{\ell}} f(\bar{r}+\bar{\ell}) = \epsilon_n (-i\nabla) f(\bar{r})$$

on the first term in the difference equation above:

$$\begin{aligned} \sum_{n\bar{\ell}} \delta_{nn'} \epsilon_{n, \bar{\ell}-\bar{\ell}'} f_n(\bar{\ell}, t) &= \sum_{\bar{\ell}} \epsilon_{n', \bar{\ell}-\bar{\ell}'} f_{n'}(\bar{\ell}, t) \\ &= [\epsilon_{n'} (-i\nabla) f_{n'}(\bar{r}, t)]_{\bar{r}=\bar{\ell}'}, \end{aligned}$$

and substituting this result we obtain

$$\begin{aligned} [\epsilon_n (-i\nabla) f_n(\bar{r}, t)]_{\bar{r}=\bar{\ell}'} + \sum_{n\bar{\ell}} U_{nn'}(\bar{\ell}, \bar{\ell}') f_n(\bar{\ell}, t) \\ = i\hbar \frac{\partial f_{n'}(\bar{\ell}', t)}{\partial t}. \end{aligned}$$

This is a set of coupled Schrödinger equations for an envelope function in each band; and these wave equations do not involve the periodic crystal potential. From this formulation we can see explicitly how band mixing comes about: to have $U_{nn'}$ appreciable with $n \neq n'$, U must be rapidly varying in space (i.e., the field must be large). There is no restriction on the size of the wave packet, as given, say, by the shape of the envelope function f . Furthermore, of course, there will be requirements imposed by energy conservation which are not reflected in the interband matrix element.

Putting band mixing aside for the moment we see that either of the equations we have deduced

$$[\epsilon_n(-i\nabla) + U(\bar{r})]\psi_n = i\hbar \dot{\psi}_n$$

$$[\epsilon_n(-i\nabla) + \sum_{\bar{r}'} U_{nn}(\bar{r}', \bar{r})]f_n = i\hbar \dot{f}_n$$

describes a quasi-particle which is called a crystal electron. Indeed the two equations are nearly identical. If U is slowly varying in space, the matrix element of U for Wannier functions on different sites is very small and one can then say $U_{nn}(\bar{r}', \bar{r}) = U_n(\bar{r})$. The equation constitutes the effective mass theorem for the crystal electron, because it provides an expression for the kinetic energy in terms of the momentum operator $\frac{\bar{p}}{\hbar} = -i\nabla$, and therefore for the Hamiltonian.

The theorem can be extended to allow for magnetic fields. It is found that in the presence of a magnetic field the Hamiltonian is

$$H = \epsilon_n \left[-i\nabla - \frac{q}{\hbar c} \bar{A}(\bar{r}, t) \right] + U$$

where the fields are found from the potentials by

$$\bar{B} = \nabla \times \bar{A}, \quad E = -\frac{1}{c} \frac{\partial \bar{A}}{\partial t} - \frac{1}{q} \nabla U.$$

Since we will not be concerned with magnetic fields, we do not derive this result here; it was first reported by Jones and Zener³⁾.

To exhibit the meaning of the theorem, we apply it to the dynamics of an electron at the bottom of a parabolic conduction band. Here $\epsilon(\bar{k}) = \hbar^2 \bar{k}^2 / 2m^*$, so

$$\epsilon(-i\nabla) = -\frac{\hbar^2}{2m^*} \nabla^2,$$

$$H = -\frac{\hbar^2}{2m} \nabla^2 + U(\bar{r})$$

and the crystal electron acts like a free electron of mass m^* . Then invoking the correspondence principle, we can say that in the proper classical limit the crystal electron will behave like a particle with the corresponding classical Hamiltonian

$$H(\bar{r}, \bar{p}) = \epsilon\left(\frac{\bar{p}}{\hbar}\right) + U(\bar{r}),$$

that is to say, its dynamics will be described by the Hamilton equations

$$\dot{\bar{r}} = \frac{\partial H}{\partial \bar{p}}, \quad \dot{\bar{p}} = -\frac{\partial H}{\partial \bar{r}}$$

These equations are, explicitly

$$\dot{\bar{v}} = \frac{\partial H}{\partial \bar{p}} = \frac{\partial}{\partial \bar{p}} \left\{ \epsilon\left(\frac{\bar{p}}{\hbar}\right) + U(\bar{r}) \right\} = \frac{1}{\hbar} \frac{\partial \epsilon(\bar{k})}{\partial \bar{k}}$$

showing that indeed $\bar{p} = \hbar \bar{k}$ is a "classical" crystal momentum, and

$$\dot{\bar{p}} = \hbar \dot{\bar{k}} = -\frac{\partial H}{\partial \bar{r}} = -\nabla U(\bar{r}) = q\bar{E}$$

where we assume that U is an electrostatic potential.

The classical correspondence holds up only if the classical particle has the same trajectory as the center of the wave packet, and preferably the wave packet should hold together fairly well. Since the Schrödinger equation is dispersive, this requires restricting the range of $\bar{k}$.

The last consideration suggests that for this classical limit it will be more convenient to construct the packet from Bloch functions than from Wannier functions. So let

$$\psi_{n\bar{k}}(\bar{r}) = e^{i\bar{k} \cdot \bar{r}} u_{n\bar{k}}(\bar{r}) \approx e^{i\bar{k} \cdot \bar{r}} u_{n\bar{k}_0}(\bar{r}) = e^{i(\bar{k}-\bar{k}_0) \cdot \bar{r}} \psi_{n\bar{k}_0}(\bar{r})$$

$$\psi = \sum_{\bar{k}} a(\bar{k}, t) e^{i(\bar{k}-\bar{k}_0) \cdot \bar{r}} \psi_{n\bar{k}_0}(\bar{r})$$

that is to say, we assume as will usually be true, that the periodic part of the Bloch function varies little with $\bar{k}$. To justify this approximation, consider the wave function from the LCAO viewpoint: if we remain confined to a small region of the Brillouin Zone, there is no reason why the linear combination of orbitals making up $\psi_{n\bar{k}}$ should change much. If $a(\bar{k}, t)$ is small except for $\bar{k}$ near $\bar{k}_0$, then

$$F(\bar{r}, t) = \sum_{\bar{k}} a(\bar{k}, t) e^{i(\bar{k}-\bar{k}_0) \cdot \bar{r}}$$

contains only long wavelengths and hence varies slowly in space. Therefore $F(\bar{r}, t)$ is nearly constant over a unit cell, and

$$\psi = F(\bar{r}, t) \psi_{n\bar{k}_0}(\bar{r}).$$

To obtain the wave equation we work out

$$\begin{aligned} \epsilon_n(-i\nabla)\psi &= \epsilon_n(-i\nabla)F(\bar{r}, t)\psi_{n\bar{k}_0}(\bar{r}) \\ &= \sum_{\bar{l}} \epsilon_{n\bar{l}} e^{i\bar{l} \cdot \nabla} F(\bar{r}, t) \psi_{n\bar{k}_0}(\bar{r}) \\ &= \sum_{\bar{l}} \epsilon_{n\bar{l}} F(\bar{r} + \bar{l}, t) \psi_{n\bar{k}_0}(\bar{r}) \\ &= \sum_{\bar{l}} \epsilon_{n\bar{l}} \psi_{n\bar{k}_0}(\bar{r}) e^{i\bar{l} \cdot (\bar{k}_0 - i\nabla)} F(\bar{r}, t) \\ &= \psi_{n\bar{k}_0}(\bar{r}) \epsilon_n(\bar{k}_0 - i\nabla) F(\bar{r}, t) \end{aligned}$$

and $H\psi = i\hbar \dot{\psi}$ becomes an effective mass equation for the amplitude F :

$$[\epsilon_n(k_0 - i\nabla) + U(\bar{r})]F(\bar{r}, t) = -i\hbar \dot{F}(\bar{r}, t).$$

Furthermore, since F is slowly varying, ∇F is small and the leading terms of a Taylor expansion of ϵ_n about $\bar{k}_0$ are adequate. It is worth noting that this is valid everywhere in the Brillouin zone, not only at a band extremum:

$$\begin{aligned}\epsilon_n(\bar{k}_0 + \bar{k}) &= \epsilon_n(\bar{k}_0) + k_i \left. \frac{\partial \epsilon_n}{\partial k_i} \right|_{k_0} + \frac{1}{2} k_i k_j \left. \frac{\partial^2 \epsilon_n}{\partial k_i \partial k_j} \right|_{k_0} + \dots \\ &= \epsilon_n(\bar{k}_0) + \hbar \bar{v}(\bar{k}_0) \cdot \bar{k} + \frac{1}{2} \hbar^2 \left(\frac{1}{m^*} \right)_{ij} k_i k_j + \dots\end{aligned}$$

and so

$$\epsilon_n(k_0 - i\nabla) = \epsilon_n(k_0) - i\hbar \bar{v}(\bar{k}_0) \cdot \nabla - \frac{1}{2} \hbar^2 \left(\frac{1}{m^*(k_0)} \right)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \dots;$$

away from a band extremum, the first order term must be included.

We can now obtain the quasi-classical equations of the crystal electron as equations of motion for expectation values. We need the normalization of F :

$$\begin{aligned}\int \psi^* \psi d\bar{r} &= 1 = \int |F|^2 |u_{n\bar{k}_0}|^2 d\bar{r} \approx \sum_{\bar{l}} |F(\bar{l}, t)|^2 \int_{\text{unit cell}} |u_{n\bar{k}_0}|^2 d\bar{r} \\ &= \frac{1}{N} \sum_{\bar{l}} |F(\bar{l}, t)|^2 = \frac{1}{V} \int |F(\bar{r}, t)|^2 d\bar{r};\end{aligned}$$

Therefore the quasi momentum has an expectation value

$$\langle \bar{p} \rangle = \frac{1}{V} \int F^* \bar{p} F = \frac{1}{V} \sum_{\bar{k}', \bar{k}''} a^*(\bar{k}', t) a(\bar{k}'', t) \int \hbar \bar{k}'' e^{i(\bar{k}'' - \bar{k}') \cdot \bar{r}} d\bar{r}.$$

Since $a(\bar{k}, t)$ is appreciable only near $\bar{k}_0 = \bar{k}_0(t)$, we can choose $\bar{k}'' = \bar{k}_0(t)$ and get, for the crystal (pseudo) momentum

$$\langle \bar{P} \rangle = \hbar k_0(t), \text{ and}$$

$$\begin{aligned} \dot{\langle \bar{P} \rangle} &= \frac{1}{\hbar} \langle [H, \bar{P}] \rangle = \frac{1}{\hbar} \langle [U, \bar{P}] \rangle \\ &= \frac{1}{\hbar v} \int F^* [U(-i\hbar \nabla) - (-i\hbar \nabla)U] F d\bar{r} \\ &= -\frac{1}{v} \int F^* (-\nabla U) F d\bar{r} = -\langle \nabla U \rangle. \end{aligned}$$

Similarly, from

$$\dot{\langle \bar{r} \rangle} = \bar{v} = \frac{1}{\hbar} \langle [H, \bar{r}] \rangle = \frac{1}{\hbar} \langle [\epsilon_n(\frac{\bar{P}}{\hbar}), \bar{r}] \rangle$$

we find from a Taylor expansion of ϵ_n about k_0 .

$$\bar{v} = \frac{1}{\hbar} \left\langle \frac{\partial \epsilon_n}{\partial k} \right\rangle$$

thus verifying $\hbar \dot{\bar{k}} = -\nabla U + \frac{q}{c} \bar{v} \times \bar{B}$.

To summarize the exposition up to this point: We have shown that the effective mass theorem, in the form in which we have used it in the theory of Zener oscillations, involves two approximations: The neglect of inter-band coupling, and a restriction of the wave function to a small region of the Brillouin Zone.

4.3.3 A Static Test of the Theorem

Before we proceed to an examination of the role played by these approximations in transport theory, it is instructive to consider a static example⁴⁾ which demonstrates that the uncritical use of the effective mass theorem can produce grossly wrong answers in an extremely simple context.

It is quite customary to use the effective mass theorem in the discussion of the hydrogenic states of a Coulomb potential, as on a shallow donor. This is a time-independent problem so $i\hbar \dot{\psi} = \epsilon \psi$, and the appropriate wave function at the center of the Brillouin zone ($\bar{k}_0 = 0$) is

$$\psi_{\text{approx}} = F(\bar{r}) \psi_{n0}(\bar{r}) = F(\bar{r}) u_{n0}(\bar{r}).$$

With a periodic crystal potential $V(r)$

$$\left[\frac{p^2}{2m} + V(\bar{r}) \right] u_{n0}(\bar{r}) = \epsilon_{n0} u_{n0}(\bar{r})$$

$$\left[\frac{p^2}{2m^*} + U(\bar{r}) \right] F(\bar{r}) = (\epsilon - \epsilon_{n0}) F(\bar{r})$$

where $U(\bar{r})$ is the coulomb potential, and $\epsilon - \epsilon_{n0}$ is the binding energy of the donor level in the effective mass approximation.

We can test the combination of eigenvalue and eigenfunction by a type of self-consistency argument. Let us postulate a potential V_{eff} such that both the eigenvalue ϵ and the eigenfunction ψ are exact:

$$\left(\frac{p^2}{2m} + V_{\text{eff}} \right) \psi_{n0} = \epsilon \psi_{n0}.$$

If the effective mass theorem were exact, it would require $V_{\text{eff}} = V(\bar{r}) + U(\bar{r})$.

We can test the theorem by evaluating

$$\Delta \equiv \langle V(\bar{r}) + U - V_{\text{eff}} \rangle$$

$$= \int d\bar{r} F^* u_{n0}^* [V + U - V_{\text{eff}}] F u_{n0}$$

This is carried out easily:

$$V_{\text{eff}} F u_n = \left(\epsilon - \frac{p^2}{2m} \right) F u_n$$

$$V(\bar{r}) F u_n = \epsilon_{n0} F u_n - F \frac{p^2}{2m} u_n$$

$$U F u_n = (\epsilon - \epsilon_{n0}) F u_n - u_n \frac{p^2}{2m^*} F$$

so

$$(V(\bar{r}) + U - V_{\text{eff}}) F u_n = - F \frac{p^2}{2m} u_n - u_n \frac{p^2}{2m} F + \frac{p^2}{2m} F u_n$$

We can work on the last term

$$\begin{aligned} p^2 F u_n &= p(p F u_n) = p[F p u_n + (p F) u_n] \\ &= (p F) (p u_n) + F p^2 u_n + (p^2 F) u_n + (p F) (p u_n) \end{aligned}$$

so

$$\frac{1}{2m} F p^2 u_n = \frac{1}{2m} [p^2 F u_n - 2(p F) (p u_n) - u_n p^2 F]$$

and

$$\begin{aligned} (V(\bar{r}) + U - V_{\text{eff}}) F u_n &= \frac{2}{2m} (p F) (p u_n) + \frac{1}{2m} u_n p^2 F \\ &\quad - \frac{1}{2m^*} u_n p^2 F \end{aligned}$$

so finally

$$\Delta = \int d\vec{r} F^* u_n^* \left[\left(\frac{1}{2m} - \frac{1}{2m^*} \right) u_n^2 F + \frac{1}{m} (pF)(pu_n) \right]$$

$$= \int d\vec{r} \left[\left(1 - \frac{m}{m^*} \right) |u_n|^2 F^* \frac{p^2}{2m} F + \frac{1}{m} (F^* p F)(u_n^* p u_n) \right].$$

Note that this gives the expected result $\Delta = 0$ for the "empty lattice" where

$m^* = m$ and $pu_{no} = -i\hbar \nabla u_{no} = 0$, but with slowly varying F

$$\int_{Xtal} d\vec{r} F^* \frac{p^2}{2m} F u_n^* u_n \approx \sum_{\text{unit cells}} F^* \frac{p^2}{2m} F \int_{\text{cell}} d\vec{r} u_n^* u_n = \int_{Xtal} F^* \frac{p^2}{2m} F d\vec{r}.$$

This asserts that the first term in Δ is just $\left(\frac{m^*}{m} - 1 \right)$ times the kinetic energy of a particle of mass m^* and wave function $F(r)$, equal to the same factor $\left(\frac{m^*}{m} - 1 \right)$ times the binding energy. The second term, to the same approximation, is zero from symmetry.

Since it is known from experimental results that the effective mass approximation is quite accurate in predicting the binding energy, this result indicates that the wave function is poor: either the assumption that a narrow range of $\vec{k}$ suffices to construct the localized orbital, or the neglect of interband terms, must be invalid even for these shallow states which are normally viewed as one of the major successes of effective mass theory.

4.3.4 Motion in a Uniform Electric Field

We now proceed to a detailed description of the collisionless motion of a crystal electron in a uniform electric field. This is a sufficiently simple problem so that it is surprising that it has ever been controversial, and that some of the controversy is apparently not completely resolved.

The physical core of the difficulty is this: When one accelerates a carrier, one might suspect that the carrier has to move some finite distance before it "finds out" what its effective mass is, i.e., it has to sample enough of the lattice to "learn" what the periodic potential is like. So one might guess that the classical motion with mass m^* takes some finite time to develop. In one rather obvious sense this crude formulation of Ehrenfest's theorem is confirmed in optical band-to-band transitions, where the interaction Hamiltonian involves m and not m^* . (The effective masses m_n and m_p appear in the transition probability because they describe the density of initial and final states, not because they are involved in the transition kinetics⁵⁾).

When the perturbation is a uniform electric field, a formal resolution of this question is obtained by writing the Schrödinger equation in a momentum representation that explicitly separates the intraband and interband parts of the Hamiltonian. We have

$$H = H_0 + H', \quad H_0 = \frac{p^2}{2m} + V(\vec{r}), \quad H' = -q\vec{E} \cdot \vec{r}.$$

The eigenfunctions of H_0 are Bloch functions

$$H_0 \phi_{n\vec{k}} = \epsilon_n(\vec{k}) \phi_{n\vec{k}}; \quad \phi_{n\vec{k}} = e^{i\vec{k} \cdot \vec{r}} u_{n\vec{k}}(\vec{r}).$$

Consider the operation

$$iqe^{i\vec{k}\cdot\vec{r}} \vec{E} \cdot \nabla_k e^{-i\vec{k}\cdot\vec{r}} \phi_{n\vec{k}} = iq\vec{E} \cdot \nabla_k \phi_{n\vec{k}} + q\vec{E} \cdot \vec{r} \phi_{n\vec{k}}.$$

This permits us to write the Hamiltonian as

$$H = H_0 + iq\vec{E} \cdot \nabla_k - iqe^{i\vec{k}\cdot\vec{r}} \vec{E} \cdot \nabla_k e^{-i\vec{k}\cdot\vec{r}}$$

and the Schrödinger equation in the (crystal) momentum representation as

$$i\hbar \dot{\psi} = [\epsilon_n(\vec{k}) + iq\vec{E} \cdot \nabla_k - iqe^{i\vec{k}\cdot\vec{r}} \vec{E} \cdot \nabla_k e^{-i\vec{k}\cdot\vec{r}}] \psi.$$

Now we allow $\vec{k}$ to vary with time, and we shall see presently that the second term on the right hand side of this equation gives only the acceleration within a band, the third term only interband effects. To demonstrate this, let

$$\psi = e^{i\vec{k}(t)\cdot\vec{r}} u_n(\vec{k}(t), \vec{r}) e^{-\frac{i}{\hbar} \epsilon_n t}$$

and consider as a first example a free electron with $V(\vec{r})=0$, $u_{n\vec{k}}(\vec{r})=1$; then

$$i\hbar \dot{\psi} = \left(\epsilon_n - \hbar \frac{d\vec{k}}{dt} \cdot \vec{r} \right) \psi = \left(\frac{\hbar^2 k^2}{2m} - q\vec{E} \cdot \vec{r} \right) \psi$$

(since the last term on the rhs $\vec{E} \cdot \nabla_k e^{-i\vec{k}\cdot\vec{r}} \psi = 0$).

Thus we can identify the acceleration term $\hbar \dot{\vec{k}} = q\vec{E}$.

For a crystal electron

$$\begin{aligned} i\hbar \dot{\psi} &= [\epsilon_n u_{n\vec{k}} - \hbar \frac{d\vec{k}}{dt} \cdot \vec{r} u_{n\vec{k}} + i\hbar \frac{d\vec{k}}{dt} \cdot \nabla_k u_{n\vec{k}}] e^{i\vec{k}\cdot\vec{r}} e^{-\frac{i}{\hbar} \epsilon_n t} \\ &= [\epsilon_n u_{n\vec{k}} - q\vec{E} \cdot \vec{r} u_{n\vec{k}} + iq\vec{E} \cdot \nabla_k u_{n\vec{k}}] e^{i\vec{k}\cdot\vec{r}} e^{-\frac{i}{\hbar} \epsilon_n t} \\ &\quad - iqe^{i\vec{k}\cdot\vec{r}} \vec{E} \cdot \nabla_k e^{-i\vec{k}\cdot\vec{r}} \psi, \end{aligned}$$

So if the last term does not make any contribution that changes $\bar{k}$, we again have $\hbar \dot{\bar{k}} = q\bar{E}$.

To establish the last point, consider the matrix element

$$\begin{aligned} \langle n, \bar{k}' | e^{i\bar{k} \cdot \bar{r}} \bar{E} \cdot \nabla_k e^{-i\bar{k} \cdot \bar{r}} | n, \bar{k} \rangle &= \\ &= \int d\bar{r} u_{n, \bar{k}'}^* e^{i(\bar{k} - \bar{k}') \cdot \bar{r}} \bar{E} \cdot \nabla_k u_{n, \bar{k}} = \bar{E} \cdot \bar{R}_{n, n}(\bar{k}). \end{aligned}$$

The last equality expresses the fact that, since $u_{n, \bar{k}}^* \nabla_k u_{n, \bar{k}}$ has the periodicity of the lattice, the matrix element is zero unless $\bar{k} = \bar{k}'$, so it only links states of the same $\bar{k}$ and does not accelerate the electron; it does induce interband transitions and polarization, the next topic to be discussed. For the sake of simplicity, we have neglected the changing kinetic energy of the electron in the wave function; we now repair this deficiency, to begin with by writing for ψ a one-band Houston function⁶⁾

$$\psi = e^{-\frac{i}{\hbar} \int_0^t \epsilon_n(\bar{k}(\tau)) d\tau} e^{i\bar{k}(t) \cdot \bar{r}} u_n(\bar{k}(t), \bar{r}).$$

When this is substituted in the momentum-representation Schrödinger equation, we again derive the acceleration theorem (omitting the interband term):

$$\begin{aligned} i\hbar \left[-\frac{1}{\hbar} \epsilon_n(\bar{k}, t) u_n + i\dot{\bar{k}} \cdot \bar{r} u_n + \dot{\bar{k}} \cdot \nabla_k u_n \right] e^{-\frac{i}{\hbar} \int_0^t \epsilon d\tau} e^{i\bar{k} \cdot \bar{r}} \\ = \left[\epsilon_n(\bar{k}_0) u_n + iq\bar{E} \cdot \left(-\frac{1}{\hbar} \right) \int_0^t \nabla_k \epsilon_n(\bar{k}) d\tau u_n + iq\bar{E} \cdot (i\bar{r}) u_n \right. \\ \left. + iq\bar{E} \cdot \nabla_k u_n \right] e^{-\frac{i}{\hbar} \int_0^t \epsilon d\tau} e^{i\bar{k} \cdot \bar{r}} \end{aligned}$$

Here on the rhs $\epsilon_n(\bar{k}) = \epsilon_n(\bar{k}_0)$ because it is the eigenvalue of the

time-independent equation $H_0 \psi = \epsilon \psi$. Using the result $\hbar \dot{\mathbf{k}} = q\mathbf{E}$ selfconsistently, this can be rewritten

$$\begin{aligned} & \epsilon_n(\bar{\mathbf{k}}(t)) u_n + \hbar \dot{\mathbf{k}} \cdot \bar{\mathbf{r}} u_n + i \hbar \dot{\mathbf{k}} \cdot \nabla_{\mathbf{k}} u_n \\ &= \epsilon_n(\bar{\mathbf{k}}_0) u_n + \frac{q\bar{\mathbf{E}}}{\hbar \dot{\mathbf{k}}} \int_0^t \frac{\partial \epsilon}{\partial \mathbf{k}} \frac{d\mathbf{k}}{dt} d\tau u_n - q\bar{\mathbf{E}} \cdot \bar{\mathbf{r}} u_n + i q \bar{\mathbf{E}} \cdot \nabla_{\mathbf{k}} u_n \\ &= \epsilon_n(\bar{\mathbf{k}}_0) u_n + [\epsilon_n(\bar{\mathbf{k}}(t)) - \epsilon_n(\bar{\mathbf{k}}_0)] u_n - q\bar{\mathbf{E}} \cdot \bar{\mathbf{r}} u_n + i q \bar{\mathbf{E}} \cdot \nabla_{\mathbf{k}} u_n \end{aligned}$$

The one-band Houston function is not an exact solution of the wave equation. If we want to evaluate the effect of the interband terms, we need an exact solution, which we can write as an expansion (for simplicity in one dimension)

$$\psi = \sum_n a_n(t) e^{\frac{i}{\hbar} \int_0^t \epsilon_n(k(\tau)) d\tau} e^{i k(t)x} u_n(k(t), x) = \sum_n a_n \psi_n.$$

Substituting this in the Schrödinger equation produces

$$\begin{aligned} i\hbar \sum_n (a_n \dot{\psi}_n + \dot{a}_n \psi_n) &= \sum_n a_n [\epsilon_n(k) + i q E \frac{\partial}{\partial k}] \psi_n \\ &- i q E e^{i k x} \frac{\partial}{\partial k} \sum_n a_n e^{-\frac{i}{\hbar} \int_0^t \epsilon_n d\tau} u_n, \end{aligned}$$

where we have cancelled the terms that we have just shown to be equal from the acceleration theorem of the one-band Houston function. Carrying out the indicated differentiations on the right hand side gives

$$\begin{aligned} i\hbar \sum_n \dot{a}_n u_n e^{-\frac{i}{\hbar} \int_0^t \epsilon_n d\tau} &= -i q E \sum_n a_n \left[\frac{\epsilon_n(t) - \epsilon_n(0)}{qE} u_n + \frac{\partial u_n}{\partial k} \right] \\ &\cdot e^{-\frac{i}{\hbar} \int_0^t \epsilon_n d\tau} \end{aligned}$$

Multiplying by u_m^* and integrating over a unit cell

$$i\hbar \dot{a}_m = -qE \sum_n a_n X_{mn} e^{-\frac{i}{\hbar} \int_0^t (\epsilon_n - \epsilon_m) d\tau}$$

where

$$X_{mn} \equiv -i \int u_m^* \frac{\partial u_n}{\partial k} dx$$

is the dipole matrix element.

If we start out with an electron in the Bloch state $|nk\rangle$ as defined in zero field, this has the usual interpretation: the field, which is turned on at $t=0$, causes rapidly varying terms to appear. The velocity $\langle v \rangle = -\frac{i\hbar}{m} \int \psi^* \frac{\partial \psi}{\partial x} dx$ also oscillates rapidly, with frequencies $(\epsilon_m - \epsilon_n)/\hbar$ about its average value

$$\bar{v} = \frac{1}{\hbar} \frac{\partial \epsilon_n}{\partial k} \bigg|_{k=k_0 + \frac{E}{\hbar} t}$$

We will come to a physical explanation of these oscillations later. For the present we note that they make it possible to satisfy Ehrenfest's theorem according to which at $t=0$ we must have $\langle \dot{v} \rangle = \frac{qE}{m}$ with m the free-electron mass; this is proved in detail below. For a Bloch initial state $|nk\rangle$ the oscillations persist. For a wave packet, the oscillations of the Bloch states constituting the packet get out of phase rapidly and their cumulative effect dies out, leaving the carrier moving with the effective mass. We can estimate the time τ needed for this dephasing from the uncertainty principle

$$\Delta \omega \tau \sim \pi, \quad \Delta \epsilon \tau \sim \pi \hbar$$

$$\tau \sim \frac{\pi \hbar}{\Delta \epsilon} = \frac{\pi \hbar 2m}{\hbar^2 (\Delta k)^2} = \frac{2\pi m}{\hbar (\Delta k)^2} ;$$

then from

$$\Delta k \Delta x \sim 1$$

we have

$$\tau \sim \frac{2\pi m (\Delta x)^2}{\hbar}$$

In cgs units the electron mass and Dirac's constant are numerically roughly equal. This means that the time required for the effective mass to be "established" depends on the localization of the carrier

$$\tau \sim 2\pi (\Delta x)^2 \quad x \text{ in cm;}$$

alternatively, if we take $\Delta \epsilon \sim kT \sim .01 \text{ eV}$, we find $\tau \sim 10^{-13} \text{ sec.}$

To work out the velocity for very short times we start with

$$i\hbar \dot{\psi} = \epsilon_n(k) \psi_{nk} + iqE \cdot \frac{\partial \psi_{nk}}{\partial k} - qE \sum_{n'} X_{n'n} \psi_{n'k}$$

where we have already rewritten the interband term

$$\begin{aligned} e^{ikx} \frac{\partial}{\partial k} e^{-ikx} \psi &= \sum_{n'k'} |n'k'\rangle \langle n'k'| e^{ikx} \frac{\partial}{\partial k} e^{-ikx} |nk\rangle \\ &= \sum_{n'} |n'k\rangle \langle n'k| e^{ikx} \frac{\partial}{\partial k} e^{-ikx} |nk\rangle = -i \sum_{n'} X_{n'n} |n'k\rangle. \end{aligned}$$

This can be integrated for a short time δt starting at $t=0$ to first order in the electric field

$$\psi(\delta t) = \psi_{nK} - \frac{qE}{i\hbar} \delta t \sum_{n'} X_{n'n} \psi_{n'K}$$

where $K = k + \frac{qE}{\hbar} \delta t$.

The velocity $\langle v \rangle = -\frac{i\hbar}{m} \int \psi^* \nabla \psi d^3x$ to first order in E is

$$\langle v \rangle = -\frac{i\hbar}{m} \int d^3x \left[\psi_{nk}^* + \frac{qE\delta t}{i\hbar} \sum_n X_{nn'} \psi_{n'K}^* \right] \left[\nabla \psi_{nK} - \frac{qE\delta t}{i\hbar} \sum_{n'} X_{n'n} \nabla \psi_{n'K} \right]$$

$$\approx v_n(K) - \frac{qE\delta t}{i\hbar m} \sum_n \left[X_{n,n} \psi_{nK}^* (-i\hbar \nabla) \psi_{n'K} - X_{nn'} \psi_{n'K}^* (-i\hbar \nabla) \psi_{nK} \right]$$

$$= v_n(K) - \frac{qE\delta t}{i\hbar m} \sum_n \left[X_{n,n} P_{nn'} - X_{nn'} P_{n'n} \right]$$

The momentum matrix element is related to $X_{nn'}$, by ⁷⁾

$$X_{nn'} = -\frac{i\hbar P_{nn'}}{m(\epsilon_n - \epsilon_n')}$$

so

$$\langle v \rangle = v_n(K) - \frac{qE\delta t}{m} \sum_n \frac{2|P_{nn'}|^2}{m(\epsilon_n - \epsilon_n')}$$

Then using the f-sum rule ⁸⁾

$$\begin{aligned} \langle v \rangle &= v_n(K) - \frac{qE\delta t}{m} \left(\frac{m}{m^*} - 1 \right) \\ &= v_n(K) - qE\delta t \left(\frac{1}{m^*} - \frac{1}{m} \right) \end{aligned}$$

and since

$$\begin{aligned} v_n(K) &= v_n(k + \frac{qE\delta t}{\hbar}) = v_n(k) + \frac{qE\delta t}{\hbar} \frac{\partial v_n}{\partial k} \\ &= v_n(k) + \frac{qE\delta t}{\hbar^2} \frac{\partial^2 \epsilon_n(k)}{\partial k^2} = v_n(k) + \frac{qE\delta t}{m^*} \end{aligned}$$

this proves that for short enough times

$$\langle v(\delta t) \rangle = v(k) + \frac{qE\delta t}{m},$$

that is to say, the carrier accelerates with the free electron mass and the effective mass theorem fails. The implication of this failure can be clarified by using a perturbation transformation due to Wannier⁹⁾ and Adams¹⁰⁾ which removes the interband terms. To first order in E this was done, in the present problem, by Adams and Argyres¹¹⁾. The appropriate multiband Houston function is

$$\phi_{nk} = \psi_{nk} + qE \sum_{n'} \frac{X_{n'n} \psi_{n'k}}{\epsilon_n - \epsilon_{n'}}$$

where the ψ_{nk} are one-band Houston functions. Now, leaving out all terms in E^2 , the Schrödinger equation is

$$\begin{aligned} i\hbar \dot{\phi}_{nk} &= i\hbar \dot{\psi}_{nk} + qE \sum_{n'} \frac{X_{n'n} i\hbar \dot{\psi}_{n'k}}{\epsilon_n - \epsilon_{n'}} = i\hbar \dot{\psi}_{nk} + qE \sum_{n'} \frac{X_{n'n} \epsilon_{n'} \psi_{n'k}}{\epsilon_n - \epsilon_{n'}} \\ &= \left[\epsilon_n - iqE \frac{\partial}{\partial k} \right] \left[\psi_{nk} + qE \sum_{n'} \frac{X_{n'n} \psi_{n'k}}{\epsilon_n - \epsilon_{n'}} \right] - qE \sum_{n'} X_{n'n} \psi_{n'k} \\ &= \left[\epsilon_n - iqE \frac{\partial}{\partial k} \right] \psi_{nk} + qE \sum_{n'} \frac{\epsilon_n X_{n'n} \psi_{n'k}}{\epsilon_n - \epsilon_{n'}} - qE \sum_{n'} \frac{(\epsilon_n - \epsilon_{n'}) X_{n'n} \psi_{n'k}}{\epsilon_n - \epsilon_{n'}} \\ &= \left[\epsilon_n - iqE \frac{\partial}{\partial k} \right] \left[\psi_{nk} + qE \sum_{n'} \frac{\epsilon_n X_{n'n} \psi_{n'k}}{\epsilon_n - \epsilon_{n'}} \right] = \left[\epsilon_n - iqE \frac{\partial}{\partial k} \right] \phi_{nk}, \end{aligned}$$

so to first order in E this multiband Houston function has $\hbar \dot{k} = qE$; furthermore

$$\begin{aligned}
\langle v \rangle &= -\frac{i\hbar}{m} \langle \phi_{nk}^* \frac{\partial \phi_{nk}}{\partial x} \rangle = \left\langle \left(\psi_{nk}^* + qE \sum_{n'} \frac{X_{nn'} \psi_{n'k}^*}{\epsilon_n - \epsilon_{n'}} \right) \frac{\partial}{\partial x} \left(\psi_{nk} + qE \sum_{n'} \frac{X_{n'n} \psi_{n'k}}{\epsilon_n - \epsilon_{n'}} \right) \right\rangle \\
&= v_{nk} + qE \sum_{n'} \frac{1}{\epsilon_n - \epsilon_{n'}} \left\langle X_{nn'} \psi_{n'k}^* \frac{\partial \psi_{nk}}{\partial x} + X_{nn'} \psi_{nk}^* \frac{\partial \psi_{n'k}}{\partial x} \right\rangle \\
&= v_{nk} + qE \sum_{n'} \frac{X_{nn'} P_{n'n} + X_{n'n} P_{nn'}}{\epsilon_n - \epsilon_{n'}} = v_{nk} = \frac{1}{\hbar} \frac{\partial \epsilon_n}{\partial k}
\end{aligned}$$

and for these modified bands the effective mass theorem is valid to first order in the electric field (Wannier has shown elsewhere¹²⁾ that the decoupling can be made exact).

We are now in a position to interpret these results. There are really three phenomena taking part simultaneously. When the field is turned on, the bound (filled band) electrons experience a force that alters their orbits; very simply, the semiconductor becomes electrically polarized. This modification, of course, happens to the wave functions of all the states, whether they are filled or empty. This is the state mixing familiar, for example, from the atomic Stark effect: the states of a hydrogen atom in an electric field can be represented as a superposition of field-free states. Even the persistent oscillations we have noted occur in this simple analog. Put in the crudest terms, when the field is turned on suddenly on an electron in, say, the ground state, it receives a "kick", its orbit is "shocked" and continues to "ring" indefinitely. Ehrenfest's theorem holds--the initial acceleration of an electron bound to a proton is qE/m (but, of course, it does not continue to move with free-electron dynamics).

The modified states ϕ_{nk} correspond to the polarized (Stark-mixed) states of the atom. They are the states in a steady electric field, presumed to have

been turned on infinitely slowly, so there are no oscillations. Unlike discrete atomic bound states the ϕ_{nk} as continuum states can experience acceleration governed by effective mass dynamics. This is the second of the dynamical phenomena.

The third phenomenon, electron tunneling, is implied by this discussion, but has not yet been mentioned explicitly. The bound states of an atom do not form a complete set; at high enough energy there is always an ionized continuum. Similarly, at high enough energy in a crystal there are empty closed or open bands.

In a uniform electric field an electron can acquire in principle an unlimited amount of energy. One way of representing this idea is by means of an energy level diagram in which the vacuum level and the band energies are functions of position - "tilted bands". Evidently levels or bands that are energetically separate in the absence of a field, become degenerate when it is present. This means that real as well as virtual band-to-band transitions become possible. This phenomenon is crucial to the possible occurrence of Zener oscillations, and we take it up quantitatively in the next section.

4.3.5 Tunneling and the Stark Ladder

Zener oscillations have not so far been observed, nor has anyone really expected to observe them. That is so because under "readily realizable" conditions the Zener oscillation period is appreciably longer than the mean time between collisions of a carrier. "Readily realizable" conditions here means fields small enough so that avalanche multiplication can be avoided. But there appears to be some question whether Zener oscillations could be observed even if the collision time could be increased well beyond the oscillation period.

The putative existence of the oscillations is based on the "theorems" $\hbar \dot{k} = qE$ and $v = \hbar^{-1} \partial \epsilon / \partial k$. We have now seen that the validity or failure of these approximations is bound up with interband transitions. It has been argued¹³⁾ that because of the band mixing terms it cannot be established that a carrier can on the average complete a full cycle of the oscillation.

The controversy is apparently not completely resolved--a paper on it has appeared as recently as 1976¹⁴⁾. The argument is usually carried on in the equivalent terms of quantized Stark levels. To follow this, it is easiest to go back to the pre-1925 quantum theory. Since the collision-free motion of a band electron is periodic, it is quantized according to the Bohr-Sommerfeld phase integral $\oint p dq = 2\pi\hbar$ or the equivalent Bohr-Wilson rule $\Delta \epsilon = 2\pi\hbar/\tau = qEa$, using τ from p 1.

This is the electron energy spectrum of a band if the interband interaction is neglected. It is called a Stark ladder, and it should be observable using available highly sensitive modulation spectroscopy methods such as electro-reflectance measurements.

It has been objected¹³⁾ that one cannot neglect the effect of the interband terms for as long as a period of the Zener oscillation; or equivalently, that the interband terms $qEX_{nn'}$, are of the same order of magnitude as the Stark ladder energy separation. This order-of-magnitude argument is plausible, because the matrix element $X_{nn'} = \int_{\text{unit cell}} u_n^* \nabla_k u_n d^3x$ might well be of the order of the lattice constant. This question can be resolved more precisely. What limits the Zener oscillation is band-to-band tunneling, and if the band structure is known, tunneling probabilities can be calculated with reasonable confidence. We shall show a model calculation based on Kane's papers¹⁵⁾.

Concern has also been expressed that the Stark ladder might not be observable because the spectrum is unstable "in the sense that the slightest change in the direction of $\vec{E}$ completely alters the level structure....". The passage in quotation marks is taken verbatim from Wannier's 1962 paper¹²⁾. It is a very surprising remark, since it was its author who originally proposed the idea of a Stark ladder, and it can be disposed of readily. The meaning of the remark is this: As the direction of $\vec{E}$ changes, the trajectory of an electron in the extended Brillouin zone also changes direction, it will traverse different regions of the zone, the motion in real space may even become aperiodic, etc.

However the effect of a small change in the direction of $\vec{E}$ on the real-space motion is small. For example, if $\vec{E}$ is tilted slightly off a principal crystal axis, the motion goes from being periodic to aperiodic; but the time taken to traverse each unit cell of reciprocal space is still nearly equal, and the aperiodicity is nothing more than a small modulation at a frequency incommensurable with the fundamental or, in terms of the spectrum, a slight broadening of the Stark levels.

4.3.6 Tunneling

We start with the time-independent Schrödinger equation in a field $\vec{E} = E\hat{x}$:

$$[\epsilon_n(\vec{k}) + iqE \frac{\partial}{\partial k_x}] b_n(\vec{k}) - qE \sum_{n'} X_{n',n} b_{n'}(\vec{k}) = \epsilon b_n(\vec{k}).$$

We can get a one-band solution neglecting interband terms immediately:

$$b_n(\vec{k}) = \frac{1}{\sqrt{\kappa}} e^{-\frac{1}{qE} \int_0^{\kappa} (\epsilon - \epsilon_n(k')) dk'_x} \delta(k_y - k_{y0}) \delta(k_z - k_{z0})$$

The k -space geometry is sketched in Figure 1. The δ -functions are normalized in the Brillouin Zone. κ is the length of the line segment in the direction of $\vec{E}$ subtended by the Brillouin zone, so it stands in a rational relation to the dimensions of the zone only when $\vec{E}$ lies along a reciprocal lattice vector. However, the incommensurability of κ and π/a , although it is mathematically troublesome, is evidently of no physical significance—this is the same "problem" as the instability of the Stark spectrum that we have just discussed in the last paragraph.

When the trajectory in k -space is periodic, the boundary condition is that the phase of the wave function $b_n(\vec{k})$ must be the same at equivalent points of the path of integration:

$$\frac{1}{qE} \int_{k_1}^{k_2} (\epsilon_m - \epsilon_n(k')) dk'_x = m \cdot 2\pi \quad m = 1, 2, \dots$$

giving a spacing of energy eigenvalues

$$\int_{k_1}^{k_2} (\epsilon_{m+1} - \epsilon_n(k')) dk'_x - \int_{k_1}^{k_2} (\epsilon_m - \epsilon_n(k')) dk'_x = 2\pi qE ;$$

but $k_2 - k_1 = \kappa$, so $\Delta\epsilon = 2\pi qE/\kappa$. This is the Stark ladder; for the simplest orientation of $\vec{E}$, $\kappa = \frac{2\pi}{a}$, so $\Delta\epsilon = aqE$.

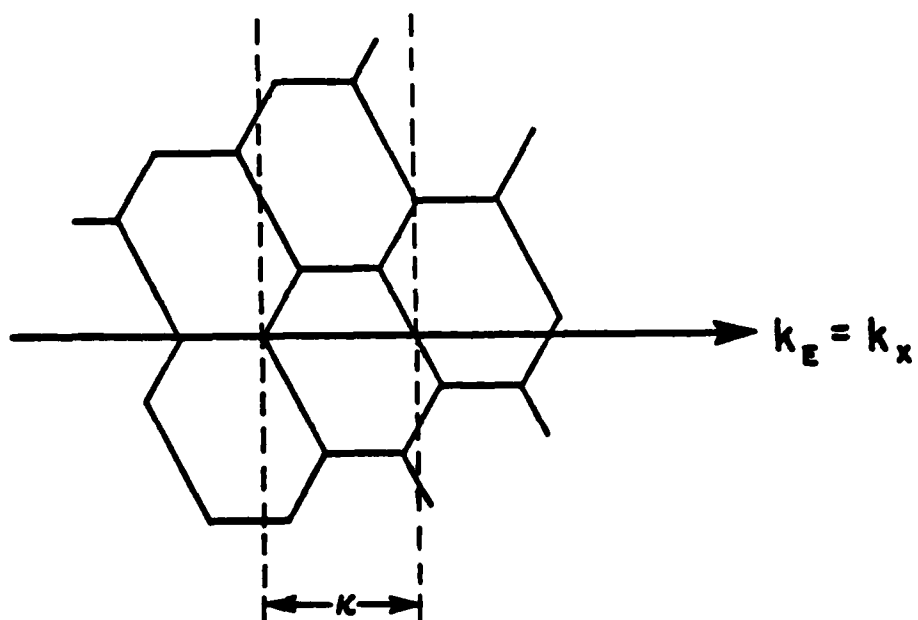


Figure 1. Section of the extended Brillouin zone

Using this wave function we can immediately get the matrix element between states of equal energy in bands n', n :

$$\begin{aligned} M_{n', n} &= -qE \int b_{n'}^*(k) X_{n', n}(k) b_n(k) dk_x \\ &= -\frac{qE}{\hbar} \int X_{n', n} \exp \left[\frac{i}{qE} \int_0^k x dk' (\epsilon_n(k') - \epsilon_{n'}(k')) \right] dk_x. \end{aligned}$$

From this point the calculation proceeds by standard perturbation theory methods. The Fermi golden rule gives a transition probability per second

$$\omega = \frac{2\pi}{\hbar} |M_{n', n}|^2 \rho(\epsilon)$$

with

$$\rho(\epsilon) = \frac{1}{\Delta\epsilon} = \frac{\kappa}{2\pi qE}.$$

The calculation will be valid only if the transition probability is small for times comparable with a Zener oscillation period.

This formulation is fairly abstract. To aid intuition, it is useful to rewrite the wave function in the more familiar coordinate representation. Since this is only for visualization purposes, we restrict the calculation to one dimension, with $\kappa = 2\pi/a$. Using $|x\rangle = \int |k\rangle \langle k|x\rangle dk$

$$q_n(x) = \sqrt{a/2\pi} \int dk e^{ikx} u_{nk}(x) \exp \left[\frac{i}{qE} \int_0^k (\epsilon - \epsilon_n(k')) dk' \right].$$

This integral can be evaluated approximately using the stationary phase method. The integral is written as $\exp(if)$ where

$$f = kx + \frac{1}{qE} \int_0^k (\epsilon - \epsilon_n(k')) dk' - i\ln u_{nk},$$

and expanded around the point of stationary phase k_s using

$$f' = 0 = x + \frac{1}{qE}(\epsilon - \epsilon_n(k_s)) - i \frac{\partial}{\partial k} \ln u_{nk} \Big|_{k_s}$$

in the form

$$f = f(k_s) + \frac{1}{2} (k - k_s)^2 f''(k_s),$$

where

$$f''(k_s) = - \frac{1}{qE} \frac{\partial \epsilon_n}{\partial k} \Big|_{k_s} - i \frac{\partial^2}{\partial k^2} \ln u_{nk} \Big|_{k_s}.$$

In the limit of "small" fields the terms involving u_{nk} can be neglected and the integral is

$$\begin{aligned} q_n(x) &= \sqrt{a/2\pi} \sqrt{2\pi/if''(k_s)} e^{ik_s x} e^{\frac{1}{qE} \int_0^{k_s} (\epsilon - \epsilon_n(k')) dk'} \\ &= \sqrt{a qE / \frac{\partial \epsilon_n}{\partial k}(k_s)} e^{\frac{1}{qE} \int_0^{k_s} (\epsilon - \epsilon_n(k') + qEx) dk'} \end{aligned}$$

This is entirely plausible: The stationary phase trajectory is the relation between x and k of the classical particle, i.e., the wave function $q_n(x)$ "accumulates phase" along the classical particle orbit. To see this explicitly, one can change variables from k to x . Along the classical orbit

$$x = \frac{1}{\hbar} \int_0^t \frac{\partial \epsilon_n}{\partial k} dt = \frac{1}{qE} \int_{k_0}^k \frac{\partial \epsilon_n}{\partial k} dk = \frac{1}{qE} [\epsilon_n(k) - \epsilon_n(k_0)]$$

so

$$\frac{dx}{dk} = \frac{1}{qE} \frac{\partial \epsilon_n}{\partial k}.$$

This permits us to work out the phase of $q_n(x)$:

$$\begin{aligned}
\int_0^{k_s} (\epsilon - \epsilon_n(k) + qEx) dk &= \int_0^{k_s} [\epsilon_n(k_s) - \epsilon_n(k)] dk \\
&= k_s \epsilon_n(k_s) - \int_0^{k_s} \epsilon_n(k) dk = \int_0^{k_s} k \frac{\partial \epsilon_n}{\partial k} dk \\
&= qE \int_0^{k_s} k \frac{dx}{dk} dk = qE \int_{x_0}^x k dx
\end{aligned}$$

and we finally have explicitly

$$q_n(x) = (aqE)^{1/2} \left(\frac{\partial \epsilon_n(k_s)}{\partial k} \right)^{-1/2} u_{nk_s}(x) e^{i \int_{x_0}^x k dx}$$

A schematic plot of this coordinate-space wave function (omitting the rapid modulation by u_{nk}) is shown in Figure 2. The wave function is shown in real space, drawn in spatial relation to a Zener tilted-energy-band diagram.

It is apparent that in this tilted-band representation all the wave functions in a band are identical except for a lateral displacement by multiples of the lattice constant. The figure also indicates how this picture fails when there is another band separated from the band labeled n by a gap: the wave function does not quite go to zero in the gap, so it can leak into the neighboring band, that is to say, the electron tunnels.

An intriguing question that has caused some argument comes from the possibility of eliminating the interband matrix element through the Wannier-Adams transformation. We used this transformation to first order in E in the discussion of the effective mass acceleration theorem, and Wannier has shown that the transformation can be carried out to all orders in E . Here it might be taken to mean that a uniform steady field cannot cause tunneling, but Kane has shown that this is not so, and indeed that the tunneling probability is only slightly affected by the transformation. The proof proceeds as follows:

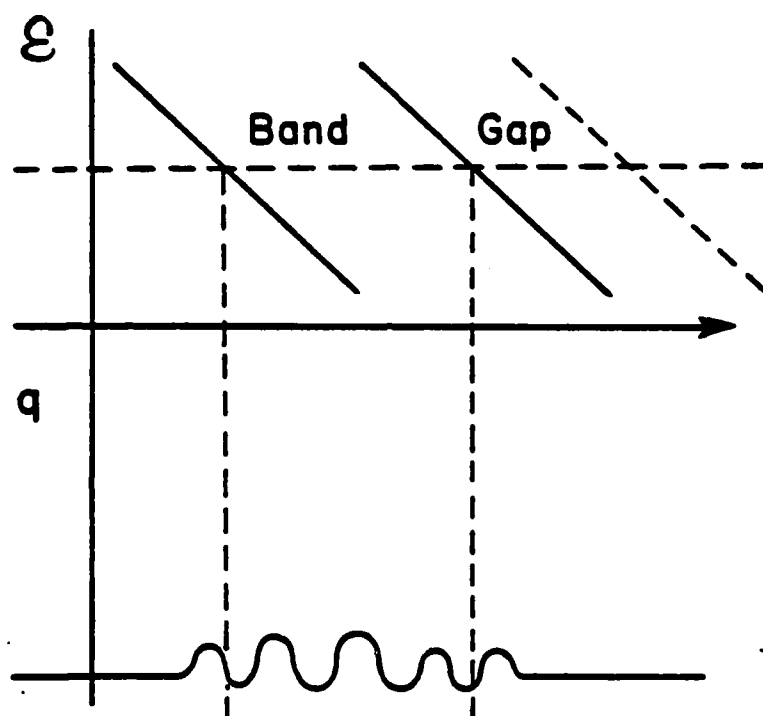


Figure 2. Band structure and wave function
in an electric field

Starting with the (for convenience one-dimensional) Schrödinger equation

$$\left[\epsilon_n(k) + i q E \frac{\partial}{\partial k} \right] b_n(k) - q E \sum_{n'} X_{n',n} b_{n'}(k) = \epsilon b_n(k)$$

we now write explicitly

$$\epsilon_n^{(1)}(k) = \epsilon_n(k) - q E X_{nn}(k)$$

which takes into account the previously neglected first order polarization of the bands by the field. Then

$$b_n^{(1)}(k) = \sqrt{\frac{a}{2\pi}} e^{-\frac{1}{qE} \int_0^k (\epsilon - \epsilon_n^{(1)}(k')) dk'}$$

and the matrix element connecting equal-energy states in different bands is

$$\begin{aligned} M_{n',n}^{(1)} &= \int dk \langle k, n' | E X_{n',n} | k, n \rangle \\ &= -\frac{q E a}{2\pi} \int dk X_{n',n} e^{\frac{1}{qE} \int_0^k dk' (\epsilon_n^{(1)}(k') - \epsilon_{n'}^{(1)}(k'))}. \end{aligned}$$

This matrix element will be used later to compute the tunneling probability.

Now the Wannier-Adams transformation is used to remove the interband terms to first order:

$$d_n(k) = e^{iT} b_n(k) \text{ with } T = -\frac{i q E X_{n',n}}{\epsilon_n^{(1)} - \epsilon_{n'}^{(1)}}.$$

When this is substituted in the Schrödinger equation, there results

$$\left[\epsilon_n^{(2)}(k) + i q E \frac{\partial}{\partial k} \right] d_n(k) + (qE)^2 \sum_{n'} Q_{n',n} d_{n'}(k) = \epsilon d_n(k)$$

where

$$\epsilon_n^{(2)}(k) = \epsilon_n^{(1)}(k) + (qE)^2 \sum_{n'} \frac{|X_{n',n}|^2}{\epsilon_n^{(1)} - \epsilon_{n'}^{(1)}}$$

and

$$Q_{n'n} = \frac{1}{2} \sum_{n'' \neq n, n'} X_{n''n} X_{n'n''} \left(\frac{1}{\epsilon_n^{(1)} - \epsilon_{n''}^{(1)}} + \frac{1}{\epsilon_{n'}^{(1)} - \epsilon_{n''}^{(1)}} \right) + i \frac{\partial}{\partial k} \frac{X_{n'n}}{\epsilon_{n'}^{(1)} - \epsilon_n^{(1)}} .$$

Since the form of the wave equation is unaltered, the interband matrix element now is

$$M_{n'n}^{(2)} = \frac{(qE)^2 a}{2\pi} \int dk Q_{n'n} e^{\frac{i}{qE} \int_0^k dk' (\epsilon_n^{(2)}(k') - \epsilon_{n'}^{(2)}(k'))} .$$

But now the first term in $Q_{n'n}$ —the term involving three bands n, n', n'' —is a higher order perturbation which was not taken into account in the calculation of the original matrix element, and accordingly should not be included now. This leaves

$$\begin{aligned} M_{n'n}^{(2)} &= \frac{i(qE)^2}{2\pi} Q \int dk \frac{\partial}{\partial k} \frac{X_{n'n}}{\epsilon_n^{(1)} - \epsilon_{n'}^{(1)}} e^{\frac{i}{qE} \int_0^k dk' (\epsilon_n^{(2)}(k') - \epsilon_{n'}^{(2)}(k'))} \\ &= - \frac{qEa}{2\pi} \int dk \frac{\epsilon_n^{(2)}(k) - \epsilon_{n'}^{(2)}(k)}{\epsilon_n^{(1)}(k) - \epsilon_{n'}^{(1)}(k)} X_{n'n} e^{\frac{i}{qE} \int_0^k dk' (\epsilon_n^{(2)}(k') - \epsilon_{n'}^{(2)}(k'))} \end{aligned}$$

where we have carried out a partial integration (the first term of which, giving the contribution of the end points, is zero because ϵ_n and $X_{n'n}$ are periodic in k -space, and the phase of d_n must match at the end points). Thus we find that the first and second order tunneling matrix elements differ only by the small second order correction of the bandgap.

This appears to conflict with Wannier's and Adams and Argyes interpretation of the properties of the modified Bloch bands, but the paradox is resolved as follows: The interband terms $X_{nn'}$, give rise to two effects, a polarization of the band states (as for example the polarization of the filled valence band) which can be ascribed to virtual transitions; and energy conserving interband

tunneling arising from real transitions. The field-modified Bloch (or Houston) functions are not bounded in either space or energy. and so they do not display the effects separately. The energy eigenfunctions do, and they show that the tunneling matrix elements are little affected by the Wannier-Adams transformation.

In the atomic hydrogen analogy we have used, the phenomenon we are describing is depicted in Figure 3, showing the states of a Coulomb potential in a uniform external field. The electron has a small probability of tunneling out the field ionization probability. This probability is nearly the same for the field-free and for the Stark shifted bound state.

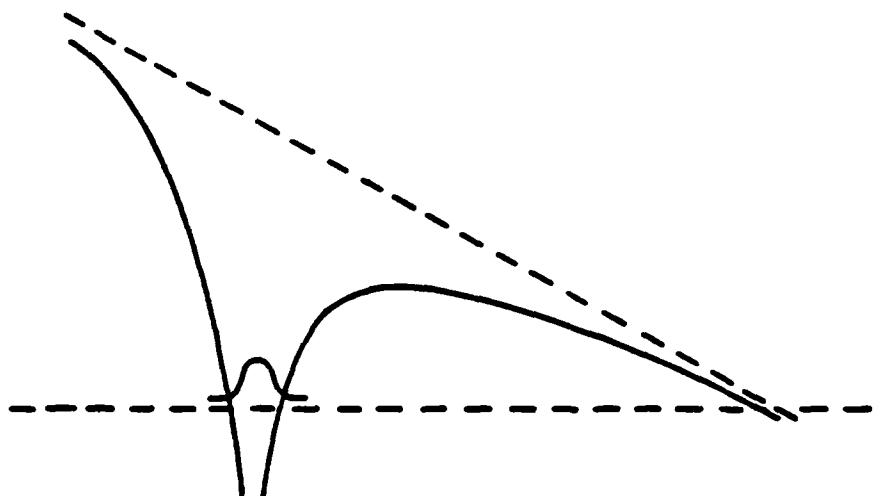


Figure 3. Field ionization as tunneling

4.3.7 Detailed Calculation of the Tunneling Probability

The simplest example of a tunneling process is illustrated in Figure 4, showing the penetration of a square barrier, and the wave functions that are involved. The calculation of the tunneling probability for this configuration is just like the calculation of energy transmission through a section of cutoff waveguide.

A more realistic representation of bandgap tunneling is shown in Figure 5. The upper half of this figure shows the conventional tilted-band representation of the potential. The lower half gives a schematic plot of the electron energy as a function of k^2 , the square of the crystal momentum. In the bandgap k^2 is negative (Tamm states), and the conduction and valence band are joined at the branch point k_B in complex k -space. If the field is small enough, so that the slope of the tilted-band diagram is sufficiently gentle, the wave function can be obtained using the WKB method:

$$\psi(x) = e^{-\int \kappa(x) dx} u_{c\kappa}(x)$$

with

$$\varepsilon(i\kappa(x)) = \varepsilon_c + \varepsilon(x).$$

To carry this out, it is necessary to know the form of $\kappa(x)$, and this has to be obtained from the band structure, as indicated schematically in Figure 5(b).

The connection of the WKB integral with the interband matrix element $M_{n,n}$ can be appreciated from the coordinate form $q_n(x)$ of the wave function:

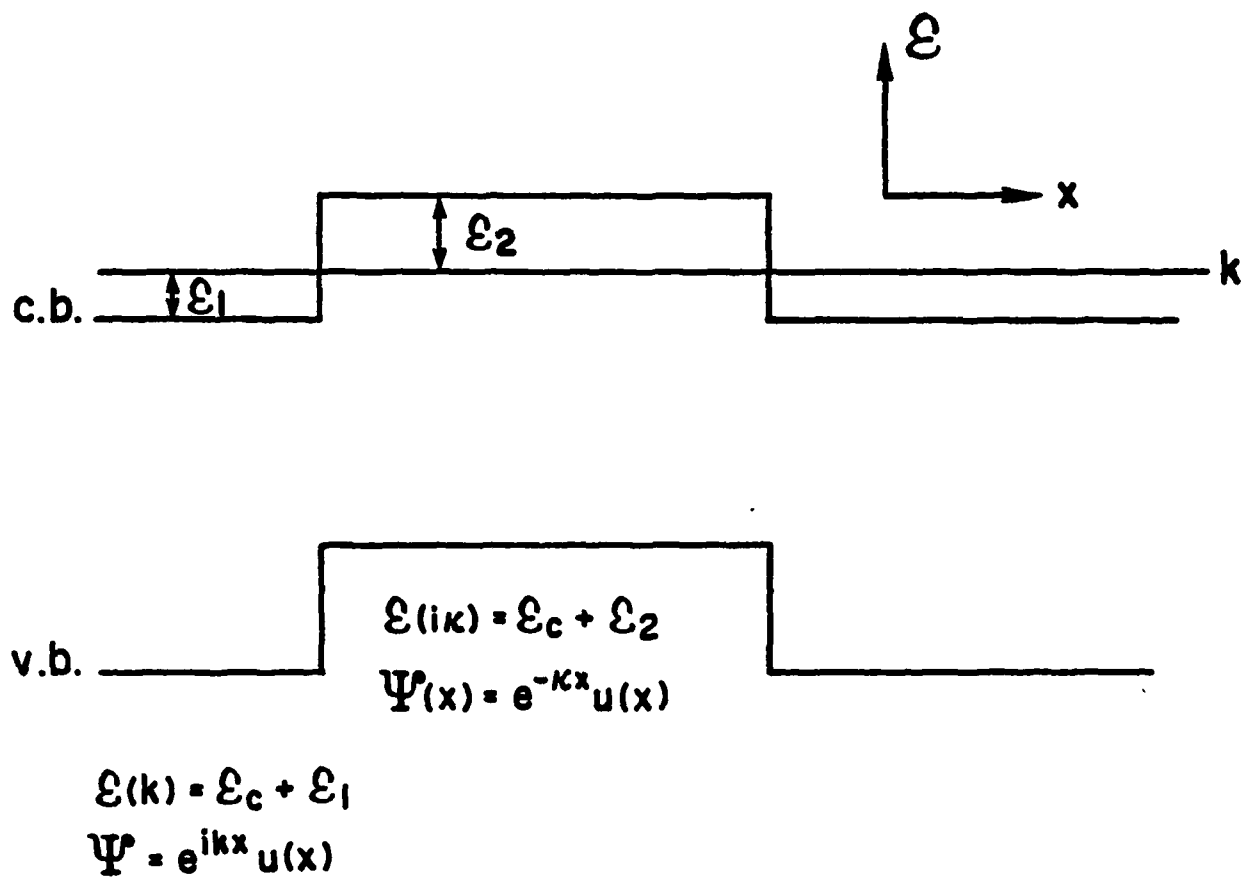


Figure 4. Energies and wave functions for a square barrier

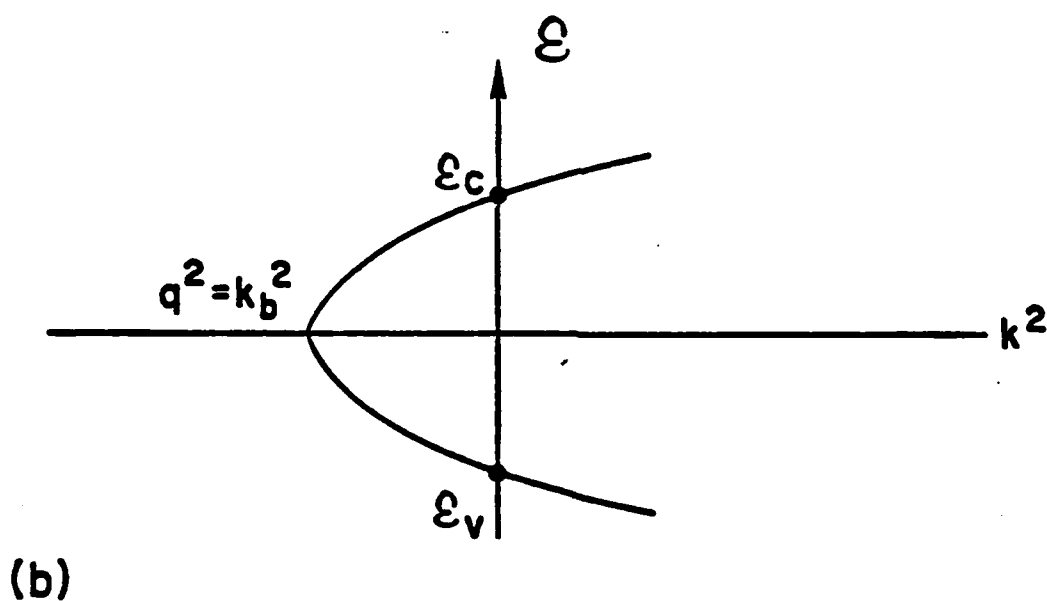
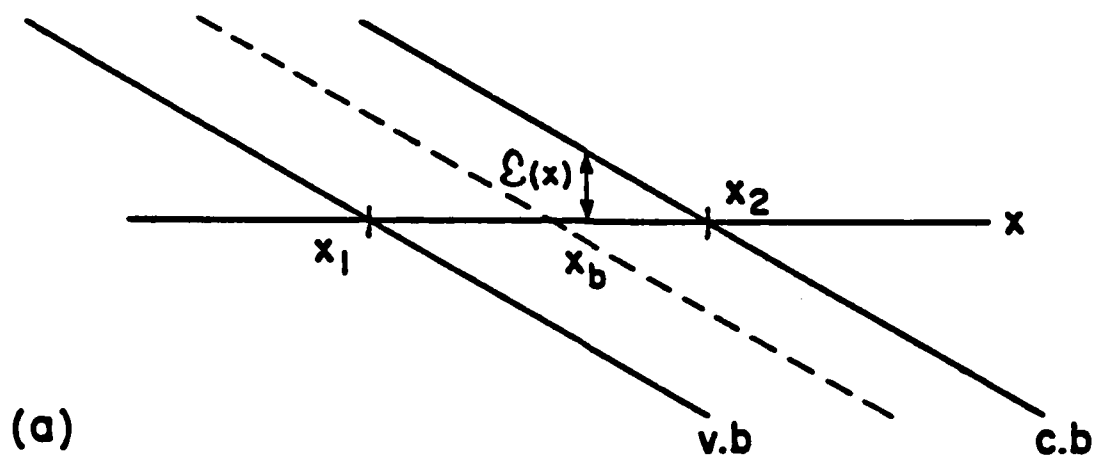


Figure 5. (a) Energy bands in a uniform electrostatic field
 (b) Energy as a function of k^2 , showing the joining of the conduction and valence band at the branch point k_b .

$$\begin{aligned}
M_{n',n} &= -qE \int q_{n'}^*(x) X_{n',n}(k) q_n(x) dx \\
&\rightarrow -aqEX_{n',n} e^{i \left[\int_{x_0}^x k dx - \int_{x'_0}^{x'} d' dx' \right]} \\
&\rightarrow -aqEX_{n',n} e^{i \left[\int_{x_0}^x k dx + \int_{x'_0}^{x'} k' dx' \right]}
\end{aligned}$$

where the path of integration and the meeting point of the bands are chosen by the method of steepest descent, using the band structure for imaginary k in the bandgap. Such a calculation was carried out by Kane, using two-band $\bar{k} \cdot \bar{p}$ perturbation theory to model the band structure of a narrow-gap direct semiconductor.

The band structure is obtained as follows. Starting with the one-band Bloch function $\psi = e^{i\bar{k} \cdot \bar{r}} u_{n\bar{k}}(\bar{r})$ of a lattice periodic potential $V(\bar{r})$ in the Schrödinger equation (with $p = -i\hbar \text{grad}$)

$$H\psi = \left(\frac{p^2}{2m} + V(\bar{r}) \right) \psi = \epsilon \psi$$

we have

$$\left(\frac{p^2}{2m} + \frac{\hbar^2}{m} \bar{k} \cdot \bar{p} + \frac{\hbar^2 k^2}{2m} + V(\bar{r}) \right) u_{n\bar{k}} = \epsilon_n(\bar{k}) u_{n\bar{k}}.$$

For fixed $\bar{k}$, the $u_{n\bar{k}}$ are a complete basis for all lattice-periodic functions, so that we can expand

$$u_{n\bar{k}}(\bar{r}) = \sum_{n'} c_{n',n}(\bar{k}-\bar{k}_0) u_{n',\bar{k}_0}(\bar{r}).$$

Defining

$$H_{k_0} = \frac{p^2}{2m} + \frac{\hbar}{m} \bar{k}_0 \cdot \bar{p} + \frac{\hbar^2 k_0^2}{2m} + V(\bar{r})$$

so that

$$H_{k_0} u_{nk_0} = \epsilon_n(k_0) u_{nk_0}$$

and

$$(H_{k_0} + \frac{\hbar}{m} (\bar{k} - \bar{k}_0) \cdot \bar{p} + \frac{\hbar^2}{2m} (k^2 - k_0^2)) u_{nk} = \epsilon_n(\bar{k}) u_{nk},$$

we substitute the expansion of u_{nk} about $\bar{k}_0$, multiply by $u_{n\bar{k}}^*$ and integrate over a unit cell. This yields

$$\sum_{n'} \left[(\epsilon_n(\bar{k}_0) + \frac{\hbar^2}{2m} (k^2 - k_0^2)) \delta_{nn'} + \frac{\hbar^2}{m} (\bar{k} - \bar{k}_0) \cdot \bar{p}_{nn'} \right] C_{n'n} = \epsilon_n(\bar{k}) C_{nn}$$

where

$$P_{nn'} = \int_{\text{unit cell}} d\bar{r} u_{nk_0}^* p u_{n'\bar{k}_0}.$$

To apply this to the conduction band and valence band of a direct-gap semiconductor at Brillouin zone center, we take the zero of energy at the top of the valence band. The Hamiltonian is

$$H = \begin{pmatrix} \epsilon_g + \frac{\hbar^2 k^2}{2m} & \frac{\hbar}{m} k p \\ \frac{\hbar}{m} k p & \frac{\hbar^2 k^2}{2m} \end{pmatrix}$$

giving energies

$$\epsilon_{\pm} = \frac{1}{2} \epsilon_g + \frac{\hbar^2 k^2}{2m} \pm \frac{1}{2} \eta, \quad \eta = \left(\epsilon_g^2 + \frac{4\hbar^2 k^2 p^2}{m^2} \right)^{1/2}.$$

The effective masses are

$$\frac{1}{m_{\pm}} = \frac{1}{\hbar^2} \frac{d^2 \epsilon_g}{dk^2} \bigg|_{k=0} = \frac{1}{\hbar^2} \frac{d}{dk^2} \left(\frac{\hbar^2 k^2}{2m} \pm \frac{1}{2} \eta \right) = \frac{1}{m} \pm \frac{2p^2}{m^2 \epsilon_g}.$$

If we use the convention that effective masses are positive, this means

$$\frac{1}{m_{\pm}} = \frac{2p^2}{m \epsilon_g} \pm m$$

and the two-band approximation is valid only if the bands repel enough to make both m_+ and m_- positive. The usual reduced effective mass is defined by

$$\frac{1}{m^*} = \frac{1}{2} \left(\frac{1}{m_+} + \frac{1}{m_-} \right)$$

so that p and m^* are related by $p = m(\epsilon_g/2m^*)^{1/2}$ and $\eta = (\epsilon_g^2 + 2\epsilon_g \hbar^2 k^2/m^*)^{1/2}$.

This provides the needed band structure, and we see that the conduction and valence bands are joined at $k_x = s$ (k_B in Figure 5) which makes $\eta = 0$:

$$s = \pm i \left(\frac{m^* \epsilon_g}{2\hbar^2} + k_y^2 + k_z^2 \right)^{1/2}.$$

The next step is to compute $X_{n'n} = i \int d\vec{r} u_{n'k} \frac{\partial u_{nk}}{\partial k}$. For this we need the linear combinations u_c, u_v that diagonalize the Hamiltonian. In terms of the zone center u_{c0} and u_{v0} these linear combinations are

$$u_c = \frac{1}{\sqrt{2\eta}} \left[(\eta + \epsilon_g)^{1/2} u_{c0} + (\eta - \epsilon_g)^{1/2} u_{v0} \right]$$

$$u_v = \frac{1}{\sqrt{2\eta}} \left[(\eta - \epsilon_g)^{1/2} u_{c0} - (\eta + \epsilon_g)^{1/2} u_{v0} \right].$$

In this model, the diagonal matrix elements vanish

$$X_{cc} = X_{vv} = 0.$$

The computation of X_{cv} is tedious but elementary

$$\begin{aligned}
X_{cv} &= i \int dr u_c^* \frac{du_v}{dk} = i \int \frac{dr}{\sqrt{2\eta}} \left[(\eta + \epsilon_g)^{1/2} u_{c0}^* + (\eta - \epsilon_g)^{1/2} u_{v0}^* \right] \\
&\quad \cdot \frac{d}{dk} \frac{1}{\sqrt{2\eta}} \left[(\eta - \epsilon_g)^{1/2} u_{c0} - (\eta + \epsilon_g)^{1/2} u_{c0}^* \right] \\
&= i \left[\frac{1}{\sqrt{2\eta}} (\eta + \epsilon_g)^{1/2} \frac{d}{dk} \frac{1}{\sqrt{2\eta}} (\eta - \epsilon_g)^{1/2} - \frac{1}{\sqrt{2\eta}} (\eta - \epsilon_g)^{1/2} \frac{d}{dk} \frac{1}{\sqrt{2\eta}} (\eta + \epsilon_g)^{1/2} \right] \\
&= \dots = \frac{1}{2\eta} \frac{d\eta}{dk} \frac{\epsilon_g}{(\eta^2 - \epsilon_g^2)^{1/2}} = - \frac{1}{2\eta} \frac{d\eta}{dk} \frac{\epsilon_g \sqrt{m^*}}{\hbar k \sqrt{2\epsilon_g}}
\end{aligned}$$

and since

$$\frac{d\eta}{dk} = \frac{1}{2\eta} \cdot \frac{4\epsilon_g \hbar^2 k}{m^*}$$

we finally have

$$X_{cv} = - \frac{i\hbar}{\eta^2} \frac{\epsilon_g^{3/2}}{\sqrt{2m^*}^{1/2}} \quad \text{and, of course, } \epsilon_c - \epsilon_v = \eta.$$

For M_{cv} we need

$$\int_0^{k_x} (\epsilon_c - \epsilon_v) dk_x = \int_0^{k_x} \eta dk_x.$$

This is of the form

$$\int_0^{k_x} x(a+bu^2) du = \frac{1}{2\sqrt{b}} \left[\sqrt{b} k_x (a+bk_x^2)^{1/2} + a \ln \frac{\sqrt{b} k_x + \sqrt{a+bk_x^2}}{\sqrt{a}} \right]$$

where

$$a = \epsilon_g^2 + \frac{2\epsilon_g \hbar^2}{m^*} (k_y^2 + k_z^2), \quad b = \frac{2\epsilon_g \hbar^2}{m^*};$$

then

$$M_{cv} = \frac{qE}{H} \int_{B.Z.} X_{cv} e^{\frac{1}{q} \int_0^{k_x} \eta dk_x'} dk_x .$$

But, especially if E is small, most of the integral must come from the vicinity of $\eta = 0$ —the path of stationary phase. Since the path lies in a region of complex k_x , M_{cv} is a contour integral, and the contour must be chosen to pass $\eta = 0$ (equivalent to $k_x = s$). So the contour must look like Figure 6, because X_{cv} has a pole at $k_x = s$. The contribution from the semicircle is half the residue at the pole. So using

$$X_{cv} = - \frac{i\hbar \epsilon_g^{3/2}}{\sqrt{2m^*}^{1/2}} \frac{1}{\eta^2} \text{ with}$$

$$\eta^2 = \epsilon_g^2 + 2\epsilon_g \hbar^2 k^2 / m^* = - \frac{2\epsilon_g \hbar^2}{m^*} (s^2 - k_x^2) \text{ so}$$

$$X_{cv} = \frac{i\hbar \epsilon_g^{3/2}}{\sqrt{2m^*}} \frac{m^*}{2\epsilon_g (s^2 - k_x^2)} = \frac{i(\epsilon_g m^*)^{1/2}}{2\sqrt{2}\hbar (s^2 - k_x^2)}$$

we have

$$\begin{aligned} - \frac{qE}{\kappa} \int_{\text{semi-circle}} X_{cv} e^{\frac{1}{qE} \int_0^{k_x} \eta dk_x'} dk_x &= \frac{i q E (\epsilon_g m^*)^{1/2}}{2\sqrt{2} \hbar \kappa} e^{\frac{1}{qE} \int_0^s \eta dk_x} \int_{\text{semi-circle}} \frac{dk_x}{s^2 - k_x^2} \\ &= - \frac{\pi q E (\epsilon_g m^*)^{1/2}}{4\sqrt{2} \hbar \kappa s} e^{\frac{1}{qE} \int_0^s \eta dk_x} \end{aligned}$$

because

$$\int_{\text{semi-circle}} \frac{dk_x}{(s+k_x)(s-k_x)} = \frac{\pi i}{2s} .$$

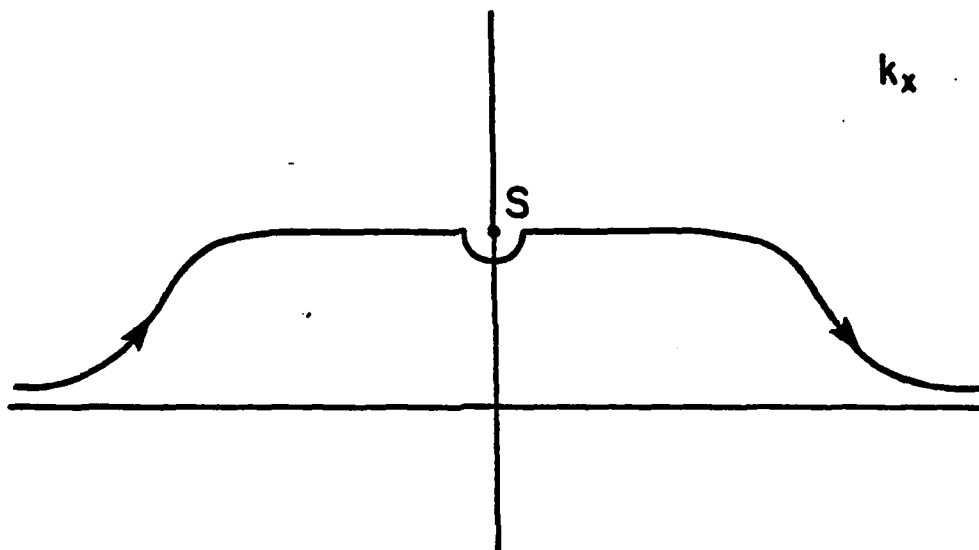


Figure 6. Contour for calculation of the tunneling matrix element

Next we get the contribution from the horizontal portion of the path by expanding η about s . Let $\sigma = k-s$, then

$$\eta = \left(\frac{2\epsilon_g}{m^*} \right)^{1/2} \hbar (k^2 - s^2)^{1/2} = 2(\epsilon_g/m^*)^{1/2} \hbar (s\sigma)^{1/2}$$

$$\eta dk = 2(\epsilon_g s/m^*)^{1/2} \hbar \cdot \frac{2}{3} \sigma^{3/2}$$

and the integral over the horizontal portion contributes

$$\begin{aligned} & - \frac{qE}{\kappa} \int - \frac{i\hbar\epsilon_g^{3/2}}{\sqrt{2}m^{*1/2}} \frac{m^*}{4\epsilon_g \hbar^2 s\sigma} e^{\frac{i}{qE} \left[\int_0^s \eta dk_x - \int_{\kappa}^{\sigma} \eta dk_x \right]} \\ & \approx \frac{i q E}{\kappa} \frac{(\epsilon_g m^*)^{1/2}}{4\sqrt{2} \hbar s} e^{\frac{i}{qE} \int_0^s \eta dk_x} \int_{-\infty}^{\infty} \frac{d\sigma}{\sigma} e^{\frac{i}{qE} \frac{4}{3} \hbar \left(\frac{\epsilon_g}{m^*} \right)^{1/2} \sigma^{3/2}} \end{aligned}$$

and using

$$\int \frac{d\sigma}{\sigma} e^{i a \sigma^{3/2}} = \frac{2i\pi}{3}$$

$$\int_{\text{hor}} = - \frac{\pi q E (\epsilon_g m^*)^{1/2}}{6\sqrt{2} \hbar \kappa s} e^{\frac{i}{qE} \int \eta dk_x}$$

We have carried this calculation out in such great detail because our result differs from Kane's by a numerical factor (Kane has $\frac{\pi}{12}$ where we find $\frac{\pi}{6\sqrt{2}}$).

Evaluating

$$\begin{aligned} \int_0^s \eta dk_x &= \frac{1}{2} \left(\frac{m^*}{2\epsilon_g} \right)^{1/2} \frac{1}{\hbar} \left[\epsilon_g^2 + \frac{2\epsilon_g \hbar^2}{m^*} (k_y^2 + k_z^2) \right] \ln i \\ &= \frac{i\pi}{4\hbar} \frac{m^{*1/2}}{\sqrt{2}} \epsilon_g^{3/2} + \frac{i\pi}{2} \frac{\epsilon_g^{1/2} \hbar}{\sqrt{2}m^{*1/2}} (k_y^2 + k_z^2) \end{aligned}$$

yields

$$M_{cv} = -\frac{5\pi qE}{12\hbar} e^{-\frac{\pi}{4qE} \left[\frac{m^{*1/2} \epsilon_g^{3/2}}{\sqrt{2} \hbar} + \frac{\sqrt{2} \epsilon_g^{1/2} \hbar}{m^{*1/2}} (k_y^2 + k_z^2) \right]}$$

(Kane's result contains the numerical factor $\frac{1}{3}$ instead of $\frac{5}{12}$). Using this matrix element, we find the tunneling probability per unit time from the Fermi golden rule

$$w = \frac{2\pi}{\hbar} \left| M_{n',n} \right|^2 \rho(\epsilon)$$

$$\text{with } \rho(\epsilon) = \frac{1}{\Delta\epsilon} = \frac{H}{2\pi qE}$$

or

$$w = \frac{25\pi^2}{144} \frac{qE}{\kappa\hbar} e^{-\frac{\pi}{2qE} \frac{m^{*1/2} \epsilon_g^{3/2}}{\sqrt{2} \hbar}} e^{-\frac{\pi}{2qE} \frac{\sqrt{2} \epsilon_g^{1/2} \hbar}{m^{*1/2}} (k_y^2 + k_z^2)}$$

This can be put in the form of a transmission coefficient

$$T = w\tau = \frac{\kappa\hbar}{qE} w.$$

The prefactor of the exponential $\frac{25\pi^2}{144}$ differs from Kane's value of $\frac{\pi^2}{9}$. The difference is not significant, since the numerical value is probably an artifact of the perturbation theory. An exact calculation for a sinusoidal crystal potential by Kane and Blount¹⁶⁾ (not using WKB) gives a prefactor unity.

It may be worth remarking that the second exponential factor in the transmission probability might be visualized as arising from the wider bandgap that has to be crossed by electrons with non-zero transverse momentum, but it must be cautioned that a naive calculation that takes only this effect (and not the detail of the bandstructure) into account yields the wrong functional dependence.

If we write the transmission coefficient as

$$T = T_0 e^{-\alpha k_{\perp}^2}$$

then in numerically convenient units with ϵ_g in electron volts, E in V/cm

$$T_0 = C e^{-4.04 \times 10^7 \left(\frac{m^*}{m}\right)^{1/2} \frac{\epsilon_g^{3/2}}{E}}$$

where C is the prefactor of order unity.

Even with rather large fields, one usually finds $T_0 \ll 1$. Examples are InSb with $\epsilon_g = 0.18$, $\frac{m^*}{m} = .013$, $T_0 \approx e^{-3.5 \times 10^5/E}$; GaAs with $\epsilon_g = 1.4$, $\frac{m^*}{m} = 1.4$, $T_0 \approx e^{-2.5 \times 10^7/E}$ etc. This is consistent with Avron's¹⁴⁾ Stark ladder calculation which views the Stark states as resonances and shows that the resonances are narrow unless the bandgaps are very small.

If Kane's transmission coefficient is recalculated from a complete model using a sinusoidal crystal potential instead of the $\bar{k} \cdot \bar{p}$ band structure, thus

$$V(r) = 2V_0 \sin \frac{\pi x}{a}, \quad \epsilon_g = 2V_0, \quad m^* = \frac{m^2 a^2 V_0}{\pi^2 \hbar^2},$$

the result is

$$T_0 = C e^{-\frac{m a V_0^2}{\pi^2 \hbar^2 E}} = C e^{-1.32 \times 10^5 \frac{a V_0^2}{E}}$$

with the lattice constant a in Angstrom units.

For an electron in the (full) valence band the probability that it tunnel to the (empty) conduction band is given by T times the probability that it is in $dk_y dk_z$:

$$d_p = \frac{dk_y dk_z}{\left(\frac{2\pi}{a}\right)^3} T_0 e^{-\alpha(k_y^2 + k_z^2)}, \quad \alpha = \frac{\pi \hbar \epsilon_g^{1/2}}{\sqrt{2} m^*^{1/2} q E}$$

so

$$p = \frac{T_0 a^2}{(2\pi)^2} \cdot 2\pi \int_0^\infty k_\perp dk_\perp e^{-\alpha k_\perp^2}$$

$$= \frac{a^2 m^{*1/2} q \epsilon_g^{1/2}}{2\sqrt{2}\pi^2 \hbar \epsilon_g^{1/2}} C e^{\frac{-\pi m^{*1/2} \epsilon_g^{3/2}}{2\sqrt{2}\hbar q E}}$$

This is the form usually quoted for the tunneling probability; in the convenience units as above, the prefactor is $0.41 \times 10^{-8} C \left(\frac{m^*}{m}\right)^{1/2} \frac{a^2 E}{\epsilon_g^{1/2}}$

The number of electrons tunneling per Zener period per cm^3 is

$$n' = (\text{number in band}/\text{cm}^3) \cdot p = \frac{2}{a^3} p = \frac{m^{*1/2} q E}{\sqrt{2}\pi^2 \hbar a \epsilon_g^{1/2}} C e^{\frac{-\pi m^{*1/2} \epsilon_g^{3/2}}{2\sqrt{2}\hbar q E}}$$

In this expression the prefactor is $0.82 \times 10^{16} \left(\frac{m^*}{m}\right)^{1/2} \frac{E}{a \epsilon_g^{1/2}}$; this is

$4.6 \times 10^{14} E$ in GaAs, $3 \times 10^{14} E$ in InSb. A necessary criterion for a Zener oscillator is that

$$n' \ll \text{number of oscillating conduction band carriers.}$$

This criterion will not be difficult to meet. One may conclude that Zener oscillations require primarily an adequate lengthening of the scattering times.

1. C. Zener, Proc. Roy. Soc. A145, 523 (1934).
2. M. Shur and L. F. Eastman, IEEE Trans. El. Dev., ED-26, 1677 (1979).
- 2a. J. R. Barker and D. K. Ferry, Sol. St. Electronics 23, 519, 531, 545 (1980).
3. H. Jones and C. Zener, Proc. Roy. Soc. A144, 101 (1934).
4. A. M. Stoneham, "Theory of Defects in Solids", Oxford 1975.
5. See for example, R. A. Smith, "Wave Mechanics of Crystalline Solids", Chapman and Hall 1961.
6. W. V. Houston, Phys. Rev. 57, 184 (1940).
7. R. A. Smith, op. cit.
8. P. T. Landsberg, "Solid State Theory", Wiley 1969.
9. G. H. Wannier, Phys. Rev. 100, 1227 (1955).
10. E. N. Adams, Phys. Rev. 107, 698 (1957).
11. E. N. Adams and P. N. Argyres, Phys. Rev. 102, 605 (1956).
12. G. H. Wannier, Rev. Mod. Phys. 34, 645 (1962).
13. J. Zak, Phys. Rev. Lett. 20, 1477 (1968); Phys. Rev. 181, 1366 (1969); Solid State Physics (H. Ehrenreich, ed.) Academic Press 1972.
14. J. E. Avron, Phys. Rev. Lett. 37, 1568 (1976).
15. E. O. Kane, J. Phys. Chem. Solids 12, 181 (1959).
16. E. O. Kane and E. I. Blount in "Tunneling Phenomena in Solids", E. Burnstein and S. Lundqvist, eds, Plenum 1969.

DISTRIBUTION LIST - TECHNICAL REPORTS
Contract N00014-79-C-0840

Office of Naval Research Code 427 Arlington, VA 22217	4	Dr. H. C. Nathanson Westinghouse Research and Development Center Beulah Road Pittsburgh, PA 15235	1
Naval Research Laboratory 4555 Overlook Avenue, S.W. Washington, DC 20375 Attn: Code 6811 6850	1 1	Dr. Daniel Chen Rockwell International Science Center P.O. Box 1085 Thousand Oaks, CA 91360	1
Defense Logistics Agency Defense Documentation Center Bldg. 5, Cameron Station Alexandria, VA 22314	12	Dr. D. Krumn Hughes Research Laboratory 3011 Malibu Canyon Road Malibu, CA 90265	1
Dr. Y. S. Park AFWAL/DHR Building 450 Wright-Patterson AFB Ohio 45433	1	Mr. Lothar Wandinger ECOM/AMSEL/TL/IJ Fort Monmouth, NJ 07003	1
ERADCOM DELET-M Fort Monmouth, NJ 07703	1	Dr. Harry Wieder Naval Ocean Systems Center Code 922 271 Catalina Blvd. San Diego, CA 92152	1
Texas Instruments, Inc. Central Research Lab M.S. 134, P.O. Box 225936 13500 North Central Expressway Dallas, TX 75265 Attn: Dr. Wisseman	1	Dr. William Lindley MIT Lincoln Laboratory F124 A, P.O. Box 73 Lexington, MA 02173	1
Dr. R. M. Malbon/M.S. 1C Avantek, Inc. 3175 Bowers Avenue Santa Clara, CA 94304	1	Commander U.S. Army Electronics Command V. Gelnovatch (DRSEL-TL-IC) Fort Monmouth, NJ 07703	1
Mr. R. Bierig Raytheon Company 28 Seyon Street Waltham, MA 02154	1	RCA Microwave Technology Center Dr. F. Sterzer Princeton, NJ 08540	1
Dr. R. Bell, K-101 Varian Associates, Inc. 611 Hansen Way Palo Alto, CA 94304	1	Watkins-Johnson Company E.J. Crescenzi, Jr./K. Niclas 3333 Hillview Avenue Stanford Industrial Park Palo Alto, CA 94304	1
Hewlett-Packard Corp. Dr. Robert Archer 1501 Page Mill Road Palo Alto, CA 94306	1		

Commandant Marine Corps Scientific Advisor (Code AX) Washington, DC 20380	1	Professor L. Eastman Phillips Hall Cornell University Ithaca, NY 14853	1
Communications Transistor Corp. Dr. W. Weisenberger 301 Industrial Way San Carlos, CA 94070	1	AIL TECH 612 N. Mary Avenue Sunnyvale, CA 94086 Attn: G. D. Vendelin	1
Microwave Associates Northwest Industrial Park Drs. F.A. Brand/J. Saloom Burlington, MA 01803	1	Professors Hauser and Littlejohn Department of Electrical Engr. North Carolina State University Raleigh, NC 27607	1
Commander, AFAL AFWAL/AADM Dr. Don Rees Wright-Patterson AFB Ohio 45433	1	Professor J. Beyer University of Wisconsin-Madison 750 University Avenue Madison, Wisconsin 53706	1
Professor Walter Ku Phillips Hall Cornell University Ithaca, NY 14853	1	General Electric Company Attn: W. Perkins Electronics Lab 3-115/B4 P.O. Box 4840 Syracuse, NY 13221	1
Commander Harry Diamond Laboratories Mr. Horst W. A. Gerlach 2800 Powder Mill Road Adelphia, MD 20783	1	Professors Rosenbaum and Wolfe Washington University Semiconductor Research Laboratory P.O. Box 1127 St. Louis, Missouri 63130	1
Advisory Group on Electron Devices 201 Varick Street 9th Floor New York, NY 10014	1		
D. Claxton MS/1414 TRW Systems One Space Park Redondo Beach, CA 90278	1		

