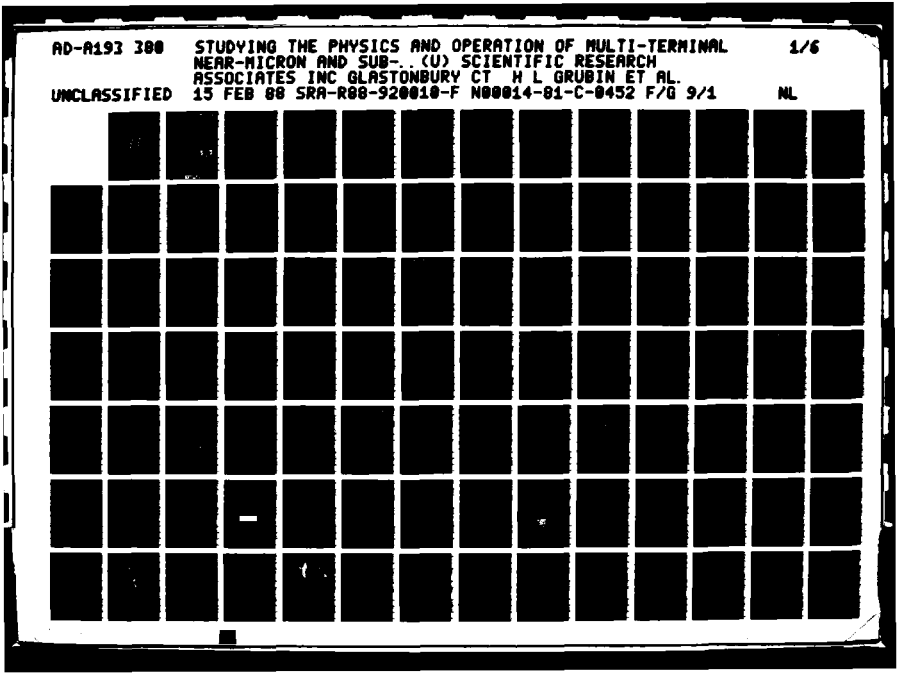
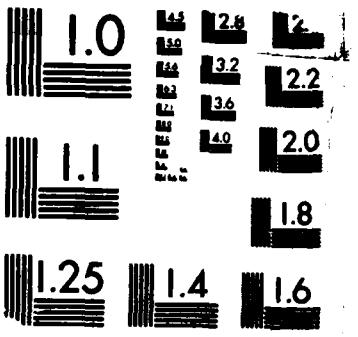

AD-A193 388 STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL 1/6
NEAR-MICRON AND SUB-.. (U) SCIENTIFIC RESEARCH
ASSOCIATES INC GLASTONBURY CT H L GRUBIN ET AL.
UNCLASSIFIED 15 FEB 88 SRA-R88-920010-F N00014-81-C-0452 F/G 9/1 NL



MICROCOPY RESOLUTION TEST CHART

BUREAU OF STANDARDS-1963-A

DTIC FILE COPY

Contract No. : N00014-81-C-0452
SRA No. : R88-920010-F

Final Report
Period: June 1981 - May 1986

STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL NEAR-MICRON AND SUB-MICRON LENGTH, HOT ELECTRON SEMICONDUCTOR DEVICES

AD-A193 380

Prepared by

H. L. GRUBIN, J. P. KRESKOVSKY, M. MEYYAPPAN, B. J. MORRISON

for

THE OFFICE OF NAVAL RESEARCH
ARLINGTON, VIRGINIA

JANUARY 1988

DTIC
ELECTE
MAR 3 0 1988
S D
E

REPRODUCTION IN WHOLE, OR IN PART, IS PERMITTED
FOR ANY PURPOSE OF THE UNITED STATES GOVERNMENT

88 3 30 074

Scientific Research Associates, Inc.
P.O. Box 1058
Glastonbury, CT 06033

This document has been approved for public release and sale; its distribution is unlimited.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) R88-920010F	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Scientific Research Associates 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) P.O. Box 1058 Glastonbury, CT 06033	7b. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER NR 619-006 1N0014-81-C-0542
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Studying the Physics and Operation of Mult-Terminal Near-Micron and Sub-Micron Length, Hot Electron Semiconductor Devices

12. PERSONAL AUTHOR(S)
H. L. Grubin, J. P. Kreskovsky, M. Meyyappan, B. J. Morrison

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM June 81 TO May 86	88 Feb 15	18 494

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Boltzmann Transport, Transients, Overshoot, GaAs, Si, InP Gallium Arsenides, Silicon

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

This document summarizes work sponsored by the Office of Naval Research under contract N00014-81-C-0452. The study encompasses a broad examination of transport in submicron and near-micron semiconductor devices through implementation of the moments of the Boltzmann transport equation and the semiconductor drift and diffusion equation. The study utilized advanced algorithms developed at Scientific Research Associates, and recommends development of a network of user based algorithms for closely combined theoretical/experimental interactions.

Keywords:

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL H. L. Grubin	22b TELEPHONE (Include Area Code) / 22c OFFICE SYMBOL

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

TABLE OF CONTENTS

		Page
	ABSTRACT	i
I.	INTRODUCTION	1
II.	INADEQUACY OF THE MOBILITY CONCEPT	2
III.	THE EQUATIONS USED IN THE STUDY	4
IV.	SUMMARY OF RECENT RESEARCH RESULTS	9

Accession For
NTIS GRA&I X
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

DTIC COPY INSPECTED 1

STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL NEAR-MICRON AND SUB-MICRON LENGTH, HOT ELECTRON SEMICONDUCTOR DEVICES

ABSTRACT

This document summarizes work sponsored by the Office of Naval Research under contract N00014-81-C-0452. The study encompasses a broad examination of transport in submicron and near-micron semiconductor devices through implementation of the moments of the Boltzmann transport equation and the semiconductor drift and diffusion equation. The study utilized advanced algorithms developed at Scientific Research Associates, and recommends development of a network of user based algorithms for closely combined theoretical/experimental interactions.

I. INTRODUCTION

During the past decade there has been a remarkable resurgence of interest in the physics of nonlinear semiconductor devices and transport. The technological reason for this interest has been the rapidly moving VLSI and VHSIC programs. These programs have been geared primarily to the development of conventionally conceived devices fabricated on a submicron scale, and have provided the primary motivation for most submicron programs. While the VLSI and VHSIC developments have provided new military and commercial gains, it was clear from the outset that potential accomplishments in this program would be limited by weaknesses in the understanding of semiconductor device physics, and by design rules based upon the mobility concept, where the carriers responded instantaneously to changes in the electric field. The scientific reason for the interest in the physics of nonlinear transport has been the increasing availability of new numerical, analytical and experimental tools, enabling a careful reexamination of the assumptions underlying semiconductor device operation.

During the tenure of this contract, workers at Scientific Research Associates, Inc. of Glastonbury, Connecticut (SRA) undertook the development of state-of-the-art algorithms to enable a suitable description of transport in micron length, submicron length and ultrasubmicron length semiconductor devices. The study involved development of the drift and diffusion equations for application to heterostructure transport, the development of the moments of the Boltzmann transport equation for application to near and submicron transport incorporating heterostructure contributions, the initiation of quantum transport algorithms for examining transport in angstrom scale devices.

Numerous papers were written and lectures given. The resulting programs were used in other government programs within the Navy, such as the Naval Research Laboratories, and within other government agencies such as: (1) the Army Research Office, (2) the Air Force Office of Scientific Research, (3) Air Force Wright Aeronautical Laboratories / Avionics Laboratory, (4) Defense Advanced Research Project Agency, (5) Defense Nuclear Agency, (6) U.S. Army, Electronics Technology and Devices Laboratory, (7) NASA / Goddard. In many cases comparisons were made to experiment to provide verification of the results. This document summarizes some of the findings of the study. However, before we proceed with a summary of the results, it is perhaps useful to discuss several specific issues associated with numerical simulation.

It is a widely accepted view, that interpretation of experiments and the design of new devices benefits from a deep understanding of the manner in which these devices operate. In the absence of any understanding, or with only marginal descriptions of phenomena, workers will proceed to design devices and develop experiments based upon intuitive notions of the results to be expected. These procedures have in the past been augmented by judicious use of numerical simulations, and there is increasing interest in enhanced use of numerical methods to deepen intuition. Unfortunately, there is a major bottleneck to efficient use of numerical methods. Presently, there are only a handful of centers developing state-of-the-art algorithms for application to submicron and ultrasubmicron devices. These algorithms are primarily research

algorithms, and are not generally accessible to large defense contractors, or to universities. Indeed, because of the research nature of the present contract much of the algorithm work undertaken during the course of the study was not developed with the idea of broad dissemination. In other words, experts are required to use these research codes. This is a weakness that must be overcome, if numerical procedures are to have the major role that they are capable of developing. One recommendation of the study, based upon the needs of workers in the field, is that procedures be established for an orderly transition of 'research-based' algorithms to 'user-based' algorithms, and that these algorithms be made available to universities for the training of young scientists and engineers, and to industry for use in engineering lines.

II. INADEQUACY OF THE MOBILITY CONCEPT

The inadequacy of the mobility concept for submicron length high speed devices was raised implicitly by Butcher and Hearn in 1968 and Rees in 1969, and by Ruch in 1972. Indeed Ruch's paper was a watershed for device transport studies insofar as it focused attention on accelerative effects of transport, and demonstrated that the response of carriers to a sudden change in electric field was considerably different from that associated with the mobility models. In fact, theoretical values of peak electron velocities nearly four times that of the steady state electron velocities of GaAs were predicted. These values emerged from transient calculations, and workers soon began to predict that electrons would be able to travel significant distances without collisions, i.e., ballistically. Experiments were reported to have measured velocity overshoot, as the excess velocity was named, and measurements of ballistic transport were purported to have been obtained. Where do we now stand with respect to the transient phenomena first reported by Ruch?

It is currently accepted by most workers involved in submicron device transport that velocity overshoot exists in all devices currently being designed. But apart from shrinking device scales, e.g., gate length and source to drain separation, and recognizing that very tiny devices are likely to lead to degradation in specific device parameters such as forward conductance, there is very little device design that has been scheduled for military or commercial use that incorporates key nonequilibrium transport phenomena. There have been many attempts including repeated overshoot structures, ballistically launched field effect transistors, the hot electron transistor, the THETA device, etc. These attempts will continue, but the devices that will likely have the most success in incorporating nonequilibrium phenomena into their design are those that pay considerable respect to one key effect absent from the early transient calculations-the influence of space charge. The influence of space charge should be apparent, because it is current that is measured in most of the structures of interest; and current = $dQ/dt = qnv$. Thus the product of carrier density and velocity enter, and the carrier density in nowhere uniform. Current studies of the hot electron transistor and of the THETA device, include examining the role of space charge on device operation, but here caution is in order. Information on operation of the device is extracted from spectral measurements, and the connection to space charge is not clear. What is needed, as an adjunct to all of these

experiments is an adjacent user-based algorithm that can address 'what if' questions. These 'what if' questions form one basis for an emergent field we have called the numerical physics of semiconductor devices.

The numerical physics of semiconductor devices serves to determine the temporal and spatial transient dependence of submicron and ultrasubmicron length semiconductor devices and to account for the environment surrounding the device. Under the present study the temporal and spatial transients have been obtained primarily through implementation of the moments of the Boltzmann transport equation. More recently Monte Carlo procedures have been implemented, as have algorithms for examining quantum transport in ultrasubmicron devices. In the discussion below, nonequilibrium transport means solutions to the moments of the Boltzmann transport equation. Equilibrium solutions mean solution to the drift and diffusion equations.

A variety of key results have emerged as a result of these studies:

- In two terminal devices, current transients are not necessarily dominated by velocity overshoot--displacement current effects can yield transients on the same time scale as velocity transients.

- In high mobility submicron devices, transport is dominated by gamma valley carriers. High speed devices require high mobility materials.

- Equilibrium saturated drift velocity is not a figure of merit for submicron devices.

- Two terminal submicron devices do not sustain negative conductance associated with electron transfer to low mobility bands.

- Nonequilibrium transport in two terminal devices leads to current drives significantly above that associated with equilibrium calculations.

- Three terminal nonequilibrium calculations demonstrate the presence of current levels that are of the order of three times that associated with steady state values. Space charge calculations are sometimes qualitatively similar to that obtained from equilibrium calculations, often they are different.

- Three terminal nonequilibrium calculations in the absence of substrate, as in the permeable base transistor, show the presence of an electron transfer induced negative conductance, as seen experimentally. Three terminal equilibrium calculations do not show the negative conductance.

In addition to the above results, the following were introduced.

- Nonequilibrium and equilibrium algorithms were modified to include the presence of heterostructure interfaces. The algorithms have been implemented.

- Algorithms were developed and implemented for solving the single particle time dependent Schrodinger equation. The equation was solved simultaneously with Poisson's equation, for tunneling devices.

. Moments of the Wigner function and of the density matrix were formulated for performing time dependent transient calculations. Algorithms for the density matrix equation were developed during the succeeding contract.

III. THE EQUATIONS USED IN THE STUDY

The specific moment equations (obtained by multiplying the Boltzmann transport equation by an arbitary function, $Q(\mathbf{k},\mathbf{r})$, and intergrating over momentum space) used in the study have been discussed in a variety of publications. Their most complete forms are discussed below.

For nonequilibrium transport, the assumption is made of a displaced Maxwellian, with parabolic, spatially dependent energy bands:

$$E = \frac{\hbar^2 \mathbf{k}\cdot\mathbf{k}}{2m^*} = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2 + k_z^2) \tag{1}$$

Additionally, the following normalization,

$$n(\mathbf{r},t) = \frac{2}{(2\pi)^3}\int_{\text{all k space}} f d^3\mathbf{k} \tag{2}$$

along with the definition, is invoked

$$n\langle Q(\mathbf{r},t)\rangle = \frac{2}{(2\pi)^3}\int Q(\mathbf{k},\mathbf{r}) f d^3\mathbf{k} \tag{3}$$

where the position dependence in <Q> is symbolic through $Q(\mathbf{k},\mathbf{r})$ and f.

The first moment equation is obtained for Q=1, and is the continuity equation:

$$\frac{\partial n}{\partial t} + \nabla_{\mathbf{r}}\cdot n\mathbf{v} = \frac{2}{(2\pi)^3}\int \left.\frac{\partial f}{\partial t}\right|_{\text{coll}} d^3\mathbf{k} \tag{4}$$

The right hand side in equation (4) represents carrier scattering. Note for a displaced Maxwellian, $\langle\mathbf{k}\rangle = \mathbf{k}_d$. The second moment is obtained for $Q = \hbar\mathbf{k}$. Then:

$$\frac{\partial}{\partial t} n\hbar\mathbf{k}_d + \nabla_{\mathbf{r}}\cdot\frac{(n\hbar^2\mathbf{k}_d\mathbf{k}_d)}{m} =$$

$$-n\nabla_{\mathbf{r}}E_c + qn\mathbf{v}\times\mathbf{B} - \nabla_{\mathbf{r}} nk_B T + \left[nE(\mathbf{k}_d) + \frac{3}{2}nkT\right]\frac{\nabla_{\mathbf{r}} m}{m} + \frac{2}{(2\pi)^3}\int \left.\frac{\partial f}{\partial t}\right|_{\text{coll}} \hbar\mathbf{k}\, d^3\mathbf{k} \tag{5}$$

where we have included the presence of a magnetic field. Gradients in E_c reflect the position dependent conduction band, gradients in effective mass reflect the position dependent effective mass. In the above equation,

$$E(k_d) = \frac{\hbar^2 k_d \cdot k_d}{2m} \tag{6}$$

It is worthwhile noting, that by setting the left side of equation (5) to zero, writing,

$$\frac{2}{(2\pi)^3} \int \left.\frac{\partial f}{\partial t}\right|_{coll} \hbar k d^3k = -\frac{n\hbar k_d}{\tau_m} \tag{7}$$

and recognizing that current density:

$$J = qn\hbar k_d/m \tag{8}$$

where

$$n\hbar k_d = n[\tau_m \nabla_r (E_c - kT \ln N) - q\tau v \times B] - m\tau_m kT \nabla_r n \tag{9}$$

we obtain current in the standard drift and diffusion formulation. A similar expression exists for hole transport, and will not be written down. The next moment is the energy moment and is obtained by considering, $Q = \hbar^2 k \cdot k/2m$

This yields

$$\frac{\partial}{\partial t} n\left(\frac{\hbar^2 k_d \cdot k_d}{2m} + \frac{3}{2}kT\right) + \nabla_r \cdot nv\left(\frac{\hbar^2 k_d \cdot k_d}{2m} + \frac{5}{2}kT\right)$$

$$= -nv \cdot \nabla_r E_c + \frac{2}{(2\pi)^3} \int \left.\frac{\partial f}{\partial t}\right|_{coll} \frac{\hbar^2 k \cdot k}{2m} d^3k \tag{10}$$

In addition, to account for non-spherical contributions, a term proportional to the gradient of electron velocity is added to the right hand side of the above equation (5), and a term proportional to the gradient of electron temperature is added to the right hand side of equation (10). These terms are:

$$+ \eta \nabla^2 v \tag{11}$$

for momentum balance, and where η is an assigned constant

$$- \kappa \text{ grad } T \tag{12}$$

for energy balance, where κ is an assigned constant

The carrier balance, momentum balance, and energy balance equations have been implemented for nonuniform field situations for two species of carriers. For uniform fields they have been implemented for five species of carriers. In addition the moment balance equations have been restructured for nonparabolic bands. The nonparabolic formulation is being prepared for publication. Some details of these results are discussed in the sections to follow.

The moment equations for nonparabolic spherical energy bands, where:

$$E(k) + \frac{E^2(k)}{E_g} = \frac{\hbar^2}{2m_1} (k_x^2 + k_y^2 + k_z^2) = \gamma(E) \tag{13}$$

take the following form:

For carrier balance, as before:

$$\frac{\partial n_i}{\partial t} + \text{div } v_i n_i = \left(\frac{\partial n_i}{\partial t}\right)_{coll} \tag{14}$$

where

$$v = \frac{\hbar k_d}{m} \frac{{}^{o}M_i^{3/2}}{{}^{o}M_o^{3/2}} \tag{15}$$

and

$$k_{M_n^m} = \int_o^\infty e^{-x_x k} \left(x + \frac{x^2}{a}\right)^m \left(1 + \frac{2x}{a}\right)^n dx \tag{16}$$

which, in the limit of infinite band gap separation, i.e., $a=\infty$, reduces to the gamma function $\Gamma(k+1/2)$. Here $a = E_g/kT$

For momentum balance:

$$\left(\frac{\partial}{\partial t} + \text{div } v_i\right) n_i \hbar k_{di} = q n_i F - \text{div } n_i k T_i \left(\overleftrightarrow{I} + \overleftrightarrow{C}\right) + \left(\frac{\partial}{\partial t} n_i \hbar k_i\right)_{coll} \tag{17}$$

where the dyadic $\overleftrightarrow{C}$ is defined implicitly as

$$n_i k_B T_e \overleftrightarrow{C}_i = - \frac{n_i \gamma}{2 M_0^{3/2}} \left[\frac{4}{5} M_{-3}^{5/2} \left(\overleftrightarrow{I} + 2\overleftrightarrow{jj}\right) - \frac{\gamma}{k_B T_e}\left(M_{-3}^{3/2} + \frac{6}{5} m_x^{5/2}\right)\overleftrightarrow{jj}\right] \tag{18}$$

For energy balance:

$$\frac{\partial}{\partial t} n_i \left(\frac{\hbar^2 k_d}{2m} + \frac{3}{2} k \frac{{}^{o}M_i^{3/2}}{{}^{o}M_o^{3/2}}\right) + \text{div } n_i \left(v\left(\frac{\hbar^2 k_d}{2m} + \frac{3}{2} k \frac{{}^{o}M_i^{3/2}}{{}^{o}M_o^{3/2}}\right)\right)$$

$$= q n_i \frac{\hbar k_d \cdot F}{m} - \text{div } n_i \left(\frac{\hbar k_d k T}{m} \left(\overleftrightarrow{I} + \overleftrightarrow{C}\right) \cdot (i + j + k)\right) + \left(\frac{\partial}{\partial t} n k T\right)_{coll} \tag{19}$$

The above equations constitute the most general form of the moment balance equations for nonparabolic bands.

The moment balance equations constitute the core of much of the work that was undertaken during the tenure of this contract, and the results are discussed within the context of reprints, which are included as part of this document.

Reprints 1 and 2 were among the first papers to address the problem of why submicron devices do not show the apparent benefits of the uniform field predictions of velocity overshoot. In examining these problems we modeled two spatially nonuniform devices, one 0.5 um long, and a second 1.0 um long. We demonstrate that overshoot, while present in these devices, provided only marginal improvements because of slowly varying electric fields.

Reprint 3 addressed the continuing issue of the presence of ballistic transport in devices, and introduced the concept of the thermal deBroglie wavelength and its implications for submicron device physics.

Reprint 4 is a broad discussion of the consequences of the intracollisional field effect.

In reprint 5, the uniform field momentum and energy balance equations are solved in the small signal limit. The noise properties of devices under hot electron conditions are examined by coupling the Langevin noise equation to the small signal results. Numerical calculations were performed for silicon and show the presence of significant contributions due to the intracollisional field effect.

In reprint 6 the extent of ballistic transport is discussed on the basis of electron correlation. The conclusion of the study is that the time duration over which the electron velocity rises in a ballistic manner is exceedingly short, perhaps only 0.004 picoseconds, even though the mean free time is much longer. The implication is that ballistic transport may occur only on a tens of angstrom spatial scale. (It is worthwhile noting that recent resonant tunneling discussions are even questioning this conclusion.)

Reprint 7 is another review paper on transport in submicron devices and was presented at the first NRL Workshop on Molecular Electronics.

Reprint 8 is a comprehensive review of the state of numerical calculations in 1981, and incorporates a discussion of the drift and diffusion equations, the moments of the Boltzmann transport equation, Monte Carlo methods, and an introduction to quantum transport through the Wigner function. The study assesses the physics of submicron transport, with particular emphasis on vewlocity overshoot.

Reprint 9 contains a discussion of the role of boundary conditions to transport in submicron devices, as well as calculations showing the dependence of scattering on applied field. Again the issue of ballistic transport is raised and shown that for applied fields of 50kv/cm, only about 35% of the carriers are unscattered at a distance of 500 angstroms. The concept of introducing a launching site to improve field effect transistor performance is raised for the first time. It should be noted, that unlike most approaches suggesting the presence of a ballistic launcher, this study suggests that the launch site be design for moderate fields only. (Recent calculations in a 1987 Army contract support the notion that strong fields associated with abrupt heterojunction launchers are inappropriate for high current levels; rather, moderate fields are desired.) In addition, scaling, based upon the concepts of Thornber, as applied to submicron devices is introduced in this paper.

Reprint 10. The role of the environment, in terms of coupled submicron structures is discussed in this paper. In addition an array of submicron devices in a lateral superlattice, and transport through a lateral superlattice using Wigner functions and Monte Carlo techniques are introduced.

In reprint 11 the effects of boundary scattering in submicron structures are investigated within the framework of balance equations arising from the Boltzmann transport equation. Integrated equations are presented in which surface dissipation can be inserted. It is shown that surface scattering strongly perturbs and distorts the distribution function with corresponding effects on transport processes.

Reprint 12 discusses the role of boundaries and interfaces on the electrical characteristics of long (greater than 10 micron) and submicron scale semiconductor devices. Here the inadequacy of using current-voltage dc measurements, in elucidating the role of the boundaries is discussed. Additionally, the role of boundaries on the noise characteristics of semiconductor devices is discussed. It is demonstrated that considerable

information concerning the nature of the nonuniform field profiles in the device can be obtained from noise measurements.

Reprint 13 includes an initial assessment of the role of experiment in measuring velocity overshoot. A significant number of questions are addressed, particularly with respect to Shanks experiments. The questions are again raised in reprint 14 where a transient calculation through solutions to the moments of the Boltzmann transport equation for nonuniform fields is performed. It is demonstrated that displacement current effects may be of the same order of magnitude as velocity overshoot effects, and that the time scales for both are similar. The results tend to add further caution to claims of measurements of velocity overshoot.

A detailed study of the role of field nonuniformities and boundaries was reviewed and constitutes part of a long paper, enclosed as reprint 15.

The moments of the Boltzmann transport equation were applied to scaling procedures. The focus of attention was the discussion of mobility versus saturated drift velocity as the figure of merit for semiconductor devices. The results of the study demonstrated that high mobility is the key factor in submicron devices, whereas saturated drift velocity is the key figure of merit in long near micron devices. This study is summarized in reprint 16.

In reprint 17 a study of InP transferred electron devices was performed with particular emphasis on the role of boundaries.

A broad review of transport and concepts of transport in semiconductor devices was presented at a NATO Summer School. This review is included as reprint 18.

More recent efforts are summarized below.

IV. SUMMARY OF RECENT RESEARCH RESULTS

HETEROSTRUCTURE TRANSPORT WITHIN THE FRAMEWORK OF A ONE DIMENSIONAL ANALYSIS

It is known that transport is governed by the properties of the boundaries, and that uncertainties in fabrication processes lead to vagaries in the electrical properties of these boundaries. One significant goal in the design of micron and submicron devices is to remove the active region of the device from such physical boundaries as the cathode (source) and anode (drain) contacts and to replace the often uncontrollable influence of thse boundaries by a controllable heterostructure interface. To initiate this study the semiconductor drift and diffusion equations were solved in one dimenison for the aluminum-gallium arsenide/gallium arsenide system. Transport in the study was normal to the interface.

The structure chosen for the study was a five micron long element, with a heavily doped n-type ($N=1.0\times10^{18}/cm^3$) AlGaAs region, 0.5 microns long, followed downstream by a 4.5 micron GaAs region ($N=5.0\times10^{16}/cm^3$). The

electrical characteristics of this device were studied as a function of bias, with the following results: The presence of the heavily doped AlGaAs resulted in charge injection at the heterostructure interface, as also expected from a heavily doped homojunction interface. The presence of the heterostructure reults in a certain amount of carrier confinement, although it is not at all clear that this confinement is significant with respect ot device operation. At low values of bias the field downstream from the interface is uniform and the current throughout the device is capable of achieving values associated with the peak steady state values of the field dependent velocity characteristics of gallium arsenide. For bias values high enough to send the carriers into the region of negative differential mobility, there is a preciptious drop in current, followed by current saturation. Accompanying this drop in current is a one cycle propagating space charge layer that comes to rest at the anode boundary. Further increases in bias results in the movement of the leading edge of the space charge layer upstream toward the heterostructure boundary. Note that the device essentially operates entirely between two steady state values of current and voltage: a prethreshold high current low voltage state, and a post threshold low current high voltage state. The switching time for device operation is transit time limited.

Note that the type of behavior exhibited above also has been predicted by the principal investigator using a simple cathode boundary field model in which the cathode field achieves values significantly below the threshold field for negative differential mobility. The result reported above may be the first theoretical manifestation of the switching characteristics using a heterostructure interface.

The above heterostructure study was expanded to include the presence of a 200nm graded heavily doped region. The results from the graded heterostructure interface were qualitatively similar to that of the abrupt interface. A dramatic change occurred when the doped graded region was replaced by a lower doped ($N^+/N^-/N$) graded region. At this point Gunn dipole layer transit time oscillation occurred. This result may represent the first simulation of transit time oscillations in heterostructure devices.

The above results are of clear interest for the development of heterostructure devices. And the design considered in the above paragraphs have been incorporated into the vertical FET structure considered by workers at Cornell. But an important word of caution is inserted here. Each of the effects considered above involve two design features, the heterostructure interface, and heavy or light doping of the spacer layer. Since heavy and light doping of homojunction interface have produced qualitatively similar results, the presence of the heterostructure for this device configuration may be to provide greater control over the behavior of the device, rather than a new operating principle.

Continued study of the role of the heterostructure interface within the context of the above one dimensional framework is suggested. The following tasks are recommended:

1. The aluminum gallium arsenide/gallium arsenide interface should be mapped out with respect to the governing boundary condition dependence of the device electrical characteristics. Thus the doping levels of the AlGaAs and the adjacent GaAs region are to be varied, as well as the length of the graded region.

2. The effects of introducing an embedded p-layer on device operation should be considered.

3. Other heterostructure combinations should be considered, such as the InGaAs/GaAs system, and the InGaP/GaAs system.

4. The role of nonequilibrium transport through solutions to the moments of the Boltzmann transport equatin and consequent velocity overshoot effect is being studied and should be continued. Implementation of solutions to the moment equations is a nontrivial matter. As discussed earlier several features enter prominantly. First, the effective mass is position dependent. Second, scattering rates are position dependent. Third, the effective fields at the heterostructure interface require generalization. In the case of the drift and diffusion equation formulation, the effective fields were obtained solely from band gap discontinuities associated with minima in the Γ valley portion of the conduction band. In multiple valley transport, conduction band, discontinuities associated with the L must also be considered. Recent conversations with Professor Karl Hess of the University of Illinois suggest that the band offset rule used for the Γ valley may find equal application for the L valley. Fourth, present studies of heterostructure transport using the drift and diffusion equations treat, e.g., the AlGaAs region and the GaAs region separately. Connection between the two is through the heterostructure fields, conductive and diffusive transport. No mechanism is introduced for the effects of scattering from one material to the second. Effects due to real space transfer should be investigated. This involves introducing scattering rates from one material to the second.

HETEROSTRUCTURE TRANSPORT WITHIN THE FRAMEWORK OF A TWO DIMENSIONAL ANALYSIS

Multi-terminal heterostructure devices such as, the HEMT, and superlattice variations of the HEMT are being studied in a number of laboratories and universities. HEMT studies performed at SRA involving solutions to the semiconductor drift and diffusion equations were reported at the 1985 WOCSEMMAD. These results displayed an unusually high level of injection into the semi-insulating GaAs. Accompanying this injection was a relatively high current density and a broad two dimensional electron gas. The results did not conform to the anticipated narrowly confined two dimensional space charge layer and caused further investigation of the assumptions. For the most part, HEMT analysis generally proceeds either analytically or through solution to the drift and diffusion equations in which the Einstein relationship is used for the carrier diffusivity. It is known that Einstein relationship is inadequate for hot carrier transport and that modifications are needed. At SRA, a field dependent diffusivity as obtained through solutions to the moments of the Boltzmann transport equation is used in the analysis. The

shape of the field dependent diffusivity has been corroborated by other studies such as Monte Carlo. Additionally, there is experimental evidence that the field dependent diffusivity bares little resemblance to the Einstein relationship. Calculations with the Einstein relationship show relatively narrow carrier confinement at the heterostructure interface as predicted by others. Calculations with the hot carrier diffusivity show a reduced carrier confinement. It must be noted that care is required in using field dependent diffusivities at zero bias conditions.

The above discussion concerning carrier confinement arises from a few computed bias points. But the variations in these results also emerged rom a comparison of calculations with the heterostructure bipolar transistors and the PBT. In all cases, there was a range of bias levels over which the assumptions associated with the diffusivity played a critical role in determining the physics of device operation and the electrical perfomrance of the device.

At this point, there are several issues of interest. First, it is necessary to map-out the effects of diffusion in determining the picture of device operation. Insofar as diffusive effects appear explicitly in the drift and diffusion equations, additional calcualtions within its context should be performed. Second, the role of carrier confinement must be explored within the framework of the Boltzmann transport equation where diffusive effects arise primarily through gradients in the electron temperature. This study should provide a good measure of the role of diffusion.

The effect of diffusivity and high field transport on HEMTs, HBTs, and PBTs were presented in a paper given at the 1986 WOCSEMMAD:

> The effects of Different Diffusivity Relationships on Transport in HEMTs, HBTs, and PBTs. M. Meyyappan, J. P. Kreskovsky, and H. L. Grubin, 1986 WOCSEMMAD.

HIGH FIELD TRANSPORT WITHIN THE FRAMEWORK OF THE MOMENTS OF THE BOLTZMANN TRANSPORT EQUATION AND MONTE CARLO

SRA has devoted a considerable amount of time and effort to the development of transport codes that solve both the semiconductor drift and diffusion equations and the moments of the Boltzmann transport equation. These algorithms have the remarkable advantage of providing reliable, space charge dependent transport device physics. Calculations were devoted to providing a basic understanding of the role of boundary conditions and material parameters on transport in GaAs one and two terminal devices. In particular, the first realistic estimate of Gunn domain transient and upper frequency limits, including space charge, were investigated. Frequencies in excess of 120GHz are possible. Additionally, two dimensional FET calculations were performed. The structure of the device is shown in figures 1 and 2, along with a plot of the current voltage characteristic at the indicated voltage levels. These calculations demonstrate the presence of overshoot. Also included in this study was the effect of a substrate. The substrate was modelled by ignoring

any scattering due to ionized or neutral impurities. As a result, the presence of the subtrate has the effect of introducing an apparent increase in total current because of a reduction in resistance. One of the interesting results is the presence of negative conductance in the source drain characteristics of a GaAs FET. Results obtained using the drift and diffusion equations which is currently the methodology performed by other researchers in the field, show no such negative conductance. The results are interesting because studies at SRA on an Air Force PBT contract, using the drift and diffusion equations, show considerable positive conductance in the source drain characteristics. This positive conductance is absent from experiment and suggests that when the MBTE is incorporated this positive conductance may be absent from the PBT calculations.

NONPARABOLIC EFFECTS IN SUBMICRON HIGH FREQUENCY TRANSPORT

Under a companion study sponsored by ARO, the moments of the Boltzmann transport equation were generalized to include nonparabolic contributions. Studies under the ARO contract were confined to the binary compounds AlAs and InAs, and the ternary compounds AlGaAs and InGaAs. The effects of nonparabolic transport on GaAs were examined under the ONR study. There were several highly significant results emerging from the GaAs study which have not appeared in any other publication concerning transport in GaAs. We remark, at the outset, that the standard low field mobility reductions associated with nonparabolic effects were corroborated in this study. The new results are summarized below.

1. Initially, the steady state calculations were performed for parabolic bands. This formed a basis for comparison. Then the nonparabolic effects were introduced. There were alterations in the scattering rates and alterations in the way the velocity was related to the momentum. Under the assumption that all coupling constants were the same for both parabolic and nonparabolic bands, it was determined that the effect of nonparabolicity was to delay the onset of electron transfer. The consequence of this was to increase the values of the nonparabolic electron velocity over the parabolic electron velocity, within the region of negative differential mobility. Very similar consequences associated with nonparabolicity were reported by Arden Sher at the DARPA Review. It is to be emphasized that the coupling coefficients for the calculation were based upon parabolic bands and the nonparabolic results at high fields offer significant differences, the coupling coefficients require alteration. The message here is that great caution must be exercised before one can be convinced of agreement between transport theory and experiment.

2. The next item of interest is the need to correct the insufficiency of velocity overshoot as the significant physical measurement. This insufficiency is not at all apparent when one is considering parabolic bands. For nonparabolic bands, use of either Monte Carlo or moment equations is predicted on determining changes in the carrier momentum, not the carrier velocity. Thus, all overshoot calculations involve momentum overshoot, not velocity overshoot. This is a nontrivial matter. For example, under uniform

fields conditions, with the exception of the scattering rates, the time rate of change of carrier density and momentum are the same for both parabolic and nonparabolic bands. However, the time rate of change of the electron temperature is different for nonparabolic bands and has the effect of altering the time to relaxation. Further, for time dependent transport, the observable is the current density rather than the momentum density. The relaxation effects associated with momentum overshoot are not linearly related to the relaxation effects associated with velocity and, hence, current density. Thus, a more general criteria for upper frequency limit of device operation is necessary.

3. It is anticipated that the moments of the Boltzmann transport equation algorithm will require generalization for nonuniform fields and nonparabolic bands.

QUANTUM TRANSPORT IN ULTRASUBMICRON DEVICES

In 1981 and 1982 the Principal Investigator published two papers concerning quantum transport in ultrasubmicron devices. The formulation of the problem was through solution to the moments of the Wigner-Boltzmann transport equation. The moment equations included scattering rates, but did not include the contributions of mixed states. Three equations were emphasized, the carrier balance equation, the momentum balance equation, and the energy equations are linearly independent, e.g., the carrier and momentum balanced Schrodinger's equation. Recently, significant advances were made in the quantum transport area.

1. Starting from the density matrix, the first three moments for an canonical ensemble of electrons were rigorously obtained. In this formulation, which incorporates mixed states implicitly, the potential seen by the carriers is the full self-consistent potential appearing in the one particle Schrodinger equation. The equations can be expressed in a completely classical framework providing a direct link to the moments of the Boltzmann transport equation. These equations can, however, be separated into a part that is purely classical in form and a part containing quantum contributions.

2. The moment equation formulation was re-expressed in three distinct representations: the energy representation, the momentum representation, and the coordinate representation. Each simplifies for specific forms of the representation dependent density matrix. In particular, displaced density matrices similar in form to the deplaced Maxwellian were examined.

3. The density matrix formulation was generalized to incorporate the effects of dissipation. In this particular formulation, the question of how many balance equations are required to obtain an approximate answer for transport in ultrasubmicron devices was formulated. This latter formulation will be the one of choice for examining quantum transport in multiple quantum wells. These equations are discussed below.

QUANTUM TRANSPORT FORMULATION

For a single particle Hamiltonian that includes scattering:

$$H = \frac{P^2}{2m} + V(x,t) + H_{sc} \tag{20}$$

the following three moment equations resulted:

carrier balance

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x} nv = \frac{i}{\hbar} < \left[H_s,\ P^{(0)}\right] >, \tag{21}$$

momentum balance

$$\frac{\partial}{\partial t} nmv + \frac{\partial}{\partial x} v(nmv) = e \frac{\partial V}{\partial x} n - \frac{1}{m} \frac{\partial}{\partial x} \Omega_{xx}$$

$$+ \frac{\hbar^2}{4} \sum \rho_{ii} n_i \frac{\partial^2 \ln n_i}{\partial x^2} + \frac{i}{h} <\left[H_s, P^{(1)}\right]> \tag{22}$$

and energy balance

$$\frac{\partial}{\partial t} \left[\Omega_{xx} + m^2 nv^2\right] + \frac{\partial}{\partial x} \left[v(\Omega_{xx} + m^2 nv^2)\right]$$

$$= 2e \frac{\partial V}{\partial x} (nmv) - 2 \frac{\partial}{\partial x} v\Omega_{xx} - 2m \frac{\partial}{\partial x} Q$$

$$+ \frac{\hbar^2}{4} \sum \rho_{ii} \left\{3mv_i \frac{\partial^2}{\partial x^2} \ln n_i + \frac{\partial^2}{\partial x^2} mv_i\right\} + \frac{i}{\hbar} <\left[H_s,\ P^{(2)}\right]> \tag{23}$$

where Ω is a stress tensor. The first two of the above equations were formulated under the ONR study and has been programed for tunnel structures under a new study.

Results of initial quantum transport formulations were presented at the Arizonia State University Workshop on Quantum Transport.

> Quantum Transport and Moment Equations in Quantum Wells Structures, H. L. Grubin, 1985 ASU Workshop on Quantum Transport.

In addition to the above studies that have been published there are a number of papers that are currently under preparation. In one paper, HEMT calculations were performed, and the role of diffusivity on the results explored. This paper has been submitted for publication. An extensive discussion of the effects of nonparabolicity on the moment equations is near completion. A study of the upper frequency limit for Gunn oscillations using the moment equations and applied to InP was completed with additional support from the Naval Research Laboratory. This study is being prepared for publication. Additionally, hole transport is included in several algorithms.

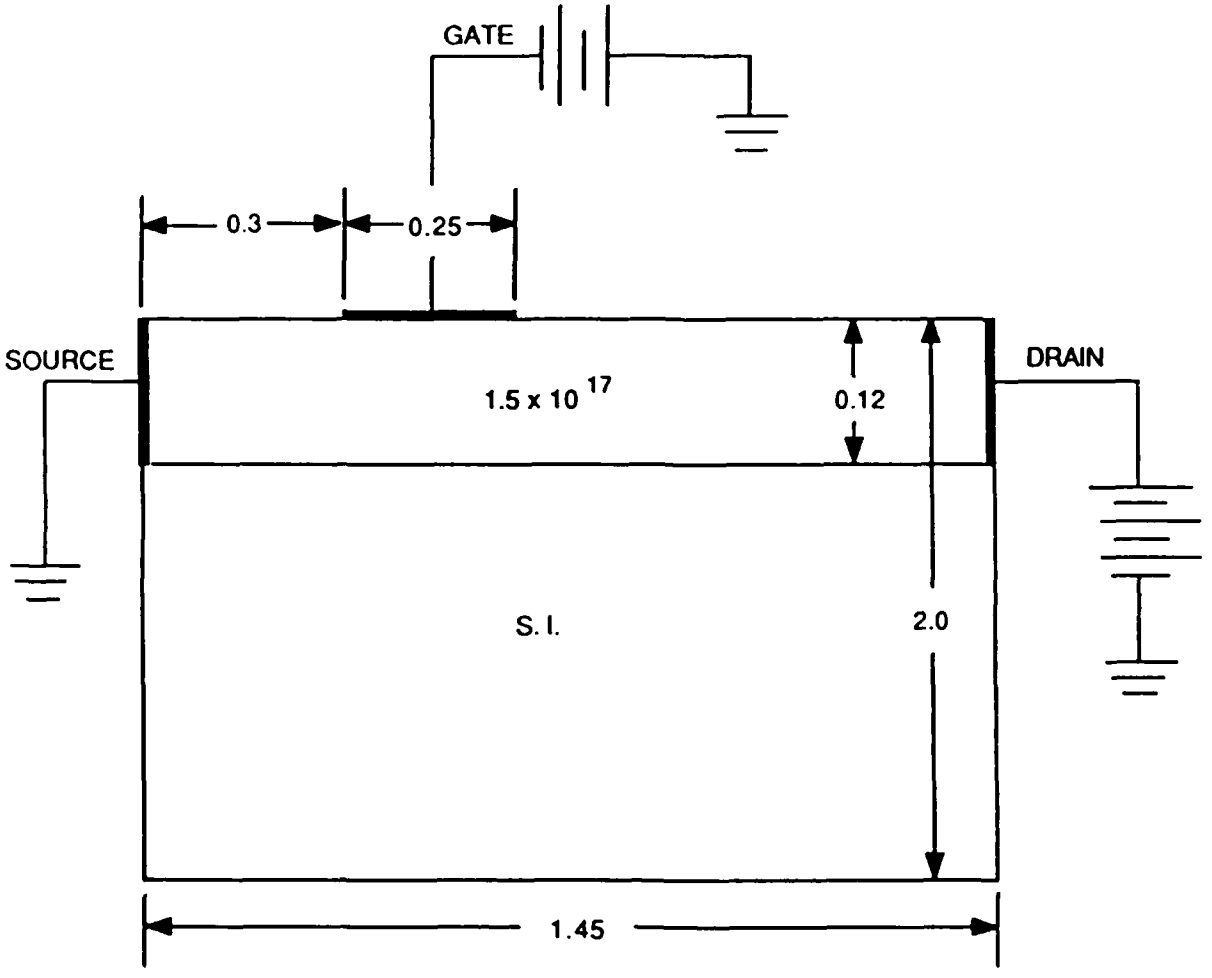


Figure 1. Schematic of GaAs FET on S.I. Substate.

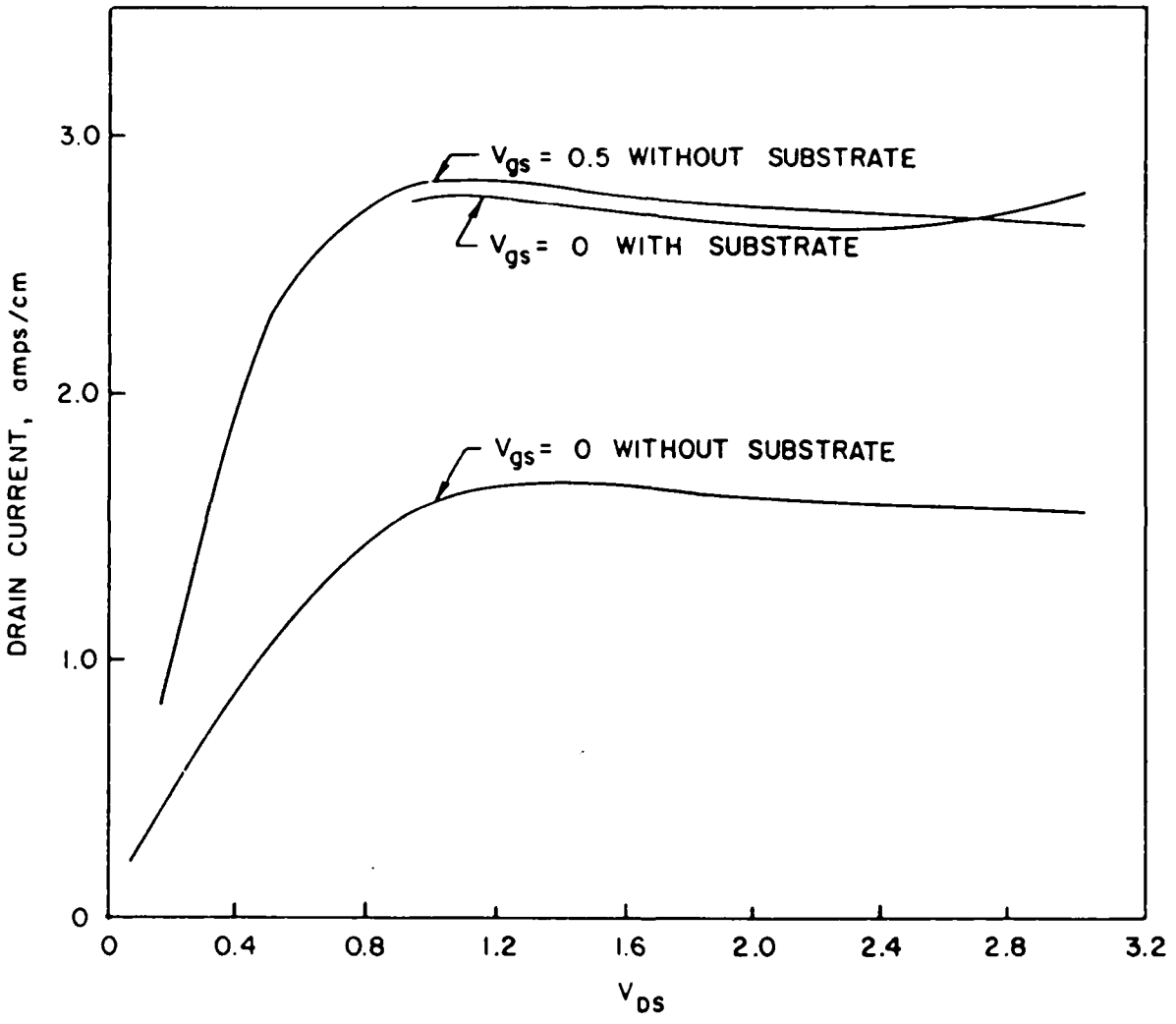


Figure 2. Current-voltage characteristics of a GaAs FET with and without consideration of the S.I. substrate.

Reprint 1

JOURNAL DE PHYSIQUE

Colloque C7, supplément au n°10, Tome 42, octobre 1981 page C7-201

HOT CARRIER SPACE AND TIME DEPENDENT TRANSIENTS IN SHORT CHANNEL GALLIUM ARSENIDE DEVICES *

H.L. Grubin, G.J. Iafrate* and D.K. Ferry**

Scientific Research Associates, Inc. P.O. Box 498, Glastonbury, CT 06033, U.S.A.

**U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey, U.S.A.*

***Colorado State University, Fort Collins, Colorado, U.S.A.*

Résumé - Nous examinons ici les phénomènes transitoires de transport dans l'arseniure de gallium lorsque les champs varient dans le temps et l'espace avec une vitesse finie. Pour des variations temporelles et des champs uniformes ; l'on montre que le pic de survitesse est sensible au temps de montée de la polarisation. Pour des variations dans l'espace la vitesse moyenne des porteurs est extrêmement sensible aux gradients et quoique l'on puisse obtenir une survitesse, celle-ci est plus faible que celle obtenue dans un champ uniforme.

Abstract - We examine transient carrier transport in gallium arsenide when the fields change temporally and spatially at a finite rate. For temporal changes and uniform fields the peak overshoot velocity is shown to be sensitive to bias rise times. For spatial changes the mean carrier velocity is extremely sensitive to gradients, and while overshoot can occur it is also below the peak uniform field value.

Introduction - Among the earliest papers to deal with transient carrier transport (TCT) and to include overshoot contributions were those of Butcher and Hearn[1] and Rees[2]. These studies were followed by Ruch's[3], whose significance was highlighted by Frey and coworkers[4] in a long series of papers. These calculations established that on short time scales (of the order of an LO phonon intravalley scattering time for GaAs) the mean velocity of an ensemble of carriers could attain values of velocity substantially greater than their steady state values. These results have since generated a great deal of scientific and technological interest. However, while experimental data[5,6] are consistent with the concept of velocity overshoot, there are difficulties in directly linking its effect to the operating properties of room temperature devices, particularly FETs. Although this absence of linkage may be partly due to limitations of resolution we think part of the problem lies in the unfinished picture of TCT. At the center of this is the fact that virtually all discussions designed to isolate overshoot phenomena are uniform field calculations in which carriers respond to fields that suddenly change from one value to another. The results of these calculations are useful in providing upper bounds for the peak transient velocities. But, they do not provide a realistic estimate of the true transient velocities within a device. The theoretical data for device design is thereby limited, and we require more representative device/circuit simulations t

* Supported by the U.S. Office of Naval Research

account for TCT and modifications arising from fields that change temporarily and spatially at a finite rate. Some spatial studies have already appeared,[7-12] but we are short of a consensus as to a phenomenology of transport within submicron devices. The purpose of this paper is to isolate some of the details of TCT for gallium arsenide when fields change at finite rates and to suggest the types of calculations that would be useful for providing a viable phenomenology. The results should be particularly relevant in the design of such devices as a submicron field effect transistor.

The vehicles for our discussion are solutions to the first three moments of the Boltzmann transport equation (MBTE), assuming a displaced Maxwellian distribution[13]. Within this framework we attach no meaning to ballistic transport, since implicit in our assumptions are strong electron-electron interactions. We examine two extreme cases: (1) uniform fields, where all spatial derivatives approach zero ($\frac{\partial}{\partial x} \to 0$) and (2) the steady state nonuniform fields, where all time derivatives are zero ($\frac{\partial}{\partial t} = 0$). For both cases the device simulated is part of a resistive circuit. The device, as represented by MBTE, is in parallel with a geometric capacitance, both of which are in series with a resistor and a dc source.

Uniform Field Temporal Transients - For uniform electric fields a host of calculations[3,4,14] have been performed in which a typical electron with zero mean initial velocity responds to an electric field of finite value. We interpret this calculation to mean that a collection of carriers in thermal equilibrium with the background lattice is suddenly subjected to a field whose value increases from zero to a finite value in zero time (or numerically in one time step). The results of this calculation are represented in Fig. 1a, where they are taken as having the significance of an upper bound. The remaining curves which were obtained for a bias turned on at a finite rate show an expected reduction in peak velocity. The results at dF_o/dT = 4kv/cm/30ps (Fig. 1b) are near those expected from a calculation in which the steady state field dependent velocity relation is used.

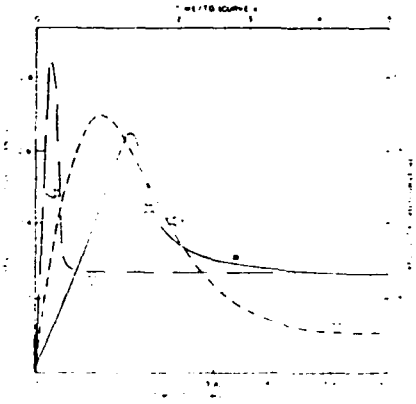


Fig. 1: Uniform v(t) calculations with bias rise time, T_b, as a parameter. The maximum bias field is 21.5kv/cm. For (a) the maximum field is reached in one time step. For (b) T_b=160 ps. (c) T_b=16ps. v_s=3.7x10^{7}cm/sec. T_o=0.32ps.

The point of Figure 1 is an extrapolated one. A field changing at a finite rate either temporally or spatially has a reduced overshoot contribution. If we artificially introduce a spatial dependence through the relation d = vx30 ps, then for $v = 2 \times 10^7$ cm/sec, d = 6 microns, and a field increasing by 4kv/cm over this distance the velocity will not exhibit significant overshoot. Similar field changes over a distance of 0.6 microns cause nonequilibrium effects to appear but the peak velocity shows only a 30% increase, much below that of curve 1a. The results are particularly relevant because these field and velocity changes are what we may expect over a major fraction of a one micron gate length FET. More rapid field changes such as that over a shorter gate region are likely to increase overshoot but here field and carrier temperature gradients will make major contributions and in some cases virtually eliminate the effects of overshoot[15].

Nonuniform Field Spatial Transients - From uniform field calculations we have become accustomed to thinking in terms of velocity versus time for transient calculations. In the absence of reactive circuit elements this scales current versus time which, in principle, is experimentally accessible. For submicron devices in which spatial contributions are significant, current no longer scales velocity. Velocity, while now difficult to measure, is still useful for phenomenology. We discuss this within the bounds of a one-dimensional calculation in which we calculate current versus time: $j(t)=j_n(x,t)+\epsilon \frac{\partial F(x,t)}{\partial t}$, with

$$j_n(x,t) = e(n_c(x,t)v_c(x,t) + n_s(x,t)v_s(x,t))$$

Through current continuity, j(t) is spatially independent.

One dimensional solutions to the Boltzmann transport equation have appeared in the literature[7-11]. What we emphasize is the micron scale space chage distribution within steady state, although we also include the time evolution to steady state. The $\frac{\partial}{\partial t} \rightarrow 0$ solutions represent the kind of information we need for FET and other device modeling.

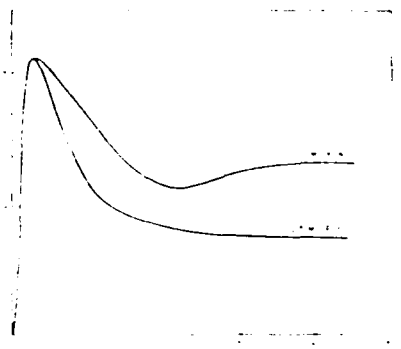


Fig. 2: Nonuniform field current versus time with $J_o = N_s e v_o$.

In examining spatial transients we become aware of significant differences from that of the uniform field calculations. In the uniform field calculations it is the field driven steady state electron temperature that determines the distribution of carriers in the central and satellite valley. For nonuniform fields where they are strong gradients in space charge, field, etc., these significantly modify the uniform field electron temperature dependence. Two distinct sets of computations illustrate the above idea. In these calculations, two elements were chosen, each with a doping level of $10^{17}/cm^3$. One element was 1 micron long, the second, 0.5 microns. For each, a doping depression of $0.9 \times 10^{17}/cm^3$ was introduced. For the 1 micron element it began at 0.1 microns downstream from the source and continued another 0.2 microns. For the 0.5 micron element the beginning was at 0.05 microns and continued for 0.1 microns. Each nonlinear element was in series with a resistor and a dc source. The magnitude of the dc source was chosen so that, should the fields be uniform, the average field across each would be equal in value. Figure 2 shows current versus time for the elements. These profiles show a remarkable similarity to the uniform velocity-time profiles, tending to camouflage a rich space charge distribution. The calculations for both cases displayed velocity overshoot during the initial transient; only part of which can be attributed to the temporal contributions. In steady state, the shorter element exhibited no spatial overshoot while the longer element did.

The differences in these results are attributed to quantitative differences in the space charge distribution as illustrated in Figures 3 and 4. By comparison, the shorter element has steeper gradients in carrier density, higher fields, carrier temperature and enhanced transfer across the notched region. The mean velocity (not shown) is near saturation even though the field across the notch is not high enough for saturation, as determined by the uniform field calculations. The question of negative differential mobility therefore arises and we examined it by extracting contributions from the derivative of electron pressure, a term which tends to behave as a diffusion current, from the total current. The remaining term is formally the familiar conduction current (some contributions remain from derivatives of the mean kinetic energy) and so should indicate the presence or absence of NDM. With the exception of extrema, NDM is marginal. Across the right hand part of the element where the field changes more gradually, there were traces of NDM, but it was not clear to what extent this result was model dependent. For comparison we refer to an earlier uniform field study where we concluded that transfer across a 0.1 micron region would not likely yield NDM[16].

For the longer element we display the mean carrier velocity $v(x,t) = j_n(x,t)/en(x,t)$, which is a more general expression than the uniform field velocity and includes diffusion implicitly. For this element point-by-point correspondence between $v(x,t)$ and $F(x,t)$ displays spatial overshoot. However, these excess values of velocity include significant contributions from gradients (See also Ref. 15). Indeed, it is more to discuss a carrier velocity through the familiar phenomenological

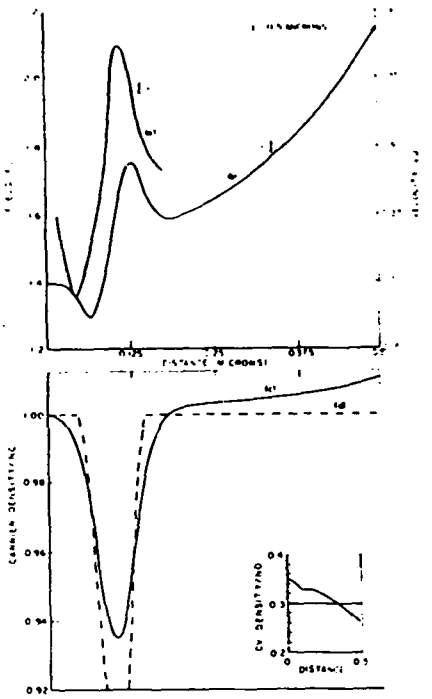


Fig. 3: Spatial distributions (a) conduction velocity, (b) internal field, (c) mobile carrier density (d) background density, (inset) central valley density.

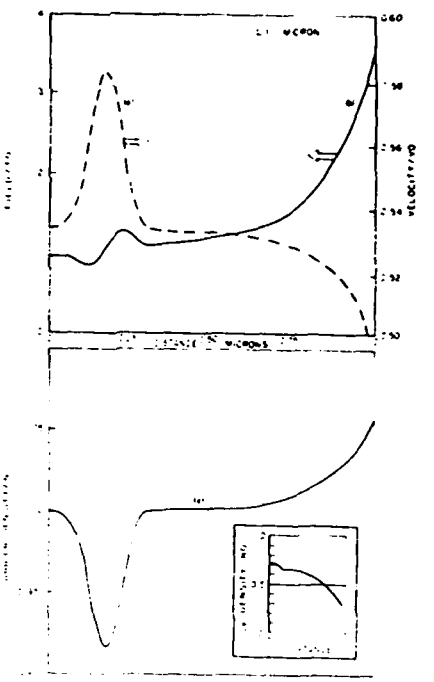


Fig. 4: Spatial distributions: (a) mean carrier velocity, (b) internal field, (c) mobile carrier density, (inset) central valley density.

steady state (ss) conduction plus diffusion relation $v = v_{ss} - \frac{D}{n}\frac{\partial n}{\partial x}$, then near a depression in n, v should be greater than v_{ss}.

Conclusion - What are these results teaching. On one level we can interpret numerical results, as above. On a second level we can infer some general patterns. Here, if we go back to the late sixties and the interpretations of transferred electron device behavior, we recall the conclusion that significant, locally confined, spatial non-uniformities control device behavior. The above results are consistent with these earlier NDM studies[17]. For here, the shorter element sustained a larger field gradient near one of the boundaries, enhanced electron transfer, saturation in carrier velocity, and a low net current, while in the longer element the field near one of the boundaries is lower, there is a more equitable sharing of potential between the resistor and the nonlinear element and the current is higher. (In sub-micron devices we may expect some space charge propagation. For the longer element discussed here there was still some residual time dependence but it was not examined.) The results also teach that the achievement of high speeds in practical devices will necessarily require the incorporation of spatial transients into device modeling.

References

1. Butcher, P.N. and C.J. Hearn, Electronics Letts. 4, 459 (1969).
2. Rees, H.D., IBM J. Res. Develop. 13, 537 (1969).
3. Ruch, J.G., IEEE Trans. Electron Devices, ED-19, 652 (1972).
4. Maloney, T.J. and J. Frey, J. Appl. Phys. 48, 781 (1977) and S. Kratzer and J. Frey, J. Appl. Phys. 49, 4064 (1978).
5. Glover, G.H., Appl. Phys. Letts. 18, 290 (1971).
6. Shank, C.V., R.L. Fork, B.I. Greene, F.K. Reinhart and R.A. Logan, Appl. Phys. Letts. 38, 104 (1981).
7. Jones, D. and H.D. Rees, J.Phys. C. 6, 1781 (1973).
8. Lebwohl, P.A. and P.J. Price, Solid St. Comm. 9, 1221 (1971).
9. Bosch, R. and H. Thim, IEEE Trans. Electron Devices, ED-21, 16 (1974).
10. Rolland, P.A., M.R. Friscourt, E. Constant and G. Salmer, J. de Phys. (In Press).
11. Grubin, H.L., D.K. Ferry, G.J. Iafrate and J.R. Barker, in Microstructure Science and Technology/VLSI, N.Einspruch, ed. Academic Press, N.Y. (In Press).
12. Kroemer, H., Solid State Electronics 21, 61 (1978).
13. Blotekjar, K., IEEE Trans. Electron Devices ED-17, 38 (1970).
14. Grubin, H.L., D.K. Ferry, J.R. Barker, M.A. Littlejohn, T.H. Glisson and J.R. Hauser, Proc. 4th Biennial Cornell University Electrical Engineering Conf. (1979).
15. Norton, D.E. and R.E. Hayes, Proc. 3rd Biennial Cornell University Electrical Engineering Conf. (1977).
16. Grubin, H.L., D.K. Ferry and J.R. Barker, Proc. IEDM, 394 (1979).
17. See, e.g., M.P. Shaw, H.L. Grubin and P.R. Solomon, The Gunn-Hilsum Effect, Academic Press, N.Y. (1979).

Reprint 2

Reprinted from INTERNATIONAL ELECTRON DEVICES MEETING, December 1981

SPATIAL AND TEMPORAL CONSTRAINTS ON TRANSPORT IN SUBMICRON GALLIUM ARSENIDE DEVICES

H. L. Grubin*
Scientific Research Associates, Inc., Glastonbury, Connecticut

G. J. Iafrate
U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey

D. K. Ferry
Colorado State University, Fort Collins, Colorado

ABSTRACT

The purpose of this paper is to demonstrate that peak overshoot effects in gallium arsenide are reduced when the fields change temporally and spatially at a finite rate. For temporal changes and uniform fields the peak overshoot velocity is shown to be sensitive to bias rise times. For spatial changes the mean carrier velocity is extremely sensitive to gradients, and while overshoot can occur it is also below the peak uniform field value.

I. INTRODUCTION

The high technological interest in submicron length devices stems from our increased ability to store and process more information on a chip of a given size. But the shrinking device introduces a host of new and exciting problems, the most direct one - how does a submicron device work? While on an ultra-small scale device-device interactions are likely to lead to synergistic behavior, where the whole behaves differently than its parts, most submicron device modeling is concerned with the constituent part - the discrete device.

For the discrete submicron device there are several critical parameters, some of which are listed below. The point of the scales is that whenever the extrinsic and bulk parameters overlap, the usual drift and diffusion modeling equations are called into question. Whenever the concept; e.g., of a wavepacket or Fermi's Golden Rule is questioned, the Boltzmann transport equation is scrutinized.

TABLE 1. Approximate spatial scales (GaAs)

Term	Value
Active region length	10^{-5}cm, 10^{-4}cm
Impurity separation ($N_o^{-1/3}$)	10^{-5}cm @ 10^{15}/cm^3 10^{-6}cm @ 10^{18}/cm^3
Mean free path	10^{-5}cm
Thermal deBroglie wavelength	2.6×10^{-6}cm

TABLE 2. Approximate temporal scales (GaAs)

Term	Value
Transit time(for v=2×10^7/cm/s)	5×10^{-13}s @ L=10^{-5}cm 5×10^{-12}s @ L=10^{-4}cm
Momentum relaxation	3×10^{-13}s @ 300K
Collision duration	2×10^{-14}s

Device modeling on time scales of the order of fractions of picoseconds has been extensive. With the exception of a few, most studies have been concerned with the limiting situation in which the carriers respond to sudden changes in electric field. These limiting situations have provided us with peak momentum-relaxation-time dominated velocities, far in excess of their steady state values. There are conceptual difficulties in relating these limiting values of velocity to the operation of present day devices. Indeed, there may not be a pressing need to do so. What appears necessary at this point is the development of a collection of results that illustrates the effects of relaxing the sudden field concept. We have begun such a collection. And when we combine our recent results with results of others, we are drawn to the conclusion that overshoot velocities are dramatically reduced when the field, as seen by the electrons, changes at a finite rate. This paper is a summary of these results.

II. FORMULATION

Our results are obtained numerically for a two valley GaAs[1,2] element through self consistent solution to the first three moments of the Boltzmann Transport Equation (BTE), which we write in the form first discussed by Blotekjar[2]:

Carrier Density (i = 1,2)

$$\frac{\partial N_i}{\partial t} + \nabla\cdot(\bar{v}_i N_i) = \left(\frac{\partial N_i}{\partial t}\right)_c$$

Momentum Density $\bar{P}_i = m_i N_i \bar{v}_i$

$$\frac{\partial \bar{P}_i}{\partial t} + \nabla\cdot(\bar{v}_i \bar{P}_i) = -eN_i\bar{F} - \nabla(N_i k T_i) + \left(\frac{\partial P_i}{\partial t}\right)_c$$

Energy Density $W_i = \frac{3}{2}N_i k T_i + \frac{m_i}{2}N_i v_i^2$

$$\frac{\partial W_i}{\partial t} + \nabla\cdot(\bar{v}_i W_i) = -eN_i\bar{v}_i\cdot\bar{F} - \nabla\cdot(\bar{v}_i N_i k T_i) - \nabla\cdot\bar{q}_i + \left(\frac{\partial W_i}{\partial t}\right)_c$$

The bracketed terms on the RHS of the above equations represent scattering integrals. And each of the above are solved simultaneously with Poissons equation and the circuit equation. In solving these equations we have found a convenient set of normalizing parameters. They appear in Table 3.

CH1708-7 81/0000-0633 $00.75 © 1981 IEEE

TABLE 3. Normalization

$t = \tau_o t'$	τ_o = scattering time
$x = v_o \tau_o x'$	$v_o = (2k_o T_o/\pi_o)^{1/2}$
$p_i = p_o p_i'$	T_o = ambient temperature
$v_i = p_i/m_i$	$p_o = m_o v_o$, $P_o = N_o p_o$
$n_i = N_o n_i'$	m_o = c.v. effective mass
$T_i = T_o T_i'$	N_o = background doping
$w_i = N_o k_o T_o w_i'$	$F_o = m_o v_o / e\tau_o$
$F = F_o F'$	$\kappa_o = k_o N_o v_o^2 \tau_o$
$\kappa = \kappa_o \kappa'$	
$m_i = m_o m_i'$	
$\left(\frac{\partial}{\partial t'} + \frac{\vec{\nabla}'}{m_i'} \vec{p}_i'\right) \equiv \frac{D^i}{Dt'}$	

$$\frac{D^i}{Dt'} n_i' = \left(\frac{\partial n_i'}{\partial t'}\right)_c$$

$$\frac{D^i}{Dt'} n_i' \vec{p}_i' = -n_i' \vec{F}' - \frac{\text{grad}' n_i' T_i'}{2} + \left(\frac{\partial p_i'}{\partial t'}\right)_c$$

$$\frac{D^i}{Dt'} w_i' = -2n_i' \vec{p}_i' \cdot \vec{F}' - \frac{\nabla_i'}{m_i'}(\vec{p}_i' n_i' T_i') + \kappa' \nabla'^2 T_i' + \left(\frac{\partial w_i'}{\partial t'}\right)_c$$

The advantage we find for the above is that the relative contribution of each term is readily accessible. The one problem offering something of a dilemma in the above is the contribution of the thermal conductivity. This term has been regarded by Blotekjar as a highly significant non-Maxwellian contribution. As with others (3) we have found this a welcome help in assuring stability of some numerical calculations. The problem is what to choose for a value for κ, aside from the fact that we have taken it to be constant. We can get some help from the Wiedmann-Franz Law $\kappa = \mathcal{L} \sigma T k^2/e^2$. Thus $\kappa' = \frac{1}{2}\mathcal{L}\langle\tau\rangle/\tau_o$ and we expect the normalized thermal conductivity to be less than unity. However, in the calculations discussed below we have virtually ignored it. Its effect will be examined later.

III. UNIFORM FIELD TRANSIENTS

The first set of calculations are for uniform fields. In this case the nonlinear element was in series with an inductor, a resistor and a dc source. A geometrical capacitance was placed across the device. For the case where the inductor was absent we computed the response of the nonlinear element to an external source that reached a maximum value $\mathcal{E}_{max}$ as a controlled rate. Two extreme cases are shown in figure 1. Curve 'a' of figure 1 yields the familiar "high" velocity overshoot results(4,5), and here the bias reached its peak value in one time step. For curve 'b' the maximum bias was reached after a time of 160 ps had elapsed. There is a dramatic reduction in peak velocity. This result is not at all surprising when we recall that overshoot is primarily a consequence of differences in the energy and momentum relaxation times. If the device voltage changes on a time scale comparable to the energy relaxation time, overshoot effects are considerably reduced.

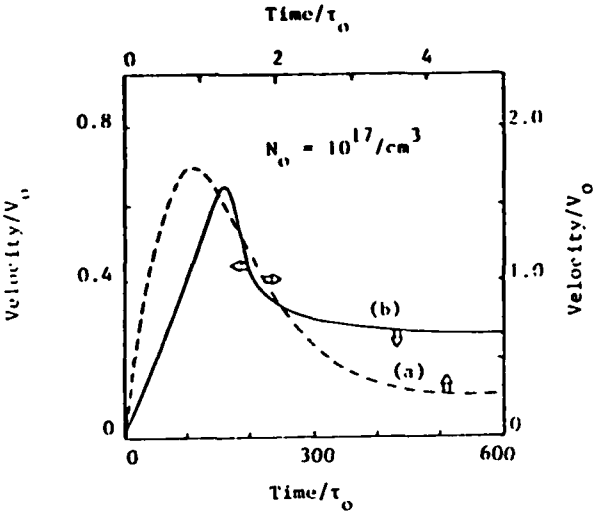


Fig. 1: Uniform v(t) calculations with bias rise time, T_b, as a parameter. The maximum bias field is 21.5kv/cm. For (a) the maximum field is reached in one time step. For (b) T_b=160 ps. v_o=3.7x10^7 cm/sec. T_o=0.32ps.

The above results are consistent with the unnoticed overshoot calculations of Ref. 6, where the GaAs element responded to a driving ac signal. We present a slight variant of that result in figure 3, where we show the results of a self-excited relaxation oscillation. The details of the self-excited oscillation have been chartered by many authors (7). It suffices to state that the circuit now contains an inductor, and when the oscillation occurs the current-voltage pattern is that of a Lissajous figure. When the semiconductor drift and diffusion equation are used the velocity versus field relation exhibits no hysteresis. When the BTE is used to compute the velocity, hysteresis and overshoot occur, as shown below. But as seen the overshoot is substantially below the peak of that of figure 1.

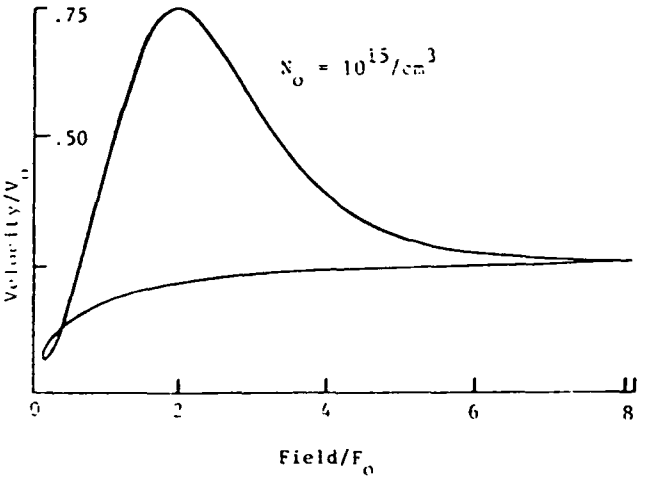


Fig. 2: Velocity versus field for an 80 GHz relaxation oscillation.

The question to ask is what can we expect at higher frequencies where the field changes more rapidly. A direct answer is somewhat difficult of this moment. We note that it takes the electrons a finite period of time to respond to any changes in electric field. During the time the carriers sense a high field, the latter may already have turned around. Further complications arise in that the way the field changes is constrained by the external circuit. In any case, preliminary studies show that increasing the rate of change of field in an oscillating circuit does not automatically result in a higher overshoot.

IV. NONUNIFORM SPATIAL TRANSIENTS

When spatial nonuniformities are present in the semiconductor the interpretation of nonequlibrium contributions becomes more complex. Essentially what is needed here is a cataloguing of results of model calculations. Some, have already appeared [8]. Below we discuss spatial transients in devices for which all temporal transients approach zero (i.e., $\frac{\partial}{\partial t} \rightarrow 0$). This is the opposite extreme to the temporal transients where all spatial gradients are zero.

The configuration we have chosen will be familiar to those involved in Gunn devices during the sixties. We assume a jellium distribution of donors with a density of $10^{17}/cm^3$ and a 10% "notch" downstream from the cathode. The jellium distribution must be viewed suspiciously for ultra-small devices (see Table 1). The lengths of the elements are 0.5 and 1.0 μm, with notched widths of 0.1 and 0.2 microns, situated, respectively, 0.05 and 0.1 um downstream from the cathode. Each nonlinear element was in series with a resistor, and a dc source whose magnitude was chosen so that should the fields be uniform the average field across each element would be the same. The initial transient was virtually indistinguishable from the uniform field case. The long time, $\frac{\partial}{\partial t} \rightarrow 0$, limit showed significant structure reminiscent of "cathode notched" supercritical amplifiers.[9] The results are shown in figures 3 and 4, where we display the magnitude of the field and the net mobile carrier density. For the longer element we display the mean carrier velocity v(x,t) which is more general than the uniform field velocity and implicitly includes diffusion. For the 0.5 micron element we display the conductive velocity which is obtained by subtracting the term containing $\frac{\partial}{\partial x}$ $\frac{\partial}{\partial x}$ (nT) from v(x,t). The conductive velocity tends to isolate negative differential mobility (NDM) effects, if any.

To examine these results we also display some supplementary calculations. Figures 5a and b display central valley (CV) results for 0.5 and 1.0 micron devices. Figure 6 shows results for a uniform field calculation where the bias field rose at a slow rate. From figures 3-5 we can extract N_c versus F and compare them to the uniform field value results. With the exception of the immediate region of the notch the results are somewhat similar. For the short element, the fields near the notch are high and cathode electrons are predominantly in the satellite valley. As a result,

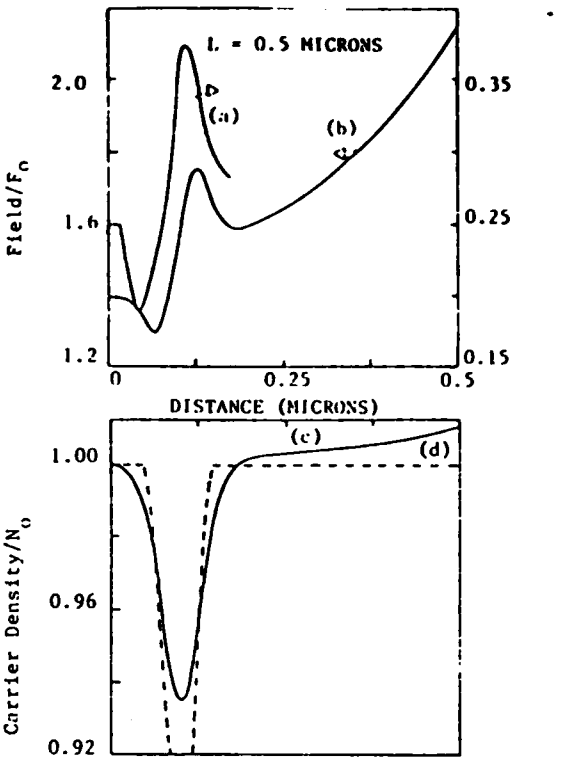


Fig. 3: Spatial distributions (a) conduction velocity, (b) internal field, (c) mobile carrier density, (d) background density.

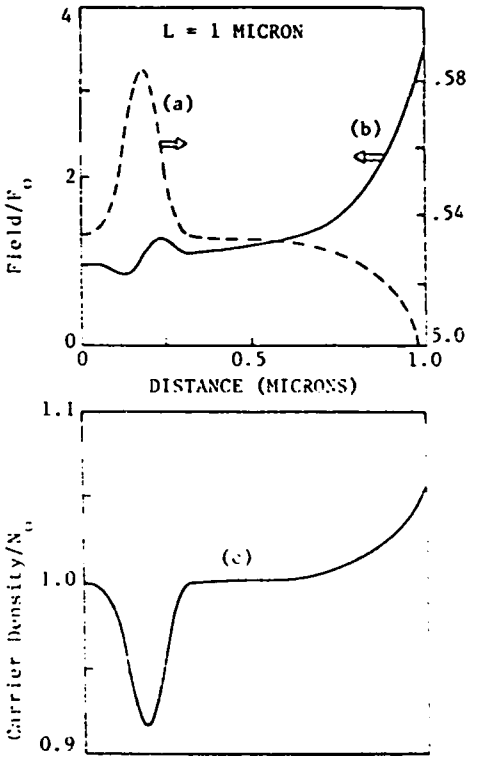


Fig. 4: Spatial distributions: (a) mean carrier velocity, (b) internal field, (c) mobile-carrier density.

apart from a marginal NDM region near the cathode there are no overshoot contributions. The longer element displays more interest. Here the cathode fields are low enough to allow a significant fraction of carriers to remain the CV. Initially the CV momentum and temperature correspond closely to the uniform field case. Decreasing, and then increasing fields (towards the anode) result in corresponding changes in energy and momentum. But these spatial changes do not completely follow the field, there is some spatial relaxation. Thus, while reductions in field near the cathode produce lower values of electron temperature, they are not as low as the uniform field results suggest. Increases in field are followed by increases in temperature, but not as great as that suggested by the uniform field analysis. As a result, the electrons are cooler than in uniform field equilibrium and overshoot occurs. We have also examined a similar set of supplementary calculations for the satellite valley. Because of the significantly higher scattering rates the results are similar to those obtained using the steady state uniform field values of velocity.

V. CONCLUSIONS

The basic conclusion of this study was stated in our introductory remarks. Reiterating, non-equilibrium overshoot effects are present in submicron devices. They occur because of spatial and temporal variations within the device, coupled to strong differences in energy and momentum relaxation times. However, to extract the maximum speed from these devices new and very careful designs will be necessary.

REFERENCES

1. Blotekjar, K., IEEE Trans. Electron Devices ED-17, 38 (1970).
2. Grubin, H.L., D.K. Ferry and J.R. Barker, Proc. IEDM, 394 (1979).
3. Bosch, R. and H. Thim, IEEE Trans. Electron Devices, ED-21, 16 (1974).
4. Ruch, J.G., IEEE Trans. Elect. Devices, ED-19, 652 (1972).
5. Maloney, T.J. and J. Frey, J. Appl. Phys. 48, 781 (1977) and S. Kratzer and J. Frey, J. Appl. Phys. 49, 4064 (1978).
6. Butcher, P.N. and C.J. Hearn, Electronics Letts. 4, 459 (1969).
7. See, e.g., M.P. Shaw, H.L. Grubin and P.R. Solomon, The Gunn-Hilsum Effect, Academic Press, N.Y. (1979).
8. Norton, D.E. and R.E. Hayes, Proc. 3rd Biennial Cornell Univ. Elect. Engn. Conf. (1977).
9. Spitalnik, R., M.P. Shaw, A. Rabier, J. Magarshack, Appl. Phys. Letts. 22, 162 (1973).

*Supported by the U.S. Navy Office of Naval Research

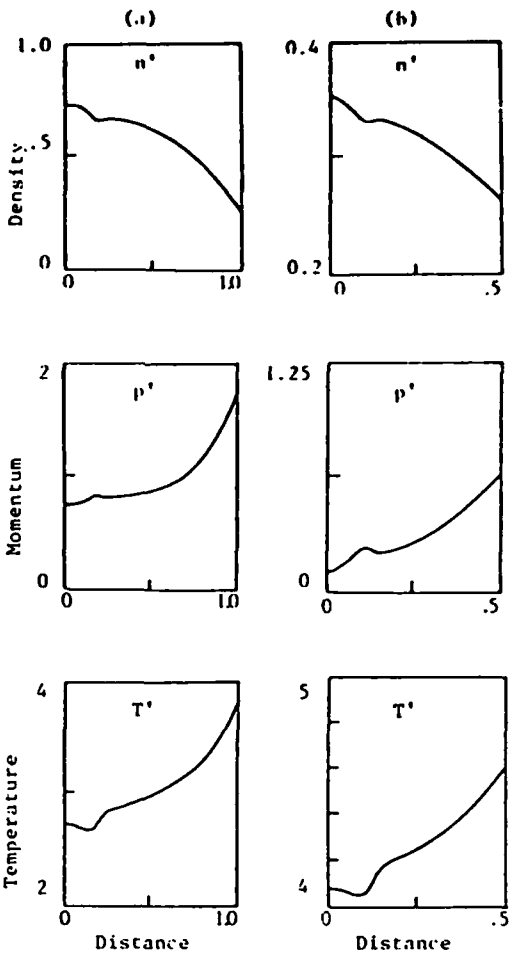


Fig. 5: Central Valley fractional density, momentum and electron temperature, for the 1.0 and 0.5 micron elements.

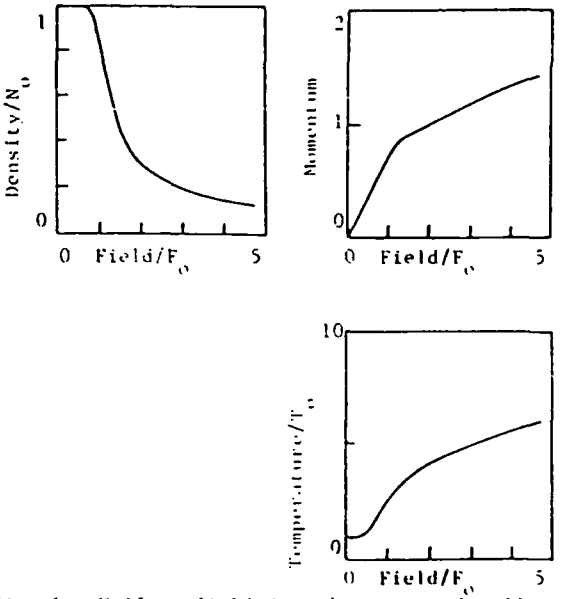


Fig. 6: Uniform field dependent central valley density momentum and electron temperature.

Reprint 3

Conceptual problems in modeling submicron device physics

H. L. Grubin

Scientific Research Associates, Inc. Glastonbury, Connecticut 06033

D. K. Ferry

Colorado State University, Fort Collins, Colorado 80521

(Received 23 March 1981; accepted 25 June 1981)

The study of high field transport in semiconductor devices has progressed very rapidly over the past decade, sustained in part by rapid technological advances in the solid state electronics industry. These have, however, generated important conceptual problems in modeling of high speed transport in nonlinear semiconductors. Some of these problems are discussed below.

PACS numbers: 72.20.Ht

I. INTRODUCTION

The study of high field transport in semiconductor devices has progressed very rapidly over the past decade, inspired directly by Gunn's observations of high field propagating domains,[1] and sustained by rapid technological advances in the semiconductor industry. Accompanying this growth are dramatic changes in the way transport properties are described. For example, most textbooks written prior to 1965 (see e.g., Smith, Ref. 2) solve the Boltzmann transport equation using the relaxation time approximation and perturbation theory; nonlinear transport is treated either phenomenologically or descriptively. Among the earliest general treatments of nonlinear transport, per se, are those of Conwell[3] and Paige.[4] Present treatments of device transport tend to combine the results of, e.g, Refs. 2–4 with the techniques of numerical simulation enabling a more complete representation of device phenomena (see, for example, Ref. 5). The types of problems currently studied stem from the technology associated with the very large scale integration (VLSI) and very high speed integrated circuits (VHSIC) programs. These programs introduce conceptual problems in modeling the device physics, problems generally ignored in long devices operated at the low gigahertz frequency scale. Some of these problems are discussed below.

II. SPATIAL AND TEMPORAL SCALES

Tables I and II identify important VLSI/VHSIC scales. Present and near term device designs are concentrating in the 0.1–1 μ regime. The lower bound is typically the order of a carrier mean free path in such compound semiconductors as gallium arsenide, and is approximately equal to $N_0^{-1/3}$ when N_0, the mean impurity concentration, is equal to $10^{15}/\text{cm}^3$. For $N_0 = 10^{18}/\text{cm}^3$ the separation is approximately 0.01 μ. In the former case the possibility of collisionless "ballistic" transport arises, a topic discussed below. The latter case implies that the concept of a semiconductor device with thermal generated majority carriers is meaningful only at high concentration. For example, a $0.1 \times 1 \times 1$ μm element with a density of $10^{15}/\text{cm}^3$ contains 100 thermally generated carriers. At $10^{18}/\text{cm}^3$ this amount increases to 10^5 carriers. We are apparently constrained to regard some submicron devices as injected devices.

Another matter related to impurity atoms is their distribution. Typical device simulations of micron and longer devices usually assume a continuous (jellium) distribution of impurities with spatial variations occurring on a macroscopic scale. As the nominal device dimensions are reduced to a submicron scale this becomes less realistic and statistics associated with impurity placement (e.g., clustering) are required.

The fourth entry in Table I is the thermal deBroglie wavelength, a useful parameter for a large collection of charged particles. The implication is that if the critical device dimensions are comparable to the deBroglie wavelength, carrier reflection effects occur, which may enhance the probability of scattering. Quantum effects have already been identified in submicron MOS structures.[6] Figure 1 displays thermal deBroglie wavelength vs effective mass. On a submicron scale quantum effects should be most prominent in the light mass semiconductors.

Table II identifies the more significant temporal scales. Prominent here is the transit time, which is 0.5 ps for a carrier traveling at 2×10^7 cm/s across a 0.1 μm region. This value is of the order of the mean time between collisions and represents a major difference between submicron and the longer multimicron length devices. For the latter, collisions are a small fraction of the characteristic time interval, and the

TABLE I. Approximate spatial scales (GaAs).

Term	Value
Active region length	10^{-5} cm, 10^{-4} cm
Impurity separation ($N_0^{-1/3}$)	10^{-5} cm @ $10^{15}/\text{cm}^3$
	10^{-6} cm @ $10^{18}/\text{cm}^3$
Mean free path	10^{-5} cm
Thermal deBroglie wavelength	2.6×10^{-6} cm

 0022-5355/81/030540-05$01.00 © 1981 American Vacuum Society

TABLE II. Approximate temporal scales (GaAs).

Term	Value
Transit time (for $v = 2 \times 10^7$ cm/s)	5×10^{-13} s @ $L = 10^{-5}$ cm 5×10^{-12} s @ $L = 10^{-4}$ cm
Momentum relaxation	3×10^{-13} s @ 300 K
Collision duration	2×10^{-14} s

concept of mobility emerges. For submicron devices where subpicosecond effects are important mobility is not a significant parameter. Without overstating the case, submicron device transport requires solutions to the Boltzmann and quantum transport equations rather than the drift and diffusion equations.

From an intuitive point of view, there is little distinction between the mean free path and the time between collisions, ie., there is overlap between the temporal and spatial scales. The third entry in Table II offers no such parity. This entry represents the fact that collision processes take place over a time interval that may be a sizable fraction of the carrier response time. Thus during a collision event the carriers may, additionally, either absorb or lose energy to a self-consistent field. Treatment of this effect requires modification of the scattering integrals of the Boltzmann transport equation.[7] Results will be illustrated below.

Implicit in the above discussion is that as devices are reduced to the order of a mean free path or smaller, their behavior becomes boundary condition intensive. No longer is it reasonable to conceive that transport is determined by, e.g., band structure, and, e.g., then modified by contacts.[8] Rather, the transient behavior of light mass submicron devices are dominated by properties of the surface, as well as the environment. While this is clear from several examples (layered real space transfer devices,[9] quantized inversions layers,[6] etc.), the study of these surface dominated submicron devices is still a new discipline, with its concomitant uncertainties. We illustrate the latter below.

III. BALLISTIC TRANSPORT: TRANSIENT CARRIER TRANSPORT

The possibility of fabricating devices with an active channel length of the order of a mean free path has led to the proposal that ballistic devices are conceptually possible.[10] We explore this briefly, emphasizing the dependence on boundary conditions.

Ballistic transport as used in vacuum tubes implies the absence of collisions, except perhaps for Coulombic interactions and their resulting space charge limitation. In a semiconductor such as GaAs, it has been suggested that microscopic ballistic transport is also possible so long as the carrier energy is below the optical phonon energy, approximately 36 meV for GaAs. This, however, ignores such things as carrier statistics. For, in an otherwise perfect crystal (ignoring the weak acoustic phonons), the mobile electrons have a well defined energy distribution with some carrier energies very near the threshold for phonon emission. As a result, even for small applied fields electron–phonon interactions will occur, and ballistic transport in the microscopic sense is not an observable. Instead ballistic transport in the *sense of the mean* is conceptually possible,

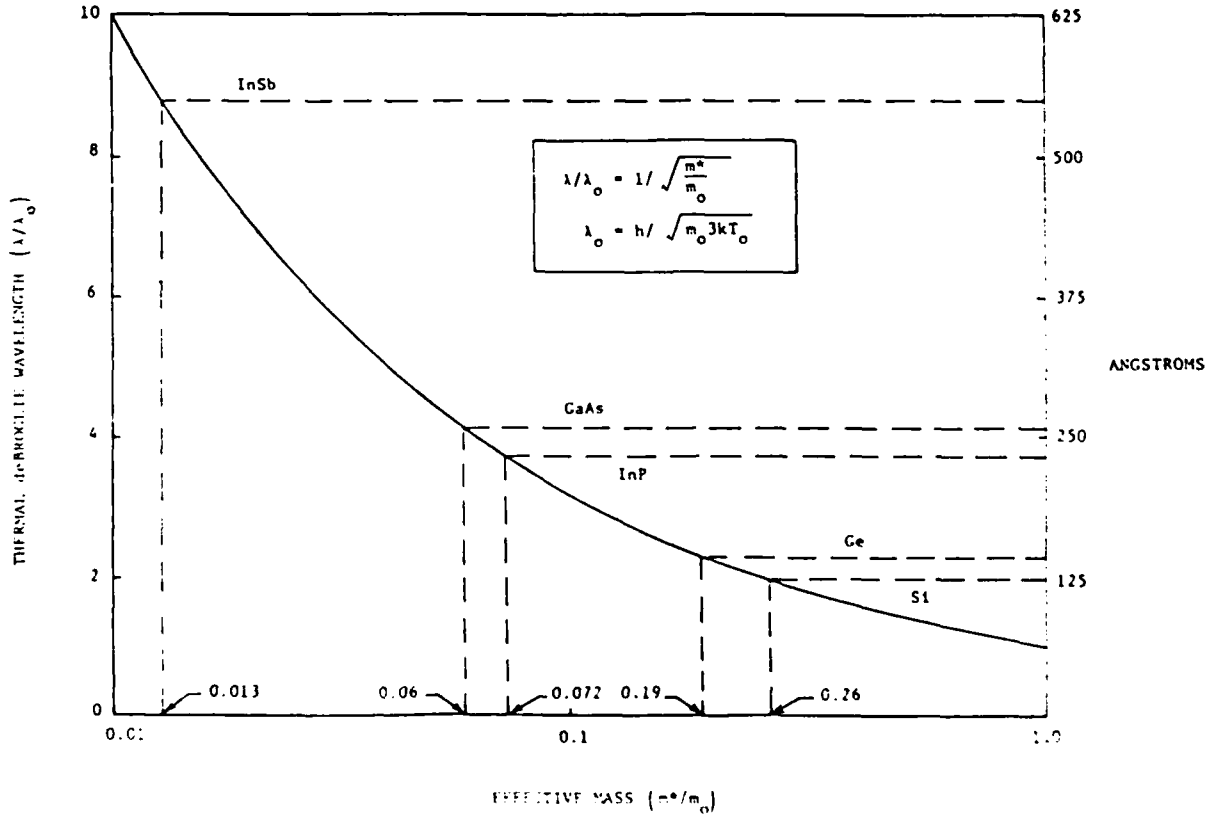


FIG. 1. Thermal debroglie wavelength vs effective mass. The free electron mass is m_0 (after Iafrate).

where on a short time scale the *mean* velocity is approximately given by

$$\langle V \rangle = qFt/m^*. \quad (1)$$

Equation (1) implies an ensemble average in which electron–electron interactions are sufficiently rapid for the carriers to be in thermal equilibrium with each other. Further, for GaAs, "t" in Eq. (1) is generally less than the relaxation time for the Frohlich interaction, i.e., $t \ll 0.3$ ps.

If we accept the conceptual possibility of *mean ballistic transport* then a number of simple results are immediate and are borrowed from the literature of vacuum tubes. In one case (neglecting diffusion!) current continuity, energy conservation, and the assumption of an infinite source of cathode electrons entering with zero velocity yields Childs law:

$$j : \phi^{3/2}. \quad (2)$$

Relaxing the unlikely zero cathode velocity condition[8] causes the current voltage characteristic to depart significantly from a "3/2" power law,[11] and reinforces the idea that for *mean ballistic transport*, the I–V characteristics are dominated by the boundary region, not by ballistics.[12]

When we examine transport on a time scale longer than that for which Eq. (1) is approximately valid, dramatic nonequi-

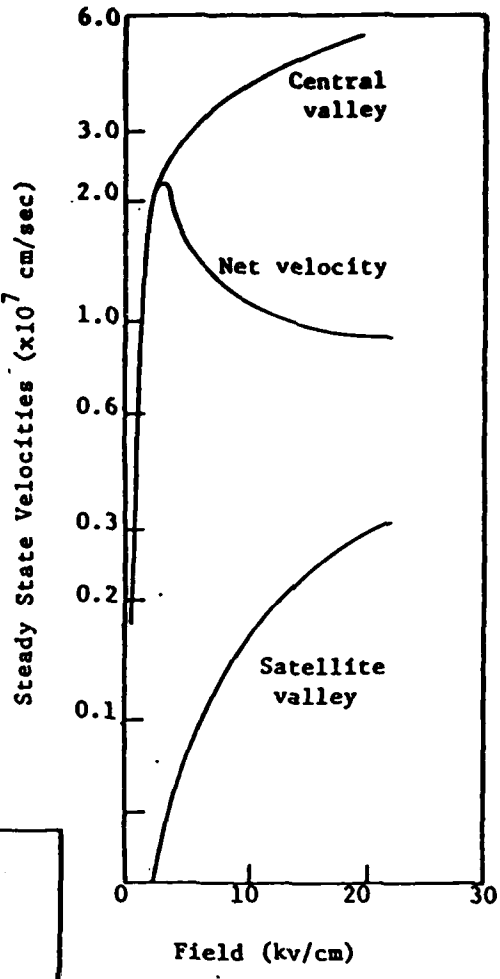


FIG. 2. Steady state velocities for the central and satellite valleys of GaAs in a Γ,X orientation. Also shown is the mean velocity including electron transfer.

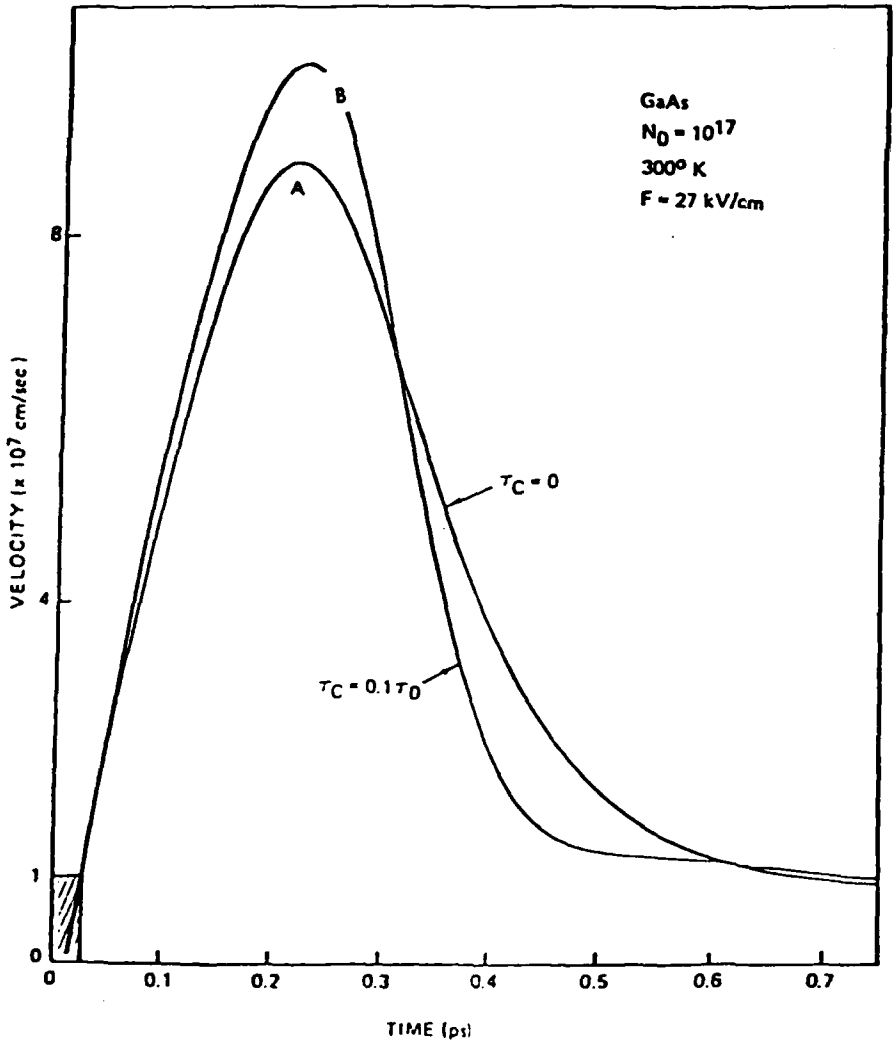


FIG. 3. Mean transient carrier velocity vs time for zero and finite collision duration. Results are obtained from the first three moments of the Boltzmann transport equation.

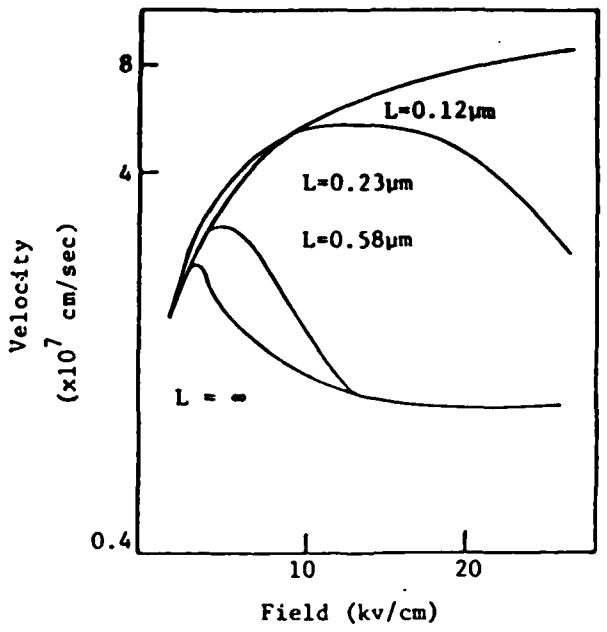


FIG. 4. Field dependent drift velocity vs distance for gallium arsenide. Lengths denote transit-distances.

librium size-dependent effects occur. The phenomena has been referred to as velocity overshoot for reasons that will be apparent. We again concentrate on gallium arsenide, which for simplicity we treat as a two-valley semiconductor.

At low values of electric field and at room temperature almost all of the carriers are in the central valley. In the absence of ionized impurities, detailed balance dominates and the mobility, a "mean" quantity, is determined by intravalley LO phonon scattering. At high values of electric field LO phonon intervalley scattering becomes the dominant scattering mechanism, and there is an increased level of spontaneous phonon emission. For the central valley, velocity saturation occurs but at values of velocity higher than the steady state velocity. The general situation is represented by steady state curves (Fig. 2) for the mean carrier velocity in the central and satellite valleys. Due to differences in the effective mass, the carrier velocity is highest for central valley electrons.

Once we recognize that a finite amount of time is necessary for carriers to transfer from the central to satellite valleys we see the immediate possibilities of high transient velocities. Typical transient curves are shown in Fig. 3: one result is for zero collision duration and a second for finite collision duration. The shaded region, where the expired time is less than LO phonon relaxation time, we refer to as the mean ballistic regime.

The calculation of Fig. 3 can be translated into a velocity vs distance transient. Then repeating the calculation at different values of field we obtain the results of Fig. 4.[13] It is clear that as the device dimensions are reduced the apparent saturated drift velocity increases while the presence of negative differential mobility is eliminated. This particular result is highly significant insofar as it suggests the possibility of engineering a size dependent device which in one case offers negative differential mobility, while in another case velocity saturation.

IV. SUMMARY

The above discussion developed from the idea that conceptual problems in modeling device transport on a submicron scale require care for their resolution. And in emphasizing the

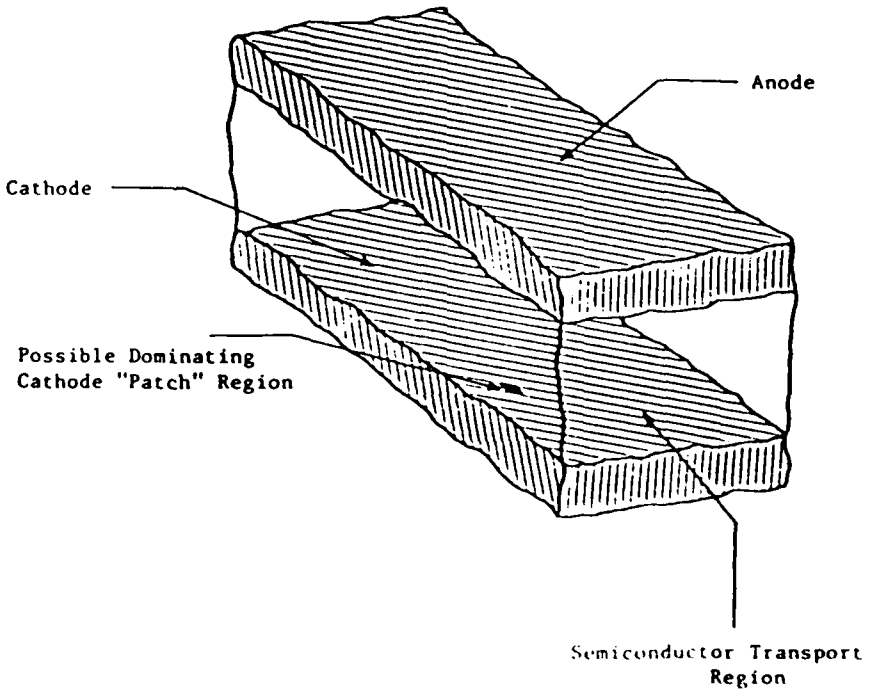


FIG 5 Representation of a three-dimensional submicron device, dominated by one section of the contact

role of the boundary, the discussion suggests that when device dimensions become sufficiently small, we may wish to regard devices as surface limited devices separated by regions whose principal contribution is to introduce phase changes.[12] But the very nature of a surface limited device also signals that one-dimensional submicron transport can be effective only in so far as it identifies isolated phenomena. Realistic modeling must incorporate lateral variations—a feature ignored in long devices where the contact region occupies a small fraction of device size. In short devices this approximation is not acceptable, as Fig. 5 suggests. In fact, it would not be surprising if one local region dominated all device behavior.

ACKNOWLEDGMENTS

The authors thank G. J. Iafrate, K. Hess and P. J. Price for their insightful comments. This work was supported by the Office of Naval Research and the Army Research Office, to whom we are grateful.

[1]J. B. Gunn, Solid State Commun. 1, 88 (1963).
[2]R. A. Smith, *Semiconductors* (Cambridge University, New York, 1959).
[3]E. M. Conwell, *High Field Transport in Semiconductors* (Academic, New York, 1967).
[4]E. G. S. Paige, *The Electrical Conductivity of Germanium* (Heywood, London, 1964).
[5]*Physics of Nonlinear Transport in Semiconductors*, edited by D. K. Ferry, J. R. Barker, and C. Jacoboni (Plenum, New York, 1980).
[6]For a general review of the quantized inversion layer, see F. Stern, Crit. Rev. Solid State Sci. 4, 499 (1974).
[7]D. K. Ferry and J. R. Barker, Solid State Commun. 30, 301 (1979).
[8]M. P. Shaw, H. L. Grubin, and P. R. Solomon, *The Gunn-Hilsum Effect* (Academic, New York, 1979).
[9]K. Hess, H. Morkoc, H. Shichijo, and B. G. Streetman, Appl. Phys. Lett. 35, 469 (1979).
[10]See, e.g., M. S. Shur, and L. F. Eastman, IEEE Trans. Electron Devices ED-26, 469 (1979).
[11]H. L. Grubin, D. K. Ferry, G. J. Iafrate, and J. R. Barker, *Microstructure Science and Engineering/VLSI*, edited by N. Einspruch (Academic, New York, 1981).
[12]D. K. Ferry, J. R. Barker, and H. L. Grubin, IEEE Electron Devices Lett. EDL-1, 209 (1980).
[13]H. L. Grubin, D. K. Ferry, and J. R. Barker, in *Proceedings of IEDM* (IEEE, New York, 1979), p. 394.

IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. ED-28, NO. 8, AUGUST 1981

Hot-Carrier Constraints on Transient Transport in Very Small Semiconductor Devices

DAVID K. FERRY, SENIOR MEMBER, IEEE, JOHN R. BARKER, AND HAROLD L. GRUBIN, MEMBER, IEEE

Abstract—Current technology has progressed rapidly and is pushing toward fabrication of submicron dimensioned devices. As this occurs, we expect that the temporal and spatial scales in these devices will become sufficiently small that the semiclassical approach to transport theory, as expressed by the Boltzmann equation, becomes of questionable validity. In developing a corrected transport equation from quantum kinetic theory, several constraints arise on the normal concepts of transport parameters. The intra-collisional field effect, concomitant nonzero collision duration, and retarded collisional interactions have pronounced effects upon the carrier transport, especially in the transient dynamic response region in small devices. The description of diffusion is also complicated by the relatively long duration of the velocity autocorrelation function. Calculations have been carried out for the velocity autocorrelation function for Si. It is found that the autocorrelation $\phi'(t)$ initially relaxes exponentially, due to momentum relaxation, goes negative and displays a local minimum, then relaxes to zero at a slower rate due to energy relaxation. This complicated behavior leads to enhanced diffusion and noise on the short-time scale.

I. INTRODUCTION

THE THRUST OF integrated electronics in recent years has been toward the realm of VLSI of IC's. From the early beginning of IC's, the complexity of the circuit increased rapidly, approximately doubling each year [1]. In fact, the growth of complexity appears to be maintaining this rate for regular arrays of devices, although slowing somewhat for random logic [2]. There are several factors to this increase in complexity, including major effects arising from increased die size, increased circuit cleverness, and reduced device size. The latter of these is significant, and current IC's involve devices of 1-μm design rules [3]. However, research MOSFET's have been fabricated with channel lengths in the 0.1-0.25-μm range [4], [5]. These developments are not restricted to MOSFET's, and experimental GaAs microwave devices have also been built with gate lengths in this range. Indeed, simple experimental structures have been fabricated in the 0.01-0.1-μm range [6]. These new developments and advancements into submicron semiconductor devices are considerably hampered by a large gap in our understanding of nonequilibrium semiconductor transport on scales intermediate to the true atomic scale (<10 Å) and the bulk solid-state macroscale (>1 μm = 10^4 Å). It is already apparent that simple down-scaling of processing, device function and performance, bulk physics, etc., is not adequate in this region, nor indeed is a straightforward upscaling of known atomic-scale phenomena [7], [8]. Indeed, whether the transport is ballistic [9] or retarded [10], [11], it is complicated by the high electric fields and the resultant hot-carrier phenomena. For example, 1 V across a 0.1-μm channel produces an average electric field of 10^5 V/cm, enough to produce hot electron effects in any semiconductor.

Hot electron effects were suggested as early as 1963 to be important in MOSFET's [12] and the role of velocity saturation was considered shortly thereafter [13]. Subsequently, the role of hot carriers in MOS and MES devices has been investigated and reviewed [14]-[20]. Much of the problems that arise in transport in submicron devices are due to the very fast time scales inherent in these small devices. For example, an electron traveling at 10^7 cm/s can cross a 0.1-μm channel in 10^{-12} s, which is a time on the scale of the appropriate relaxation times for momentum, energy, and charge. On this time scale, the electrons encountering a high-field region of this dimension, such as the pinchoff region in a MOSFET, do not have adequate time to establish any sort of equilibrium distribution, a point made rather pointedly earlier by Ruch [21], and by Maloney and Frey [22]. Additional complications arise from the fact that the collision duration is no longer negligible on this time scale and strongly affects the transport dynamics [10], [23].

In modeling of semiconductor devices, the major physical effects are dominantly tied up on the manner in which the charge fluctuations and current response are coupled to the local electric field, formally related through the continuity equation. To accurately determine the current response, one must solve an appropriate transport equation and it is in these transport equations that many of the major modifications arising from these short-time scales occur.

In the following sections, we examine the consequences of changes that the short-time scales play on the transport of carriers in these small devices. First, we look at the effect of the nonnegligible collision duration and the concommitant intra-collisional field effect. Then we turn to its effect on the transport equations, especially with regard to transient dynamic response. In this latter regard, the concept of diffusion and the velocity autocorrelation function must be carefully examined. Throughout we concentrate on silicon devices.

Manuscript received October 27, 1980; revised April 1, 1981. This work was supported in part by the Office of Naval Research, the Army Research Office, the North Atlantic Treaty Organization, and by the Science Research Council (U.K.).

D. K. Ferry is with the Colorado State University, Ft. Collins, CO 80523.

J. R. Barker is with Warwick University, Coventry CV4 7AL, England.

H. L. Grubin was with United Technologies Research Center, East Hartford, CT 06108. He is now with Scientific Research Associates, Inc., Glastonbury, CT 06033.

0018-9383/81/0800-0905$00.75 © 1981 IEEE

II. The Intra-Collisional Field Effect (ICFE)

The Boltzmann transport equation (BTE) has long been the basis for semiclassical transport studies in semiconductors and other materials. Its utility also stems from the fact that it is readily transformable into a path-variable form which can be adapted to numerical solutions for complicated energy-dependent scattering processes [24], [25]. In this form, the BTE is often referred to as the Chambers-Rees path integral equation, and serves as the basis for Monte Carlo and iterative calculations of transport. However, the BTE is valid only in the weak coupling limit under the assumptions that the electric field is weak and slowly varying at most, the collisions are independent, and the collisions occur instantaneously in space and time. Each of these approximations can be expected to be violated in future submicron dimensioned semiconductor devices. We have previously shown that in such devices, the time scales are such that collision durations are no longer negligible when compared to the relevant time scale upon which transport through the device occurs [23]. In this situation, even for time-independent fields, the quantum kinetic equations are nonlocal in time and momentum. It may be recalled that the BTE can be rigorously derived from the density matrix Liouville equation formulation of quantum transport [26], [27] under the above amplifying conditions. In this approach, the collision terms are derived under the assumption that the collisions occur instantaneously, which is a reasonable approximation when the mean time between collisions is large. At high fields, such as will occur in very small devices, the collision duration is significant and correction terms must be generated (for the BTE) and to account for the actual nonzero time duration of each collision. If the instantaneous collision approximation that leads to the BTE is relaxed, an additional field contribution appears as a differential super-operator term (see, e.g., the discussion in [7]) in the collision integrals evaluated in the momentum representation, resulting in an ICFE [28], [29].

The ICFE can be partially understood by the following simplified model. In the Boltzmann case, the collision occurs instantaneously, so that the carrier enters the collision sphere at one point and instantaneously exits at a second point, called b for reference. However, the collision does not occur instantaneously, but requires a nonzero collision duration τ_c. In this case, it can now be accelerated by the field *during the collision*. Thus it exits not at b, but at b' some time $\Delta t = \tau_c$ later. The points b and b' differ by a modification of the momentum conservation relations due to field acceleration during the collision. When τ_c begins to become comparable to τ, the mean time between collision, this ICFE will have a significant effect on the transport dynamics, particularly in the transient response region.

The mathematical details of the ICFE have previously been given [27], so we shall not go into these details here. Rather, we shall merely cite some of the supportive evidence for the observability of the effect. In very large fields, such as can occur in SiO_2 near breakdown, the ICFE can indeed by very significant [30]. Two major modifications of the scattering integral occur as a result of this intra-collisional process. First, the total energy-conserving δ-function is broadened by the presence of the electric field. Second, the threshold energy required for the emission of an optical phonon is modified, which causes a shift (in energy) of the δ-function. This latter process is easily understood in physical terms. The argument of the energy conserving δ-function is just

$$E_f - E_i \pm \hbar\omega_0 = E(\vec{p}_f) - E(\vec{p}_i) \pm \hbar\omega_0 \tag{1}$$

but the initial and final momenta evolve during the collision as

$$\vec{p}(t') = \vec{p} - \int_{t'}^{t} e\vec{E}(t'')\,dt'' \tag{2a}$$

$$\vec{p}'(t') = \vec{p}' - \int_{t'}^{t} e\vec{E}(t'')\,dt''. \tag{2b}$$

In the emission of an optical phonon, where the electron is scattered against the electric field, the field will absorb a portion of the electron's energy during the collision, and hence a reduction in energy loss to the lattice will be favored. The opposite effect, an enhancement in energy loss to the lattice, occurs for emission along the electric field. These effects can be incorporated into the appropriate scattering integrals used in solutions to the Boltzmann equation, and this has been carried out for the transport of electrons through SiO_2. For electric fields above $(5\text{-}6) \times 10^6$ V/cm, the broadening and shift of the scattering resonances produce a noticeable effect upon the velocity-field relationship, and this reduction in threshold can be further observed in the impact ionization rates in SiO_2. Only with these modifications does the calculated ionization rate compare with the rates measured by Solomon and Klein [31].

Although the ICFE is exceedingly large in SiO_2 because of the polar nature of the phonons here, it is also significant in the case of Si. In Fig. 1, we plot the collision duration versus field and compare it with the momentum relaxation time $\bar{\tau} = m^*\mu e$ (m^* here is the conductivity mass for Si and is obtained by a sum over the multivalley band structure), where

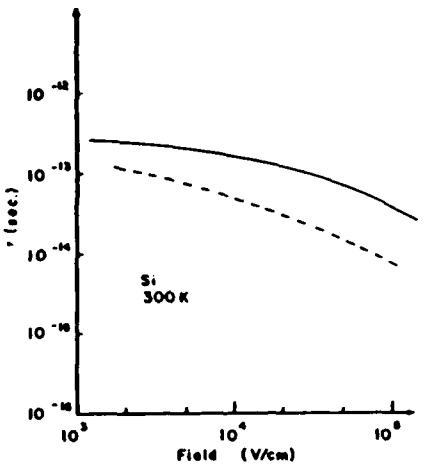


Fig. 1. Variation of the effective collision duration (dashed curve) and mean freetime for Si (solid curve). The latter quantity is calculated from the effective mobility as $\bar{\tau} = m^*\mu/e$.

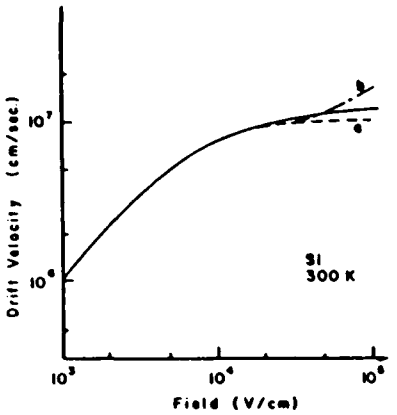


Fig. 2. The velocity-field curve for electrons in Si as calculated by a drifted Maxwellian approach. At high fields, the field weakens the collisions (dot-dashed curve *b*) causing an increase of velocity although the collisional retardation stops this effect (solid curve). The low-frequency curve normally seen (for $\tau_c = 0$) is the dashed curve *a*.

$\mu = v_d/F$ is the mobility. The effect this has on the velocity-field curve is shown in Fig. 2. Also shown is the countering effect of collision retardation discussed in the next section. The ICFE is especially noticeable at high fields, where it essentially eliminates the scattering by the low energy intervalley phonon [32]. This phonon is already weakly coupled since it scatters through a first-order interaction [33] but is normally an effective scatterer at high fields.

III. The Retarded Transport Equation

When the ICFE is included as a modification of the lowest order kinetic equations, a high field quantum kinetic equation, which replaces the BTE, is found as [27], [28]

$$\frac{\partial f_s(\vec{p},t)}{\partial t} + e\vec{F}(t)\cdot\nabla_p f(\vec{p},t) = \int_0^t dt' \sum_{\vec{p}'} \{S(\vec{p},\vec{p}';t,t')\, f(\vec{p}',t') - S(\vec{p}',\vec{p};t,t')\, f(\vec{p},t')\} \tag{3}$$

where the momenta $\vec{p},\vec{p}'$ are explicit functions of the retarded time t' on the right-hand side through the relationship of (2), and the transition terms S take the form, for inelastic phonon scattering

$$S(p,p';t,t') = \mathrm{Re}\,\frac{2\pi}{\hbar}\sum_{\vec{q}}\left(\frac{1}{\pi\hbar}\right)\exp\left(-\frac{t-t'}{\tau_\Gamma}\right)\cdot\left(N_q + \frac{1}{2} + \frac{1}{2}\eta\right)|V(q)|^2\,\delta_{\hat{p},\hat{p}'+\eta\hat{q}}\cdot\exp\left[-i\int_t^{t'}\frac{dt''}{\hbar}\beta(\vec{p},\vec{p}';t'')\right] \tag{4}$$

where β is the argument of the normal δ-function and is given by (1). The two exponential factors in (4) are related to the joint spectral density function, which reduces to an energy conserving δ-function in the instantaneous collision low-field limit, $\epsilon(\vec{p})$ is the quasiparticle renormalized energy, $\hbar/\tau_\Gamma$ is the joint linewidth due to collisional broadening of the initial and final states, and η takes the +1, -1 for phonon emission or absorption, respectively, in the in-scattering. For the out-scattering term, the roles of $\vec{p}$, $\vec{p}'$ are interchanged although this does not upset detailed balance in the equilibrium sense.

In small semiconductor devices, where the dimensional scale is of the order 0.3 μm or less, the carrier concentration will in general be relatively high. We use the drifted Maxwellian approach to developing a set of coupled balance equations, using (3) instead of the BTE. With this approach, a hierarchy of moment equations can be generated, from which the various parameters can be determined [10], [11]. These moment equations include first-order effects arising from the nonzero time duration of the collisions and general retardation effects [34]-[36]. Starting from the density-matrix developed form of the BTE, we have shown previously that collision terms derived in the normal case, but modified for intra-collisional field effects, must be convolved with a decay effect over an effective collision duration. Thus the balance equations are modified in a straightforward fashion, although the details are much more complicated. This latter follows from the role of the intra-collisional field effects, which both broaden and shift the resonances, effectively lengthening the effective collision duration and weakening the effect of the collision itself. If (3) is Laplace transformed, the moment equations can be developed by multiplying by an arbitrary function $\phi(\vec{p})$, integrating over the momentum, so that the moment equations are developed in the transform-domain, and then retransforming. This yields [11]

$$\left\langle\frac{\partial\phi}{\partial t}\right\rangle - e\vec{E}\cdot\langle\nabla_p\phi\rangle = \frac{-1}{\tau_c}\int_0^t \exp(-t'/\tau_c)\,\langle\ddot{\phi}_c(t-t')\rangle\,dt' \tag{5}$$

where

$$\langle\ddot{\phi}_c\rangle = \int_0^t \phi(t-\tau)\frac{d}{d\tau}[\Gamma_\phi(\tau)]\,d\tau \tag{6}$$

and Γ_ϕ is the relaxation rate for ϕ, so that $\dot{\phi}_c = \phi\Gamma$ is the time-rate of change of $\phi(\vec{p})$ due to collisions. The result in (5) is particularly interesting, in that it allows $\langle\dot{\phi}_c\rangle$ to be evaluated in the case of instantaneous scattering, and this result to be averaged over an effective collision duration τ_c, weighted by the function $\exp(-t/\tau_c)$. If, as is the case at low fields, $\langle\dot{\phi}_c\rangle$ does not change during the collision duration τ_c, the normal result ($t >> \tau_c$) is obtained. However, in large fields, where the intra-collisional field effect is important, the variation of $\langle\dot{\phi}_c\rangle$ during the collision becomes important. These effects will also be important in high-frequency transport where the collision duration becomes comparable to the relaxation times and the period of the wave [37]. From the form of (5), we note the right-hand side (RHS) is such that the nonzero collision duration must be combined with the normal non-Markovian nature of transport on these short-time scales [34]-[36], so that the momentum relaxation time must be con-

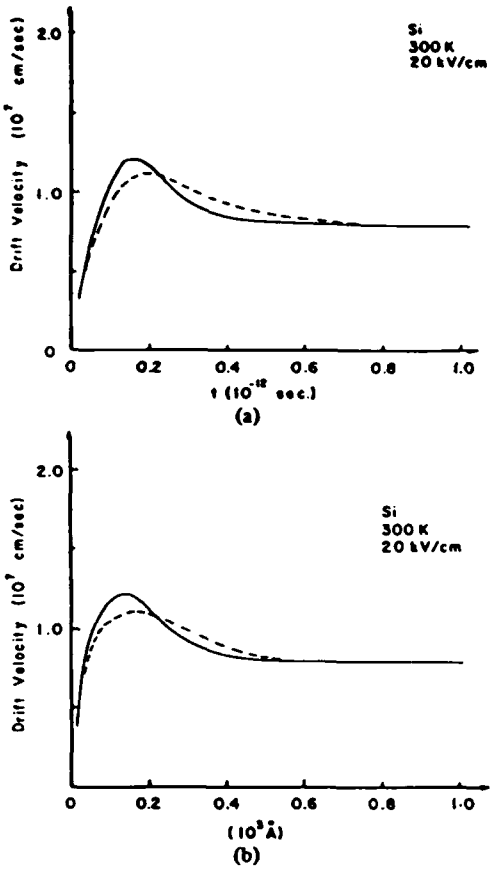


Fig. 3. The velocity response to a step, homogeneous field of 20 kV/cm in Si is shown as a function of time (a) and of distance into the semiconductor (b). The fully retarded collisional interaction is used for the solid curve, while an instantaneous collision model yields the dashed curve.

volved as in (5) and the result re-convolved with the momentum. The details of this will be presented elsewhere, but the resulting velocity-field curve for Si, calculated for $f(E)$ as a retarded and drifted Maxwellian, is shown in Fig. 2, for $E \parallel \langle 111 \rangle$.

Calculations of the transient response have also been made for Si. The details of the coupling constants and phonon parameters are those normally accepted. In Fig. 3, the transient response for a steady homogeneous field of 20 kV/cm, applied at $t = 0$, is shown. The response for a retarded collisional interaction rises quicker and settles faster than that of the unretarded case. The quicker rise follows from the retarded momentum relaxation effects, while the faster settling occurs due to retarded energy relaxation effects which causes an overshoot to occur in the temperature as well.

The collisional retardation speeds up the transient process primarily due to the effect of slowing down changes in the effective momentum and energy via collisional relaxation as well. The small-signal ac mobility is extremely sensitive to the energy distribution function, so it is extremely important that any simulation technique be very efficient in yielding this portion of the distribution function.

IV. Diffusion and the Autocorrelation Function

One of the most fundamental parameters required for modeling semiconductor devices is the diffusion coefficient $D(E, \omega)$, where E is the electric field and ω is the frequency. Not only is the diffusion coefficient necessary for evaluating operating characteristics and high-frequency characteristics, it provides a fundamental characterization of velocity fluctuations in the system and their contribution to noise in the device [38], [39]. If diffusion is relatively well understood for low fields, this situation does not carry over to the case of high electric fields [39], [40]. The general case for high-field transport in semiconductors differs in that relaxation of the velocity fluctuations is to a nonequilibrium steady-state [41] and the process is nonlinear [42], [43].

The fluctuation response in general is complicated due to the many physical processes involved, but the velocity fluctuation can be considered as having two main contributions: $v' = v(t) - \langle v \rangle = v'' + u'$. The first of these u', is the velocity fluctuation arising from a fluctuation in carrier energy: $u' = u(\epsilon + \Delta\epsilon) - u(\epsilon)$; while the second v'' arises from velocity fluctuations about u' [43], [44]. These various factors can be observed by studying, not the diffusion coefficient itself, but rather by studying the velocity autocorrelation function $\phi'(t)$, which is the inverse Fourier-cosine transform of $D(E, \omega)$. If we define $\phi'(t)$ as

$$\phi'(t) = \langle [v(t + t_0) - \langle v \rangle] [v(t_0) - \langle v \rangle] \rangle \tag{7}$$

then it is found that for high electric fields $\phi'(t)$ decreases initially as an exponential, becomes negative, passes through a minimum, and relaxes finally to zero [43]. This process is basically related to the fact that, in general, energy relaxation is slower than momentum relaxation, and the above behavior can be expected to occur via the same processes which lead, e.g., to velocity overshoot [21], [22]. The detailed behavior of $\phi'(t)$ assumes more than academic interest as semiconductor devices begin to assume submicron dimensions. In Si, for example, the time duration of $\phi'(t)$ can be of the order of a picosecond. So as channel lengths drop below say 0.3 μm, correlated electron motion and enhanced noise in the devices can be expected to occur.

In general, the diffusion coefficient $D(\omega, E)$ depends upon the velocity fluctuations in the electron system and is related to the noise power spectral density $S_v(\omega)$ associated with these fluctuations. These are related as (for longitudinal diffusion)

$$D(\omega, E) = S_v(\omega)/4 = \int_0^\infty \phi'(\tau) \cos(\omega\tau)\, d\tau$$

$$= \int_0^\infty \langle [v(\tau + t) - \langle v \rangle] [v(t) - \langle v \rangle] \rangle \cos(\omega\tau)\, d\tau \tag{8}$$

where all velocities are understood to be longitudinal. The principal difficulty in calculating transport parameters, and particularly $\phi'(t)$, in these systems lies in the complicated energy dependence of the many scattering processes. In the past few years, however, ensemble Monte Carlo techniques have been developed which can be used to calculate these transport parameters with high resolution, and this has been used to calculate the correlation function for the total velocity, $\phi(t) = \langle v(t + t_0)\, v(t_0) \rangle = \phi'(t) + \langle v \rangle^2$ (all calculations

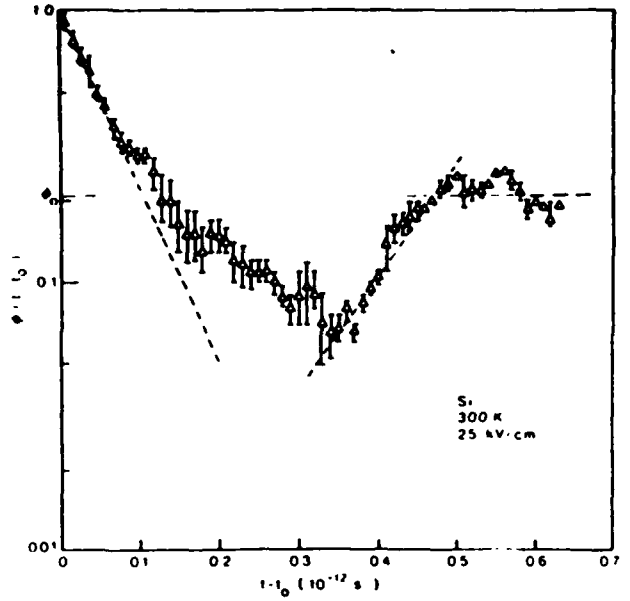


Fig. 4. The total velocity correlation function $\phi(t - t_0) = \langle v(t)\, v(t_0)\rangle$ as a function of $t - t_0$ for an applied field of 25 kV/cm. The curves are normalized to $\langle v^2\rangle$ and ϕ_0 is $\langle v\rangle^2/\langle v^2\rangle$, the final value. The individual parts of the curve are discussed in the text.

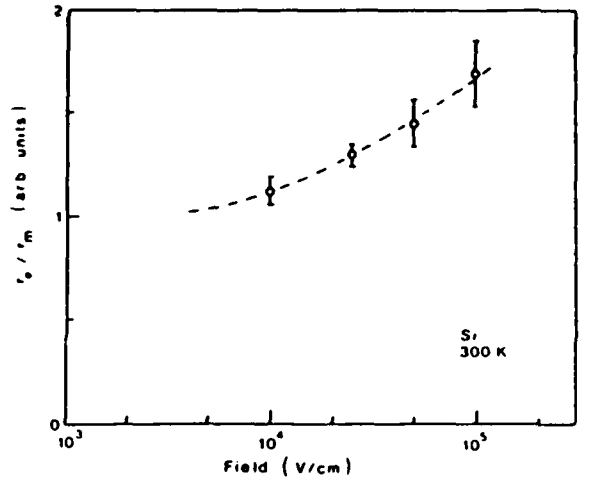


Fig. 5. The deviation of τ_0 from τ_m. The velocity autocorrelation function $\phi(t)$ decreases initially as exp $(-t/\tau_0)$. Although this initial fall corresponds to momentum relaxation, $\tau_0 > \tau_m$.

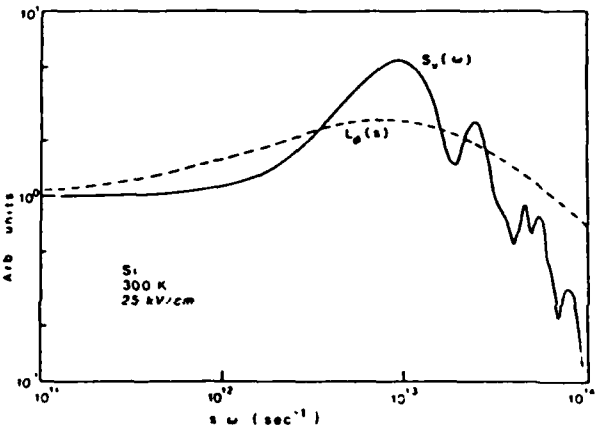


Fig. 6. The Laplace transform $L\phi(s)$ and Fourier-cosine transform $S_v(\omega)$ (noise spectral density) of $\phi'(t) = \langle[v(t + t_0) - \langle v\rangle]\cdot[v(t_0 - \langle v\rangle]\rangle$.

shown in the figures have $\phi(t)$ normalized to $\langle v^2\rangle$ [40]. The ensemble of electrons was initialized as a Maxwellian at the lattice temperature 300 K and was assumed to reside at $x = 0$ at $t = 0$. A homogeneous electric field was applied at $t = 0$, and the ensemble allowed to evolve in time. After a reasonable period of time, the ensemble was in a pseudo-equilibrium with the field and had a steady drift velocity. After this pseudo-equilibrium was achieved, the longitudinal velocity autocorrelation function $\phi(t) = \langle v(t_0 + t)\, v(t_0)\rangle$ was calculated for several initial times t_0. The stationarity of the system, therefore, is verified as well, and the averaging process was carried out over the ensemble as well as over various initial times. In Fig. 4 is shown the variation of $\phi(t)$ as a function of time t, for an electric field of 25 kV/cm. The initial fall of $\phi(t)$ is primarily due to momentum relaxation, with the local minimum and subsequent rise due to energy relaxation as suggested by Price [43]. The error bars indicate the spread of data points from the calculations and averaging procedures.

The initial exponential decay portion is significant. The time constant of this portion of the decay of $\phi(t)$ is closely related to, and slightly larger than, the momentum relaxation time τ_m associated with the chordal mobility $\mu = v_d/E$, rather than the differential mobility dv_d/dE (here we define the effective or average $\tau_m = m^*\mu/e$). The latter quantity has been suggested as the appropriate quantity for longitudinal diffusion. At 25 kV/cm, the velocity is becoming very nearly saturated, so that the differential mobility is more than an order of magnitude smaller than the linear mobility. This difference is readily distinguished from the data in Fig. 4. The decay of $\phi(t)$ varying as exp $(-t/\tau_0)$ at short times and at 25 kV/cm, for example, is best fit with a τ_0 of 7×10^{-14} s, while $\tau_m \sim 5.4 \times 10^{-14}$ s. The results of the decay constant of the exponential portion of $\phi(t)$ being slightly larger than τ_m appears to be a general result, as it was checked at several other values of electric field. In Fig. 5, we plot the variation of τ_0/τ_m with electric field. It is observed that τ_0/τ_m increases with the field, although not quite linearly.

In Fig. 6, the Laplace transform and Fourier-cosine transform of $\phi'(t)$ are shown for a field of 25 kV/cm. Contrary to linear transport, these functions are not simple monotonically decreasing functions for large ω. Rather, they exhibit peaks at high frequency. The origin of these peaks lies in the enhanced high-frequency conductivity [37] in regions where the energy relaxation process can no longer follow the ac field. Thus these peaks have their origin in the same processes that lead to velocity overshoot. The oscillations in $S_v(\omega)$ at high frequency appear to be related to the oscillations at long time on $\phi'(t)$. While these oscillations may not be real, their presence and the shape of $S_v(\omega)$ has also been observed by Grondin in GaAs [45]. From this figure, it is apparent that enhanced noise will appear in Si devices at frequencies above $\approx 10^{11}$ Hz and that correlated carrier motion can be expected for times on the order of a picosecond.

V. Conclusions

In the above sections, we have examined the important time and distance scales that are important in small semicon-

ductor devices. We find that we are already approaching the time and distance scales that raise serious objections to usage of a semiclassical approach to transport theory within the device. As a consequence, major modifications are required to both the collision interactions in the BTE, due to intra-collisional field effects arising from a breakdown of the independent perturbation approximation, and to the current response equations within the device. In this last case, effects arise not only due to spatial and temporal relaxation of the carriers and charge density, but also due to retardation effects within the current response. As a result, frequency and power limitations in small devices can be expected to be significantly different from predicted values based upon extrapolations from large devices. As an example, the increase in drift velocity expected from overshoot and intra-collisional field effects and the increased breakdown field in the small device both act to raise the estimated figure of merit.

In the previous sections, we have also pointed out that the velocity autocorrelation function shows marked structure, due to the strong nonlinearities of the hot electron transport problem, and that much of this structure can be understood in terms of the competition between momentum and energy relaxation processes. These effects lead to a local minimum, which is negative, which causes $D(t)$ to exhibit overshoot characteristics in its evolution to the long-time result. This overshoot behavior, associated with the competition between momentum and energy relaxation, is associated with that portion of the Green's function of the transport equation which leads to velocity overshoot.

The detailed nature of $\phi(t)$ also leads to an enhanced noise spectral density at high frequencies. This noise, and the fact that $\phi(t)$ extends over a major fraction of a picosecond, indicate that the noise properties and transport properties of submicron semiconductor devices, where the transit time can be less than a picosecond, will differ from their properties in large devices. In particular, correlated electron motion can be expected to play a significant role, leading to a breakdown of traditional semiconductor equations.

References

[1] G. Moore, "Progress in digital integrated electronics," in *Proc. Intern. Electron Dev. Mtg.* New York: IEEE Press, 1975, p. 11.

[2] P. M. Russo, "VLSI impact on microprocessor evolution, usage, and system design," *IEEE Trans. Electron Devices*, vol. ED-27, pp. 1332-1341, Aug. 1980.

[3] H-N. Yu, A. Reisman, C. M. Osburn, and D. L. Critchlow, "1 μm MOSFET VLSI technology: An overview," *IEEE Trans. Electron Devices*, vol. ED-26, pp. 318-324, Apr. 1979.

[4] M. T. Elliott, M. R. Splinter, A. B. Jones, and J. P. Reekstin, "Size effects in E-beam fabricated MOS devices," *IEEE Trans. Electron Devices*, vol. ED-26, pp. 469-475, Apr. 1979.

[5] W. Hunter, T. C. Holloway, P. K. Chatterjee, and A. F. Tasch, Jr., "A new edge-defined approach for submicrometer MOSFET fabrication," *IEEE Electron Device Lett.*, vol. EDL-2, pp. 4-6, Jan. 1980.

[6] A. N. Broers, J. M. E. Harper, and W. W. Molzen, "250 Å linewidths with PMMA electron resist," *Appl. Phys. Lett.*, vol. 33, pp. 329-394, 1978.

[7] J. R. Barker and D. K. Ferry, "On the physics and modeling of small semiconductor devices-I," *Solid-State Electronics*, vol. 23, pp. 518-530, June 1980.

[8] D. K. Ferry and J. R. Barker, "On the physics and modeling of small semiconductor devices-II. The very small device," *Solid-State Electronics*, vol. 23, pp. 531-544, June 1980.

[9] M. S. Shur and L. F. Eastman, "Ballistic transport in semiconductor at low temperatures for low-power, high speed logic," *IEEE Trans. Electron Devices*, vol. ED-26, pp. 1677-1683, Nov. 1979.

[10] D. K. Ferry and J. R. Barker, "Balance equations for high field transport in the finite collision duration regime," *Solid State Commun.*, vol. 30, pp. 361-363, 1979.

[11] —, "Balance equation formulation and memory effects in retarded high field semiconductor transport," *J. Phys. Chem. Solids*, vol. 41, pp. 1083-1087, 1980.

[12] J. Grosvalet, C. Motsch, and R. Tribes, "Physical phenomena responsible for saturation current in FETs," *Solid-State Electronics*, vol. 6, pp. 65-67, Jan. 1963.

[13] F. N. Trofimenkoff, "Field-dependent mobility analysis of the field-effect transistor," *Proc. IEEE*, vol. 53, pp. 1765-1766, 1965.

[14] F. F. Fang and A. B. Fowler, "Hot electron effects and saturation velocities in silicon inversion layers," *J. Appl. Phys.*, vol. 41, pp. 1825-1831, Mar. 15, 1970.

[15] R. W. Coen and R. S. Muller, "Velocity of surface carriers in inversion layers on silicon," *Solid-State Electronics*, vol. 23, pp. 35-40, Jan. 1980.

[16] K. Hess and C. T. Sah, "Hot carriers in silicon surface inversion layers," *J. Appl. Phys.*, vol. 45, pp. 1254-1257, Mar. 1974; also "Warm and hot carriers in silicon surface inversion layers," *Phys. Rev. B.*, vol. 10, pp. 3375-3386, Oct. 1974.

[17] D. K. Ferry, "Hot-electron effects in silicon quantized inversion layers," *Phys. Rev. B.*, vol. 14, pp. 5364-5371, Dec. 15, 1976.

[18] K. Lehovec and R. Zuleeg, "Voltage-current characteristics of GaAs J-FETs in the hot electron range," *Solid-State Electronics*, vol. 13, pp. 1415-1429, 1970.

[19] H. Kroemer, "Hot electron relaxation effects in devices," *Solid-State Electronics*, vol. 21, pp. 61-68, Jan. 1978.

[20] H. Grubin, "Hot electron effects in semiconductor devices," in *Physics of Nonlinear Transport in Semiconductors*, D. K. Ferry, J. R. Barker, and C. Jacoboni, eds. New York: Plenum, 1980, pp. 311-326.

[21] J. G. Ruch, "Electron dynamics in short channel field-effect transistors," *IEEE Trans. Electron Devices*, vol. ED-19, pp. 652-654, May 1972.

[22] T. J. Maloney and J. Frey, "Transient and steady-state electron transport properties of GaAs and InP," *J. Appl. Phys.*, vol. 48, pp. 781-787, Feb. 1977.

[23] D. K. Ferry and J. R. Barker, "Physics and modeling of small semiconductor devices-III. Transient response in the finite collision duration regime," *Solid-state Electronics*, vol. 23, pp. 545-549, June 1980.

[24] P. J. Price, "Calculation of hot electron phenomena," *Solid-State Electronics*, vol. 21, pp. 9-16, Jan. 1978.

[25] D. K. Ferry, "Fundamental aspects of hot electron phenomena," in *Handbook of Semiconductors*, Vol. I., W. Paul, Ed. Amsterdam, the Netherlands: North-Holland (in press).

[26] W. Kohn and J. M. Luttinger, "Quantum transport theory of electrical transport phenomena," *Phys. Rev.*, vol. 108, pp. 590-611, Nov. 1, 1957.

[27] J. R. Barker, "Quantum transport theory of high-field conduction in semiconductors," *J. Phys. C: Solid State Phys.*, vol. 6, pp. 2663-2684, 1973.

[28] —, "High-field collision rates in polar semiconductors," *Solid-State Electronics*, vol. 21, pp. 267-272, Jan. 1978.

[29] K. K. Thornber, "High-field electronic conduction in insulators," *Solid-State Electronics*, vol. 21, pp. 259-266, Jan. 1978.

[30] D. K. Ferry, "Electron transport and breakdown in SiO_2," *J. Appl. Phys.*, vol. 50, pp. 1422-1427, Mar. 1979.

[31] P. Solomon and N. Klein, "Impact ionization in silicon dioxide at fields in the breakdown range," *Solid State Commun.*, vol. 17, pp. 1397-1400, 1975.

[32] For a discussion of scattering in Si, see, e.g., D. L. Long, "Scattering of conduction electrons by lattice vibrations in silicon," *Phys. Rev.*, vol. 120, pp. 2024-2032, Dec. 15, 1960.

[33] D. K. Ferry, "First order optical and intervalley scattering in semiconductors," *Phys. Rev. B.*, vol. 14, pp. 1605-1609, Aug. 15, 1976.

[34] J. G. Kirkwood, "The statistical mechanical theory of transport processes. I. General theory," *J. Chem. Phys.* vol. 14, pp. 180-201, Mar. 1946.

[35] R. Zwanzig, "Memory effects in irreversible thermodynamics," *Phys. Rev.*, vol. 124, pp. 983-992, Nov. 15, 1961.

[36] R. Kubo, "The fluctuation-dissipation theorem," *Rep. Progress Physics*, vol. 29, pp. 255-284, 1966.

[37] P. Das, D. K. Ferry, and H. Grubin, "Hot-carrier microwave con-

ductivity in the non-zero collision duration regime," *Solid State Communications*, in press.

[38] A general review of diffusion and noise is J. P. Nougier, "Noise and diffusion of hot carriers," in *Physics of Nonlinear Transport in Semiconductors*, D. K. Ferry, J. R. Barker, and C. Jacoboni, Eds. New York, Plenum, 1980. The current status of understanding of the hot electron diffusion problem is best summarized in [39].

[39] R. Fauquembergue, J. Zimmermann, A. Kaszynski, E. Constant, and G. Microondes, "Diffusion and the power spectral density and correlation function of velocity fluctuation of electrons in Si and GaAs by Monte Carlo methods," *J. Appl. Phys.*, vol. 51, pp. 1065–1071, Feb. 1980.

[40] D. K. Ferry and J. R. Barker, "Generalized diffusion, mobility, and the velocity autocorrelation function for high-field transport in semiconductors," *J. Appl. Phys.*, in press.

[41] See, e.g., M. Lax, "Classical noise III: Nonlinear markoff processes," *Rev. Modern Phys.*, vol. 8, pp. 359–379, Apr. 1966.

[42] See, e.g., N. G. Van Kampen, in *Fluctuation Phenomena in Solids*, R. E. Burgess, Ed. New York: Academic, 1965, pp. 139–179.

[43] P. Price, "Noise theory for hot electrons," *IBM J. Res. Develop.*, vol. 3, pp. 191–193, Apr. 1959; Also [42].

[44] V. L. Gurevich, "Current fluctuations in semiconductors near a non-equilibrium stationary state," *Sovi. Phys.-JETP*, vol. 16, pp. 1252–1259, May 1963.

[45] R. Grondin, private communication.

Reprint 5

JOURNAL DE PHYSIQUE

Colloque C7, supplément au n°10, Tome 42, octobre 1981 *page C7-227*

HOT ELECTRON NOISE PROPERTIES OF SEMICONDUCTORS IN THE NON-ZERO COLLISION DURATION REGIME*

P. Das**, D.K. Ferry° and H. Grubin°°

Rensselaer Polytechnic Institute, Troy, NY 12181, U.S.A.

°Colorado State University, Fort Collins, CO 80523, U.S.A.

°°Scientific Research Associates, Glastonbury, CT 06033, U.S.A.

Résumé.- Pour caractériser le bruit de composants en régime d'électrons chauds, on résoud l'équation de Langevin en incluant les effets du temps de relaxation de l'énergie et du moment dans l'approximation d'une distribution maxwellienne déplacée. On tient aussi compte de la durée de collision qui peut être une fraction non négligeable du temps de relaxation de l'énergie et du moment. On sait que la densité spectrale de puissance de bruit et par suite le coefficient de diffusion (en petits signaux) et la fonction d'autocorrélation de la vitesse sont reliées à la conductivité microondes des semiconducteurs (pour les petits signaux). Les composantes transversales et longitudinales de ces grandeurs sont calculées en incluant leur dépendance de la durée de la collision. Un calcul numérique utilisant les constantes du silicium a été développé et montre la contribution importante des durées de collision non nulles.

Abstract.- To obtain the noise properties of devices under hot electron condition, Langevin's equation is solved including the effects of energy and momentum relaxation time in the displaced maxwellian distribution approximation. The collision duration, which can be a significant fraction of energy or momentum relaxation time for very small structured devices, is also included. It is known that the noise power spectral density and thus the small signal diffusion coefficient and velocity auto-correlation are related to the small signal microwave conductivity of semiconductors. Both the transverse and longitudinal components of these quantities are calculated including their dependence on the magnitude of collision duration. Numerical calculation using the constants of silicon has been performed and it shows the significant contribution of the non-zero collision duration.

I. INTRODUCTION

It is well-known that the microwave conductivity of semiconductors varies as a function of frequency and that this functional dependence becomes quite complicated when hot electron transport is included [1]. This occurs because one must consider not only the momentum relaxation time of the carriers but also the energy relaxation time and any consequent differential repopu-

* Partially supported by ONR, NSF and AFOSR

** On leave at the Electrical Engineering and Computer Science Department, University of California at San Diego, U.S.A.

lation in many valley systems [2-4]. In particular, enhanced conductivity arises at high frequencies due to the process of velocity overshoot. In small semiconductor devices, the time scale of carrier transport through the device, with the expected high fields present, is such that the device dynamics may well be dominated by the transient response characteristics of the carrier velocity and distribution function [5]. However, if this becomes the case, major modifications are required to the Boltzmann transport equation [6] and to the current response equations within the devices [7,8]. The fact that the relaxation times are varying, due to the evolving of the average energy, on the same time scale appropriate to the velocity response and that a finite, non-zero collision duration exists both lead to a complicated, multiply-convolved form for the transport balance equations [8].

It is also known that the noise properties of semiconductors are related to the hot electron microwave conductivity [9,10,11,12]. For an example, the noise power spectrum is proportional to the real part of the microwave conductivity. As the noise power spectrum is related to the velocity auto-correlation through Fourier transform and also to diffusion co-efficient, all these qantities can be obtained from the knowledge of microwave conductivity as a function of frequency. Recently this microwave conductivity has been calculated in the non-zero collision region [13], applicable for a single valley semiconductor. The purpose of the present paper is to extend the calculations to multivalley case and relate them to the noise power spectrum, velocity auto-correlation and diffusion co-efficient through the solution of Langevin's equation.

II. MICROWAVE CONDUCTIVITY

i) Parallel polarization case

Let us consider that the semiconductor is under the influence of a d.c. bias electric field, F_o, in addition to the microwave field given by $F_1 e^{j\omega t}$ where $F_1/F_o << 1$ and ω is the angular frequency of the microwave field. For the parallel polarization case we assume that the direction of F_1 and F_o is collinear. Using the formulation outlined in ref. 13 it is easy to show the following equations for the perturbed drift velocity $v_1 e^{j\omega t}$ in the presence of a d.c. value v_o. Before we write down the equations we note that the inclusion of a non-zero collision time, τ_c, does not change the final equilbrium value, v_o. Also the only change in the momentum and energy balance equations imposed by τ_c is to change the momentum relaxation time τ_m by $\tau_m(1+j\omega\tau_c)$ and the energy relaxation time, τ_e by $\tau_e(1+j\omega\tau_c)$. Thus for a single valley case, one obtains

$$j\omega v_1 = -\frac{m v_1 \Gamma_{mo}}{1+j\omega\tau_c} - \frac{m v_o \Gamma'_m \tau_1}{1+j\omega\tau_c} + \frac{eF_1}{m} \qquad \text{(momentum balance)}, \qquad (1)$$

$$j\omega k T_1 = -\frac{kT_1\Gamma_{eo}}{1+j\omega\tau_c} - \frac{kT_o\Gamma'_e}{1+j\omega\tau_c} + \frac{2e}{3}\,(v_o F_1 + v_1 F_o) \quad \text{(energy balance)} \quad (2)$$

m is the effective mass of the carriers, k is the Boltzmann's constant, T is the effective temperature in the displaced Maxwellian distribution approximation and

$$\frac{1}{\tau_m} = \Gamma_m = \Gamma_{mo} + \Gamma'_m\, T_1 e^{j\omega t}$$

$$\frac{1}{\tau_e} = \Gamma_e = \Gamma_{eo} + \Gamma_e'\, T_1 e^{j\omega t}$$

The subscript "o" denotes the equilibrium value and the prime denotes the differentiation with respect T taken at $T = T_o$. Both τ_m and τ_e are effective relaxation times which include the effects of all possible forms of scattering mechanisms relevant for the case (e.g. acoustic, optical, intervalley phonons, impurity scattering etc.).

For the two-valley case, following ref. 14, one obtains the following modified equations,

$$j\omega v_{1i} = -\frac{mv_{1i}\Gamma_{moi}}{1+j\omega\tau_{ci}} - \frac{mv_o\Gamma'_{mi}T_{1i}}{1+j\omega\tau_{ci}} + \frac{eF_1}{m_i} \quad \text{(momentum balance)} \quad (3)$$

$$j\omega k T_{1i} = -\frac{kT_{1i}\Gamma_{eoi}}{1+j\omega\tau_{ci}} - \frac{kT_{oi}\Gamma'_e T_{1i}}{1+j\omega\tau_{ci}} + \frac{2e}{3}\,(v_o F_1 + v_1 F_o)$$

$$+ \frac{n_{jo}\, k\theta\, \Gamma_{ijo}}{n_{io}(1+\omega\tau_{cj})}\left[\frac{n_{j1}}{n_{jo}} - \frac{n_{i1}}{n_{io}} + \frac{\Gamma'_{ij}T_{1j}}{\Gamma_{ijo}}\right] \quad (4)$$

(energy balance)

$$j\omega n_{i1} = -n_{i1}\Gamma_{ni} - n_{io}\Gamma'_{ni} + n_{ji}\Gamma_{nj} + n_{jo}\Gamma'_{nj} \quad (5)$$

$$n_{11} + n_{21} = 0. \quad \text{(number balance)}$$

The subscripts i, j = 1, 2 and denotes the various parameters for the two valleys, τ_{ni} and τ_{ij} denotes the relaxation times associated with the number balance and intervalley phonon respectively,

$$\Gamma_{ni} = \frac{1}{\tau_{ni}}$$

$$\Gamma_{ij} = \frac{1}{\tau_{ij}}\,.$$

The quantities Γ_{ij}, Γ_{nj}, Γ'_{ij} and Γ'_{nj} are functions of T_j only. $k\theta$ is the characteristic phonon temperature used in equ. (4) only for convenience.

Equations (3-5) can be solved to obtain an analytical expression for the effective microwave conductivity, μ_{11}. However, it is rather cumbersome although straightforward and will not be given here. For numerical calculations it is better to tackle the equations directly. The above equations

can also be extended in a rather straightforward manner for the multi-valley case having more than two valleys.

For the single valley case, however, one obtains

$$\mu_{11}(\omega) = \mu_o \frac{1+j\omega\tau_c-\eta(\omega)}{1+j\omega\tau_{mo}(1+j\omega\tau_c)+\eta(\omega)} \tag{6}$$

where $\mu_o = \frac{e}{m\Gamma_{mo}}$

$$\eta(\omega) = (1+j\omega\tau_c)\frac{T_{eo}}{\gamma(\omega)T_{mo}}\Gamma_m'$$

and

$$\gamma(\omega) = 1 + \frac{T_{eo}}{T_{eo}}\Gamma_e' + \frac{3}{2}j\omega\tau_{eo}(1+j\omega\tau_c)$$

ii) Perpendicular polarization case.

If F_1 is normal to F_o then the perturbation in effective temperature is zero for first order calculations. Thus for this case equ. (1) simplies to

$$j\omega v_1 = -\frac{v_1\Gamma_{mo}}{1+j\omega\tau_c} + \frac{eF_1}{m} \tag{7}$$

From this we immediately obtain

$$\mu_\perp(\omega) = \mu_o \frac{1+j\omega\tau_c}{1-\omega^2\tau_{mo}\tau_c+j\omega\tau_{mo}} \tag{8}$$

For the two valley case one obtains for each individual valley, an expression for mobility given by equ. (8). The microwave conductivity for the two valley case is given by

$$\mu_\perp(\omega) = \frac{n_{10}\mu_{\perp 1}(\omega)+n_{20}\mu_{\perp 2}(\omega)}{n_{10}+n_{20}} \tag{9}$$

III. HOT ELECTRON NOISE PROPERTIES.

To obtain hot electron noise properties one need to solve the langevin's equation. For this purpose one assumes that a bias electric field is applied across the semiconductor in addition to the small fluctuating random electric field. Thus the difference between the problem solved in the last section and that of langevin's equation is that one should consider F_1 as the spectral component of the fluctuating electric field. As the equations (1)-(6) are linear equations around a bias electric field, the solution of the langevin's equation is equivalent to that discussed in references [9 and 12]. The noise spectral density for the equivalent noise voltage $\phi(\omega)$ for the single valley case is given by for the parallel polarization case

$$\phi(\omega) = 4kT(n_o e)\,\mathrm{Re}\,\mu_{11}(\omega) \tag{10}$$

where A includes the dimensional factors. For the perpendicular polarization case $\mu_{||}$ is replaced by the $\mu_{\perp}$. Applying Weiner-Khintchin theorem one obtains the velocity auto-correlation, $< v(t)v(t+\tau)>$, to be given by

$$<v(t)v(t+\tau)> \propto F^{-1}[\mu_r(\omega)] \tag{11}$$

where F^{-1} denotes inverse Fourier transform and $\mu_r(\omega)$ is the real part of the microwave mobility. As the diffusion co-efficient $D(\omega)$ is related to the Fourier transform of velocity auto-correlation, one obtains its value from the extended Einstein relationship

$$R_e D(\omega) = \frac{kT}{q} \mu_r \tag{12}$$

where R_e denotes the real part.

Numerical results and discussion

The small signal microwave conductivity was calculated for Si for an applied d.c. field at $30 KvCm^{-1}$. The scattering mechanisms and coupling constants are those used previously for Si [15-17]. The real and imaginary part of μ are defined as

$$\mu = \mu_r + j\omega\mu_I \tag{13}$$

and are plotted in fig. 1 and 2 for the parallel and perpendicular polarization cases respectively. The peaking at high frequencies is more pronounced in the presence of the non-zero collision duration, a result expected from calculations of overshoot velocity-itself.

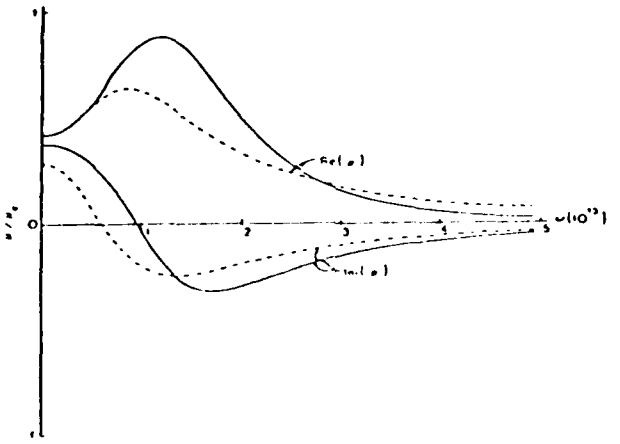


Fig. 1. Parallel Polarization: The real (a) and imaginary (b) parts of the a.c. small signal mobility for an applied d.c. field of 30 kV cm^{-1} in Si. The solid curve includes the effect of a finite, non-zero collision duration, while the dashed curve ignores this effect.

The peaking observed leads to an interesting shift of the apparent plasma edge in the semiconductor. In fig. 3 we plot the reflectivity |R| defined below for the parallel polarization case.

$$|R = \left|\frac{\epsilon^{\frac{1}{2}} - 1}{\epsilon^{\frac{1}{2}} + 1}\right| \tag{14}$$

where

$$\epsilon = \epsilon_s\left[1 - j\,\frac{ne^2\mu(\omega)}{\omega\epsilon_s\epsilon_0}\right]$$

and ϵ_s is the static dielectric constant of the semiconductor and ϵ_0 is vacuum permittivity. Three curves are shown in fig. 3. The

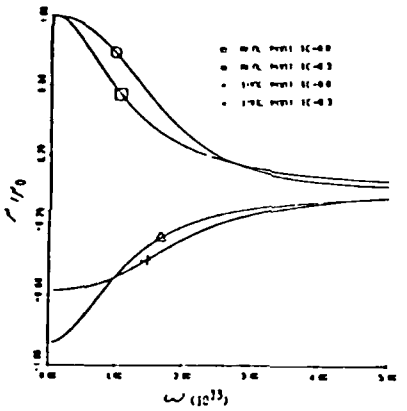


Fig. 2 Perpendicular Polarization: The real (a) and imaginary (b) parts of the a.c. small signal mobility for an applied d.c. field of 30 kV cm^{-1} in Si.

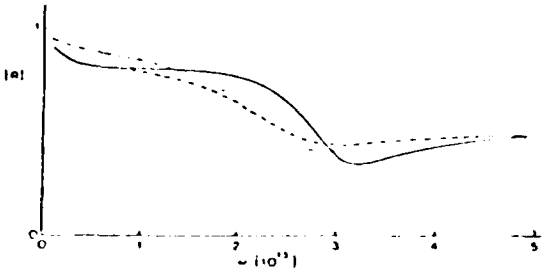


Fig. 3. The reflectivity of Si in a high electric field. The dotted curve assumes cold carriers with mobility μ_0, but a different chordal mobility. The solid curve includes the effect of the finite collision duration. A doping of 10^{18} cm^{-3} is assumed.

dotted curve assumes cold carriers with mobility μ_o while the dashed curve assumes hot carriers with the same μ_o but a different chordal mobility. The solid curve includes the effect of the finite collision duration. A doping of 10^{18} cm^{-3} is assumed. It is observed that the presence of hot carriers merely serves to smooth the apparent plasma edge. However, for $\tau_c \neq 0$, the minimum shifts significantly to higher frequencies.

The velocity auto-correlations obtained by directly Fourier transforming $\mu_r(\omega)$ is shown in figs. 4 and 5 for the two polarization cases. As expected, it is observed that the inclusion of non zero collision duration contributes significantly to the negative swing of the velocity auto-correlation.

In conclusion, the small signal microwave conductivity of semiconductors in the hot electron condition has been obtained in the non-zero collision regime which in turn has been used to obtain noise properties.

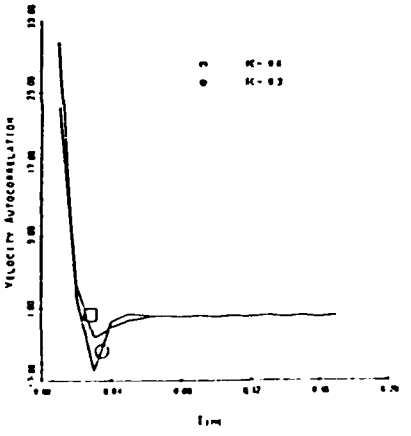


Fig. 4. Velocity auto-correlation for the parallel case. Units are arbitrary.

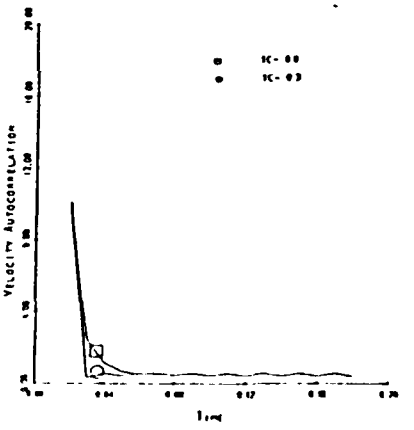


Fig. 5. Velocity auto-correlation for the perpendicular case. Units are arbitrary.

REFERENCES

[1] P. Das & D. K. Ferry, Solid State Electron, 19, 851 (1976).

[2] D. K. Ferry & P. Das, Solid State Electron, 20, 355 (1977).

[3] H. D. Rees, IBM J. Res. Develop, 13, 537 (1969).

[4] H. L. Grubin, D.K. Ferry & J. R. Barker, Proc. IEDM, p. 394 (1979).

[5] D. K. Ferry & J. R. Barker, Solid State Electron, 23, 545 (1980).

[6] J. R. Barker & D. K. Ferry, Solid State Electron, 23, 519 (1980).

[7] D. K. Ferry & J. R. Barker, Solid State Commun. 30, 361 (1979).

[8] D. K. Ferry & J. R. Barker, J. Phys. Chem. Solids (in press).

[9] A. Van Der Ziel, "noise in Solid State Devices", in Advances in Electronics and Electron Physics, Ed. L. Harton, Vol. 46, Academic Press, (1978).

[10] J. P. Nougier, D. Sodini, M. Rolland, D. Gasquet and Gasquet and G. Lecoy, Solid State Electron, 21, 133 (1978).

[11] A. Van Der Ziel, Solid State Electronics, Vol. 23, pp. 1035-1036 (1980).

[12] C. W. Helstrom, "Markov Processes and applications", in Communcation Theory", Ed. A. V. Balakrishnar, McGraw Hill, (1968).

[13] P. Das, D. K. Ferry and H. Grubin, Solid State Commun., to be published.

[14] P. Das, Appl. Phys. Letters, Vol. 11, pp-386-388 (1967).

[15] D. K. Ferry, Phys. Rev. B14, 1605, (1976).

[16] D. K. Ferry, Surf. Sci., 57, 218 (1976).

[17] D. K. Ferry, Phys. Rev. B.14, 5364 (1976).

Limitations to Ballistic Transport in Semiconductors*

D. K. FERRY, J. ZIMMERMANN**, P. LUGLI, AND H. GRUBIN

Abstract—**Limitations to the range of ballistic transport in semiconductors are discussed on the basis of electron correlation within an ensemble. It is shown that transport equations, correct in the fast transient regime, include these correlation effects.**

IN recent months, it has become necessary to talk about ballistic transport in very short semiconductor devices [1-3]. Usually, it is assumed that if the device dimension is smaller than an average free path, based upon a value interpreted from the low-field mobility, collisions will not be important. The basis for this is usually placed on the often-used semiempirical transport equations [1,4].

$$m^* \frac{dv_d}{dt} = qF - mv_d/\tau_m , \tag{1}$$

$$\frac{d}{dt}\bar{E} = qv_dF - (\bar{E} - E_o)/\tau_e , \tag{2}$$

where E and E_o are the average and zero-field energies, and τ_m and τ_e are the empirical momentum and energy relaxation times, respectively. However, it has been pointed out that these equations neglect boundary conditions [5,6] and indeed neglect spatial inhomogeneities which can be dominant [6-8]. Ballistic transport is usually treated by seeking the response to (1)-(2) of a pseudoparticle, the so-called average electron, as it evolves under the field. However, even without considering the complications due to spatial inhomogeneities, these equations are incorrect for treating the fast, transient response of carriers in semiconductors.

A conduction electron in a semiconductor is not a free electron. Within the effective mass approximation, the conduction electron is a quasiparticle whose effective mass describes an averaged (and renormalized) interaction with the atoms and the bound electrons. It should be apparent that the correlation between electrons is of importance. This is more evidenced when it is recognized that the kinetic equation, such as the Boltzmann transport equation, for the average single-particle distribution function is an approximation to the full many-body problem of large numbers of conduction electrons [9,10]. While it is possible to project such single-particle equations, these must be cast so description of the correlations between the electrons – the correlation functions – is recoverable. Indeed, it was just this view that led Kubo to formalize transport theory entirely in terms of the correlation functions [11, 12]. We propose to show how proper treatment of the initial transient response limits the range of ballistic transport.

In a semiconductor subject to an applied electric field, the carriers respond to this field as well as to a random force which leads to velocity fluctuations as well as to the concept of temperature (thermal fluctuations). The response to the applied field (applied at $t = 0$) can be written as [12]

$$v_d(t) = (eF/m^*) \int_o^t \phi_v(t',0)\,dt', \tag{3}$$

where $\phi_v(t',0)$ is the non-stationary, two-time velocity autocorrelation function whose amplitude is normalized to unity at $t' = 0$. Although Kubo obtained (3) for the equilibrium situation, its validity has also been established for the non-equilibrium, high-field case [13]. A linear increase of the velocity ($v_d = eFt/m^*$) can only occur so long as ϕ_v is constant in time. To demonstrate this period, we have calculated the transient dynamic and velocity autocorrelation function for silicon, using an ensemble Monte Carlo approach used previously [14]. Figure 1 shows the initial decay of $\phi_v(t', t_o)$ $(t' > t_o)$ for three different values of t_o including $t_o = 0$. What is evident from this figure is that the time duration over which ϕ_v is constant, and for which the velocity rises in a ballistic manner, is exceedingly short, perhaps only 0.004 psec, *even though the mean free time (~.16 psec at t = 0) is much longer.* The decay of ϕ_v represents a decay in the correlation of the velocity fluctuations as well as an increase in the dissipation because of collisions. This is due to the fact that collisions start to break up the correlation as soon as they occur, but the relaxation time is a hydrodynamic average over the ensemble rather than a value for actual collisions.

Equations such as (1) and (2), although incorrect for time-scales of the order $t \leq \tau_m$, have proven themselves extremely useful in device modeling. Equivalent versions, correct even on the short-time scale, have been obtained recently [15-17]. These are given by

$$m^* \frac{dv_d}{dt} = qE - m^* \int_o^t dt'\, X_v(t')\, v_d(t - t'), \tag{4}$$

$$\frac{d\bar{E}}{dt} = qFv_d(t)\,[1 - \phi_v(t,0)] - \int_o^t dt'\,[\bar{E}(t - t') - E_o]\, X_e(t'), \tag{5}$$

where X_v and X_e are decay functions related to the velocity

Manuscript received June 8, 1981; revised July 20, 1981.

D. K. Ferry, J. Zimmermann, and P. Lugli are with the Department of Electrical Engineering, Colorado State University, Fort Collins, CO 80523. H. Grubin is with Scientific Research Associates, Glastonbury, CT 06033.

*Work supported in part by the Army Research Office.

**Permanent address: C.H.S. and Greco Microondes, Universite de Lille, France.

0193-8576/81/0900-0228$00.75 © 1981 IEEE

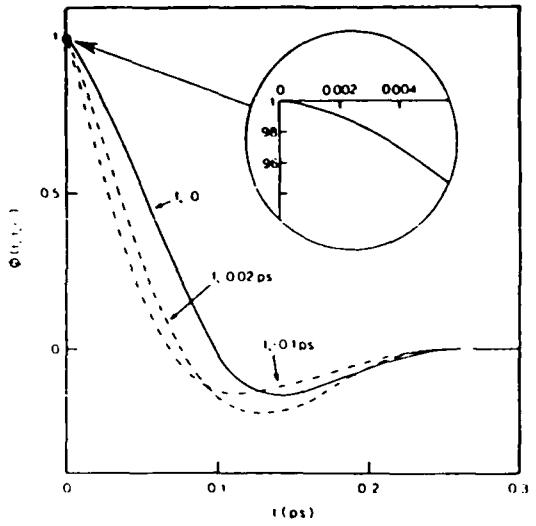


Fig. 1. Normalized velocity autocorrelation function for electrons in silicon, $\phi_v(t - t_O)$, for three initial times t_O. The function ϕ_v is constant only for a very short time.

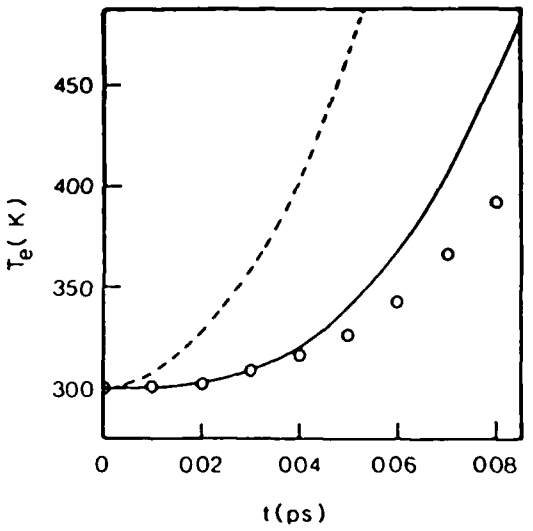


Fig. 2. Response of the average energy of an ensemble of electrons in silicon for $E = 50$ kV/cm applied at $t = 0$. The data points are calculated by an ensemble Monte Carlo technique.

and energy correlation functions. Indeed, the time integrals of X_v and X_e give the relaxation rates $1/\tau_m$ and $1/\tau_e$ in the limit as $t \to \infty$. In this regard, (4) is particularly consistent with the correlation function approach of (3) as it is an ensemble average of a generalized, retarded Langevin equation [13]. Except for the details of the correlation function and the convolution in the relaxation term, (4) differs little from (1). Such is not the case for the energy equation (5) because of the presence of the memory function in the driving term.

One problem in treatment of so-called ballistic transport has been in handling the energy equation. In these, $\bar{E}$ has been treated as the *drift* energy. In actual fact, the energy is dominated by the random (thermal) motion of the carriers, and it is the *total* energy which must be treated in equations such as (2) or (5). In this regard, the form of (5) becomes obvious. The dominant energy is the thermal energy and the memory function represents a prolonged correlation of the velocity fluctuation spectrum with the equilibrium state. When the electric field is switched on at $t = 0$, the ensemble responds instantaneously with a shift in momentum space corresponding to the field and time. Thus, the velocity begins to rise instantaneously. However, spreading of the ensemble (increase of fluctuations) that leads to an increase in temperature does not begin instantaneously. Rather, the spread of the ensemble is characteristic of the thermal motion of the individual carriers. As a consequence, the temperature cannot begin to evolve until collisions begin to break up the correlations with the equilibrium state. This is shown in Fig. 2, where the silicon results were obtained using the ensemble Monte Carlo calculation mentioned above. The solid curve represents only the forcing terms from (5), while the dashed curve represents (2). Clearly, the memory function is playing a role. Even more clearly, the collisional decay terms are already evident for $t = 0.04$ psec.

It would be desirable to separate drift and thermal energies as is done in plasma physics [14]. This procedure is untenable one[illegible] consequence, the transient regime must be treated so that the important role of correlation among electrons can be included. Indeed, these results indicate that ballistic response occurs on time scales shorter than in silicon and that any high velocities greater than the steady-state must arise not from ballistic effects, but from overshoot effects caused by the details of energy and momentum relaxation [18,19]. The importance of contacts and boundary conditions is also clearly evident from (3), as these determine the "equilibrium" state of the carriers at $t = 0$.

REFERENCES

[1] M. S. Shur and L. F. Eastman, "Ballistic Transport in Semiconductor at Low Temperatures for Low-Power, High-Speed Logic," *IEEE Trans. Electron Dev.*, ED-26, pp. 1677-83, November 1979.

[2] W. R. Frensley, "High-Frequency Effects of Ballistic Electron Transport in Semiconductors," *IEEE Electron Device Letters*, EDL-1, pp. 137-9, July 1980.

[3] R. Zuleeg, "Possible Ballistic Effects in GaAs Current Limiters," *IEEE Electron Device Letters*, EDL-1, pp. 234-5, November 1980.

[4] H. L. Grubin, D. K. Ferry, G. J. Iafrate, and J. R. Barker, "The Numerical Physics of Micron-Length and Submicron-Length Semiconductor Devices," in *Microstructure Science and Technology/VLSI*, vol. 3, N. Einspruch, Ed., New York: Academic Press, in

[5] J. R. Barker, D. K. Ferry, and H. L. Grubin, "On the Nature of Ballistic Transport in Short-Channel Semiconductor Devices," *IEEE Electron Device Letters*, EDL-1, pp. 209-10, October 1980.

[6] J. J. Rosenberg, F. J. Yotta, and M. I. Nathan, "Importance of Boundary Conditions to Conduction in Short Samples," *IEEE Trans. Electron Dev.*, ED-25, August 1981.

[7] K. Hess, "Ballistic Electron Transport in Semiconductors," *IEEE Trans. Electron Dev.*, ED-25, August 1981.

[8] R. K. Cook and J. Frey, "Ensemble Carrier Effects and 'Ballistic Transport'," *IEEE Trans. Electron Dev.*, ED-28, August 1981.

[9] N. N. Bogoliubov, in *Studies in Statistical Mechanics*, vol. 1, J. deBoer and G. E. Uhlenbeck, Eds., Amsterdam: North-Holland, 1962.

[10] H. J. Kreuzer, *Nonequilibrium Thermodynamics and its Statistical Foundations*, Oxford: Clarendon Press, 1981.

[11] R. Kubo, "Statistical-Mechanical Theory of Irreversible Processes," *J. Phys. Soc. Japan*, vol. 12, pp. 570-586, June 1957.

[12] R. Kubo, "Response, Relaxation, and Fluctuation," in *Transport Processes*, vol. 3[illegible] G. Kirczenow and J. Marro, Eds., Bel[illegible] Verlag, 1974, pp. 75-125.

[13] J. Zimmermann, P. Lugli, and D. K. Ferry, "Non-Equilibrium Hot Carrier Diffusion Phenomenon in Semiconductors," *J. de Physique*, in press.

[14] D. K. Ferry and J. R. Barker, "Generalized Diffusion, Mobility, and the Velocity Autocorrelation Function for High-Field Transport in Semiconductors," *J. Appl. Phys.*, vol. 52, pp. 818-24, February 1981.

[15] D. K. Ferry and J. R. Barker, "Balance Equation Formulation and Memory Effects in Retarded High Field Semiconductor Transport," *J. Phys. Chem. Solids*, vol. 41, pp. 1083-7, 1980.

[16] D. K. Ferry, "Transport in Submicron Devices," *J. de Physique*, in press.

[17] D. K. Ferry, P. Lugli, and J. Zimmermann, "The Ensemble Monte Carlo Method and Generalized Transport Equations for Transient Dynamic Response in Semiconductors," submitted for publication.

[18] J. G. Ruch, "Electron Dynamics in Short Channel Field-Effect Transistors," *IEEE Trans. Electron Dev.*, ED-19, pp. 652-4, May 1972.

[19] T. J. Maloney and J. Frey, "Transient and Steady-State Electron Transport Properties of GaAs and InP," *J. Appl. Phys.*, vol. 48, pp. 781-7, February 1977.

Reprint 7

CHAPTER XIX

TRANSPORT IN SUBMICRON DEVICES*

D. K. Ferry
Department of Electrical Engineering
Colorado State University
Fort Collins, CO 80523

H. L. Grubin
Scientific Research Associates
Glastonbury, CT 06033

J. R. Barker
Warwick University
Coventry CV4 7AL, U.K.

INTRODUCTION

Over the last two decades, the electronics industry has been involved in an ongoing revolution in digital large-scale integration (LSI). This digital revolution, spawned in the late 1960's, is leaving a permanent imprint on all aspects of life today, especially as the implementation of microelectronics has spread to the consumer industry. Fueled by the drive to less expensive, but more complex and sophisticated, integrated systems, the growth of LSI has in fact been phenomenol. The complexity of these circuits, in terms of the number of individual devices on a chip, has approximately doubled each year over this time span. This is shown in Fig. 1. There are, of course, several factors which contribute to this increase in complexity, including major effects arising from increased die size, increased circuit cleverness, and reduced device size. This latter factor, reduction of the individual feature size in a device, is of paramount importance and dimensions of laboratory systems are currently down to the sub-micrometer range. Indeed in Fig. 2, the leukemia virus is overlaid over a modern integrated circuit in order to emphasize the smallness of individual devices today. Progress in the microelectronics area is tied inevitably to the ability to continue to put ever increasing numbers of smaller devices on a chip; i.e.--the continual move to very-large-scale integration (VLSI) will be of paramount importance to this continued progress. It is apparent that extrapolation of today's technology will produce individual devices whose dimensions are of the order of 0.1-0.3 micrometers [3-7]. On the other hand, the advent of high-resolution electron, X-ray, molecular, and ion beam lithography is leading us toward an era in which individual feature sizes might well be fabricated on the molecular scale

*This work was supported in part by ONR and ARO.

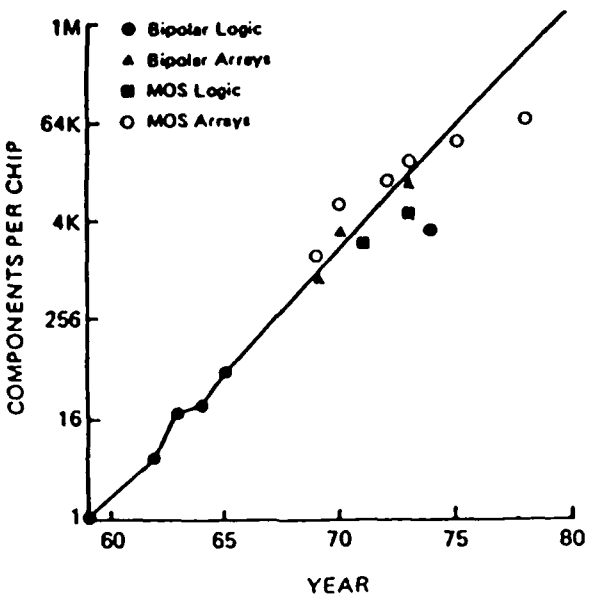


Figure 1. Growth of circuit complexity in large-scale integrated systems according to Moore [1].



Figure 2. Projection of a leukemia virus cell onto a modern integrated circuit [2]. The outline of the devices is the size of the virus cell.

balance tells us that the electron temperature increases with increasing electric field and departs significantly from room temeprature when the electric field exceeds a threshold value. The effect of the increasing electron temperature is to decrease the average collision time and to decrease the steady-state velocity, given by (3). If the momentum and energy scattering times are similar in value, then both momentum and energy will follow changes in electric field at approximately the same rate and the solid curve of Fig. 3 describes the approach to steady-state. On the other hand, if the energy scattering time is significantly longer than the momentum scattering time, the average velocity of the carriers will find its value continually corrected until steady-state in the energy distribution is reached. The velocity will then relax in the manner shown by the dashed curve. A similar situation may be expected when the electric field is decreased, for here it also takes a finite time for the electric field to decrease and for the electron temperature to decrease.

Most field-dependent velocities and values for the saturated velocity assume that steady-state conditions are reached. Clearly, this is not the case in very-small devices. In Fig. 4, we show the average velocity as a function of distance for electrons in Si seeing an electric field of 50 kV/cm. Also seen is the small change induced by retardation of the transport. It is clear that the relaxation rates τ_m^{-1} and τ_e^{-1} are evolving on the same time scale as the velocity response itself and that correlated motion needs to be considered. Indeed, the correlation function for electrons in Si is shown in Fig. 5 and 6 The vertical scales are shifted to allow ease of plotting. What is clear here is that $\phi(t)$ (the stationary quantity) lasts for a time fully comparable to the transient portion of the velocity response. In fact, it can be shown that [13]

$$v_d(t) = \frac{eE}{mI} \int_0^t \phi(o,t')\, dt' \quad , \tag{4}$$

where $I = \langle v^2(o) \rangle$ and $\phi(t",t')$ is the general non-stationary two-time correlation function [$\phi(t",t') \to \phi(t)$ for $t' \to t" + t$ and $t" \to \infty$]. The proper treatment of transport in the transient regime thus requires inclusion of memory functionals which are beyond the channel Boltzmann equation approach [14]. This leads to modifications of the relaxation terms in (1) and (2). The results for Si are also shown in Fig. 4. It is evident however, that if the channel were only 200 Å long, the velocity would never reach steady-state. While this is a very short distance, the effect is pronounced in the III-V materials and results

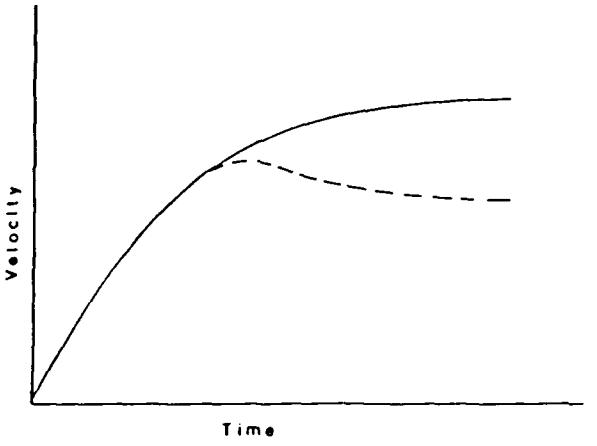


Figure 3. Approach to steady-state of the velocity response with and without overshoot effects.

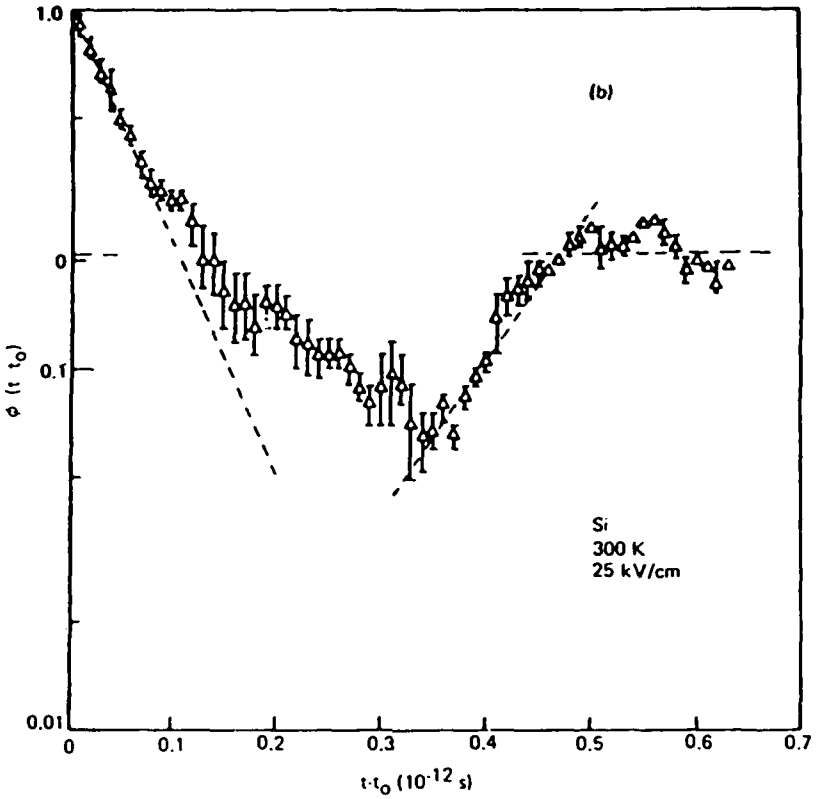


Figure 5. The correlation function for electrons in Si at 25 kV/cm.

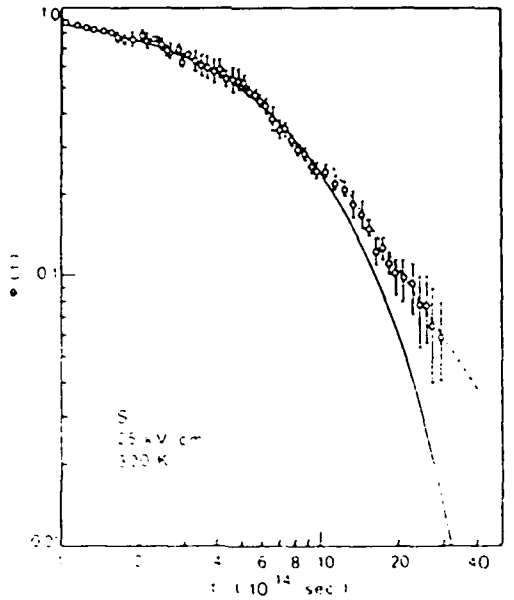


Figure 6. The correlation function of Figure 5 on a log-log plot to clearly show the long-time $t^{-3/2}$ decay of the initial fall of $\phi(t)$.

ning to dominate the interconnection capacitance at a particular node. As a result, the above simplification is likely to be seriously in error for submicron configured VLSI systems, where the isolation of one device from another (and by generalization, from the surrounding environment of insulating and conducting regions) will be difficult to achieve.

The possible device-device coupling mechanisms are numerous and include such effects as capacitance coupling, the line-to-line parasitic effect mentioned above, and wave-function penetration (tunneling and charge spill-over) from one device to another. Formally, however, one may describe these effects on system and individual device behavior by assuming the simplest form of inter-device coupling -- nearest neighbor coupling -- and using the formally exact Liouville-von Neumann density matrix equation for the full system to unravel how the system behaves. This task has been described previously from various different viewpoints [16,17], and the basic ideas are reviewed below in order to provide a baseline to which the system approach can be compared.

In the present context, a VLSI system is defined as a net of N spatially delineated structures, e.g. contacts, devices, interconnects, isolation/insulator regions, etc. Control over the system is exercised via a set of applied fields (or voltages) and input and output currents. These are labeled F^i_{ext} $(i = 1,\ldots,N)$ where F^i_{ext} is the set of generalized applied forces acting on the i-th element. The applied generalized forces give rise to local applied forces F^i which are the self-consistent solutions of the appropriate macro-equations, e.g., Poisson's equation and current continuity equation. The latter depend upon the dynamical variables of the elements concerned, and these variables are completely specified as quantum statistical expectation values of the individual device/element density matrix $\rho(i;t)$. If there is no coupling between devices, we have simply the set of N Liouville-von Neumann equations of motion (using units such that $\hbar = 1$)

$$i \frac{\partial}{\partial t} \rho(i;t) = \hat{H}(i)\rho(i;t) \quad , \tag{5}$$

where $\hat{H} = [h,\ldots,]$ is the commutator generating Hamiltonian super-operator. $H(i) = H(i,F^i(t))$ is the Hamiltonian for device i and is assumed to be time dependent through the coupling to the generalized time dependent forces F^i. If the device-environment coupling occurs on a time-scale fast compared to processes within the device, (5) can be reduced to a single $\rho_d = \rho_i$ which satisfies [16]

$$i\hbar \frac{\partial \rho_d}{\partial t} = \hat{H}_d \rho_d + \hat{H}_{ed} \rho_d + \hat{\zeta}(o)\rho_d \quad , \tag{6}$$

where $\hat{H}_d$ is the Hamiltonian for the single device, $\hat{H}_{ed}$ is the renormalization term for the effective real part of the device-environment interaction, and $\hat{\zeta}(o)$ is a dissipative term for losses to the environment, such as surface-roughness scattering in an MOS system. The term in $\hat{H}_{ed}$ is critical, in that regularity in the replication of the devices, such as an array, can lead to complete renormalization of the energy structure and super-lattice behavior.

Recently, Bate [18] has proposed a surface super-lattice structure that formally is similar to a charge-coupled-device (CCD) array. The device dimensions required were ~ 1000 Å spacings for the array, which is beyond the current VLSI technology, but within the limits of research efforts in electron- and ion-beam lithography. Although such lateral super-lattices are interesting in their own right, since they easily allow full three-dimensional quantization within a quantized inversion layer, they are especially interesting as they should also

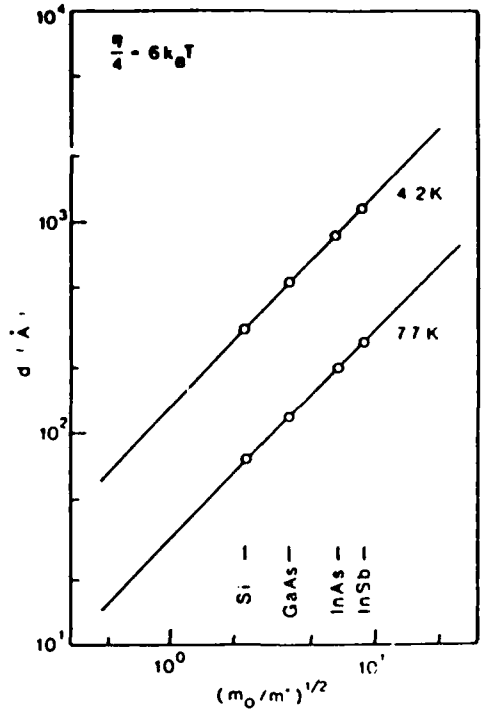


Figure 10. (a) Value of gate center-to-center spacing required for a super-lattice to form with n/4=6kBt, where n is the reduced energy L22L.

the mini-gap. These effects are well known in bulk semiconductor materials, leading, for example to correlation contributions [23] to band-gap narrowing [24]. Whereas a coulomb contribution to the potential of a few millivolts is small when compared to a band-gap of 1-2 volts, it can be a dramatic effect on a mini-gap whose total value is only 10 millivolts or so. This effect was carefully examined by Kroemer [19] for the case of a one-dimensional super-lattice, such as can occur in an organic chain molecule or polymer. The basic idea is of course that charge fluctuations across the mini-gap lead to a coulomb potential of $q \sim g \sim 2k_F$, so that the conditions for a Kohn anomaly [25] are already satisfied. While we have been concerned with optical pumping of the electrons, it should be pointed out that optical pumping of the phonons can lead to phase transitions as well [26].

If a population is induced by some technique, there exists the possibility of zeroes appearing in the dielectric response. In this case, a non-zero value of V_o can exist without the application of the external field V_a. This means that the super-lattice potential can be set up by a charge instability under the gates themselves. While it was initially pointed out that this charge effect on ϵ was similar to a Kohn anomaly, the instability itself is a classic example of a Peierl's instability. The charge instability spontaneous leads to a charge-density wave which creates the lattice potential. This behavior is illustrated in Fig. 11.

If suitable materials are selected, it appears that lateral super-lattices can be fabricated by today's technology, or indeed can be expected to arise in device VLSI arrays. The fact that these structures can show synergetic behavior under conditions of population inversion suggest that new functional performance

13. J. Zimmermann, P. Lugli, and D. K. Ferry, to be published.
14. D. K. Ferry and J. R. Barker, J. Phys. Chem. Sol. 41, 1083 (1980).
15. T. J. Maloney and J. Frey, J. Appl. Phys. 48, 781 (1977).
16. J. R. Barker and D. K. Ferry, Sol.-State Electron. 23, 531 (1980).
17. J. R. Barker and D. K. Ferry, in "Proc. Intern. Conf. on Cybernetics and Society," IEEE Press, New York, 1979, p. 762.
18. R. T. Bate, Bull. Am. Phys. Soc. 22, 407 (1977).
19. H. Kroemer, Phys. Rev. B 15, 880 (1977).
20. H. Haken, Synergetics (Springer-Verlag, Berlin, 1978).
21. N. David and N. Flesher, Naval Ocean Systems Center, unpublished.
22. D. K. Ferry, to be published.
23. See, e.g., J. C. Inkson, J. Phys. C 9, 117 (1976); V. Heine and J. A. Van Vechten, Phys. Rev. B 13, 1622 (1976); and references therein.
24. H. Y. Fan, Phys. Rev. 82, 900 (1951); C. Haas, Phys. Rev. 125, 1965 (1962); H. C. Casey, Jr., D. D. Sell, and K. W. Wecht, J. Appl. Phys. 46, 250 (1975).
25. W. Kohn, Phys. Rev. Letters 2, 393 (1959).
26. S. A. Bulgadaev and I. B. Levinson, Sov. Phys.-JETP 44, 1043 (1976).

DISCUSSION

Dr. Sandman: GTE - I am trying to understand precisely what you meant by the term superlattice as you applied it apparently in the array of devices. There is a well-known class of semiconductor superlattices, gallium arsenide and aluminum arsenide, and I think you meant something different than that. Could you specify in some detail precisely what you meant by a superlattice in the context that you used it?

Prof. Ferry: You are referring to the extreme technique of laying down precise alternate monolayers of aluminum arsenide and gallium arsenide. This superlattice, of layered aluminum arsenide and gallium arsenide gives a narrow energy well and a classic one-dimensional Kronig-Penney model, and it works exceedingly well when you compare the results of that calculation to the experiments. Now the concepts of superlattices are, of course, much broader than that. What you should think of is that the reason you get superlattice effects is the fact you have a periodic array of energy wells.

Dr. Sandman: In an individual chip?

Prof. Ferry: Think of this in the following way: under the gate of a MOS FET, you have a potential well due to band bending. If you have a regular array of gates, you will have a periodic potential well induced along the surface in the context of a lateral superlattice. Now the spacing of those wells determines the potential and you have a two-dimensional Kronig-Penney model. This two-dimensional lattice is worked out in Brillouin's book on periodic structures, but you can have superlattice effects in the two-dimensional case. Now the challenge arises if you make this on top of a layered structure by MBE; because you can then have a three-dimensional superlattice effect.

Dr. Sandman: This is the context in which you were using it?

Prof. Ferry: This is the context, right.

Dr. Carter: NRL - Thank you for a very well presented and nicely balanced talk. I appreciate your effort. I have two questions. You have mentioned the quantum effects of an array of devices, and so my first question deals with the phase effect in Josephson's junctions which is a macroscopic quantum effect. Is that the sort of array that you are talking about? Have

people in the optical sciences and also that is using interference methods for storing the processing information. In Josephson's junctions, this has also been done in so-called SQUIDS where you do have well-known macroscopic interference effects, but what I think is the really interesting question is what happens to the theory of information in a non-local situation where you are not trying to transfer a single localized byte but the byte is spread over both an amplitude and phase over a large area, and I think there are people in the brain business who have some strong ideas that this is already taking place inside our head--some of us.

Prof. Ferry: Last week in Phoenix, I learned there is a group at Cal. Tech. which is trying to apply the microscopic concept of spin glasses to the way in which the brain handles information. I mention that only in passing because I also learned at the same time that there is a group at IBM which is trying to apply the theory of spin glasses to the way in which systems operate. If you tie the two together, it might be interesting.

Dr. Buot: Cornell Univ. - I was wondering if you have thought of using BiSb alloys as a matrix on which to lay the small metal dots? (Added in editing - Two things conspire to increase the change of practically observing quantum cooperative behavior in an array of normal metal dots or islands, namely: 1) small effective electron mass in each metal dot, and 2) a small potential barrier between metal dots).

Prof. Ferry: I have enough trouble handling the III-V's. I have tried to forget about bismuth antimonide. That is from the days when I was doing instability experiments.

Dr. Buot: I was wondering if anybody has successfully grown bismuth antimonide semiconducting crystal films.

Prof. Ferry: The last person I remember working on bismuch antimonid was at Bell; it was some years ago.

Dr. Cooper: There was somebody at Illinois, I think Joe Green.

Prof. Ferry: I don't know if he was doing bismuth antimonide There is a classic rule of thumb which has been put forward by Cyril Hilsum at RSRE, which is that to bring any new material to the level at which you can begin thinking about making devices requires something like a man century of effort. I don't know that there is the intense effort on bismuth antimonide yet which would bring it to the level that people could think about using it. Silicon has at least a kiloman century work of effort on it just since the war.

Reprint 8

VLSI ELECTRONICS: MICROSTRUCTURE SCIENCE, VOL 3

Chapter 6

The Numerical Physics of Micron-Length and Submicron-Length Semiconductor Devices

H. L. GRUBIN

Scientific Research Associates, Inc.
Glastonbury, Connecuticut

D. K. FERRY

Colorado State University
Fort Collins, Colorado

G. J. IAFRATE

U.S. Army Electronics Technology and Devices Laboratory
Fort Monmouth, New Jersey

J. R. BARKER

Physics Department
University of Warwick
Coventry, England

I Introduction 198
II. The Semiconductor Equations 202
A Introduction 202
B One-Dimensional Analysis of GaAs Two-Terminal Devices 204
C Two-Dimensional Analysis of GaAs FETs 212
D Two-Dimensional Analysis of a Silicon MOSFET 216
III The Boltzmann Transport Equation 220
A Introduction 220
B Displaced Maxwellian Equation 223

Copyright © 1988 by Academic Press, Inc.
All rights of reproduction in any form reserved.
ISBN 0-12-234103-1

C. Semiempirical Transport Equation Method 225
D. Numerical Techniques 228
E. Velocity Transients 233
F. Device Simulation from the Boltzmann Transport Equation 238
G. Ballistic Transport 257
IV. Quantum Transport Theory 261
A. Introduction 261
B. Quantum Transport Formulation 263
C. Synergetic Effects from Device–Device Interactions 268
V. Diffusion 273
A. Introduction 273
B. Diffusion Formalism 274
C. Correlation Functions for Hot Electrons 276
D. Transient Diffusion 286
Appendix A. Derivation of the Balance Equations 288
Appendix B. The Wigner Distribution Function 290
References 296

I. INTRODUCTION

A turning point in the study of semiconductor devices occurred with the publication of the special January 1966 issue of the *IEEE Transactions on Electron Devices*. Here papers dealing with the numerical simulation of the space- and time-dependent behavior of the charge distribution within solid-state devices implied, either explicitly or implicitly, that these types of calculations could be used to develop the intuition needed to explain device behavior.

Today, numerical simulations are regarded as part of a device physicist's tools and are routinely used (1) when the device transport is nonlinear and the device differential equations do not admit to exact solutions, (2) as surrogates for laboratory measurements that are costly and/or not feasible, and (3) in computer-aided device design. Specifically, the kinds of devices most often treated numerically either require multidimensional concepts for their operation or are dominantly hot-carrier dependent. Transferred electron devices are examples of hot-carrier devices, but an MOS or MESFET requires two dimensions for conceptual operation and is often modified by hot-carrier contributions. In any case, through a heirarchy of differential equations, we are becoming increasingly able to simulate the operation of devices such as these and to numerically represent the effects of

(i) material properties (e.g., doping variations),
(ii) physical boundaries (e.g., contacts, surface states), and

(iii) the environment (e.g., adjacent devices, circuits) on device operation.

This numerical ability, coupled with heightened interest in device physics and generated in large measure by the current VLSI and VHSIC programs, is also forcing a long-needed reassessment of the assumptions used in device simulations. For example, most modeling of single-species transport represents the current response by the equation

$$\mathbf{J} = Ne\mu\mathbf{F} + eD\ \mathbf{grad}\ N + \epsilon\ \partial F/\partial t, \tag{1}$$

where N is the carrier concentration, e the electronic charge, F a field, μ the mobility, D the diffusivity, ϵ the dielectric constant, and t the time. This equation is based on the Boltzmann transport equation (BTE) for a distribution function only slightly modified by a self-consistent field [1]. For high-field, nonlinear, hot-carrier effects, the distribution function is strongly dependent on F and the usual approach replaces μF by a nonlinear $v(F)$ curve, and D by a field-dependent diffusivity [2]. This approach, although useful, avoids the problem of finding solutions to nonlocal transport equations, which include spatial and temporal field-dependent relaxation. Such solutions are especially important in GaAs, where recent history has shown that concepts like velocity overshoot[1] [3–5] profoundly affect transport in short-channel devices and the more conventional devices operated at high frequencies. To account for these effects, the BTE, rather than Eq. (1), must be solved.

The BTE is the cornerstone of semiclassical transport. Its central concept is the idea of a *single* carrier-distribution function $f(r, p, t)$, which may be used to compute expectation values for macroscopic current flow. When we examine this idea from a quantum-mechanical viewpoint, we see some necessary averaging if f is to be regarded as a simultaneous function of position and momentum. From a device perspective, care must therefore be exercised if f and/or the critical device dimension L are smaller than an electron wavelength Λ. For a quasi-particle with an effective mass m^*,[2]

$$\Lambda^* = h/m^* v_{\text{therm}}. \tag{2}$$

[1] The effects of overshoot are implicit in the results of [4].

[2] Note that on an ultra-short time scale, electrons accelerate with the free, rather than the effective, mass [6]. Also, although the constraint $L > \Lambda^*$ is a safe one, it does not imply that the BTE is invalid in a range

$$\Lambda < L < \Lambda^*, \tag{3}$$

for the range of validity of the BTE has not yet been delineated. Here $\Lambda = h/m^* v_{\text{therm}}$.

where Λ^* is the thermal de Broglie wavelength, h Planck's constant, and v the thermal velocity. For GaAs central valley electrons, $\Lambda^* \sim 270$ Å (see Table I).

In addition to spatial considerations, from a temporal perspective, the BTE describes irreversible phenomena and it is assumed that collisions occur on a scale short compared with an observation time. Thus, for example, relaxation times should be short compared to transit times. This assumption will be routinely violated in submicron devices (see Table I); and again, caution must be exercised in using the BTE.

It is clear that from a device-modeling viewpoint we are faced with serious problems. On the one hand, the BTE overcomes many objections to the use of equations like (1). On the other hand, the preceding arguments suggest abandoning the BTE for ultrasmall devices and replacing it with a quantum transport formulation. While the latter may be necessary, the current approach is to retain the BTE as long as useful quantum effects can be incorporated, even when first-order quantum effects occur, as, for example, the transport in quantized inversion layers [7].

The critical choices that must be made in examining the physics of semiconductor devices is the principal reason for this chapter. In the discussion that follows we take a detailed look at some of the physical assumptions underlying transport in semiconductor devices. In Section II we discuss the semiconductor equations and the picture they provide for the numerical simulation of GaAs and Si devices. We illustrate the discussion with simulations of the transient behavior of two- and three-terminal

TABLE I

Critical Boltzmann Transport Parameters for GaAs

Parameter	Variable	Value
Active region length	L	10^{-5} cm, 10^{-4} cm
de Broglie wavelength	$\Lambda^* = h/m^*v$	2.7×10^{-6} cm at $v = 4.5 \times 10^7$ cm/sec, $m^* = 0.067\, m_e$
Transit time	L/v	5×10^{-13} sec at $L = 10^{-5}$ cm, $v = 2 \times 10^7$ cm/sec 5×10^{-12} sec at $L = 10^{-4}$ cm, $v = 2 \times 10^7$ cm/sec
Mean time between collisions (momentum relaxation)	τ_m	3×10^{-13} sec at 300 K, 5×10^{-14} sec at 3000 K
Collision duration	τ_c	2×10^{-14} sec

GaAs devices. For silicon we consider the role of ion implantation in altering the charge distribution within an *n*-channel MOSFET. In Section III, we examine the Boltzmann transport equation. Two aspects are considered: (1) What are the assumptions, limitations, and methods of solution? When the limits are exceeded, what quantum modifications are possible? (2) Where does the device physics using the BTE show significant departures from that using the equations of Section II. In the first instance, modifications are introduced to account, for example, for finite collision-duration effects. Here, on a very short time scale, collisions cannot be regarded as instantaneous. Instead, during collisions, the carrier may absorb or lose energy to the self-consistent field. In the second case, transient transport calculations emphasizing velocity overshoot show significant BTE departures.

The discussion of velocity transients naturally leads into problems associated with collisionless transport in submicron devices. These problems are, at first glance, trivial. However, this notion should be dispelled by the unusually long list of questions asked by others in connection with vacuum-tube transport, such as: (1) How do contact properties affect the current–voltage relation? (2) How will the distribution of injected carriers and subsequent carrier–carrier interaction, for example, affect device transport? (3) How is one to interpret diffusionlike terms, such as grad NT (where T is an electron temperature) when diffusion in the sense of an Einstein relation is not a viable concept? The presence of these questions for collisionless transport does not lessen their importance for transport with collisions. In both cases the answers have not been found and represent current work in progress.

The discussion of Section III is primarily concerned with solutions to the BTE. The numerous corrections to it serve as a reminder of the extent to which new physics forces a close scrutiny. Most of this reexamination is done by comparing approximate evaluations of the one-particle quantum-density matrix with its classical analog, the distribution function. The connection between the two is reviewed in Section IV, where we formally introduce von Neumann's density matrix and show its relation to measurements of statistical averages. From the density matrix, a fully quantum-mechanical distribution function, the so-called Wigner distribution is introduced. The Wigner function is particularly intriguing insofar as its equation of motion is very close to the BTE, subject to the constraints of the uncertainty principle. Approximate solutions to the equation of motion and its connection to the results of Section III are given.

Analysis of semiconductor devices relies on the assumption that current flow is by drift and diffusion. We show in Section III that the formula-

tion of drift on a submicron scale is not at all obvious and that proper use of diffusion currents is even less certain. Diffusion is related to the spatial spreading of an ensemble of carriers with time, as the ensemble responds to both applied drift forces and random forces such as are generated by collisions. For examining transient diffusion processes, a Fokker–Plank equation can be generated whose solution is the transistion probability for a particle at X_0, t_0 to transition to X, t. On the short time scale over which relaxation processes occur, this equation does not reduce to the normal diffusion equation. We review diffusion processes on a short time scale by an ensemble Monte Carlo method, highlighting differences with the semiclassical description.

In summary, semiconductor device physics separates into three overlapping categories, identified by a particular type of equation:

(1) the semiconductor mobility equation,
(2) the Boltzmann transport equation, and
(3) the quantum transport equation

We shall discuss the modeling of transport in nonlinear semiconductor devices by solving, when available, or at least by examining each of these equations.

II. THE SEMICONDUCTOR EQUATIONS

A. Introduction

It has long been recognized that a proper understanding of today's devices requires detailed computer modeling of the space- and time-dependent charge distribution within the device. Simple, two-terminal configurations are usually represented by one-dimensional equations,[3] but the more important designs require a full account of at least two directions—one along the channel length and a second normal either to the metal–semiconductor interface (MESFET, for example) or the

[3] For one-dimensional devices, see Scharfetter and Gummel [8a]. This paper deals with the large-signal simulation of a silicon Read diode oscillator. Simulation of a TRAPATT oscillator is discussed by DeLoach and Scharfetter [8b]. For transferred electron devices, see Shaw *et al.* [8c]. For two-dimensional simulations, see Barnes and Lomax [8d]. This paper deals with finite-element methods. Nonisothermal carrier flow in a two-dimensional bipolar transistor under reactive circuit conditions is discussed by Turgeon and Navon [8e]. The phenomenon of avalanche breakdown in MOSFETs is discussed by Toyabe *et al.* [8f]. Yamaguchi *et al.* [8g] discuss the two-dimensional simulation of GaAs FET. See also Grubin and McHugh [8h].

oxide–semiconductor interface (MOSFET, for example). Many two-dimensional computer codes have been developed for this purpose [8a–8h]. Each of these codes involves obtaining self-consistent solutions of

(i) Poisson's equation,

$$\nabla^2\phi = -\rho/\epsilon, \tag{4}$$

where ϕ is the scalar potential and ρ the charge density;

(ii) the continuity equation,

$$\text{div} \cdot \mathbf{J}_c + (\partial\rho/\partial t) = G - R, \tag{5}$$

where J_c is the carrier current and the quantities on the right-hand side represent the possibility of local generation (G)–recombination (R) events; and

(iii) the "semiconductor equation,"

$$\mathbf{J}_c = qN\mathbf{v} - qD \,\text{grad}\, N, \tag{6}$$

where q is the electric charge.

Equation (5) is for single-species transport, and each of these equations is constrained by the external circuit.

As indicated in the introduction, the transport picture associated with Eq. (6) is based on the approximation that the carrier velocity v responds instantaneously to changes in electric field. For electrons:

$$\mathbf{v}(F) = -\mu(F)\mathbf{F}, \tag{7}$$

where

$$\mathbf{F} = -\text{grad}\, \phi. \tag{8}$$

Thus any spatial or temporal dependence in velocity arises because of such dependencies on field. This extremely important assumption breaks down at high frequencies [9]. In gallium arsenide, serious discrepancies appear at 10 to 20 GHz. The significance of this becomes clear with a simple example.

Consider solutions to Newton's equation of motion for a single carrier subjected to instantaneous change in field and scattering centers:

$$m^* \frac{dv}{dt} + \frac{m^* v}{\tau_p} = -eF. \tag{9}$$

For $v(t = 0) = 0$,

$$v(t) = (-e\tau_p/m^*)F[1 - \exp(-t/\tau_p)]. \tag{10}$$

Thus, on a temporal scale, instantaneous response implies that the observation or measurement takes place over a time interval that is long compared to the momentum relaxation time τ_p. For GaAs with scattering by intravalley–central-valley phonons, $\tau_p \simeq 0.3 \times 10^{-12}$ sec, and for a measurement over a time interval $\Delta t \simeq 3\tau_p$, the velocity is within 95% of the steady-state v ($t = \infty$) value. Note that $(3\tau_p)^{-1} \simeq 1100$ GHz, but in a solid containing a sufficiently high density of carriers (and here the solid must also be large enough to contain a sufficient number of carriers), the response of the devices is more closely related to the time required for the collection of carriers, as represented by their distribution function, to relax. For gallium arsenide, the distribution of carriers responds sluggishly to changes in field, so well before the 1100-GHz limit the "instantaneous" mobility model breaks down. The question arises of just how serious the breakdown is at lower frequencies. Although this question is a subject of current research, the indication (see Section III) is that, at least for gallium arsenide and frequencies below 20 GHz, qualitative device behavior will not be altered by the nonlocal contributions.

For the remainder of this chapter, we shall discuss, through the use of examples, the development of a numerical physics using the semiconductor equations. Three examples will be discussed:

(i) one-dimensional analysis of GaAs two-terminal devices,
(ii) two-dimensional analysis of GaAs FETs, and
(iii) two-dimensional analysis of a silicon MOSFET.

The GaAs devices are chosen because they are the most thoroughly studied transferred-electron-effect devices and exhibit the most dramatic spatial and temporal behavior. The family of GaAs devices is also among the leading candidates for future VLSI applications. The silicon devices are chosen because they are the most widely used and studied devices.

B. One-Dimensional Analysis of GaAs Two-Terminal Devices

1. *Introduction*

The standard one-dimensional simulation of GaAs two-terminal devices seeks to replicate Gunn's original observations [10] and the subsequent variations. Briefly, Gunn observed that when a GaAs sample fitted with two contacts was subjected to a sufficiently high bias, spontaneous and coherent microwave frequency current oscillations appeared (see Fig. 1). The oscillations had specific characteristics in that the period was closely related to the time it took the majority carrier (electrons) to transit between the contracts. The explanation, as we know it today, is due to

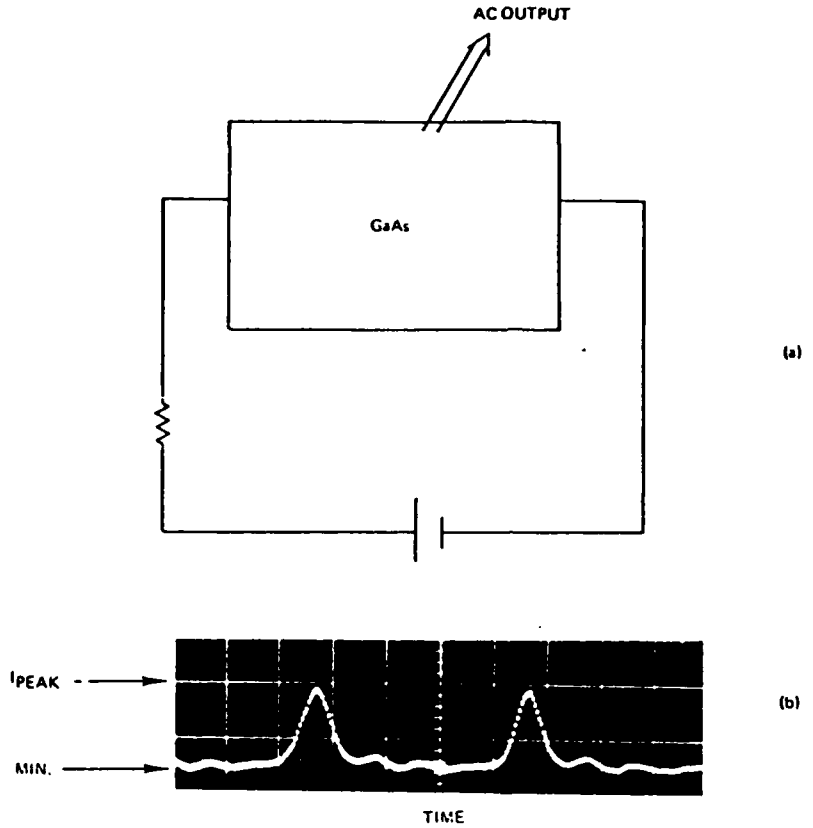


Fig. 1. (a) Schematic circuit for the appearance of Gunn oscillations. (b) One cycle of current versus time for a GaAs device sustaining dipole oscillations. Nucleation and extinction occurred during the "spiked" portion of the oscillation. Dipole transit is associated with the flat portion of the oscillation.

Kroemer [11], who proposed that when the oscillations occured the electric field within the gallium arsenide sample became highly nonuniform. This nonuniformity would manifest itself in the appearance of regions of high electric field surrounded by regions of low electric field, as in Fig. 2. and these high-field regions, often called *domains*, would move toward one end of the specimen. The oscillation was then a consequence of the following sequence of events. First, a domain would nucleate at one of the contacts (the cathode). At this point the current would drop, as in Fig. 1. Then the domain would leave the cathode region and travel down the sample toward the second contact (the anode). Here, the current would be approximately constant. Finally, the domain would be extinguished at the anode contact and the current would rise. The oscillation period T was determined mainly by the time it took the domain to travel between the

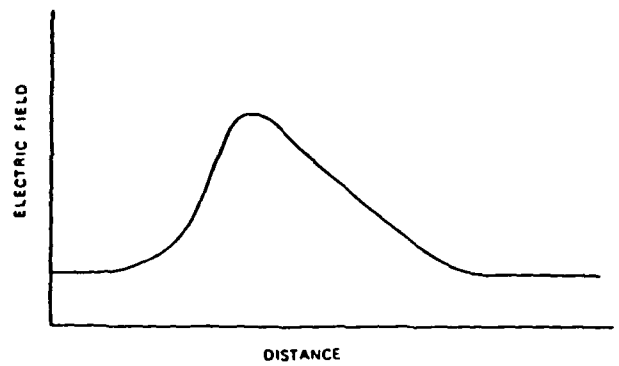


Fig. 2. Sketch of moving high electric field domain associated with the oscillation of Fig. 1.

cathode and anode contacts, and this was determined by the speed v of the moving electrons. Thus $T \simeq L/v$, where v is about 10^7 cm/sec and L the length of the specimen. Note that for $L \simeq 10^{-2}$ cm, $T \simeq 10^{-9}$ sec and $1/T = 1.0$ GHz.

Kroemer's explanation of Gunn's experiments required that GaAs sustain a region of negative differential mobility, a fact demonstrated by Hilsum [12] in 1962. Here, by explicit calculation, it was shown that gallium arsenide, through electron transfer, would exhibit negative differential mobility, where for a range of increasing electric field, the electron velocity decreases rather than increases (Fig. 3).

Typically, simulations (see Fig. 4) of the behavior of negative differen-

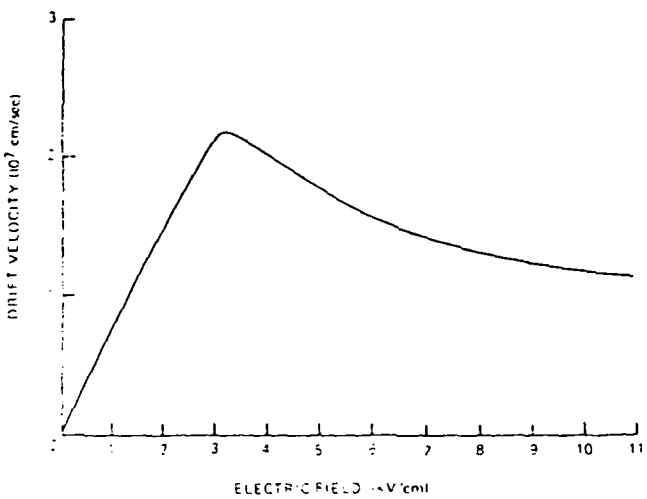


Fig. 3. Drift velocity versus electric field for GaAs. From Butcher [2], with permission of The Institute of Physics.

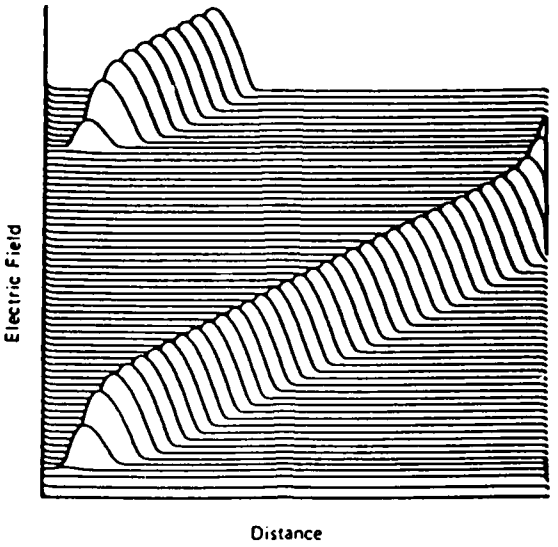


Fig. 4. Simulation of cathode-nucleated domain and subsequent transit-time oscillations. The NDM element is part of the resistive circuit, and the computations display $F(x,t)$ at successive instants of time. From Shaw *et al.* [8c], with permission.

tial mobility devices involve breaking down Eqs. (5), (6), and Poisson's equation and solving the following two sets of equations:

$$J_0(T) = v(F)\left(eN_0(x) + \epsilon\frac{\partial F(x,t)}{\partial x}\right) - D(F)\left(\epsilon\frac{\partial^2 F(x,t)}{\partial x^2} + e\frac{\partial N_0(x)}{\partial x}\right) + \epsilon\frac{\partial F(x,t)}{\partial t} \tag{11}$$

$$\Phi_B = \left(\mathscr{L}C_p\frac{d^2\Phi}{dt^2} + RC\frac{d\Phi}{dt} + \Phi\right) + \left(\mathscr{L}\frac{dI_0}{dt} + RI_0\right), \tag{12}$$

where $N_0(x)$ represents a spatially dependent donor density. Equation (12) is a representative circuit equation (see e.g., ref. 8c) with the device in parallel with a capacitor C_p, both connected serially to an inductor $\mathscr{L}$, resistor R, and bias Φ_B. Also,

$$\Phi = \int_0^L F(x,t)\,dx \tag{13}$$

and

$$I_0 = J_0(t)A, \tag{14}$$

where A is the cross-sectional area.

Equation (11) is the equation for total current density through the non-

linear semiconductor. The first collection of terms is the conduction-current contribution, the second and third contributions are the diffusion and displacement contributions, respectively. In the preceding equations, $v(F)$ represents the velocity–field relation for the carrier and $D(F)$ the diffusion–field relation. Within the framework of the mobility model $v(F)$ is represented by Fig. 3 for GaAs and $D(F)$ is given in Fig. 5.

Equation (11) is a partial differential equation rich in possibilities. Before discussing these, there are several points to be made with regard to the placement and use of the diffusion coefficient. First, note that D is to the left of the derivative. Within the framework of the Boltzmann transport theory, this is not entirely correct. For those situations where the distribution function departs slightly from equilibrium, Stratton [13] has demonstrated that the diffusion contribution should more generally appear as

$$\partial(Dn)/\partial x. \tag{15}$$

For highly nonequilibrium situations where such methods as the displaced Maxwellian are used, it is generally difficult to unequivocably identify a diffusivity. However, when it is defined, it usually emerges as a generalized Einstein relation. For multivalley semiconductors

$$D = \frac{1}{e}\sum \frac{\eta_i v_i T_i}{F}, \tag{16}$$

as in Fig. 5. Here v_i and T_i are, respectively, the carrier drift velocity of the ith valley, T_i its electron temperature, and η_i its fractional occupation. There are difficulties with this definition and we refer to Cheung *et al.* [14] for a more complete description.

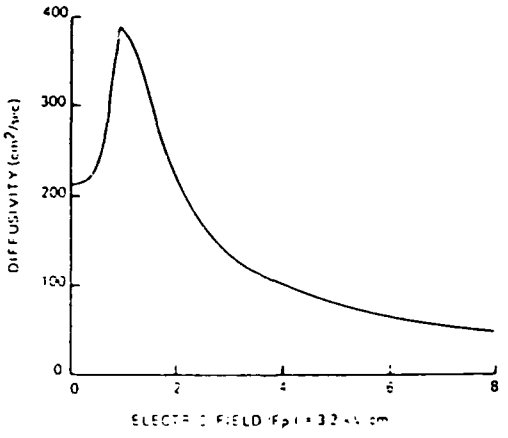


Fig. 5. Diffusivity versus electric field used in GaAs simulations. From Butcher [2], with permission of The Institute of Physics.

We also point out that Eq. (11) is written in "standard form" and retains the feature of treating electrons as positively changed particles. As long as we are treating one species of carrier, this does not present any real difficulties, but when electron and hole conduction occurs, care must be exercised. Among the difficulties to be aware of is the fact that the velocity–field curve in Fig. 3 represents the magnitude of the drift velocity since electrons move in a direction opposite to that of the electric field.

2. *Properties of Eq. (11)*

a. Length Dependence. Equation (11) is a two-point boundary-value problem and allows length-dependent effects to be studied. These effects are pronounced in GaAs and have their basis in the following: If an ostensibly homogeneous GaAs specimen is biased into the NDM (negative differential mobility) region, then ever-present space-charge fluctuations will grow. These growing fluctuations will simultaneously propagate and the amplitude of the disturbance in the simplest of cases [15] will be given by

$$G(x) = \exp(-x/v\tau), \tag{17}$$

where

$$\tau = \frac{\epsilon}{N_0 e\, dv/dF} < 0. \tag{18}$$

Significant gain will result in an instability when $G \gtrsim 1$ or

$$N_0 L > \left| \frac{\epsilon v}{e\, dv/dE} \right|. \tag{19}$$

For GaAs, $N_0 L$ is typically $\sim 10^{12}/\text{cm}^2$.

From another viewpoint, high field-domain propagation in devices greater than 50 μm is generally independent of the cathode and anode region and can be analyzed as though the device had infinitely separate boundaries. For $N_0 = 10^{15}/\text{cm}^3$ and devices approximately 10 μm in length high field domains may fill a large percentage of the devices and proximity effects occur [8c].

b. Doping Variations. In the early stages of GaAs device fabrication, it was rare to find a uniformly doped device. The effects of these nonuniformities were generally accounted for through assumed variations in $N_0(x)$. For example, a significant drop in N_0 over a small distance was shown to lead to local high field-domain nucleation with subsequent propagation [16]. In another situation, Kroemer established, by numerical simulation [17], that the presence of small statistical variations in N_0 would generate transient nonuniformities and prevent accumulation-layer propagation in

long samples. In short devices, ~10 cm long, these fluctuations are not as effective and accumulation-layer propagation is possible [8c].

c. Contact Effects. It has been established that the interface at the two-terminal active region dominates the subsequent behavior of the device. These contact effects have been treated by doping variations, mobility variations, and phenomenological interfacial fields. In many cases, through the combined use of simulations and experiments, one can determine whether doping variations or other contributions are responsible for device behavior [8c]. The effects of contact fields may, for example, be treated as boundary conditions to Eq. (11), where $F(x = 0, t)$ and $F(x = L, t)$ are specified [8c] or as a solution to a cathode differential equation, e.g.,

$$J_0(t) = \mathcal{J}(F) + \epsilon\, dF/dt \qquad \text{at} \quad x = 0, \tag{20}$$

where $\mathcal{J}$ is a phenomenological cathode current density. Our experience

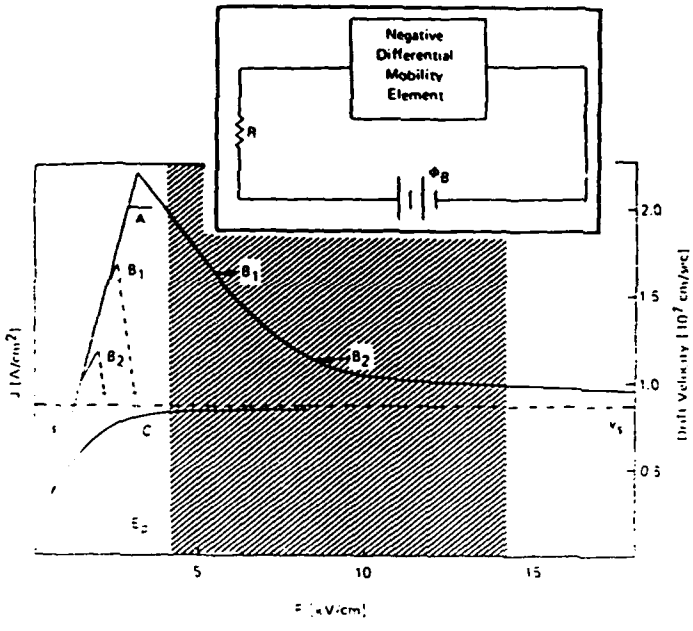


Fig. 6. The steady-state $v(F)$ curve and the simulated current density versus average field relation for various fixed values of carbode F_c. For the simulation, the NDM element is in the circuit shown in the inset. For curve A, $F_c = 0$, and the NDM element is quiet until the bulk field exceeds the NDM threshold. An instability then forms, and for relatively moderate doping, nonuniformities may damp. The result is a quiet element with a current level somewhat below the peak current density. The cathode fields for cuves B_1 and B_2 are indicated, and here cathode-nucleated domains result in distinct transit-time oscillations. For curve C, F_c = 24 kV/cm, and the NDM element is electrically stable. From Shaw *et al.* [8c] and related references therein.

indicates that the important contact to model is the cathode, and our results [8c] for a fixed cathode model are summarized in Fig. 6. More ambitious models are generated as the need arises, as for example with Eq. (20), which was generated to explain the unusual high-efficiency oscillations associated with InP [18,19].

In spite of the care one might exercise in solving Eq. (11), transient electrical instabilities require solutions to the circuit equations. For the device in a circuit containing reactive elements [e.g. Eq. (12)], the subsequent oscillation (see Fig. 7) may bear no resemblance to a propagating domain [18,19]. Another reason for insisting on the external circuit constraint is provided by the methods of characteristics and the field direction technique [20,21]. These show that there are an infinite number of time-independent solutions associated with a given value of J_0. It is the circuit constraint that picks out the correct distribution.

The preceding discussion has ignored the effects of generation–recombination on transient behavior of propagating high field regions. This is based on the assumption that instabilities develop on a time scale

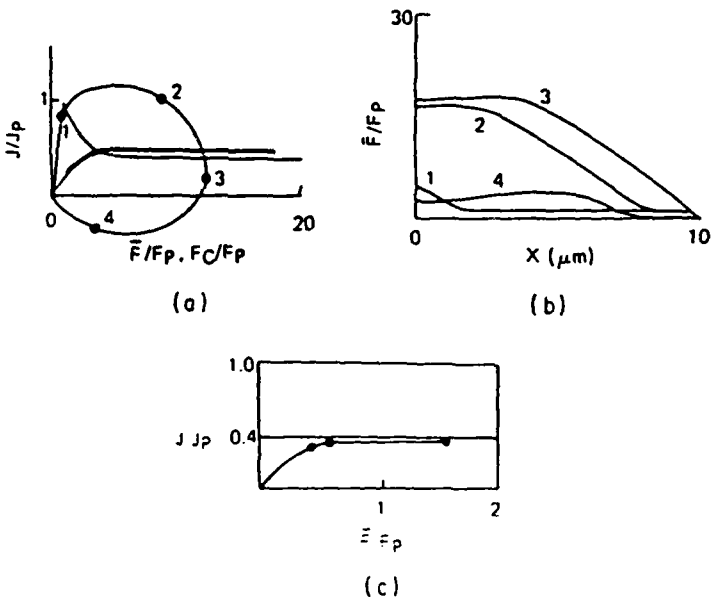


Fig. 7. Simulations using a time-dependent cathode condition [Eq. (20)]. The NDM element is in a circuit [see Eq. (12)]. It is in parallel with a package capacitor and in series with an inductor, resistor, and dc source. (a) Simulation of current versus average field for a conduction current curve represented by the bold line. Superimposed on the bold line is the $v(F)$ curve for GaAs, scaled to current and voltage. (b) Electric field versus distance curves at four instants of time, keyed to part (a). (c) Preinstability current versus voltage. Filled circles denote computed points. From Shaw *et al.* [8c] and references therein.

that is short compared to the generation–recombination process. A recent modification of Eq. (11) was effected to include space- and time-dependent impurity ionization contributions expected if carrier generation–recombination becomes important [22].

C. Two-Dimensional Analysis of GaAs FETs

1. Introduction

Two-dimensional analyses of GaAs devices are almost exclusively devoted to simulations of three-terminal FETs in a variety of configurations. The second dimension in these simulations is transverse to the principal direction of current flow and its magnitude is often the determinant of whether a propagating instability will occur [23]. The need for simulations of two-dimensional problems is more generally accepted than that of one-dimensional ones because of the unavailability of widespread intuitive equations.

The FET in its simplest form is a semiconductor slab with three terminals. Two of these are usually low-resistance contacts, while the third is either a Schottky contact or a p–n junction with an accompanying region of charge depletion. For a unipolar conduction device, operation is based on modulation of the depletion region, which is usually accomplished by changes in the gate bias. Small and large signal gain are possible. Figure 8 is a sketch of the device and the connecting lumped elements.

The equations describing this device are Eq. (4), with

$$\rho = -e[N(x,y,t) - N_0(x,y,t)] \tag{21}$$

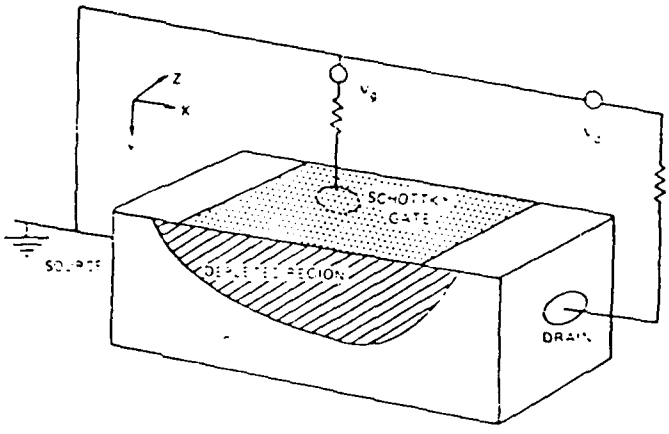


Fig. 8. Schematic representation of a Schottky gate field-effect transistor.

and

$$\mathbf{F}(x,y,t) = -\mathbf{grad}\ \phi(x,y,t), \tag{22}$$

and Eq. (5) with $G - R = 0$ and

$$\mathbf{J}_c(x,y,t) = -eN(x,y,t)v(|F|) + eD\ \mathbf{grad}\ N(x,y,t). \tag{23}$$

Note that Eq. (5) requires point-by-point divergence-free total current density

$$\mathbf{div}\ \mathbf{i}(x,y,t) = 0, \tag{24}$$

where

$$-i(x,y,t) = \mathbf{J}_c(x,y,t) + \epsilon\ \partial \mathbf{F}(x,y,t)/\partial t. \tag{25}$$

The boundary conditions to these equations necessarily approximate the physical and electrical characteristics of the outer periphery of the semiconductor structure. In most GaAs FET simulations, the exposed surfaces are assumed to be ideal insulators and no current is permitted normal to these boundaries. Along the free surfaces and in the limit of zero permittivity for the space surrounding the device [24,8g],

$$\hat{n} \cdot \mathbf{grad}\ \phi = \hat{n} \cdot \mathbf{grad}\ N = 0, \tag{26}$$

where $\hat{n}$ is a unit vector normal to the free surface of the semiconductor.

The low-resistance source and drain contacts and the Schottky gate contact are approximated by equipotential surfaces with a prespecified charge density [24,8g]. For the calculations illustrated, the low-resistance source and drain contacts are neutral and located sufficiently far from the active region of the device that for the current levels involved they have no influence on the electrical properties of the device. For the sample calculations,

$$N_S = N_D = N_0 \tag{27}$$

and

$$N_G = N_0 \exp - [e\phi_{Bi}/k_B T_0] \tag{28}$$

where ϕ_{Bi} is the "built-in" potential [25] and N_S, N_D, and N_G indicate the source, drain, and gate, respectively.

To include circuit effects, the potentials on the contacts are needed. Generally we set the source potential to zero [8h] while the gate potential ϕ_G and drain potential ϕ_D are determined by simultaneous solution of the semiconductor equations and the circuit equations.

For the three-contact device of Fig. 8, the circuit equations are

$$\phi_S = 0, \tag{29}$$

$$\phi_{BD}(t) = I_D(t)Z_D + \phi_D(t), \tag{30}$$

$$\phi_{BG}(t) = I_G(t)Z_G + \phi_G(t), \tag{31}$$

$$I_S(t) = I_G(t) + I_D(t), \tag{32}$$

where ϕ_S is the source potential, ϕ_G the gate potential, ϕ_D the drain potential, ϕ_{BG} and ϕ_{BD} the drain and gate bias potentials, respectively, I_S the source current, I_G gate current, I_D drain current, Z_G the gate impedance, and Z_D the drain impedance. The external circuit impedances are modeled as lumped linear elements. The current passing across a contact is the integral over the contact area of the component of current density normal to the contact. For a device of dimension l_z in the z direction and a contact extending over a region l,

$$I(t) = l_z \int_l \hat{n} \cdot \mathbf{i}(x,y,t)\, dl'. \tag{33}$$

When simulating a GaAs FET, the important features are shape, dimension, and doping profile. Typically, effects of semi-insulating substrates may be examined by variations in $N_0(x,y,t)$ [24,26], alloyed contacts may be examined by increasing the doping level under the source and drain regions, and semi-insulating gate regions [27] may also be examined. We shall ignore these in the succeeding illustrations, concentrating instead on *homogeneous doping profiles*. Our intent is to illustrate the role of numerical simulations in classifying GaAs FET behavior.

2. GaAs FET Classification

Numerical simulations and experiment [28] demonstrate that the GaAs FET can be placed in one of two groups as determined by the ratio

$$K = \frac{\text{gate voltage at cutoff}}{\begin{pmatrix}\text{drain voltage at the onset of current}\\ \text{saturation for zero gate voltage}\end{pmatrix}} \tag{34}$$

Devices with $K > 1$ sustain current oscillations, the origin of which lies in the presence of negative differential mobility in the semiconductor. Those with $K \leq 1$ are electrically stable. Devices with $K > 1$ are generally wider in the transverse dimension than those with $K \approx 1$.

The electrical behavior associated with this classification is summarized in Fig. 9, where we sketch the simulated current–voltage relation for two GaAs FETs with a 10-μm source–drain separation. For reference, we have drawn the velocity–field relation for GaAs scaled to the current and voltage parameters. The first point to note about these results is that the current levels do not approach the peak current associated with

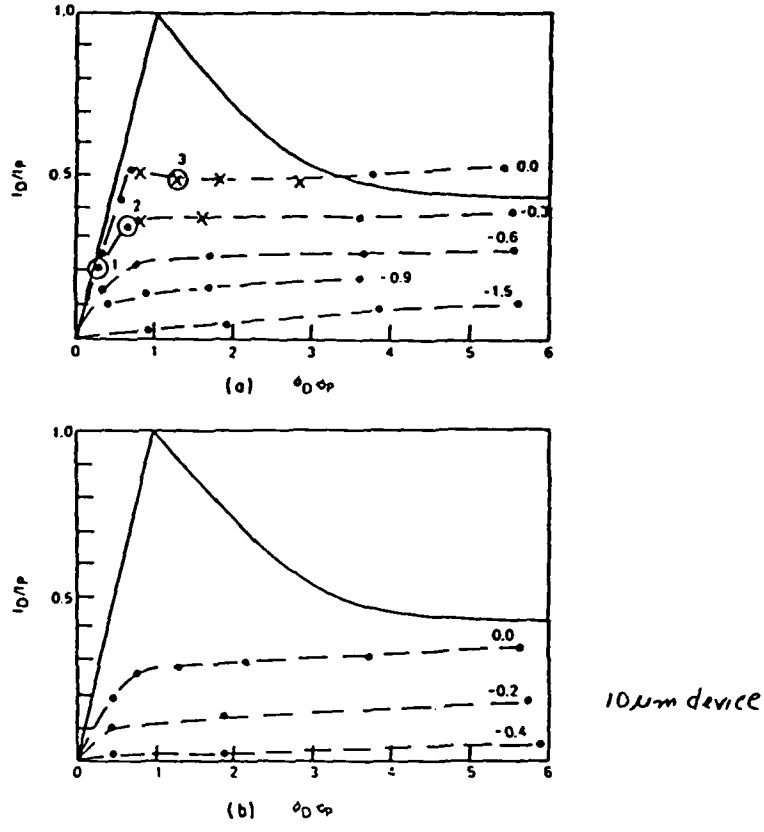


Fig. 9. Drain current–drain voltage relation: (a) $K > 1$ and channel height $l_y = 2.2$ μm; (b) $K = 1$ and $l_y = 1.2$ μm. Filled circles denote computed points; ×s denote averages of current during an oscillation. The solid curve is the GaAs $v(F)$ relation scaled to current and voltage. Reprinted with permission from *Solid State Electron.*, 23, H. L. Grubin, D. K. Ferry, and K. R. Gleason, "Spontaneous Oscillations in Gallium Arsenide Field Effect Transistors." Copyright 1980, Pergamon Press, Ltd.

GaAs. This is a consequence of the additional resistance supplied by the gate region as well as the velocity limitation. The second point is that for the wider-channel device an instability occurs. The instability is represented by the dashed I–V curves of Fig. 9a. The ×s in the diagram represent average current and voltage values for the instability, and the presence of negative conductance is due to the dynamic propagating domain. The filled circles in Fig. 9a represent stationary, time-dependent points, and we note that where there is an instability it is surrounded by regions of nonzero gate and drain bias for which there is no time-dependent behavior. The third point is that for the wider-channel device, the normalized current density at zero gate bias levels exceeds the sustaining current [8c], which is the minimum current necessary for stable domain prop-

agation in two-terminal devices. The sustaining current plays a similar role in three-terminal devices, and in these wider-channel devices, domain propagation occurs. For the narrower-channel device, the current is below the sustaining current and no instability occurs. For this narrow device, and at sufficiently high drain bias levels, nonuniform domains form during an initial transient where they absorb most of the voltage and force a reduction in current level below the sustaining current. These results are consistent with the thickness dependence previously discussed [23].

We illustrate the internal distributions of charge and current at three points, 1, 2, and 3, of Fig. 9. Additional information is available in Grubin and McHugh [8h,26] and Grubin *et al.* [28]. The internal distribution of charge and current associated with the current and potential levels 1 and 2 of Fig. 9 are shown in Fig. 10 [8h]. Parts (a) and (b) show current-density streamlines through the device, with the length of each proportional to the magnitude of the vector's current density at that point. The maximum length of the individual x and y components before overlap is $J_p = N_0 e v_p$, where v_p is the peak carrier velocity. We note that in both cases the current density is greatest under the gate contact as required by current continuity. For the higher bias, the current density under the gate region is at least as great as J_p and velocity limitation introduces carrier accumulation. The density of charge particles in the FET is generally nonuniform and parts (c) and (d) of Fig. 10 are a projection of this distribution as it relates to the current-density profile of parts (a) and (b). We note that the particle density increases in the downward direction. Part (a) shows a

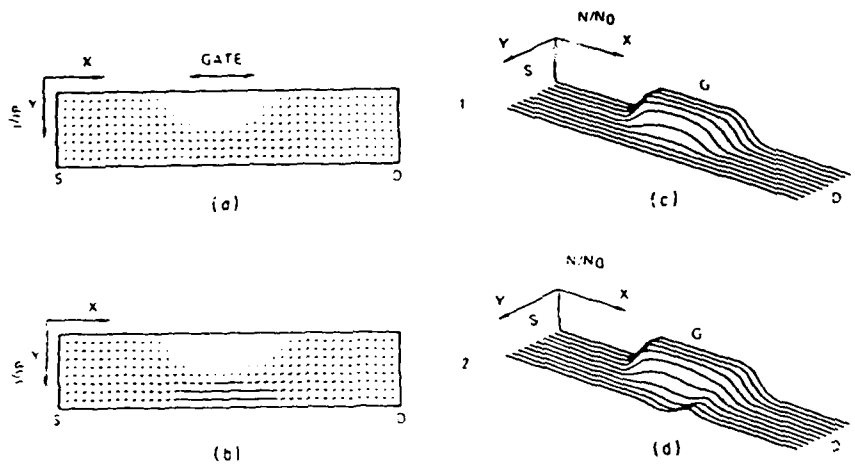


Fig. 10. The internal distribution of current and charge corresponding to the current and voltage levels represented by numbers 1 and 2 of Fig. 9. Note that the particle density surrounding the nonuniform distribution is uniformly distributed within the source–gate region and the gate–drain region. Reprinted with permission from *Solid State Electron.*, 21, H. L. Grubin and T. M. McHugh, "Hot Electron Transport Effects in Field Effect Transistors" Copyright 1978, Pergamon Press, Ltd.

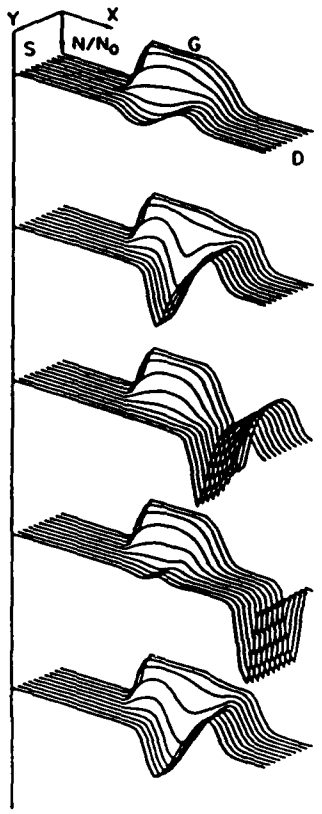


Fig. 11. Projection of the time-dependent particle density when an instability occurs. The parameters at which this occurs are represented by the number 3 in Fig. 10. Reprinted with permission from *Solid State Electron.*, **21**, H. L. Grubin and T. M. McHugh, "Hot Electron Transport Effects in Field Effect Transistors." Copyright 1978, Pergamon Press, Ltd.

region of charge depletion directly under the gate contact. Part (b) shows the formation of a weak stationary dipole layer under the gate contact. Here the x component of electric field has reached the NDM threshold field value at the drain edge of the gate contact.

We now consider the presence of an instability. We recall that for two-terminal devices the instability is determined by the value of the electric field at the cathode boundary [8c] and that the threshold current density for the instability is anywhere between J_s and J_p, where J_s is the current density associated with the high-field saturated drift velocity [8c]. In contrast, for three-terminal devices, the initiation of a domain instability generally occurs under the gate contact and at a local value of current density approximately equal to J_p. Figure 11 illustrates the sequence of events associated with an instability. Domain growth under the gate is accompanied by an increase in potential across the device. A corresponding

decrease in current occurs throughout the device and circuit, as constrained by the dc load line. As the current decreases carriers with velocities below that of the peak velocity enter the accumulation layer, which subsequently begins to detach. The domain speeds as it leaves the gate region and settles into a value of current density somewhat in excess of that associated with the saturated drift velocity of the electrons. Prior to reaching the drain contact, the domain dynamics appear to be one-dimensional.

D. Two-Dimensional Analysis of a Silicon MOSFET

Although a large body of gallium arsenide work is still regarded as research and, consequently, only a relatively moderate amount of two-dimensional simulations have appeared, this is not the case with silicon devices. Here VLSI requirements introduce an urgency into the studies, and MOS simulations, involving two types of carriers, are often used as an alternative to experimental investigations.

For this case, the basic semiconductor equations are generalized so that the charge density in Eq. (1) becomes

$$\rho = e[P(x,y,t) - N(x,y,t) + N_D(x,y,t) - N_A(x,y,t)]. \tag{35}$$

The continuity equation is generalized so that

$$\text{div } \mathbf{J}_n - e\,\partial[N(x,y,t) - N_D(x,y,t)]/\partial t = G - R, \tag{36a}$$

$$\text{div } \mathbf{J}_p + e\,\partial[P(x,y,t) - N_A(x,y,t)]/dt = -(G - R), \tag{36b}$$

and the current-density equation is

$$\mathbf{J}_n = -eN(x,y,t)\mathbf{v}_n + eD_n \text{ grad } N. \tag{37a}$$

$$\mathbf{J}_p = +eP(x,y,t)\mathbf{v}_p - eD_p \text{ grad } P. \tag{37b}$$

The boundary conditions, device shape, and material parameters specify the problems to be studied. Figures 12–14 summarize the results of a recent calculation [29] designed to show the dependence of the internal charge distribution on the impurity concentration of an MOS transistor. The device configuration is shown in Fig. 12, where the line segment *BE* represents the semiconductor–oxide interface. Within the oxide region,

$$\nabla^2\phi = 0, \tag{38}$$

and at the interface,

$$\epsilon_{ox}\left.\frac{\partial\phi}{\partial y}\right|_{ox} = \epsilon_{sem}\left.\frac{\partial\phi}{\partial y}\right|_{sem}. \tag{39}$$

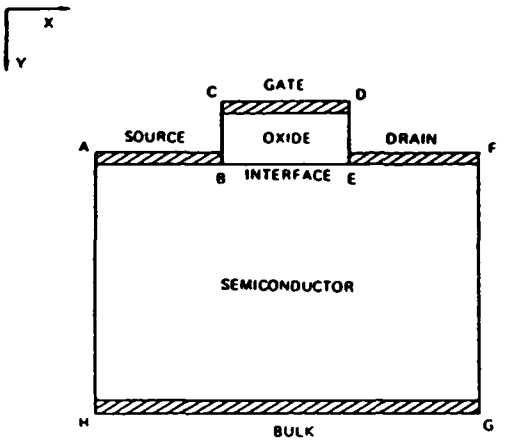


Fig. 12. Geometry for the MOSFET simulation. The regions under *AB* and *EF* are generally heavily doped n^{++} regions. From Selberherr *et al.* [29], with permission.© 1980 IEEE.

As in the MESFET calculation, the carrier densities at the source and drain contacts are set equal to the doping concentration. No current is permitted to flow across the interface *BE*, the bulk control *HG*, or the exposed surfaces *AH* and *FG*.

The doping profiles to which the mobile electrons respond are shown in

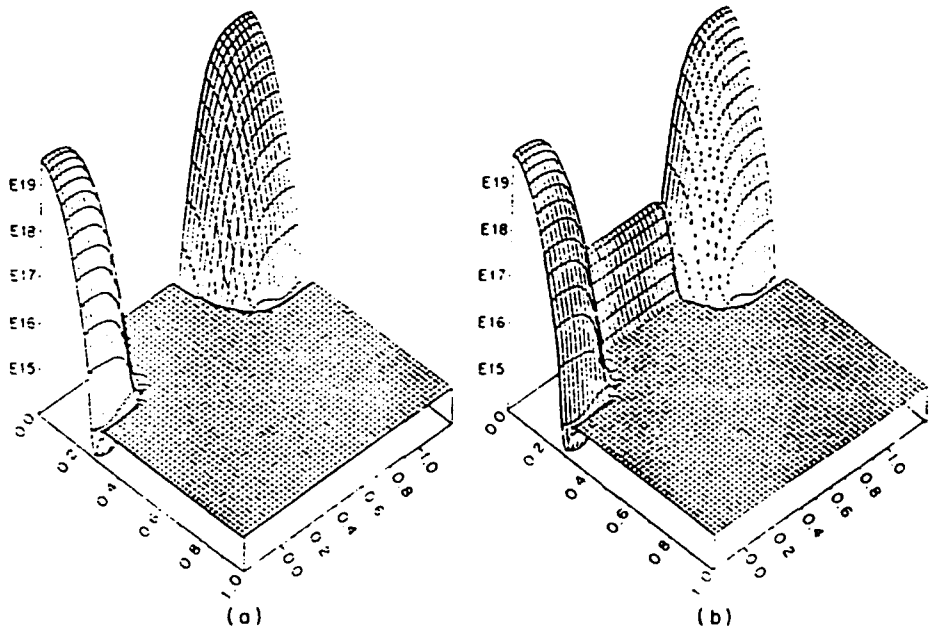


Fig. 13. Doping concentration of silicon MOSFET. Axes are lengths in microns. From Selberherr *et al.* [29], with permission.© 1980 IEEE.

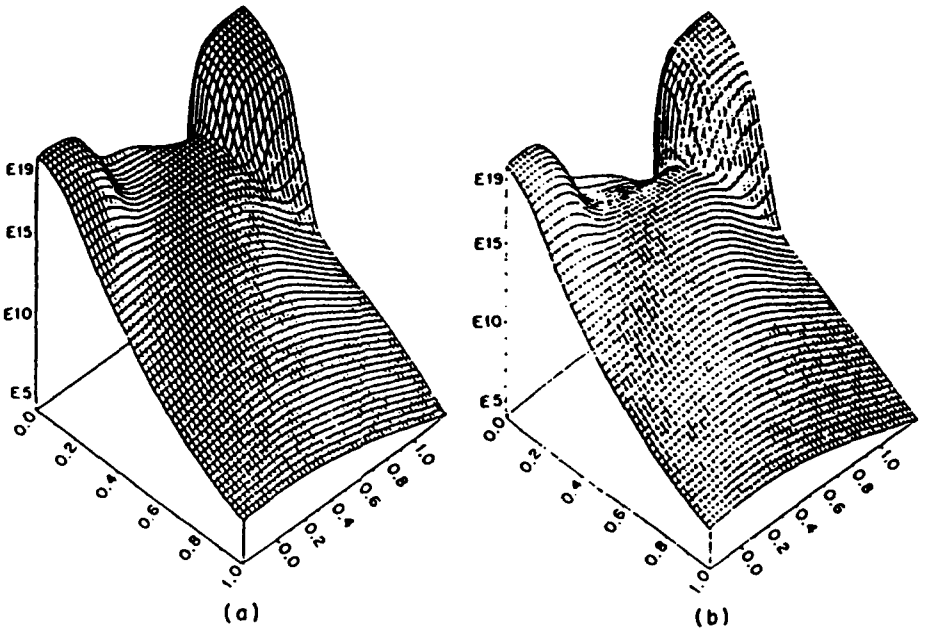


Fig. 14. Electron concentration of silicon MOSFET under strong inversion. Axes are lengths in microns. From Selberherr *et al.* [29], with permission. © 1980 IEEE.

Fig. 13. In Fig. 13a, the p-type dopant is uniformly $10^{15}/cm^3$, and in Fig. 13b, it increases at the oxide interface to approximately $5 \times 10^{16}/cm^3$. The surface concentration of the source and drain regions is $5 \times 10^{19}/cm^3$ and the depth of the $p-n$ junction under the source and drain is approximately 3000 Å. Figure 14a shows the electron distribution for the first device under strong inversion. The surface concentration in the channel is high. Figure 14b shows the ion-implanted device (Fig. 13b) again under strong inversion. As expected, the surface concentration has decreased.

III. THE BOLTZMANN TRANSPORT EQUATION

A. Introduction

In general, the free carriers in a semiconductor gain energy from the electric field. This energy must be relaxed to the lattice through electron–phonon interactions. For vanishingly small values of the electric field, the energy gained from the field is negligible in comparison with the mean energy of the carriers, whether this latter quantity is represented by a thermal energy or by the Fermi energy in degenerate semiconductors.

For larger values of the electric field, the average energy of the carriers is increased by the field, and the carriers are said to be "hot." Because the average energy of the carriers increases in the field, the net rate of phonon emission must also increase to yield the energy loss.

The last factor of increased emission of phonons to balance the energy gain from the field is a consequence of detailed balance. Consider a nondegenerate distribution characterized by an electron temperature T_e. Phonon emission and absorption processes connect states at energy E with those at $E + \hbar\omega_0$. The rate of absorption of phonons out of the states at E is given by

$$AN_q \exp(-E/k_B T_e), \tag{40}$$

where N_q is the Bose–Einstein occupation factor of the phonons of wavelength q and k_B Boltzmann's constant; the rate of emission from the states at $E + \hbar\omega_0$ is given by

$$A(1 + N_q) \exp[-(E + \hbar\omega_0)/k_B T_e]. \tag{41}$$

Now $(1 + N_q) = N_q \exp(\hbar\omega_0/k_B T_0)$, where T_0 is the lattice temperature, so that emission and absorption processes establish a detailed balance when $T_e = T_0$. If additional energy is supplied to the electron gas by an external electric field, then the gain in average energy is characterized by a rise in T_e and the emission processes increase over the absorption mainly by the factor $\exp\{[\hbar\omega_0(1 - T_0/T_e)]/k_B T_0\}$, provided that the distribution function remains Maxwellian. The amount of rise in electron temperature T_e is governed by the energy gains and losses of the electron gas, and a detailed balance is established when the average rate of energy loss to the lattice $-\langle dE/dt \rangle$ equals the rate of energy gain from the electric field, or

$$e\mu F^2 = -\langle dE/dt \rangle. \tag{42}$$

The problem of evaluating Eq. (42) often reduces to one of determining the carrier-distribution function. At very low electric fields, the distribution function can be described as merely a small drift term superimposed on the thermal distribution, as

$$f(E,T_0) = f_0(E,T_0) + f_1, \qquad f_1 \ll f_0, \tag{43}$$

where $f_0(E,T_0)$ is the Maxwell–Boltzmann (or Fermi–Dirac) distribution at the lattice temperature T_0. When the electric field becomes sufficiently large, however, that the procedure of (43) can no longer be used, we are said to be in a high-field regime and hot-electron (or hole) effects must be considered [9]. The overriding theoretical concern in high-field transport is one of discerning the form that the distribution function takes in the

presence of the electric field. In semiclassical transport, this distribution function is given as a solution to the Boltzmann transport equation (BTE). Because of the very complicated nature of the BTE—the fact that it is a nonlinear, integrodifferential equation—it is usually not possible to solve it analytically and several possible assumptions can be made. This aspect was recognized early, and much work was done in the mid-1930s by several Russian authors [30,31]. Since then a considerable amount of work has followed, both theoretical and experimental. One reason for this lies in the manner in which investigations of the high-field transport yield information on the details of the electron–phonon interaction and the interaction of the carriers with themselves and with impurities. Another major reason is the role played by hot-electron behavior in the operation of electron devices.

In spite of the many studies of hot-electron (and hot hole) behavior, the central problem underlying the entire area remains that of trying to understand the manner in which the distribution function of the electrons is modified by the presence of the electric field. This is true whether we are dealing with a bulk material or the current response in a device. It is also a formidable experimental problem [32]. In general, the Boltzmann transport equation can be expressed in its most general form as

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} - e\mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} \approx \int d\mathbf{p}'[f(\mathbf{p}')W(\mathbf{p}',\mathbf{p}) - f(\mathbf{p})W(\mathbf{p},\mathbf{p}')], \quad (44)$$

where $f(\mathbf{r},\mathbf{p},t)$ is the carrier-distribution function. Generally for the steady-state response, the first two terms on the left are ignored and the distribution is a function of the carrier pseudomomentum $\mathbf{p} = \hbar\mathbf{k}$ and the energy. It is important to note that the Boltzmann transport equation assumes that the collisions are instantaneous in both space and time and that the field and scattering are different perturbations. In spatially varying problems, the addition of the position vector $\mathbf{r}$ brings about an effective lowering of the symmetry of the problem and complicates the solution of the Boltzmann equation (or an equivalent formulation). One can usefully classify the various phenomena or solutions upon the level of this symmetry. Where hot-carrier behavior is not considered, a single dimension, the electron energy E, is sufficient. If an axis of rotational symmetry exists for the hot-electron problem, perhaps along the electric field direction, then two variables, $\mathbf{p}$ (or E) and θ, are all that is required. In many cases the problem is more complicated, however. But, in some circumstances, an analytical form can be assumed for the distribution function and if this form depends only on a small number of parameters, then the zero-dimensional case results. In this latter case, simple equations for the evaluation of these parameters can be found, usually from moments of the

BTE [33]. In other cases, no approximations can be made, and detailed numerical techniques must be used. In modern transport theory, approximation techniques based on Legendre expansions for $f(\mathbf{p},t)$ no longer find much usage. Modern computers allow relatively rapid solution of detailed equations, and so the two approaches just discussed receive extensive usage. Which of the two methods is to be preferred depends on the rate of intercarrier energy exchange, discussed by Hearn [34].

B. Displaced Maxwellian Equation

Fröhlich [35] first pointed out that the isotropic part of the carrier-distribution function is Maxwellian provided that the carrier concentration exceeds a certain critical concentration; i.e., provided that the rate of intercarrier energy exchange is sufficiently large. Under conditions for which the anisotropic terms can be taken as small, Fröhlich and Paranjape [36] pointed out that a displaced Maxwellian distribution

$$f(E) = A \exp[-(E - \mathbf{v}_d \cdot \mathbf{p})/k_B T_e], \tag{45}$$

containing the electron temperature T_e and drift velocity $\mathbf{v}_d$ as parameters, could be utilized. This is then the hot-electron distribution in the approximation for which the critical carrier concentration is exceeded, and the parameters are then determined from balance equations, which, in turn, are obtained from the Boltzmann equation. Following the approach of Price [37], Eq. (44) is multiplied by any function $\phi(\mathbf{p})$ and integrated over p. For electrons within a single valley, this gives

$$\frac{\partial}{\partial t}\langle\phi\rangle + \frac{\partial}{\partial \mathbf{r}} \cdot \langle\phi\mathbf{v}\rangle = -\left\langle e\mathbf{F} \cdot \frac{\partial}{\partial \mathbf{p}}\phi\right\rangle + \left\langle \int d\mathbf{p}' \,[\phi(\mathbf{p}') - \phi(\mathbf{p})]W(\mathbf{p},\mathbf{p}')\right\rangle. \tag{46}$$

The last term on the right arises from an interchange of variables involved in the double integration. The terms on the left-hand side vanish for the homogenous steady state. For ϕ equal to p, the first term on the right is just the force F and (46) is the momentum-balance equation. For ϕ equal to the energy E, the first term on the right is $\mathbf{v}_d \cdot \mathbf{F}$ and (46) is the energy-balance equation. In particular, the factor within the angle brackets in the last term on the right serves to define the average rate of energy loss to the lattice by collisions

$$\left\langle \frac{dE}{dt}\right\rangle_{\text{coll}} = \int d\mathbf{p}' \,[E(\mathbf{p}') - E(\mathbf{p})]W(\mathbf{p},\mathbf{p}') \tag{47}$$

and the average rate of momentum loss due to collisions

$$\left.\frac{dp}{dt}\right|_{\text{coll}} = \int d\mathbf{p}'\,[\mathbf{p}' - \mathbf{p}]W(\mathbf{p},\mathbf{p}'). \tag{48}$$

Although Eq. (46) has been developed for a single valley, the procedure is readily extended to sets of nonequivalent valleys. This is done in Appendix A.

As an example, consider a material such as indium arsenide, where the scattering in the central valley is dominated by the polar optical phonon. In this case, the energy-balance equation becomes

$$ev_d F = A_1[\exp(y - x) - 1]x^{1/2}\exp(x/2)K_0(x/2), \tag{49}$$

where $A_1 = eF_0(2\hbar\omega_0/\pi m^*)^{1/2}/[\exp(y) - 1]$, $x = \hbar\omega_0/k_B T_e$, $y = \hbar\omega_0/k_B T_0$, and F_0 is an effective electric field describing the coupling between the electrons and the phonons. Similarly, the momentum-balance equation becomes

$$\begin{aligned} -eF = {} & (A_1 m^* v_d/3\hbar\omega_0)\{[\exp(y - x) + 1]K_1(x/2) \\ & + [\exp(y - x) - 1]K_0(x/2)\}x^{3/2}\exp(x/2) \\ & + 8v_d(2\pi m^* k_B T_e)^{1/2}(3\pi l_{ac})^{-1} \end{aligned} \tag{50}$$

where the last term on the right takes into account the momentum loss to the elastic scattering by acoustic modes. Now, Eqs. (49) and (50) can be solved simultaneously to yield T_e for a given electric field F, and then this result can be used in Eq. (46) to find v_d and, hence, μ. It should be pointed out that this is an exceedingly simplified model, and factors such as intervalley transfer and band nonparabolicity should be considered. However, the results given here are useful as an illustrative example of the application of the displaced Maxwellian technique.

The accuracy of the balance equations obtained from (46) ranges, in its applications to solving transport problems, from very good to exceedingly poor, the latter in cases in which the preceding assumptions are just not valid. Perhaps the most easily violated condition is the critical carrier concentration required. In the cases in which the balance equations are good, one can use them to *infer* energy and momentum relaxation times as

$$-e\tau_E = d\langle E\rangle/d(\mathbf{v}_d \cdot \mathbf{F}). \tag{51}$$

$$-e\tau_p = d\langle \mathbf{p}\rangle/d\mathbf{F}. \tag{52}$$

It should be pointed out that although these definitions appear in the balance equations, their validity goes beyond the displaced Maxwellian. Equations (47) and (48) can be averaged over any distribution $f(E)$ to define effective energy and momentum relaxation times, but the connection

of this to the electric field, as in Eqs. (51) and (52), must be used carefully. Since these times are based on the balance equations, the validity of their definitions is inherently tied into the results, and although any distribution function can be used, it must be done judiciously. One finds from these considerations of the relaxation times that the energy relaxation time τ_E is considerably longer than the momentum relaxation time τ_p, so that the drift velocity responds to a change in F faster than the electron temperature responds. This can lead to overshoot effects in the velocity, in which v_d rises to a value corresponding to an electron temperature at the starting time just before the change in F, then changes further as the temperature and distribution relax to the new temperature [3,38,39]. This time-dependent behavior and the steady-state ac cases can be treated by using the time-dependent form of (46), using, for example, a large, steady, dc F_0 and a small, sinusoidal, ac field F_1 superimposed upon it. In this case, one finds that the ac mobility includes terms like $(1 + i\omega\tau_p)$ and $(1 + i\omega\tau_E)$, reflecting the two time scales in the problem [40]. More complicated time variations have been examined by Butcher and Hearn [4], and spatial variations of the displaced Maxwellian have been considered by Bosch and Thim [41]. We shall expand considerably on these discussions, with particular attention paid to device simulation.

C. Semiempirical Transport Equation Method

In many cases, the BTE cannot be solved utilizing a displaced Maxwellian, but a set of balance equations is desirable for simplicity of application, especially in device modeling. An approach has built up that achieves this simplified goal and is called (by us) the *semiempirical transport equation method* (SETEM). Equation (46) can be rewritten as

$$\frac{\partial\langle\phi\rangle}{\partial t} + \nabla\cdot\langle\phi\mathbf{v}\rangle = -\langle e\mathbf{F}\cdot\nabla_p\phi\rangle - \frac{\langle\phi\rangle}{\tau_\phi}, \tag{53a}$$

where $1/\tau_\phi$ is an effective relaxation rate for $\langle\phi\rangle$. Of course, this quantity can be calculated directly from the collision integrals (47) and (48) if the distribution function is known. The approach of the SETEM is similar to that of the semiempirical tight-binding method in band-structure calculations, where the simple approach is used as an interpolation method. Relatively exact calculations of $\mathbf{v}(\mathbf{F},t)$ and $\langle E(\mathbf{F},t)\rangle$ can be made using the numerical techniques to be discussed. Then the set of equations given by (53) for different functions $\phi(\mathbf{p})$ can be fit to the exact calculations by using τ_p, τ_E, . . . as adjustable parameters that are functions of the field F. Once the relaxation times are known, the balance equations (53) can be

used for device modeling (see e.g. Shur [42]), and we shall illustrate this next.

Carnez *et al* [43] applied the SETEM to the modeling of submicron FETs. They obtained simultaneous solution of Poisson's equation, the equation of continuity, and the time-dependent, but spatially independent, momentum and energy equations. For the latter, the SETEM equations are

$$d[m^*(E)v]/dt = -eF - [m^*(E)v/\tau_p(E)], \tag{53b}$$

$$dE/dt = -eFv - [(E - E_0)/\tau_E(E)], \tag{53c}$$

where $\tau_p(E)$ and $\tau_E(E)$ were obtained from the steady-state values of velocity and energy

$$-\tau_p(E) = m^*(E)\langle v(E)\rangle/e\langle F\rangle, \tag{53d}$$

$$-\tau_E(E) = (E - E_0)/e\langle F\rangle\langle v(E)\rangle. \tag{53e}$$

Here $\langle F\rangle$ relates F to energy, as in Eq. (53g). For a two-valley semiconductor,

$$m^*(E) = [1 - N_s(F)/N_0]m_c^* + [N_s(F)/N_0 m_s^*] \tag{53f}$$

and

$$E(F) = E_c(F)[1 - N_s(F)/N_0] + [E_s(F)N_s(F)/N_0]. \tag{53g}$$

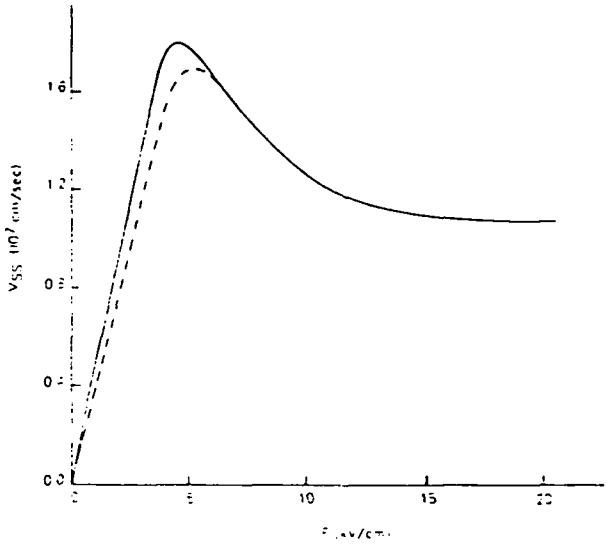


Fig. 15. Steady-state drift velocity versus electric field; (—) $N_1 = 10^{17}$ cm^{-3}, (---) $N_2 = 3 \times 10^{17}$ cm^{-3}. From Carnez *et al.* [43], with permission.

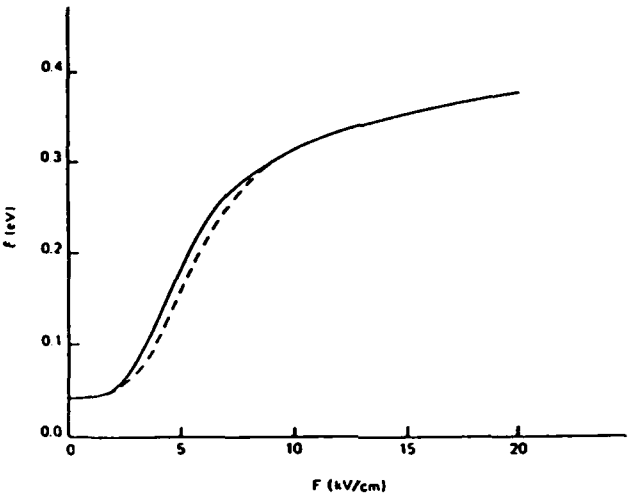


Fig. 16. Steady-state mean energy versus electric field: (—)$N_d = 10^{17}$ cm^{-3}, (---)$N_d = 3 \times 10^{17}$ cm^{-3}. From Carnez *et al.* [43], with permission.

Here $N_s(F)$, $E_c(F)$, and $E_s(F)$ and the brackets $\langle\ \rangle$ denote steady-state relations.

Equations (53b)–(53g) contain the essential philosophy of SETEM as currently used; namely, nonequilibrium conditions can be obtained entirely from the steady-state parameters. The calculations require steady-state values for the field dependence of energy, velocity, and effective mass, which may be obtained from Monte Carlo method (see, e.g., Littlejohn *et al.* [44]). An illustration of this is given in Figs. 15–17.

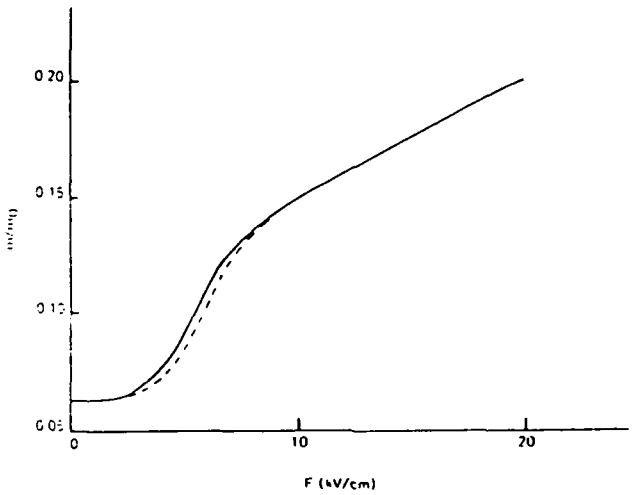


Fig. 17. Steady-state effective mass ratio versus electric field: (—)$N_d = 10^{17}$ cm^{-3}, (---)$N_d = 3 \times 10^{17}$ cm^{-3}.

An important question concerning the use of Eqs. (53b)–(53g) is: How good is the approximation? Clearly, as used by Carnez *et al.* [43] and others, only two relaxation mechanisms, energy and momentum, are considered. An extremely important relaxation mechanism is intervalley carrier relaxation. While this, in principle, can be accounted for in a general SETEM treatment, this has not been addressed by Eqs. (53b)–(53g). In addition, all spatial relaxation has been ignored. In submicron devices where the fields are highly nonuniform, spatial relaxation may be the dominant transient contribution. Should this occur, Eqs. (53b)–(53g) would be inadequate. Nevertheless, as applied by Carnez *et al.*, surprisingly good results can be obtained (see Figs. 18 and 19).

D. Numerical Techniques

The truncation of the expansion for the distribution function to the first two terms, as is done in the displaced Maxwellian, is, in general, not a valid approach. It usually is justified only in the rare cases that eFl is small compared with the energy range over which f_0 varies appreciably, where l is a composite mean free path. The rapid variation of f_0 in the region about the optical phonon energy limits this truncation to small values of the field or to cases of very rapid energy exchange via carrier–carrier scattering. The problem is to derive the distribution function for free electrons in a semiconductor from a knowledge of the various scattering processes and the applied fields. Although it is possible to justify various numerical solutions of the electron-transport problem without specific reference to the

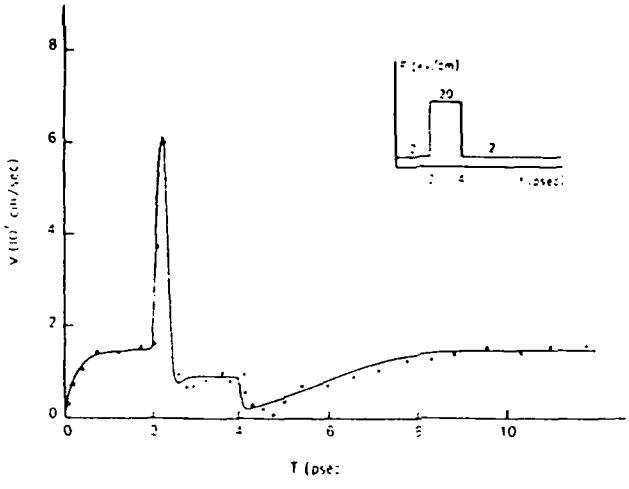


Fig. 18. Drift velocity versus time when applying an electric field pulse; (—) analytic formulation, (· · ·) Monte Carlo calculations. From Carnez *et al.* [43], with permission.

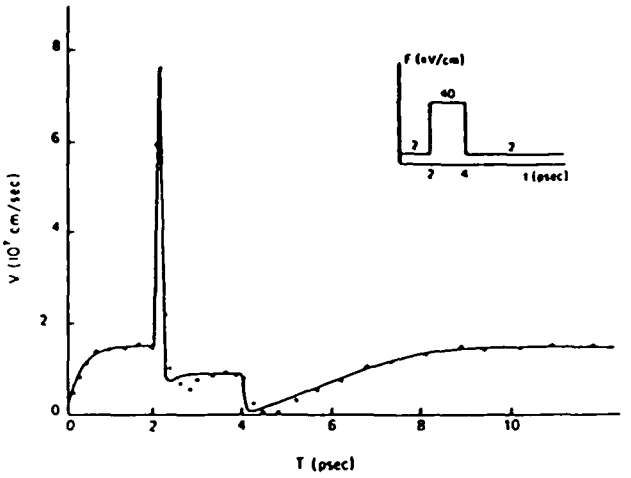


Fig. 19. Drift velocity versus time when applying an electric field pulse; (—) SETEM formulation, (· · ·) Monte Carlo calculations. From Carnez *et al.* [43], with permission.

integrodifferential form of the Boltzmann equation, it is more illustrative to take this as a starting point and, thus, to clarify the basic properties of the numerical methods. For a spatially uniform electron system in an external field **F**, the time-dependent Boltzmann equation is given from (44) as

$$\left[\frac{\partial}{\partial t} - e\mathbf{F}\cdot\frac{\partial}{\partial \mathbf{p}} + \lambda(\mathbf{k})\right] f(\mathbf{k},t) = \int W(\mathbf{k}',\mathbf{k}) f(\mathbf{k}',t)\, d\mathbf{k}', \qquad (54)$$

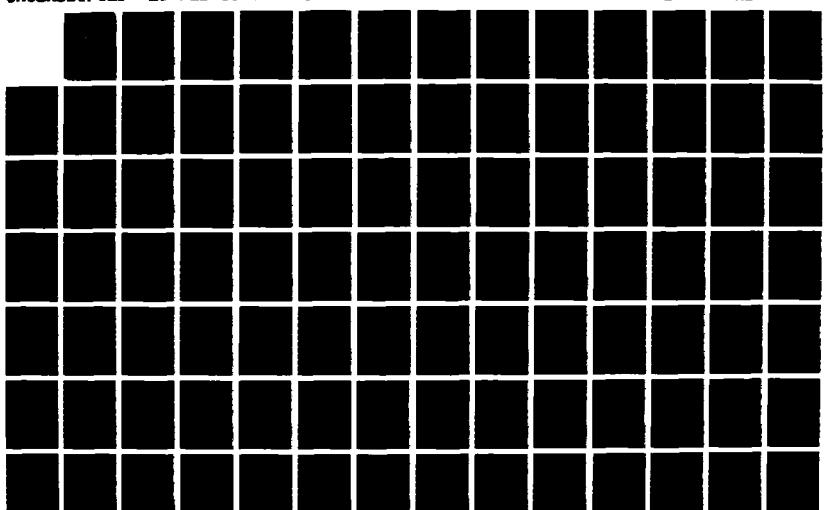
where $\mathbf{p} = \hbar\mathbf{k}$ and $\lambda(\mathbf{k})$ is the total out-scattering rate ($\lambda = 1/\tau$) and represents the second term on the right-hand side of (44). The term on the right-hand side of (54) represents just the in-scattering contributions. However, the definitions in $\lambda(\mathbf{k})$ are relatively incomplete since part, or all, of the $\lambda(\mathbf{k})f(\mathbf{k},t)$ term could be absorbed into the term on the right. The precise definition of $\lambda(\mathbf{k})$ and, consequently, of $W(\mathbf{k},\mathbf{k}')$ will usually depend on the particular calculation to be undertaken, but the formal theory is independent of these considerations. We shall therefore proceed as just defined.

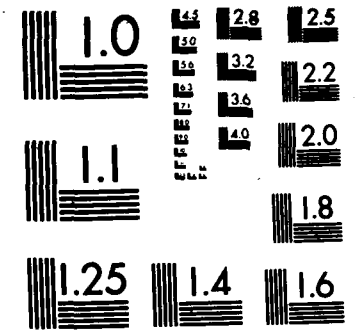
The inverse of the differential operator (54) is just an integral operator. Budd [45] points out that the integration is a generalization of the Chambers [46] path integral, and the result can be written as

$$f(\mathbf{k},t) = \int_0^\infty ds \int d\mathbf{k}\; f\left(\mathbf{k}', t - s\right) W\left(\mathbf{k}', \mathbf{k} - \frac{e\mathbf{F}s}{\hbar}\right)$$
$$\times \exp\left\{-\int_0^s \lambda\left(\mathbf{k} + \frac{e\mathbf{F}}{\hbar}\right) ds\right\}$$

AD-A193 380 STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL NEAR-MICRON AND SUB-.. (U) SCIENTIFIC RESEARCH ASSOCIATES INC GLASTONBURY CT H L GRUBIN ET AL. 2/6

UNCLASSIFIED 15 FEB 88 SRA-R88-920010-F N00014-81-C-0452 F/G 9/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The transformation of the BTE into this form, when supplemented by numerical techniques on a grid of points in k space, as may be peformed on modern digital computers, provides an extremely powerful technique for solving the Boltzmann equation. It is demonstrated, quite elegantly, by Rees [47,48] that an iterative approach can be utilized to great advantage. Moreover, he also points out that the iterative solution is a surrogate for the time evolution of $f(\mathbf{k},t)$, a point we shall discuss further.

An important technical innovation introduced by Rees [47] is the concept of self-scattering, a fictitious scattering process, which does not alter the physics, but allows a great simplification of the mathematical detail. To each side of (55) we add a term of the form

$$\Gamma(\mathbf{k})\,\delta(\mathbf{k}-\mathbf{k}'). \tag{56}$$

In particular, the simplification arises if we define this term as

$$\Gamma(\mathbf{k}) = \Gamma - \lambda(\mathbf{k}), \tag{57}$$

with Γ sufficiently large that $\Gamma(\mathbf{k}) > 0$ for all k, although adequate results can often still be obtained if this condition is relaxed. Then, (55) becomes

$$f(\mathbf{k},t) = \int_0^\infty ds \int d\mathbf{k}'\, f(\mathbf{k}',t-s) W^*\left(\mathbf{k}',\mathbf{k}+\frac{e\mathbf{F}s}{\hbar}\right) \exp(-\Gamma s), \tag{58}$$

where

$$W^*(\mathbf{k}',\mathbf{k}) = W(\mathbf{k}',\mathbf{k}) + \Gamma(\mathbf{k})\,\delta(\mathbf{k}-\mathbf{k}'). \tag{59}$$

The iteration presented by Rees [47,48] consists of two distinct parts, or steps. The first part of each iteration is represented in the time domain by the evaluation of an intermediate function $g_n(\mathbf{k},t)$ from the nth iterate $f_n(\mathbf{k},t)$ according to

$$g_n(\mathbf{k},t) = \int d\mathbf{k}'\, f_n(\mathbf{k}',t) W^*(\mathbf{k}',\mathbf{k}). \tag{60}$$

The second part of the iteration generates the $(n+1)$st iterate $f_{n+1}(\mathbf{k},t)$ as the causal solution of (44), recognizing that $g_n(\mathbf{k},t)$ can be the right-hand side and $f(\mathbf{k})$ is replaced by $f_{n+1}(\mathbf{k},t)$. This causal solution is then

$$f_{n+1}(\mathbf{k},t) = \int_0^\infty ds\, g_n\left(\mathbf{k}+\frac{e\mathbf{F}s}{\hbar}, t-s\right) \exp(-\Gamma s). \tag{61}$$

Physically, this last integral represents integration along the trajectory (the path integral) and the exponential factor is just the probability that no scattering has occurred during the traverse of the path. The appeal to the stability of the steady state gives the result that the final distribution func-

tion arises from

$$f(\mathbf{k}) = \lim_{n\to\infty} f_{n+1}(\mathbf{k},t), \tag{62}$$

and since the scattering factors and path variables shape $f(\mathbf{k})$, the initial guess for $f_1(\mathbf{k})$ is not critical. There is the further useful result that with $\Gamma(\mathbf{k})$ defined as in (57),

$$\lim_{n\to\infty} f_n(\mathbf{k}) = f(\mathbf{k},n/\Gamma). \tag{63}$$

Rees [47], in arriving at this result, shows that each iteration is equivalent to a time step of $1/\Gamma$. However, whatever value of Γ is chosen, the resulting steady state $f(\mathbf{k})$ is the same. The time-development capability, however, allows a description of the approach to equilibrium from any given initial function to be ascertained, and by varying the field between iterations, the time-dependent response can be found.

It can be readily observed that the iterative approach represents a chain of integrations, representing alternate applications of a path integral and a scattering integral. The entire chain operates on an initial trial function. This chain suggests an alternative approach to the solution of (54), the Monte Carlo evaluation of the integrals. In this latter case, a two-step iteration is also followed. The first step is a path traversal, terminated at a time t selected on a random value of the function $\exp(-\Gamma t)$. The second step involves scattering from the state resulting at the end of this traverse to a new state. The new state is governed by the type of scattering process used and this latter quantity is randomly selected from those present. That is, a typical electron is considered at $t = 0$ to be accelerated to t_1, where $t_1 = \ln(R_1)/\Gamma$ and R_1 is a random number. At t_1, the electron has been accelerated to a state $(\mathbf{k}_1,E_1)$. At this point, the relative probability of the ith scattering event is $\lambda_i(E_1)/\Gamma$, where $\lambda_1 + \lambda_2 + \cdots + \lambda_n = \Gamma$, including self-scattering. The particular scattering event is selected by a second random number R_2 as the kth process when $\lambda_1 + \cdots + \lambda_{k-1} < \Gamma R_2 < \lambda_1 + \cdots + \lambda_{k-1} + \lambda_k$. The properties of this scattering event are then used to determine the final state $(\mathbf{k}_2,E_2)$, which is used as the new $t = 0$ state and the process repeated. If the states $(\mathbf{k}_1,E_1)$ are tabulated in a k-space grid, then their distribution becomes a representation of $f(\mathbf{k})$. An estimator for the physical variable ϕ, such as the velocity, is generated as $\Sigma\phi(\mathbf{k}_1)/(\text{number of scatterings})$, where the sum runs over all of the scatterings. The validity of such an average lies in the ergodicity of the physical process, providing that a sufficiently large number of iterations has been used. The basic Monte Carlo technique was first put forward by Kurosawa [49], but its full capabilities were not evident until the introduc-

tion of self-scattering by Boardman *et al.* [50]. The application of Monte Carlo to time-dependent phenomena is restricted to cases where $\omega \ll \Gamma$, because of the need to establish equilibrium among the states.

Comparison of Techniques

In Fig. 20, a velocity–field curve is shown for electrons in the central valley of gallium nitride as found from two methods of calculation. The curve calculated from a Monte Carlo technique by Littlejohn *et al.* [53] is contrasted with that calculated by a displaced Maxwellian by Ferry [54]. The difference between the two lies in the anisotropy introduced by the polar optical phonon scattering. For low carrier concentrations, this anisotropy, due to the small angle of scattering in this situation, leads to streaming of the carriers. The streaming effect is generated by a spike in k space directed along the electric field. To model this adequately by a Legendre expansion would require an exceedingly large number of terms to be retained. At high densities, however, carrier–carrier scattering will cause the spike to be dissipated. Which of the two curves is correct depends on the degree to which the carrier–carrier interaction is significant in the development of the distribution function.

In Fig. 21, the opposite case is shown. Here the velocity–field curve of Si at $N_0 = 10^{17}$ cm^{-3} is shown. In this case, the nonpolar optical phonons are relatively isotropic scatterers and carrier–carrier scattering is large. Here, the two methods of calculation, Monte Carlo and drifted Maxwel-

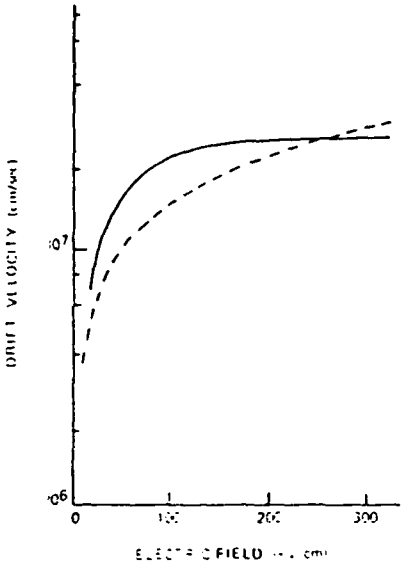


Fig. 20. The drift velocity in GaN at 300 K as calculated by a Monte Carlo technique [51] (—) and by a displaced Maxwellian [52] (– – –). The differences are discussed in the text.

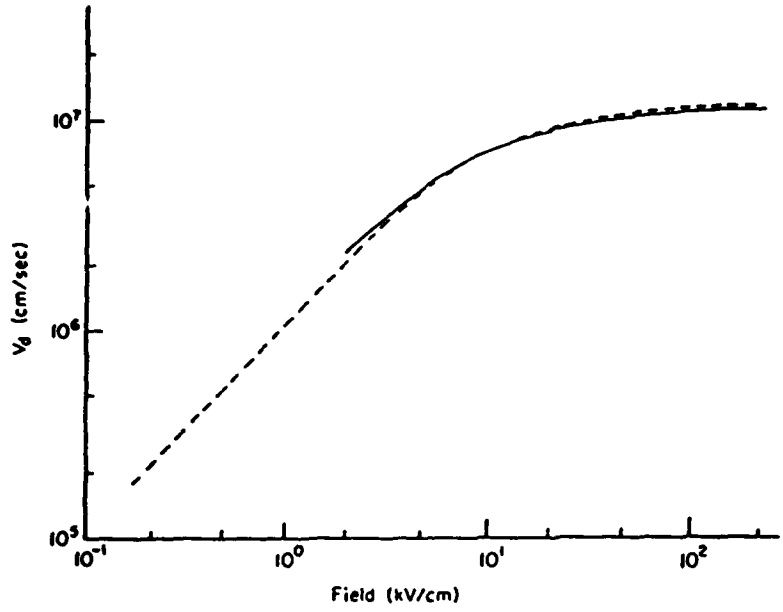


Fig. 21. The drift velocity in Si at 300 K as calculated by a Monte Carlo technique [55] (—) and by a displaced Maxwellian (---).

lian, agree quite well in this high-field region. We might also expect the two methods to agree well for GaAs, if the carrier concentration is high, such as 10^{17} cm^{-3}, and the scattering is dominated by the intervalley phonons. This is shown in Fig. 22 for the transient velocity in GaAs at 10 kV/cm. This figure also shows the importance of having $\Gamma \gg \omega_{max}$, as previously mentioned, where ω_{max} is the highest-frequency component in the transient response.

The explicit representation of the actual distribution function, such as occurs in the iterative solution and to a lesser extent in the Monte Carlo approach, has natural advantages. Effects that are nonlinear in $f(E)$, such as carrier–carrier scattering and degeneracy of $f(E)$, can readily be incorporated into the calculations. The details of the scattering are fully tied up in the terms $W^*(k,k')$ and $\Gamma(k)$, so no conceptual difficulty arises in incorporating nonphonon scattering events such as impact ionization [51], optical carrier generation [52], or even cyclotron-resonance-type transitions.

E. Velocity Transients

In traditional semiclassical approaches to solutions of the BTE, it is assumed that the response of the carrier to the applied force is simultaneous with the applied force, even though the system may undergo subsequent

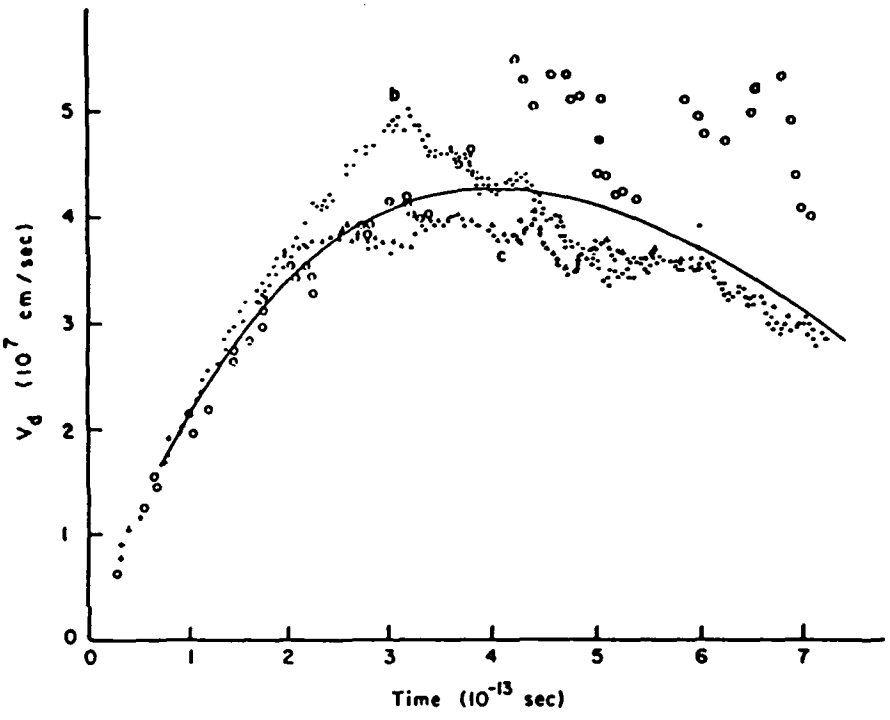


Fig. 22. Comparison of the transient dynamic response of electrons in GaAs for a field of 10 kV/cm at 300 K. The solid curve is the result of a displaced Maxwellian [59], while the data curves labeled a, b, and c are Monte Carlo results for $\Gamma = 1 \times 10^{14}$, 3×10^{14}, and 5×10^{14}, respectively [55].

relaxation. However, on the short time scale of the velocity transient, a truly causal theory introduces memory effects that lead to convolution integrals in the transport coefficients [55–62].

In this section, we shall provide the detailed derivations of this memory or retardation, showing in complete form both the momentum and energy-balance equations for a drifted Maxwellian approach. The energy-balance equation involves retardation of the collisional energy relaxation by the nonzero collision duration only. However, the collisional momentum relaxation is retarded, not only by this effect, but also through the normal memory effect. In order to allow completeness, we shall also include here the finite collision duration discussed in the next section.

The Boltzmann transport equation is valid in the weak coupling limit in which the collisions can be treated independently and perturbatively. It can readily be developed from more exact density matrix equations, in which the field contribution also appears as a differential superoperator

in the momentum representation [55]. In the case where the field is large and/or the collision duration is significant, the collision integral in the BTE generalizes to a form involving the replacement of the energy-conserving δ function factors in the Golden Rule transition rates by a path integral over the time t into a collision [63–65], and the BTE becomes

$$\frac{\partial f}{\partial t} - e\mathbf{F}\cdot\nabla_{\mathbf{p}} f = \int_0^t dt' \left[\sum_{\mathbf{p}'} S(\mathbf{p},\mathbf{p}',t') f(\mathbf{p}',t-t') - \sum_{\mathbf{p}'} S(\mathbf{p}',\mathbf{p},t') f(\mathbf{p},t-t')\right]. \tag{64}$$

The right-hand side of (64) is defined as the collision integral, and the details of the transition terms take the form, for inelastic phonon scattering, of

$$S(\mathbf{p},\mathbf{p}';t,t') = \frac{1}{\pi\hbar}\,\mathrm{Re}\{W(\mathbf{p},\mathbf{p}')\exp\left(-\frac{t-t'}{\tau_\Gamma}\right) \times \delta_{\mathbf{p},\mathbf{p}'+\eta\mathbf{q}} \exp\left[-i\int_t^{t'}\frac{dt''}{\hbar}\beta(\mathbf{p},\mathbf{p}';t'')\right], \tag{65}$$

where

$$\beta(\mathbf{p},\mathbf{p}';t'') = E[\mathbf{p}(t'')] - E[\mathbf{p}'(t'')] + \eta\hbar\omega_{\mathbf{q}} \tag{66}$$

and the term in $\exp(-t/\tau_\Gamma)$ weights the time variation with the lifetime of the quasi-practical states and $W(\mathbf{p},\mathbf{p}')$ in $2\pi/\hbar$ times the square of the perturbing matrix element. The momenta $\mathbf{p},\mathbf{p}'$ are explicit functions of the retarded time t on the right-hand side through the relationships.

$$\mathbf{p}(t') = \mathbf{p} - \int_{t'}^{t} e\mathbf{F}(t'')\,dt'', \tag{67a}$$

$$\mathbf{p}'(t') = \mathbf{p}' - \int_{t'}^{t} e\mathbf{F}(t'')\,dt''. \tag{67b}$$

The two exponential factors in Eq. (65) are related to the joint spectral density function, which reduces to an energy-conserving δ function in the instantaneous collision, low-field limit. Here $E(p)$ is the quasi-particle renormalized electron energy, $\hbar/\tau$ the joint linewidth due to collisional broadening of the initial and final states, and η takes the values $+1, -1$ for phonon emission or absorption, respectively, in the in-scattering term. For the out-scattering term, the roles of p,p' are interchanged, although this does not upset detailed balance in the equilibrium sense.

In small semiconductor devices, where the dimensional scale is of the

order of 1.0 μm or less, the carrier concentration will, in general, be relatively high.[4] Under these conditions, the anisotropic terms in the distribution function are small and a parameterized distribution function, the displaced Maxwellian, can be utilized. With this assumption, a hierarchy of moment equations can be generated from the BTE, from which the various parameters can be determined. A normal expansion of the displaced Maxwellian appears as Eq. (45), but for a form required for the memory functions, we specify $f(E)$ by the ansatz

$$f(E) \propto \int_0^t d\tau \left[\delta(t-\tau) - \mathbf{v}_d(t-\tau)\frac{d}{d\tau}\left(\mathbf{p}\frac{d}{dE}\right)\right] f_0(E,\tau), \tag{68}$$

where the two derivatives are required, respectively, to preserve the dimensionality and to bring β_e ($=1/k_B T_e$) into the second expression in the bracket. In the limit of long times Eq. (68) reduces to (45). Here $f_0 = \exp(-\beta_e E)$.

We now take the Laplace transform of (64) as follows: assuming a single spherical isotropic conduction band,

$$sf(\mathbf{p},s) - f(\mathbf{p},0) - e\mathbf{F}\cdot\nabla_p f(\mathbf{p},s) = \frac{2}{\hbar^2}\sum_{\mathbf{p}'}\{W(\mathbf{p},\mathbf{p}')f(\mathbf{p}',s) - f(\mathbf{p},s)W(\mathbf{p}',\mathbf{p})\}G(s) \tag{69}$$

where

$$G(s) = \mathrm{Re}\int_0^\infty \exp(-st)\exp\left[i(E-E'\pm\hbar\omega_0)\frac{t}{\hbar} + i\frac{e\mathbf{F}\cdot(\mathbf{p}-\mathbf{p}')}{2m^*\hbar}t^2 - \frac{t}{\tau_\Gamma}\right]dt \tag{70}$$

and the time dependence in (67) has been explicitly expanded. The second exponential in the integral is a very rapidly oscillating function and we can generate a first-order estimate by an asymptotic approximation to $G(s)$. We do this within the spirit of the method of stationary phase. The first exponential can be brought outside the integral and evaluated at τ_c, the time for which the phase in the second exponential is zero. We expect then that $s\tau_c$ will be small, the more so because we generally must take the limiting case of s small to assure that the collisions are complete. Under these assumptions, we can make the approximation

$$\exp(-s\tau_c) \simeq 1 - s\tau_c \simeq (1+s\tau_c)^{-1}. \tag{71}$$

The remaining integral is just the field-shifted and -broadened joint spec-

[4] See, for example the arguments on size scaling in Hoeneisen and Mead [66].

tral density function [63–65], which, in the limit of small field and large τ_Γ, reduces to $\pi\,\delta(E - E' \pm \hbar\omega_0)$, the normal form. While the exact form of the integral is not easily analyzed in physical terms, a simple perturbation expansion allows us to model all but the irrelevant fast oscillatory behavior by approximations [67].

Using these forms for the function $G(s)$, we can rewrite (69) as

$$sf(\mathbf{p},s) - f(\mathbf{p},0) - e\mathbf{F}\cdot\nabla_{\mathbf{p}} f(\mathbf{p},s) = \frac{2\pi}{\hbar}\sum_{\mathbf{p}'}\left[W(\mathbf{p},\mathbf{p}')f(\mathbf{p}',s) - W(\mathbf{p}',\mathbf{p})f(\mathbf{p},s)\right]\frac{\delta_E(\mathbf{p}',\mathbf{p})}{1 + s\tau_c}. \quad (72)$$

We can now multiply by an arbitrary function $\phi(\mathbf{p})$ and eventually sum over all final states $\mathbf{p}$. Then a general approach, such as in Appendix A for developing moment equations, can be followed using the transformed version of Eq. (68). These balance equations are developed in the transformed domain, however, and must be inverted to obtain the time response. The results are: for the energy equation,

$$\frac{\partial}{\partial t}\left(\frac{3}{2}k_B T_e\right) = -e\mathbf{F}\cdot\mathbf{v}_d(t) - \frac{1}{\tau_c}\int_0^t \exp\left(\frac{-\tau}{\tau_c}\right)\langle\dot{\Gamma}_E(t-\tau)\rangle\,d\tau, \quad (73)$$

where we have recognized that $\langle\phi\rangle = 3k_B T_e/2$ when $\phi = E$, and $\dot{\Gamma}_E(t)$ is

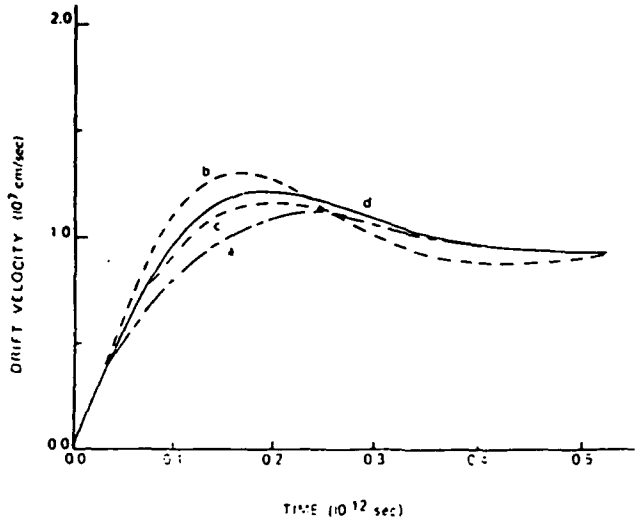


Fig. 23. Transient dynamic response of carriers in silicon at 300 K. It is assumed the carriers see a homogeneous field of 20 kV/cm that is applied at $t = 0$. Curve a shows the response neglecting the short-time effects but including the weakening of the collision in the high field. Curve b includes the effects of retardation due to the nonzero collision duration alone. Curve c includes the memory effects alone. Curve d includes all effects according to Eqs. (73) and (74).

now defined by (47); and for the momentum-balance equation,

$$m^* \frac{\partial v_d}{\partial t} = -eF - \frac{m^*}{\tau_c} \int_0^t v_d(t-\tau) \int_0^\tau \langle \dot{\Gamma}_m(\tau - \tau') \rangle \exp\left(\frac{-\tau'}{\tau_c}\right) d\tau' \, d\tau. \quad (74)$$

If $\tau_c \to 0$ and v_d is slowly varying, then (74) reduces to the normal form (46) and Γ_m is recognized as the rate of momentum relaxation (48). The various convolutions introduce a significant temporal retardation in the rate of momentum relaxation. Although Eqs. (73) and (74) include significant corrections on a short time scale, there is reason to believe that further corrections still need to be included.

In Fig. 23, the transient response of the electrons in Si to a steady homogeneous field of 20 kV/cm is shown. Although the retardation due to the nonzero collision duration acts to speed up the response, the dominant factor is the memory effect of $\dot{\Gamma}_m$ and v_d. The collisional retardation speeds up the process, however, primarily because of the effect of slowing down changes in the effective energy via collisional relaxation.

F. Device Simulation from the Boltzmann Transport Equation

Based on the preceding discussion, it is clear that mobility models will have limited usefulness in submicron and high-frequency devices. The extent to which this is true is material-dependent and is most likely to be revealed by device simulations. Complete one- and two-dimensional simulations involving solutions of the BTE are only sparsely available. Rather, bits of the problem are treated. Here Monte Carlo, iteration, and moment methods are used and in the following we shall illustrate some results, drawing on GaAs, as the nonlocal spatial and temporal contributions are dramatically enhanced by the transferred electron effect. Our approach will emphasize the displaced Maxwellian moment equations, primarily because of intuitive advantages and for the relative ease with which space-charge contributors can be handled.

The device-moment equations [68] for a displaced Maxwellian of the form given by Eq. (A-1), are

$$(\partial N_i/\partial t) + \nabla \cdot (\mathbf{v}_i N) = (\partial N_i/\partial t)_c, \quad (75)$$

$$(\partial \mathbf{P}_i/\partial t) + \nabla \cdot (\mathbf{v}_i \mathbf{P}_i) = -eN_i\mathbf{F} - \nabla(N_i k_B T_i) + (\partial P_i/\partial t)_c, \quad (76)$$

$$(\partial W_i/\partial t) + \nabla \cdot (\mathbf{v}_i W_i) = -eN_i \mathbf{v}_i \cdot \mathbf{F} - \nabla \cdot (\mathbf{v}_i N_i k_B T_i) - \nabla \cdot \mathbf{q}_i + (\partial W_i/\partial t)_c. \quad (77)$$

where the subscripts identify a particular valley. Here

$$\mathbf{P}_i = m_i N_i \mathbf{v}_i, \quad (78)$$

$$W_i = \tfrac{3}{2} N_i k_B T_i + \tfrac{1}{2} m_i N_i v_i^2, \tag{79}$$

m_i is the effective mass of electrons in the ith valley, and T_i is the electron temperature of the ith valley, and W_i is regarded as the average total kinetic energy density.

These equations have simple physical interpretations. For Eq. (75), the increase of electron density plus the outflow of electrons equals the increase of density due to collisions. For Eq. (76), the left-hand side is the rate of change plus the outflow of momentum density of the ith valley. The right-hand side represents the forces exerted by the electric field and by the electron pressure $n_i k_B T_i$ and the rate of momentum density gained in collisions. For Eq. (77), the left-hand side contains the rate of change plus outflow of total kinetic energy density. The right-hand side represents the energy supplied by the electric field, the work performed by the electron pressure, the divergence of the heat flow $\mathbf{q}_i$, and the rate of change of total kinetic energy density due to collisions.

When the transport equations are derived by integration of the Boltzmann equation, the heat-flow vector $\mathbf{q}_i$ appears as a third moment of the distribution function. With the assumption of displaced Maxwellian distributions (or any symmetric distributions), these moments vanish. In spite of this, following Blotekjar [68], we may allow for heat conduction by assuming

$$\mathbf{q}_i = -K_i \nabla T_i. \tag{80}$$

where K_i is the heat conductivity of the electron gas in the ith valley. This term is believed to represent the most important effect of a non-Maxwellian distribution function [68].

The $(\partial/\partial t)_c$ terms in Eqs. (75)–(77) represent scattering integrals. They may be given in approximate form by Eq. (49), for example, for polar optical scattering or in exact form for the displaced Maxwellian. In the following simulations we chose the latter and use the integrals summarized by Butcher [2] (see also Blotekjar and Lunde [69]). The scattering integrals are then represented as

$$(\partial N_i/\partial t)_c = (-N_i/\tau_{N_{ij}}) + (N_j/\tau_{N_{ji}}). \tag{81}$$

$$(\partial \mathbf{P}_i/\partial t)_c = -\mathbf{P}_i/\tau_{P_i}, \tag{82}$$

$$(\partial W_i/\partial t)_c = -\tfrac{3}{2}(N_i k_B T_i/\tau_{E_{ij}}) - \tfrac{3}{2}(N_j k_B T_j/\tau_{E_{ji}}). \tag{83}$$

We illustrate a set of scattering curves for a Γ–X orientation in Fig. 24. The parameters used in the calculations are given in Table II. These results should be compared to those of Bosch and Thim [41]. We retain the Γ–X orientation because most of the early moment equations were evalu-

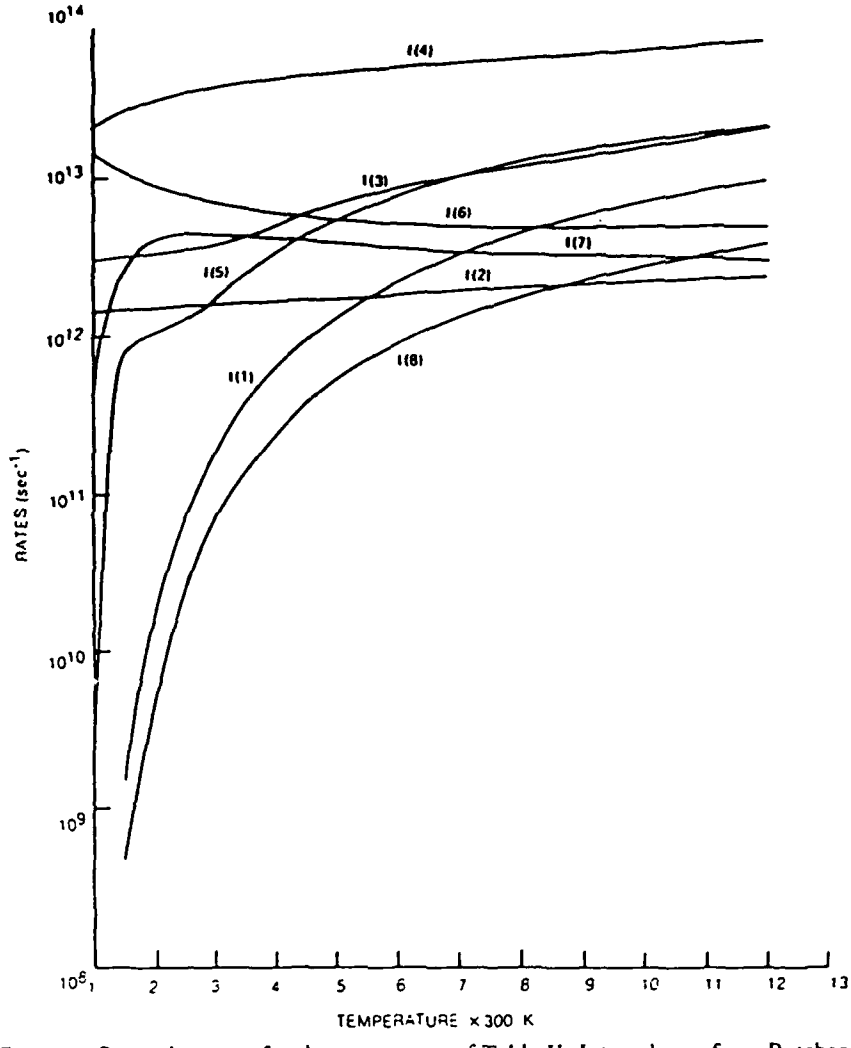


Fig. 24. Scattering rates for the parameters of Table II. Integrals are from Butcher [2]. Curves shown are I(1), CV particle; I(2), SV particle; I(3), CV momentum; I(4), SV momentum; I(5), CV energy; I(6) SV/CV energy; I(7), SV energy; I(8) CV/SV energy.

ated for this ordering. Changing to a Γ–L orientation (see, e.g., Littlejohn *et al.* [44], would offer only quantitative differences and might obscure some of the discussion.

1. Recovery of the Semiconductor Equations. Comparison to Nonlocal Equations

In analyzing transport from the nonlocal balance equation, we are sometimes interested in recovering the ordinary semiconductor equations

TABLE II
Parameters Used in Calculation

Parameters	Γ(000)	X(100)	Common
Number of equivalent valleys	1	3	
Effective mass (m_e)	0.067	0.40	
Γ–X Separation (eV)			0.36
Lattice constant (Å)			5.64
Density (gm/cm³)			5.37
Polar optical scattering			
Static dielectric constant			12.53
High-frequency dielectric C			10.82
LO phonon (eV)			0.0354
Γ–X *Scattering*			
Coupling constant (eV/cm)			0.621×10^9
Phonon energy (eV)			0.0300
X–X Scattering			
Coupling constant (eV/cm)		$1.0.64 \times 10^9$	
Phonon energy (eV)		0.0300	
Acoustic Scattering			
Deformation potential (eV)	7.0	7.0	
Acoustic velocity (cm/sec)			5.22×10^5

and, hence, mobility concepts. For a single parabolic conduction band, we can obtain the usual semiconductor equations through judicious neglect of select space and time derivatives. For example, in the case of continuity, neglect of avalanching allows us to set $(\partial N/\partial t)_c = 0$. We then obtain current continuity for one valley:

$$(\partial N/\partial t) + \nabla \cdot (\mathbf{v}N) = 0. \tag{84}$$

For the momentum term, we set $(\partial \mathbf{P}_i/\partial t)_c = -\mathbf{P}_i/\tau_p$, neglect the terms $\partial \mathbf{P}_i/\partial t$, and the contribution from $\nabla \cdot \mathbf{vP}$ and obtain

$$N\mathbf{v} = \mathbf{P}/m = -N\mu\mathbf{F} - (\mu/e)\,\nabla(Nk_BT), \tag{85}$$

where $\mu = e\tau_p/m$. Then using the Einstein relation $D(F) = \mu k_B T/e$, we obtain

$$N\mathbf{v} = -N\mu\mathbf{F} - \nabla ND. \tag{86}$$

We see that within the framework of the Boltzmann picture even the use of the "semiconductor" equation (86) is suspicious. Further complications arise in multivalley calculations where the mobility is taken as a

weighted average

$$\langle \mu(F) \rangle = \frac{N_c \mu_c(F) + N_s \mu_s(F)}{N_c + N_s} \tag{87a}$$

and the diffusion coefficient is given, with limited validity, by

$$D(F) = \frac{k_B}{e}\left(\frac{N_c \mu_c(F) T_c + N_s \mu_s(F) T_s}{(N_c + N_s)}\right). \tag{87b}$$

Since we have a more general set of transport equations than the semiconductor equations, we are in a position to isolate differences between the two approaches, even on such a relatively direct topic such as domain propagation. Cheung and Hearn [70] examined this question by considering the mobility and scattering rates to be unique functions of electron temperature rather than of field. They are unique functions of field only in steady state. In their study, the particle current flux is given by a multivalley version of Eq. (85):

$$J_i = -N_i \mu_i(T_i)\left\{F + \frac{k_B}{eN_i}\frac{\partial}{\partial X}(N_i T_i)\right\}, \tag{88}$$

$$\frac{\partial N_i}{\partial t} = \frac{-\partial J_i}{\partial x} + \frac{N_j}{\tau_{N_{ji}}(T_j)} - \frac{N_i}{\tau_{N_{ij}}(T_i)}, \tag{89}$$

with

$$\frac{\partial}{\partial t}\left(\frac{3}{2} N_i k_B T_i\right) = -eJ_i F + \frac{3}{2}\left(\frac{-N_i k_B T_i}{\tau_{\varepsilon_{ij}}} + \frac{N_j k_B T_j}{\tau_{\varepsilon_{ji}}}\right) - \frac{\partial}{\partial x}\left(\frac{5}{2} J_i k_B T_i\right). \tag{90}$$

We note that Eq. (88) neglects inertial terms and hence overshoot effects. Figure 25 shows a comparison of domain size using the semiconductor equations and the BTE. We can see significant differences.

We consider some of these inertial effects more closely by solving the full set of equations, (75)–(77), for a one-dimensional 5000-Å GaAs element with a donor distribution as shown in Fig. 26. The element is part of a resistive circuit. The carrier dynamics are examined at two instants of time. (Note: Serious objections can be raised to the use of a "jellium" distribution, insofar as any combination of decreased donor density or size reduction will necessarily introduce effects due to the discrete nature of the donors. This will be ignored in the following discussion.

As the bias is turned on, there is an increase in potential across the device and a corresponding increase in current and field. The field is computed self-consistently and its slope reflects any incomplete screening of inhomogeneities, the mobile carriers. For the device in the schematic configuration of Fig. 26, as the field is rising, energy relaxation is incomplete

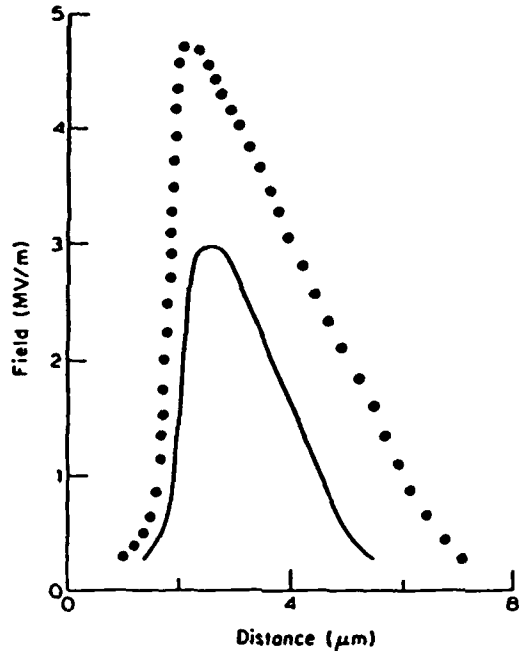


Fig. 25. Field profiles for a high-field-propagating domain. Dotted curve is obtained from the mobility equations. Solid curve is detained using Eqs. (88)–(90). From Cheung and Hearn [70], *with permission.*

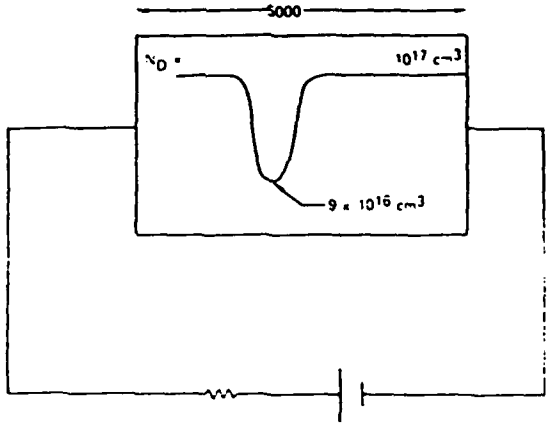


Fig. 26. Schematic representation of device and circuit configuration for submicron homogeneous field profiles. The inhomogeneous doping profile is treated as a "jellium" distribution. The beginning and end of this inhomogeneity are displayed in Fig. 28c. The circuit equation for this calculation is $\phi_B = \phi_0 + IR$. We set R to the low-field resistance of the element and $\phi_B = \phi_B F_0 L$, where $F_0 = 4.3$ kV/cm and $L = 5000$ Å. Generally, ϕ_B is turned on at a finite rate. For these calculations, $\phi_B = 3.0$.

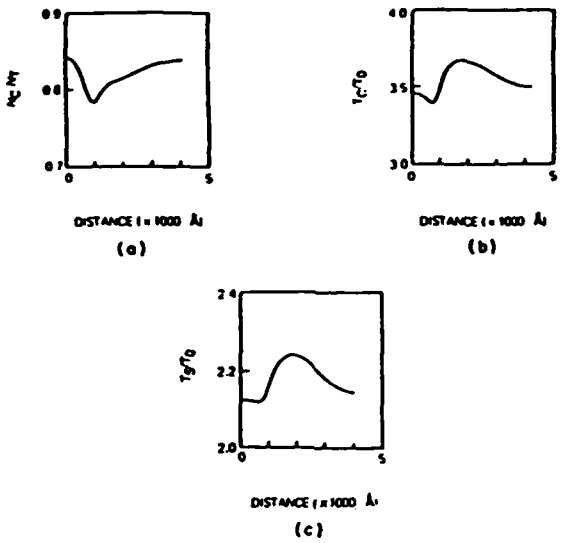


Fig. 27. (a) Fractional central valley population versus distance; (b) Central valley temperature versus distance; (c) satellite valley temperature versus distance. Computation occurs at time $t/\tau_0 = 4$, where τ_0 is the LO phonon intervalley scattering time and equals 0.32 psec [see Fig. 24, curve I(3)] and T_0 is room temperature. The gradient of N is zero at the boundaries. All nonuniformities are due to the notch.

and velocity overshoot contributions are dramatic (see Figs. 27 and 28). Here velocity is computed from the equation

$$\langle v(x,t)\rangle = \frac{P_c/m_c + P_s/m_s}{N_T(x,t)}. \tag{91}$$

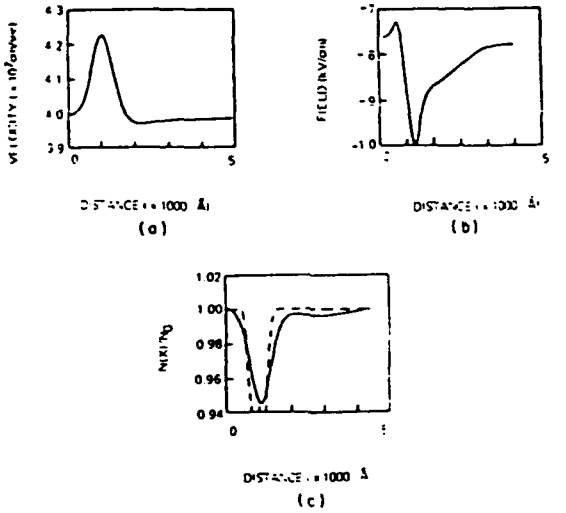


Fig. 28. (a) Mean velocity versus distance; (b) field versus distance; (c) free carrier density versus distance at time $t/\tau_0 = 4$. Dashed curve denotes background density.

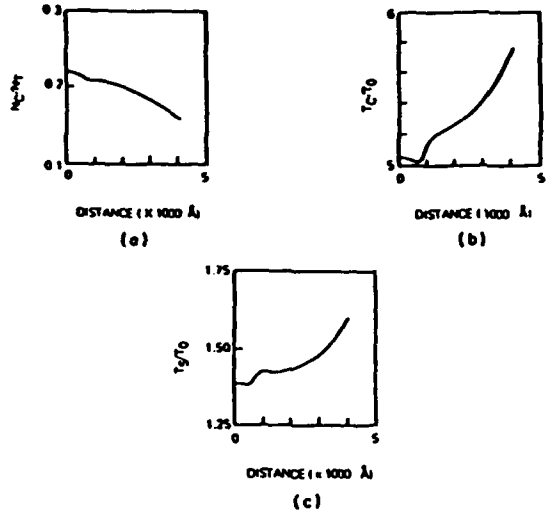


Fig. 29. As in Fig. 27, but at $t/\tau_0 = 16$.

An important point to make here is that when spatial gradients occur, the carrier velocity can overshoot its equilibrium value even though all time derivatives ($\partial/\partial t$) are zero. This was emphasized by Kroemer [71], who estimated substantial overshoot when $\partial F/\partial x \gtrsim F_{th}/\text{mean free path}$, where the subscript "th" designates the NDM threshold field. Thus spatial overshoot should be most pronounced in devices biased near the NDM threshold. The results in Figs. 29 and 30 are for bias fields substantially higher than the NDM threshold. In these figures, the fields are high enough to accommodate almost complete transfer. Here the carrier tem-

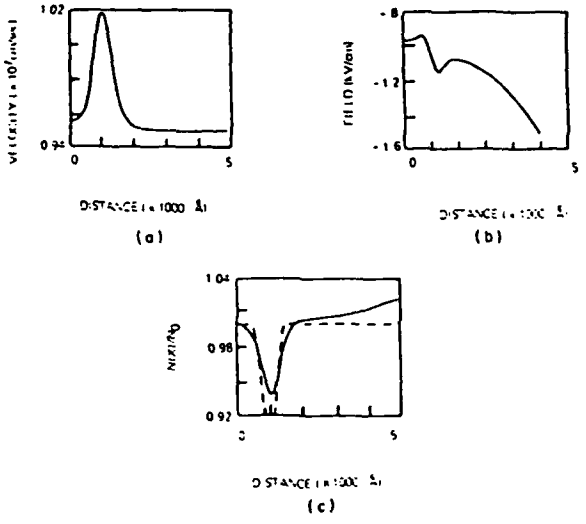


Fig. 30. As in Fig. 28, but at $t/\tau_0 = 16$.

perature is reduced, the energy relaxation rates are shorter, and there is virtually no overshoot. These results are virtually the same as we would obtain using the steady-state curves.

The nonuniform field calculations are the clear order of business in future numerical simulations but a catalog or results have yet to be produced. Until this is done, we are forced to rely heavily on uniform field analysis.

2. Uniform Field Transients

For uniform fields, the transport equations undergo considerable simplifications:

$$\frac{a_i\, dN_i}{dt} = \frac{-N_i}{\tau_{N_{ii}}} + \frac{N_j}{\tau_{N_{ji}}}, \tag{92}$$

where a_i denotes the number of equivalent ith valleys. For $\mathbf{P}_i = \mathbf{P}_i/N_i$,

$$\frac{\partial N_i \mathbf{p}_i}{\partial t} = -eN_i\mathbf{F} - \frac{N_i \mathbf{p}_i}{\tau_{p_i}}, \tag{93}$$

$$\frac{\partial}{\partial t}\left(\frac{N_i p_i^2}{2m_i} + \frac{3}{2} N_i k_B T_i\right) = \frac{-eN_i\mathbf{F}\cdot\mathbf{p}_i}{m_i} \frac{-3N_i k_B T_i}{2\tau_{\varepsilon_{ii}}} + \frac{3N_j k_B T_j}{2\tau_{\varepsilon_{ji}}}. \tag{94}$$

It is instructive at this time to rewrite the momentum-balance equation as

$$\frac{dp_i}{dt} + p_i\left[\frac{1}{\tau_{p_i}} + \frac{d\log N_i}{dt}\right] = -eF. \tag{95}$$

In this form, the effect of intervalley transfer enters as an additional transient momentum scattering term [72].

As previously discussed, dynamic overshoot effects are consequences of differences in momentum and energy relaxation times. In multivalley semiconductors, the overshoot contributions therefore appear in the *momentum*, as well as the velocity, computations. We shall illustrate this for both the central and satellite valleys and for the situations where electrons starting with zero drift velocity are subjected to a sudden change in electric field.

For the central valley and at low values of bias, the electron temperature is approximately equal to room temperature and the ordinary time-dependent dynamic behavior occurs. At elevated bias levels, the electron temperature is substantially increased and the momentum-relaxation time, due to strong intervalley coupling, decreases with increasing temperature (Fig. 31). Thus, we see overshoot, in that the final momentum is below the peak momentum (Fig. 32a). (We point out that above moderate temperature increases, LO phonon intravalley and ionized impurity scat-

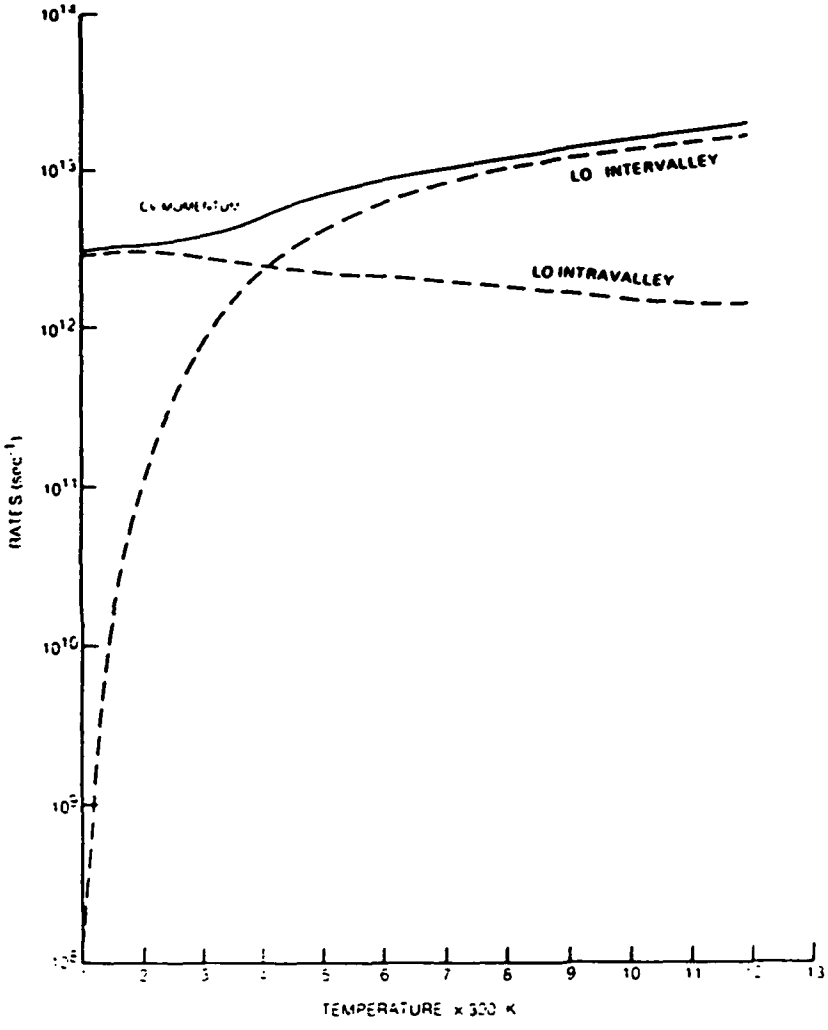


Fig. 31. Central valley momentum scattering rates versus electron temperature.

tering do not provide a momentum-relaxation time that decreases with energy. Intervalley phonons are required. Indeed, for ionized impurity scattering, the relaxation time increases with energy.[5]) During this same time interval, the increasing central valley temperature (Fig. 32b) results in electron transfer, and the momentum density $N_c p_c$ shows an even greater overshoot (Fig. 32c). We now examine the contribution of the term $d(\log N)/dt$ appearing in Eq. (95). For the central valley, where at $t = 0$, $N_s \simeq 0$, and $p_c = p_s = 0$, this term is approximately zero. However,

[5] See, for example, Smith [1, Chapter 5].

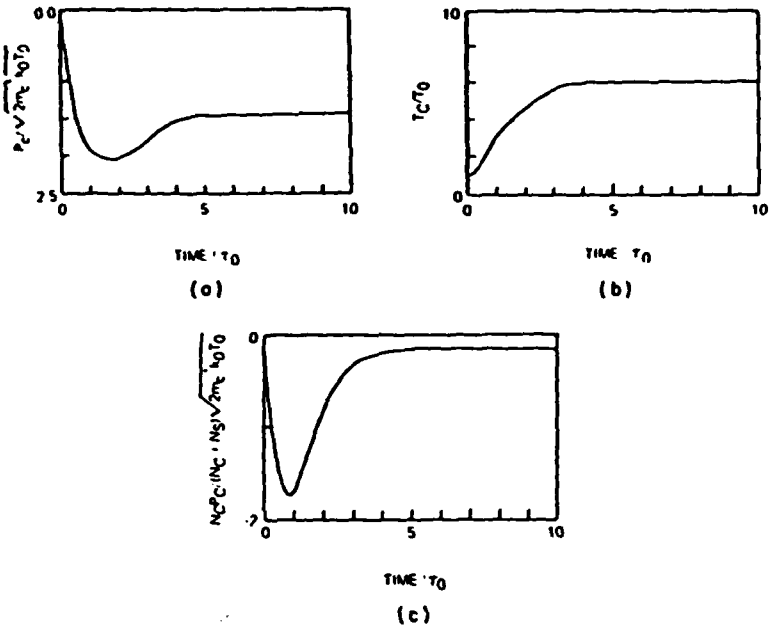


Fig. 32. Transient central valley behavior: (a) momentum versus time; (b) temperature versus time; (c) momentum density versus time. The circuit in this calculation is the same as that of Fig. 26. Here, because the field is uniform, we write the circuit equation as $F_B = F_D + R/R_0 i$, where $i = I/N_0 e v_p A$ with $v_p = \sqrt{2m_c^{-1}k_B T_e} = 3.7 \times 10^7$ cm/sec. For this calculation N_0 is $10^{17}/\text{cm}^3$ and we have included geometrical capacitance. For this calculation, the normalized bias F_B is turned on suddenly to the value $F_B = 5$, which corresponds to an average field of 21.5 kV/cm ($\tau_0 = 0.32$ psec).

when we consider the transient behavior of the satellite valley, the time derivative of log N_s is important because the change in the satellite population, relative to the original number present, is quite large. In addition, we note that the satellite valley momentum-relaxation contribution is almost an order of magnitude larger than that of the central valley. Thus, in a time considerably shorter than that associated with the central valley, the satellite valley momentum reaches the value

$$p_s = -eF(t)\left[\frac{1}{\tau_{ps}} + \frac{d \log N_s}{dt}\right]^{-1},$$

where, because the satellite temperature remains close to room temperature for large changes in field [73], the scattering rates may often be taken as approximately constant. Combining both scattering contributions, we see some "overshoot" due to differential repopulation but none as dramatic as that associated with the central valley. Figure 33 illustrates this point. For part (a), 1 corresponds to the increasing momentum prior to

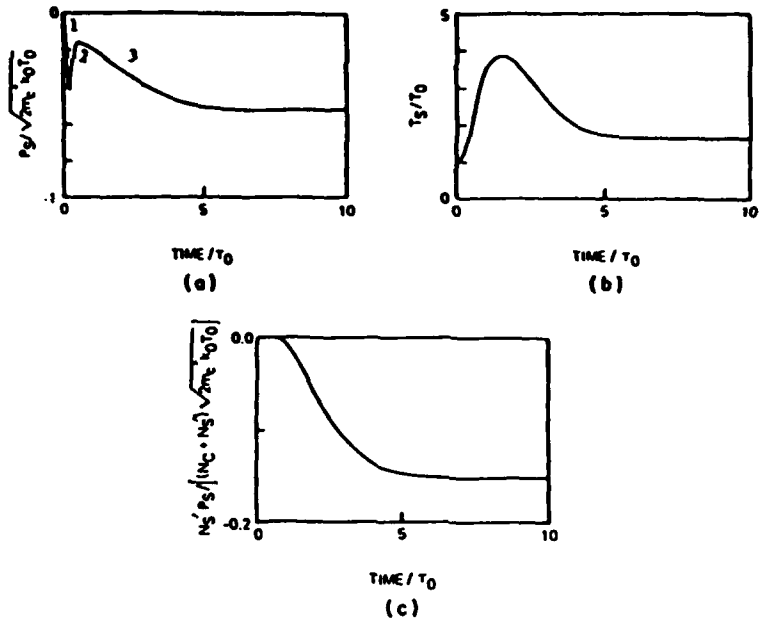


Fig. 33. As in Fig. 32, but for satellite valley behavior.

any significant electron transfer; 2 represents the scattering rate due to repopulation; 3 is the steady-state value. In Fig. 33c, we show the product of $N_s' P_s/(N_c + 3N_s')$. Here the prime indicates population of a single satellite valley ($N_s = 3N_s'$). We see that as far as the contribution to velocity overshoot is concerned, the electron transfer tends to wipe it out.

The average drift velocity versus time is shown in Fig. 34 and is computed from

$$v = \frac{(N_c p_c/m_c) + (N_s p_s/m_s)}{N_c + N_s}. \tag{96}$$

Perhaps the most remarkable aspect of this result is the very large peak velocity prior to steady state (see, e.g., Ruch [3]). This result has been one among many that has led some to suppose that narrow-channel devices will yield higher carrier velocities. To some extent, high overshoot velocities are illusory, as they are very sensitive to rise time. We shall illustrate this for a sequence of trapezoidal bias pulses, each with a varying rise time (see Fig. 35). (See also [74].)

The first set of results is for a relatively slow rise time and the dynamic curves come very close to the steady-state curves (see Fig. 36). A more significant departure from steady state occurs for the somewhat steeper rise time. In Fig. 37, we see some asymmetry in the time dependence of the central valley population and temperature and an increase in the peak

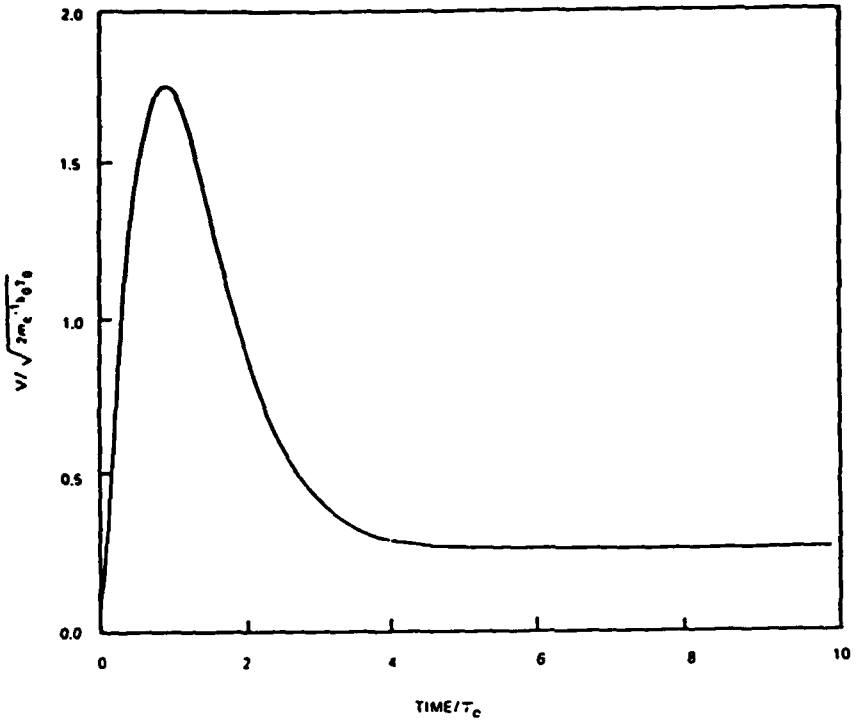


Fig. 34. Mean carrier velocity, from Figs. 32 and 33.

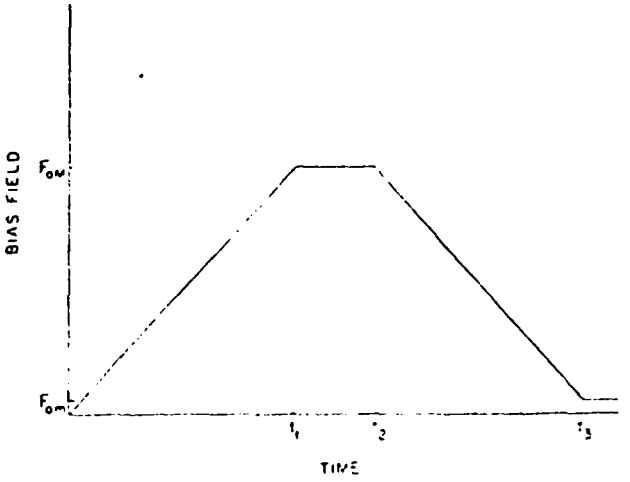


Fig. 35. Time-dependent bias field used to show rise-time dependence of transient velocity. For all of these calculations, $F_{om} = 5$ (21.5 kV/cm) and $F_{om} = 0.1$ (0.43 kV/cm).

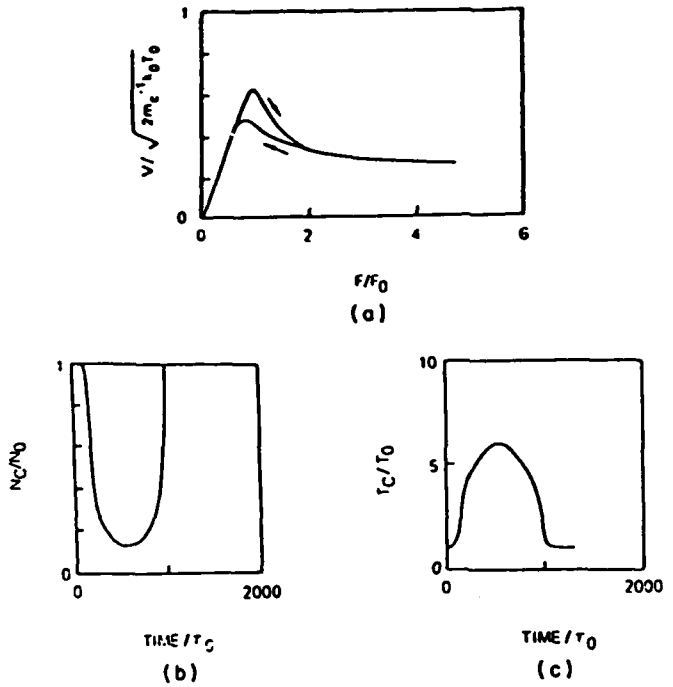


Fig. 36. For this pulse, $t_1 = 500/\tau_0$ (see Fig. 35), $t_2 = 600/\tau_0$, and $t_3 = 1200/\tau_0$. (a) Mean velocity versus field ($F_0 = 4.3$ kV/cm); (b) central valley populations versus time; (c) central valley temperature versus time.

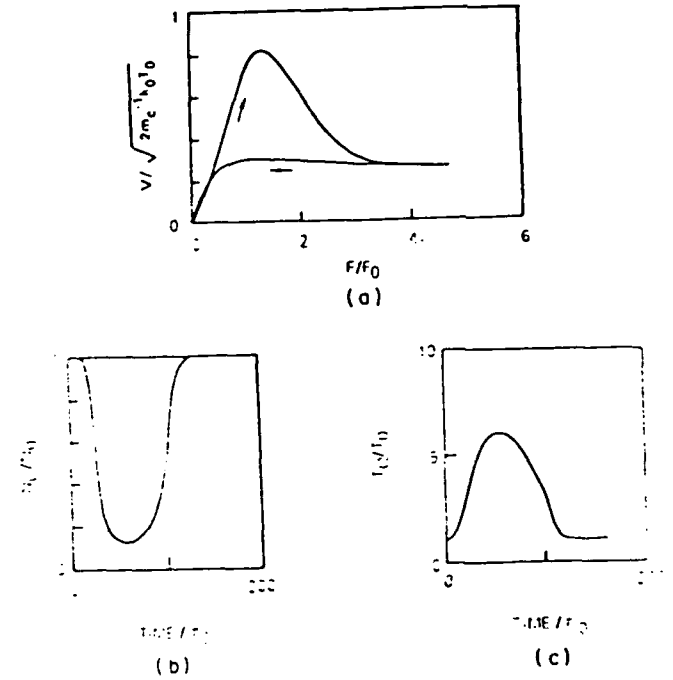


Fig. 37. As in Fig. 36, but here $t_1 = 50/\tau_0$, $t_2 = 60/\tau_0$, and $t_3 = 110/\tau_0$.

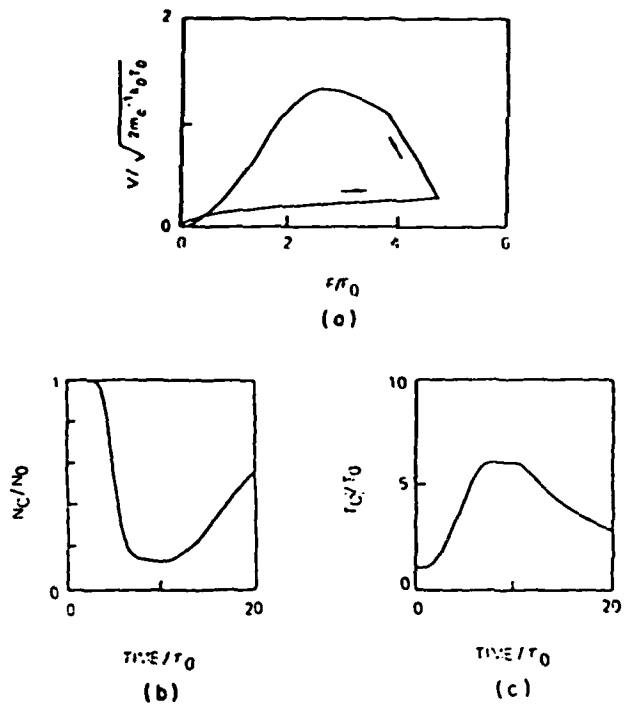


Fig. 38. As in Fig. 36, but here $t_1 = 5/\tau_0$, $t_2 = 10/\tau_0$, and $t_3 = 15/\tau_0$.

velocity. In the final sequence, we show results for a very short rise time. We see a dramatic increase in the peak velocity and clear asymmetry in the carrier dynamics (Fig. 38). Indeed, the final point of approximately zero field and velocity is not an equilibrium state. Rather, we have a dramatic example of velocity undershoot. A longer time is needed for the electron temperature to approach equilibrium. There are strong implications here for upper-frequency limits of device operation.

3. *Determination of Maximum Frequency for Small-Signal, Large-Signal, and Self-Excited Oscillations*

Perhaps the earliest attempt to examine the upper-frequency limit for large-signal oscillation was that of Butcher and Hearn [4]. Using a set of displaced Maxwellian electron distributions for each valley, the set of differential equations [see Eqs. (92)–(94) for the time-dependent electron temperatures, drift velocities, and valley populations was solved for a dc bias field plus rf field. The results of their study are shown in Fig. 39.

In Fig. 39, the mean drift velocity and satellite population are shown as functions of a total field consisting of a dc field of 15 kV/cm and a 60 GHz rf field with an amplitude of 13.1 kV/cm. The arrows indicate the

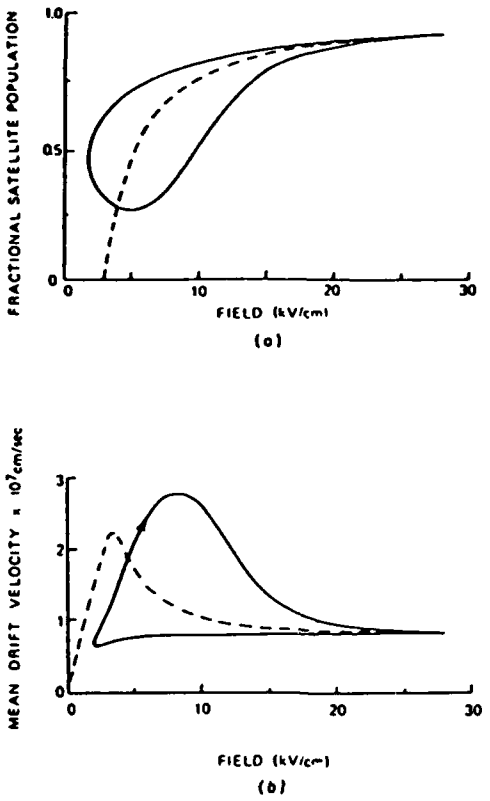


Fig. 39. Plots against field: (a) fractional satellite population; (b) mean drift velocity for a 60-GHz rf field of 13.1 kV/Vcm superimposed on a dc field of 15 kV/cm. The dashed curves give the static values. From Butcher and Hearn [4], with permission of IEE.

direction of increasing time and the dashed lines represent the static relationships. Near the maximum field of 28.1 kV/cm, the satellite population (Fig. 39a) approximates well the static value because the field is stationary and the variation of the population is nearly saturated. The curve is qualitatively similar those of Fig. 37. We see again the higher satellite population on the downswing and the absence of negative differential mobility. We point out that these curves very likely constitute the earliest attempt at including overshoot contributions. For dc bias levels of around 10 to 20 kV/cm and ac levels of 13.1 to 18.6 kV/cm, they obtained an upper-frequency limit of 100 GHz.

From the point of view of a device physicist, a driven oscillator probably lies somewhere between a small-signal and a large-signal self-excited oscillator. The self-excited oscillator is perhaps the most interesting of the three because it highlights the tenacious balance between electron

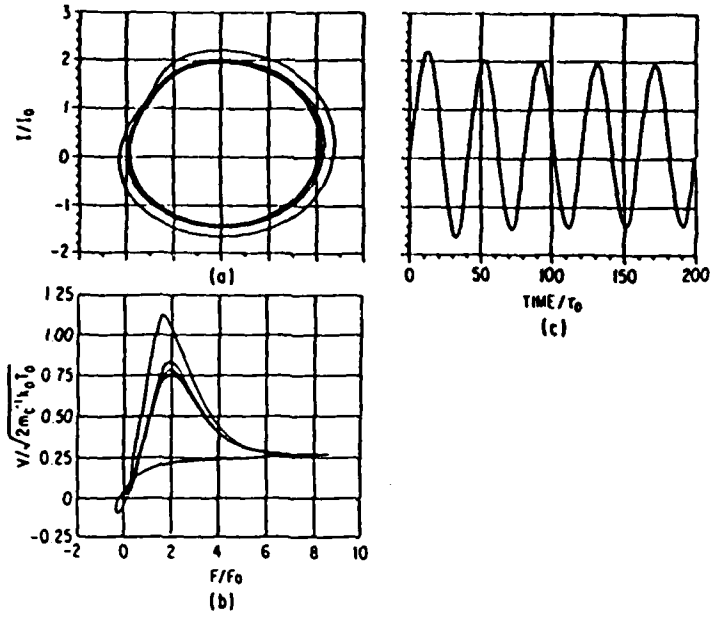


Fig. 40. Self-excited oscillation for the circuit displayed in Fig. 41, where $N_0 = 10^{15}/cm^3$. Oscillation frequency is 78 GHz.

transfer and sustained oscillations. It is extremely sensitive to contact, space-charge, and circuit conditions [8c]. We examine the upper-frequency limit for the device in a circuit with reactive elements. Figure 40 shows the oscillation.

The circuit differential for this oscillation is given by Eq. (11). For nonlinear devices, this equation offers difficulties in interpretation. For interpretative ends, we replace the nonlinear element by a nonlinear resistor with the current–voltage $I_0(V)$ relation. Then, Eq. (12) is conceptually replaced by [8c]

$$\phi_B = \omega_0^{-2}\frac{d^2\phi}{dt^2} + \frac{Z_0}{\omega_0}\frac{dI_0}{d\phi}\frac{d\phi}{dt} + RC\frac{d\phi}{dt} + \phi + RI_0,$$
$$\omega_0 = (LC)^{-1/2}, \qquad Z_0 = (L/C_T)^{1/2}, \tag{97}$$

where $dI_0/d\phi$ represents resistance. When $dI_0/d\phi > 0$, Eq. (97) yields damped oscillations. When $dI_0/d\phi < 0$, the oscillations grow in amplitude. As $dI_0/d\phi$ changes sign during each cycle, the correct set of circuit, bias, and device parameters can yield sustained circuit-controlled oscillations [8c].

Several aspects of a self-excited oscillation are displayed in Fig. 40. The current through the load resistor is displayed in part (b); the dynamic voltage and I vs ϕ, obtained by eliminating time between current and device is

shown in part (b). We also display the mean velocity (dynamic conduction current). The details of the oscillation will be discussed next.

As the field across the device increases and exceeds a threshold, value transfer begins to occur. However, because the field changes more rapidly than the electron temperature, more carriers are retained in the central valley, and with higher momenta than steady state would dictate. This effect is responsible for the higher peak conduction current. But if the increasing electric field sustains high fields for a sufficient duration, enough carriers will transfer to result in negative differential mobility (NDM), which must be of sufficient magnitude for sustained self-excited oscillations.

On the downswing, the field again changes more rapidly than the electron temperature and more carriers are retained in the satellite valley; i.e., we achieve transient undershoot. Now, although NDM is not necessary on the downswing [8c], enough carriers must be returned to the high-mobility valley for transfer on the upswing and NDM to occur. This means that the field must change slowly enough to allow a dump of carriers from the satellite to the central valley. If the field changes too rapidly, too many carriers are retained in the satellite valley and NDM is too weak to sustain steady-state oscillations. In Fig. 41, we plot the maximum frequency of self-excited oscillations as a function of dc bias.

The large variations in field and carrier temperature for both self-excited and large-signal-driven oscillations should result in quantitatively different upper-frequency limits than that obtained for small-signal oscil-

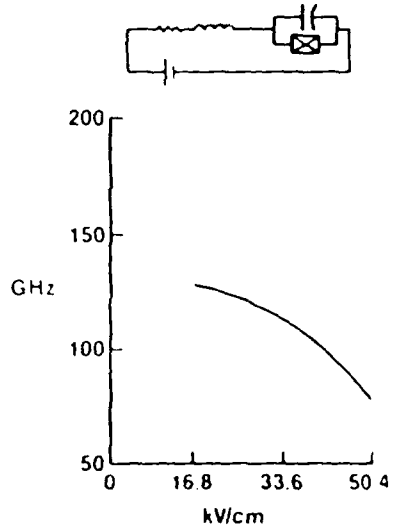


Fig. 41. Maximum frequency for self-excited oscillations

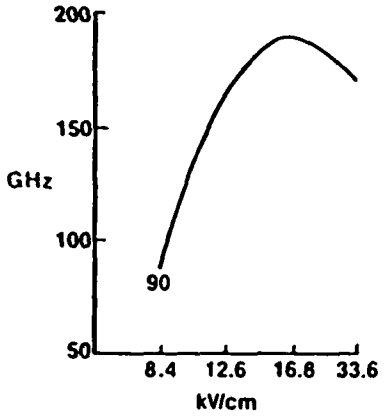


Fig. 42. Maximum frequency for small-signal oscillations

lations. We have examined the latter by perturbing a nonlinear element in steady state. The perturbation is a square wave voltage pulse and the resulting response is then Fourier-analyzed. The results are shown in Fig. 42, where we plot the maximum frequency of small-signal negative conductance as a function of dc bias. The most significant feature here is that f_{max} for small-signal operation is significantly greater than that for large-signal operation. We note that when similar calculations with the nonlinear element driven by an ac source of controlled amplitude and frequency are performed the results bridge the small- and large-signal oscillation calculations.

The explanation for differences in the large- and small-signal results lies in the energy scattering rates (see Fig. 24). For large-signal oscillations, the transient temperature in both the central and satellite valleys oscillates over a larger range than that for the small-signal oscillations and sample lower scattering rates. This gives rise to the reduced maximum frequency for the large-signal oscillation.

Detailed analysis of the frequency limitation of the transferred electron effect in GaAs using the SETEM approach was recently discussed by Rolland *et al.* [75]. They were able to show that reasonable efficiencies could be obtained with frequencies up to 150 GHz and offered the uniform field mode as an alternative means for circumventing the drastic size restrictions usually associated with multimeter wave devices.

4. *Length Dependence of Negative Differential Mobility*

We have been discussing the upper-frequency limit of transferred electron devices from the circuit viewpoint and the transfer and return of electrons between the central and subsidiary valleys. In this analysis we have

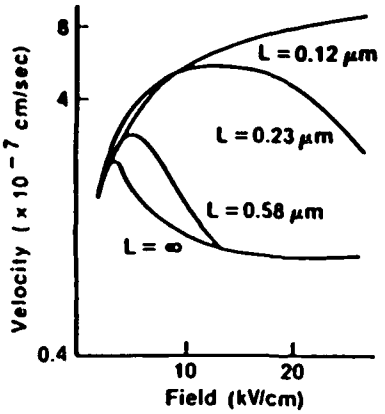


Fig. 43. Length dependence of negative differential mobility.

not explicitly considered the following problem: When an electron enters the active region of a device, it accelerates in the presence of an applied electric field. If the initial drift velocity of the carrier is low, is the transit length sufficient to cause electron transfer and negative differential mobility? The answer lies in earlier calculations. If the carrier experiences a sudden change in field, the mean initial transient ($t \ll$ LO phonon scattering time) will increase approximately linearly with time, followed by a region where v will approach $-e\tau F/m^*$ for a single valley. If the transit time is short enough to prevent significant transfer, NDM will be weak, if at all. Figure 43 summarizes where, for uniform fields, device length is a derived quantity [74,76]. Here

$$L = \int_0 v(t)\, dt. \tag{98}$$

Now the velocity versus field curves and the velocity versus time curves of the type shown here provide an indication of why there is interest in submicron devices. The possibility exists for achieving very high velocities over very short distances. But, again, a word of caution. The calculations of Fig. 42 are for carriers subjected to sudden changes in field. As we have seen, finite rise time dramatically reduces this peak, so the results will be somewhat less important.

G. Ballistic Transport

On the basis of the discussion associated with Fig. 42, there exists an interesting conceptual possibility of a "mean" ballistic transport, where a "typical" electron may "apparently" travel without scattering. (In the

context of semiconductor carrier dynamics, this concept of an electron lucky enough to escape collisions was first suggested by Shockley [77] in studying impact ionization phenomena.)

Although the physical concepts associated with ballistic transport have not been completely discussed, one point should be emphasized. The equations used tend to treat the mean particle as if it were an isolated electron uncorrelated with any other electrons. In actual fact, the correlation among electrons is very important and governs the transient response. As a consequence, ballistic transport based on mean electrons may not be appropriate to semiconductors. With this caveat, we nevertheless discuss some aspect of this transport. We neglect statistics and assume that the number of electrons involved has a monoenergetic distribution and that electron–electron scattering has not contributed to the broadening of the distribution (see Hess [78]). Under these conditions, the following three equations are necessary: the conduction current equation,

$$J = Nev, \tag{99}$$

the Poisson equation, and the equation for electron velocity (conservation of energy),

$$\tfrac{1}{2}m^*v^2(x) - e\phi(x) = \tfrac{1}{2}m^*v^2(0) - e\phi(0), \tag{100}$$

where $v(0)$ corresponds to some initial velocity.

It is clear that these equations yield a set of current–voltage characteristics for any spatially dependent doping level. The simplest case to consider is that for which the charge density injected at the cathode is considerably greater than the background doping density, which is assumed to be uniform. In this case N_0 is ignored, and solutions are borrowed from the analyses of electrical phenomena in gases [79]. The point we shall emphasize here is the role of the cathode on the resulting current–voltage characteristics.

First we take the special case in which the cathode is an inexhaustible source of electrons, i.e., $N(x = 0) \to \infty$. The $v_0 = 0$, and we obtain Child's law (see, e.g., Shur *et al.* [80]),

$$J = (4\epsilon/9)(2e/m)^{1/2}[\phi(x)^{3/2}/x^2]. \tag{101a}$$

(Note: Assuming energy loss by collisions, Eq. (101a) no longer applies. In the simplest case where $v = \mu F$, where μ is constant, the current–voltage relation is the same as that for unipolar flow in dense gases[79]:

$$J = (9\epsilon/8)\mu(\phi(x)^2/x^3), \tag{101b}$$

[Ref. 79] for space-charge-limited currents.

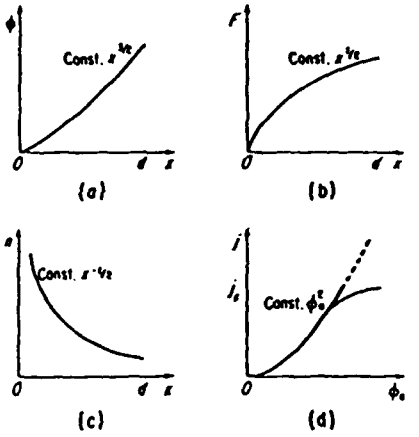


Fig. 44. Characteristics of space-charge-limited current with $v = \mu F$ and zero cathode velocity: (a) voltage versus distance; (b) field versus distance; (c) carrier density versus distance; (d) current versus voltage. Dashed line signifies possible thermionic-limited cathode. See also Papandor [81].

where x is the distance from the source. Figure 44 [81] illustrates the potential and carrier distribution for this case.) In general, we may expect the electrons to be emitted with a finite velocity. In this case, $n(x = 0)$ is finite. Among the consequences of this are that the field at the cathode is no longer zero. Then, if a potential minima ϕ_m occurs at a small distance from the origin, as shown in Fig. 45, for example, beyond this minima the system behaves as if there were a potential difference $\phi_{(L)} + \phi_m$ (see Fig. 45) between the anode and a virtual cathode at X_m. Thus [79],

$$J = \frac{\epsilon}{9}\left(\frac{2e}{m^*}\right)^{1/2}\left(\frac{\phi(L) + \phi_m}{d - X_m}\right)^{3/2}. \tag{102}$$

Since we may expect the cathode velocity and, hence, X_m to depend on the current, the $J(\phi)$ characteristics may be expected to depart significantly from a "$\frac{3}{2}$" relation.

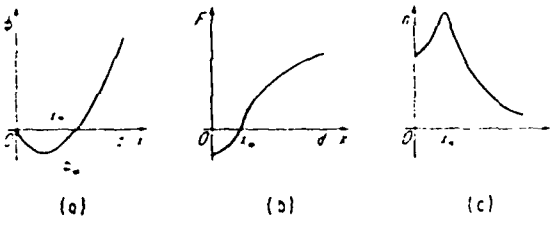


Fig. 45. Space-charge-limited current flow, showing the influence of finite cathode velocity. (a) voltage versus distance, (b) field versus distance; (c) carrier density versus distance. See also Papondar [81]

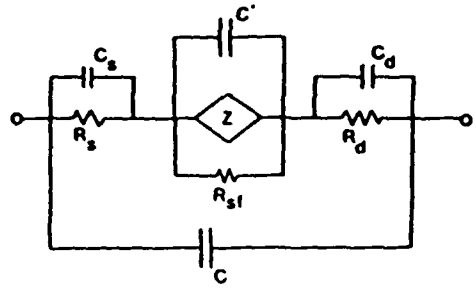


Fig. 46. Suggested simple model for a semiconductor device showing the role played by the contact, space-charge, and transport properties. The parameters are discussed in the text. From Ferry *et al.* [82], with permission.

The situation is even more complicated than just stated. The cathode may be thermionically (see Fig. 44) limited, there may be two-dimensional variation in the cathode structure, the distribution of carriers at the cathode may vary, etc. What this means is that for collisionless transport the $I-V$ characteristics are dominated by boundary conditions and not by ballistics. From a systems viewpoint, the ballistic device has an equivalent circuit like Fig. 46 [82]. Here R_d and C_d define the drain characteristics, R_s and C_s define the source characteristics, C is the total device and package capacitance, C' is a space-charge capacitance, R_{sf} is the resistance due to surface scattering, and Z is the carrier dynamic impedance, which is mostly resistive in long channels but will have an inductive contribution in the ballistic regime due to carrier retardation (see also [83]). Here, however, ac contact impedance may camouflage this contribution. The impedance Z is further complicated in ultrashort channels where transport is likely to be dominated by size-quantization effects [7,84]. It is unlikely that Z can be observed in dc measurements, but the inductive nature should be observable in microwave experiments [83], or time-of-flight [85] techniques can be used to observe the change of $v(F)$ characteristics to the characteristic $\phi^{1/2}$.

Since the possibility of ballistic transport offers interesting device applications, care must be exercised to avoid oversimplifying the criteria for its existence. As indicated, the simplest condition for ballistic transport is that the channel length d be smaller than the bulk mean free path l_b. However, a number of effects conspire to make the problem more complicated. First, if the overall device dimensions are such that $d = l_b$, carriers injected at a nonzero angle to the channel may traverse trajectories whose path length exceeds l_b and will, therefore, scatter. This will be particularly important for wide devices. Second, if the overall device dimensions are less than l_b, some carriers injected at nonzero angles to the channel direc-

tion may intersect the device boundaries and undergo surface elastic or inelastic scattering. Depending on the critical dimension l_b, surface scattering may extend through a significant portion of the device volume. Consequently, l_b should be replaced by l_s, which depends on the finite geometry, the electric field strength, and the surface/environ scattering. Third, if intradevice scattering is negligible, the current flow may be space-charge-limited because of Coulomb scattering. As we have seen, if space-charge exists, it is almost impossible to unfold the ballistic efforts from the space-charge-limited current. A fourth problem relates to devices configured to dimensions comparable to the de Broglie wavelength; carrier reflection effects may then become important and carry an enhanced probability of scattering. In this case we need to examine quantum transport theory (QTT) and apply it to submicron devices. Generally, QTT involves solving the equation of motion of the density matrix. We shall introduce this briefly in Appendix B (p. 290).

IV. QUANTUM TRANSPORT THEORY

A. Introduction

Boltzmann transport theory (BTT) is an ideal theory. It has the twin virtues of conceptual and mathematical simplicity. Quantum transport theory (QTT) [60,86–91] enjoys no such status; it is neither conceptually nor mathematically simple and often reduces to the Boltzmann picture [92,93] only after considerable labor. Even if there were no outward experimental manifestation of quantum transport phenomena, QTT would still be necessary to explain how the phenomenological BTT picture and its related concepts actually arise from the underlying framework of reversible quantum statistical mechanics. Thus QTT is necessary as an explanatory and supportive theory for the Boltzmann picture (where such a picture is applicable), for setting confidence limits for the application of BTT, and for developing the novel concepts and transport kinetics necessary for describing manifest quantum transport phenomena (those effects that depend explicitly on the quantum-mechanical nature of the electron as well as those processes for which the local Boltzmann description fails).

Boltzmann transport theory models the conduction electrons as an approximately independent-particle dilute gas in which the electronic states are nearly stationary and free-electron-like with a well-defined momentum k [92]. Nonstationarity arises from the assumption that perfect

crystal periodicity is violated by imperfections, impurities, and phonons. These sources of crystal imperfection are assumed to cause weak, infrequent scattering of the electrons among the states {k}. The applied electric field serves to accelerate carriers through the momentum states without distorting the states or interfering with the scattering process. It then seems meaningful to describe the carriers by a classical distribution function $f(\mathbf{r},\mathbf{k},t)$ over a vaguely defined phase space [r,k] in which r, the position coordinate for the carrier, is not too well defined in lieu of the uncertainty principle [94].

Usually, one deals with macroscopic systems so that the intuitive concepts of τ_d, the transit time through the channel length; τ, the mean free time between collisions; and τ_c, the atomic duration of a collision, are assumed to satisfy the inequality

$$\tau_c \ll \tau \ll \tau_d. \tag{103}$$

Transport processes are thus conventionally viewed on a coarse-grained time scale $\tau \gg \tau_c$ so that many independent collisions are assumed to occur in the passage of a carrier through a channel length. Moreover, each collision event is treated as an irreversible process that occurs locally in space, locally in time (instantaneously), and independently of any driving fields and other scattering processes. Given these assumptions, the time evolution of the distribution function f is governed by

$$(\partial f/\partial t) + (\partial f/\partial t)_{\text{diff}} + (\partial f/\partial t)_{\text{fields}} = -(\partial f/\partial t)_{\text{coll}}. \tag{104}$$

The left-hand side of Eq. (104) is time-reversible, but the equation overall is irreversible due to the gain–loss structure of the collision integral on the right-hand side.

From the preceding discussion, it is clear that the central concept of classical transport physics is the assumption that a single carrier-distribution function exists that may be used to compute statistical expectation values for macroscopic current flow. In the quantum formulation of transport physics, the concept of a distribution function that depends on position and momentum of the particle is not possible inasmuch as the Heisenberg uncertainty principle precludes simultaneous specification of position and momentum. From the quantum viewpoint, it is therefore necessary to conceptually view phase space as coarse-grained if a distribution is to be regarded as a simultaneous function of carrier momentum and position. The conceptual difficulties introduced by the noncommutativity of the carrier position and momentum operators are not too critical for large devices (channel length greater than 1 μm), for, in this case, it is generally assumed that spatial variations in the distribution function occur over distances that are large compared to the de Broglie wavelength of the carrier.

The temporal development of the distribution function is also an important facet of the transport picture. When considering carrier transport in large devices, the temporal scale of the distribution function is very long compared to the collision duration but very short compared to the total transit time of the carriers, in which case the response of the distribution function to fields and collisions is instantaneous. For medium-sized devices (channel length of about 2500 Å), the temporal scale of the distribution function is of the order of the collision time but short compared to the transit time of the carriers, so that retardation due to the finite duration of collisions becomes possible. For very small devices (channel lengths less than 250 Å), the collision time is of the same order of magnitude as the transit time, so that the applicability of a classical distribution function is questionable.

In order to continue the distribution-function approach, it must be possible to define such a distribution function over suitable momentum, spatial, and temporal variables. The concept of the carrier-distribution function is expected to retain a useful role in medium-sized device transport, but with major modification in the Boltzmann transport equation. However, for "very small device" transport, where the analog of a distribution function certainly exists, an alternative approach to BTT, which emphasizes the role of device environment, size quantization, and fluctuations seems to be necessary. There has been some success in establishing the existence of distribution functions by employing the concepts of the statistical density matrix and the Wigner- density matrix (see Appendix B). In the remainder of this section, the density matrix formalism and its relation to quantum transport will be briefly discussed; an illustrative example of the usefulness of quantum transport will be given for an array of very small devices.

B. Quantum Transport Formulation

Features that are neglected in BTT provide the warning signs for the failure of the semiclassical approach. These features will now be cited and discussed.

(1) *Nonlocality of scattering processes.* Each collision event is actually extended in space and time. If the spatial and temporal variations (described by wavevector q and frequency ω) of the applied driving forces approach the microscopic scale, then collisions will be only partially completed. Normally, BTT assumes $\omega\tau < 1$ and $qL < 1$, where L is the mean free path, i.e., many collisions are completed in one cycle of the applied forces. If, however, $\omega\tau_c \gtrsim 1$ and $q\bar{v}\tau_c \gtrsim 1$ (where τ_c is estimated by $\hbar/E$, E is a characteristic carrier energy, and $\bar{v}\tau_c$ is the de Broglie wave-

length), appreciable quantum effects can arise and the irreversible character of completed collisions will be lost. Interband effects may also occur if $\hbar\omega, \hbar^2q^2/m^* \gtrsim E_b$, where E_b is the vertical energy separation to the band controlling m^*. Very high frequencies correspond to a quantal behavior more reminiscent of optical response [95–97]. The elementary treatment of collisions must also be reconsidered if the mean free time becomes comparable to τ_c; multiple scattering involving at least two scatterers is then possible.

(2) *Strong driving forces.* Once the extended nature of a collision is recognized, it becomes obvious that applied fields can transfer energy and momentum to the carrier during the collision; an interference, or intracollisional, field effect results [63,65,98]. The effect will be very large if $eF v \tau_c = E$. The reverse effect may also occur; scattering can forestall the instantaneous accelerative effect of the driving field (this effect also occurs for low fields). In general, the driving and scattering terms in BTT cannot be independent.

(3) *Strong scattering.* Strong scattering magnifies the previous problems and weakens the assumption that the electronic states are of long lifetime and are free-electron-like. Polaron and cooperative effects are typical consequences.

(4) *Dense systems.* Many-body effects and a single-carrier description fail.

(5) *Small systems.* Size quantization or surface-limited transport effects become important [99]. Ultimately, the condition implicit in Eq. (103) breaks down.

(6) *Nonclassical influence of driving fields.* Sufficiently strong electric and magnetic fields lead to Stark or Landau quantization of the electronic states [99].

These features and the concepts discussed therein are intuitive but can be given precise meaning within QTT, to which we shall now turn.

Quantum transport theory is generally based on the Liouville–von Neumann equation for the statistical density matrix $\rho(t)$, which is given as

$$i\,\partial\rho(t)/\partial t = \dot{H}_F\rho(t), \tag{105}$$

$$\dot{H}_F\rho \equiv [H_F,\rho] \equiv H_F\rho - \rho H_F, \tag{106}$$

where $H_F = H - \mathcal{F}$, and the Hamiltonian H describes the full system in the absence of the coupling $\mathcal{F}$ to the externally applied driving forces. The usual boundary condition is that $\rho = \rho_0(H)$ for $t < 0$, where ρ_0 is a thermal equilibrium solution (e.g., the grand canonical density matrix). The driving perturbation is initiated at $t = 0$. This starting point needs modification for the description of small systems embedded in an interac-

tive environment [59]; this aspect of QTT will be discussed later in this section. (Note: The units are chosen so that $\hbar = 1$ throughout.)

All observable properties, e.g., the current or charge densities, labeled generically by J_t may be evaluated as quantum-statistical expectation values determined by $\rho(t)$ as

$$\langle J_t(t)\rangle = \mathrm{Tr}[J_t\rho(t)] \equiv \sum_{\lambda,\lambda'} \langle\lambda|J_t|\lambda'\rangle\langle\lambda'|\rho|\lambda\rangle = \sum_{\lambda,\lambda'} J_t^{\lambda\lambda'}\rho_{\lambda'\lambda}. \tag{107}$$

Here $\{|\lambda\rangle\}$ is any complete set of states. Usually, the Hamiltonian H_F is partitioned into "free" carrier, "free" scatterer, carrier–scatterer interaction, and driving force components as

$$H_F = H_e + H_s + V + \mathcal{F} \equiv H_0 + V + \mathcal{F}. \tag{108}$$

This partitioning is not unique. The component H_e might describe small polaron states by incorporating part of the electron–phonon interaction, or H_e might include the coupling to a magnetic field and describe Landau states. The basis states $|\lambda\rangle$ are then chosen to diagonalize $H_0 = H_e + H_s$, usually via $|\lambda\rangle \equiv |e\rangle|S\rangle$, where $\{|e\rangle\}$, $\{|s\rangle\}$ diagonalize H_e and H_s, respectively. The choice of representation $\{|\lambda\rangle\}$ decides the character and interpretation of the subsequent transport theory.

If the current–density operator J depends only on electronic variables and commutes with H_e (which is true for extended-state, homogeneous transport in zero magnetic fields), the observable response is

$$\langle \mathbf{J}(t)\rangle = \sum \mathbf{J}^{ec} f(e). \tag{109}$$

where

$$f(e) = \langle e|\sum_s \langle s|\rho|s\rangle|e\rangle \equiv Tr_s\langle e|\rho|e\rangle \equiv \langle\langle e|\rho|e\rangle\rangle_s \tag{110}$$

defines a real, time-dependent, generalized, electron-distribution function over the free-carrier states $\{|e\rangle\}$. A transport equation for $f(e)$ may then be constructable from the Liouville equation. It may turn out to have Boltzmann-like form, although the quantum nature of the states $|e\rangle$ will be reflected in the detailed forms for the collision rates. However, for inhomogeneous transport and for transport in quantizing magnetic fields, for example, J is not necessarily diagonal and we must consider the off-diagonal matrix elements of the electron density matrix $f = \langle\rho\rangle$. Various methods exist [99] for expressing $f(e)$ in terms of $f(e,e')$, but the subsequent transport theory will not have a Boltzmann-like form. In gen-

eral, a closed equation of motion for $f(t)$ can only be obtained for the special case of independent carrier transport in a stationary scattering system (i.e., where the scatterers remain in thermal equilibrium at all times). This situation has been extensively studied for homogeneous, nonlinear transport.

Wigner [100] has shown that QTT can get quite close to the classical concept of a phase-space distribution function (see Appendix B). The Wigner one-electron distribution function, for example, is defined for free-carrier states by

$$\begin{aligned} f_\sigma(\mathbf{k},\mathbf{r},t) &\equiv \int d^3\mathbf{y}\, \exp(-i\mathbf{k}\cdot\mathbf{y})\, \mathrm{Tr}\{\rho(t)\psi_\sigma^\ddagger(\mathbf{r}-\tfrac{1}{2}\mathbf{y})\psi_\sigma(\mathbf{r}+\tfrac{1}{2}\mathbf{y})\} \\ &\equiv \int d^3\mathbf{y}\, \exp(-i\mathbf{k}\cdot\mathbf{y}) f_\sigma(\mathbf{r},\mathbf{y},t), \end{aligned} \tag{111}$$

where $\psi_\sigma^\ddagger(\mathbf{r})$ and $\psi_\sigma(\mathbf{r})$ are the second quantized creation and annihilation operators, respectively, for a carrier of spin σ at location $\mathbf{r}$. Here ρ is second quantized and the trace is a many-body trace. Similarly, a phonon Wigner distribution may be defined by

$$N_\alpha(\mathbf{k},\mathbf{r},t) = \sum_{\mathbf{K}} \exp i\mathbf{K}\cdot\mathbf{r}\langle b_\alpha^\ddagger(\mathbf{k}-\tfrac{1}{2}\mathbf{K}) b_\alpha(\mathbf{k}+\tfrac{1}{2}\mathbf{K})\rangle \tag{112}$$

where $\langle\cdots\rangle \equiv \mathrm{Tr}(\ldots)$ and $b_\alpha^\ddagger$ and b_α are creation or annihilation operators for a phonon of type α and momentum $\mathbf{k}$. The Wigner construction utilizes a partial, generalized, Fourier transform over the full (off-diagonal) density matrix and is easily generalized to other basis states (e.g., Bloch states or Landau states). One may easily prove that f_σ and N_α are real-valued, generalized distributions that give the correct statistical expectation values, e.g., the carrier density and current density in an inhomogeneous system are given by

$$\langle n(\mathbf{r},t)\rangle = \sum_\sigma \int \frac{d^3\mathbf{k}}{(2\pi)^3} f_\sigma(\mathbf{k},\mathbf{r},t), \tag{113}$$

$$\langle \mathbf{J}(\mathbf{r},t)\rangle = \sum_\sigma \int \frac{d^3\mathbf{k}}{(2\pi)^3} e\mathbf{v}(\mathbf{k}) f_\sigma(\mathbf{k},\mathbf{r},t), \tag{114}$$

where $\mathbf{v}(\mathbf{k}) \equiv \nabla_{\mathbf{k}} E(\mathbf{k})$ is the (c-number) group velocity of the electron momentum state $|\mathbf{k}\rangle$. Homogeneous systems are translationally invariant [$f_\sigma(\mathbf{k},\mathbf{r})$ is independent of $\mathbf{r}$], which implies that

$$f(\mathbf{k},\mathbf{K},t) \equiv \int \exp(-i\mathbf{K}\cdot\mathbf{R}) f_\sigma(\mathbf{k},\mathbf{R}) d^3\mathbf{R}$$

is independent of $\mathbf{K}$ and the equivalent electron-density matrix is diagonal in momentum space.

The are some difficulties with interpreting Wigner distributions as probability densities; they are not necessarily positive-definite (usually a sign of strong quantum interference effects). However, the Wigner method does allow one to construct a gauge-invariant transport theory [97]. For independent carriers, f_σ reduces to $\langle r - \frac{1}{2}y|\langle\rho\rangle_s|r + \frac{1}{2}y\rangle$ where ρ is now a functional of the one-electron Hamiltonian. Quantum transport theory has been extensively developed for this case [88].

We shall now sketch the general features of QTT. The electron-density matrix $f(t)$ is determined from the full-density matrix by $f(t) = \mathrm{Tr}_s[\rho(t)]$. For the case of stationary phonon and impurity distributions, we factorize approximately the initial thermal-equilibrium density matrix as

$$\rho(t = 0) = f_0(H_e + V)\Omega_s(H_s),$$

where f_0 will be taken as the Maxwellian equilibrium form and Ω_s describes the equilibrium distribution of scatterers. Let us now Laplace-transform the Liouville equation (105), rearrange the terms using projection calculus, and then retransform back to the time domain to obtain a general master equation in the form

$$\frac{\partial f}{\partial t} + i[H_e f] + i[\mathcal{F} f] = i\int_0^t d\tau\, \hat{C}_F(\tau) f(t - \tau) + M_F(t). \qquad (115)$$

The left-hand side of (115) describes the collision-free diffusion and acceleration of the carriers. For homogeneous systems, f is a function of momentum only and if $H_e = p^2/2m^*$, the term $[H_e, f]$ vanishes (for inhomogeneous transport it gives rise to a term $\mathbf{v}(\mathbf{k})\,\partial f/\partial \mathbf{r}$; $[H_e, f]$ is also nonvanishing if H_e includes coupling to a quantizing magnetic field). With our previous model assumption, the coupling to the electric field is simply $\mathcal{F} \equiv -eF \cdot r$, and in the momentum representation, $i[\mathcal{F} f]$ reduces to $eF\,\partial f/\partial \mathbf{k}$. The right-hand side is proportional to the scattering interaction and describes collision effects via C_F and memory effects via M_F.

Boltzmann transport theory may be exactly recovered under the following conditions [63]: (1) weak, infrequent scattering; (2) point collisions; (3) translational invariance of the scattering system; and (4) asymptotic time scale $t \gg \tau_c$ [actually related to (104)]. Following these assumptions, it can be shown that the right-hand side of Eq. (115) reduces to

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \sum_{\mathbf{k}'} [R(\mathbf{k},\mathbf{k}')f(\mathbf{k}') - R(\mathbf{k}',\mathbf{k})f(\mathbf{k})]. \qquad (116)$$

where the $R(\mathbf{k},\mathbf{k}')$ are the usual second-order perturbation theory scattering rates. Thus, under these conditions, the master equation of Eq. (115) reduces to the Boltzmann equation.

The master equation of Eq. (115) contains a memory term $M_F(t)$, which

is one consequence of the interference between the electric field and the scattering processes and is a function of the initial equilibrium state $f_0(H_e + V)$. It represents the correction to the otherwise instantaneous accelerative effect of the field due to the field having to break up the correlations in the electron states induced by scattering processes. Indeed, M_F may be interpreted as a renormalization of the driving-force term in the kinetic equation [99], because its nonvanishing part is proportional to F.

The influence of the electric field within a collision event has been called an intracollisional field effect (ICFE). The mathematical details of the ICFE have previously been given [88], so we shall not go into detail here. Two major modifications of the scattering integral occur as a result of this intracollisional process. First, the total energy-conserving δ function is broadened by the presence of the electric field. Second, the threshold energy required for the emission of an optical phonon is modified, which causes an energy shift of the δ function. This latter process is easily understood in physical terms. The argument of the energy-conserving δ function is just

$$E_f - E_i \pm \hbar\omega_0 = E(\mathbf{p}_f) - E(\mathbf{p}_i) \pm \hbar\omega_0, \tag{117}$$

but the initial and final momenta evolve during the collision as

$$\mathbf{p}(t') = \mathbf{p} - \int_{t'}^{t} eF(t'')\,dt'', \tag{118a}$$

$$\mathbf{p}'(t') = \mathbf{p}' - \int_{t'}^{t} eF(t'')\,dt''. \tag{118b}$$

In the emission of an optical phonon, where the electron is scattered against the electric field, the field will absorb a portion of the electron energy during the collision, and, hence, a reduction in energy loss to the lattice will be favored. The opposite effect, an enhancement in energy to the lattice, occurs for emission along the electric field.

C. Synergetic Effects from Device–Device Interactions

Preliminary theoretical studies [59] of the quantum-mechanical operation of an array of very small devices suggest that synergetic effects are possible when the individual feature size decreases below 1000 Å. In this case, novel device possibilities become available.

Let us summarize some previous findings [59]. Small semiconductor devices (<0.1-μm channel length) are controlled by: short spatial scales, very fast temporal response, very high fields (≤ 600 kV/cm), high carrier densities ($\leq 10^{18}$ cm^3), and strong size-related effects (coupling to the

environment of contacts, interfaces/boundaries/surfaces, interconnects, and other devices). The ultrafast transit times ($\lesssim 1-2$ psec) within any one device plus the collision-quenching intracollisional field effect preclude any significant dissipation of energy within the device volume; traditionally, this is the so-called ballistic regime). However, studies of the device-density matrix equations [59], which take into account coupling to the device environment, show that dissipation will predominantly occur over the extended region surrounding the device. This mechanism admits a second-order device–device correlations interaction, provided the spatial extent of interconnect regions becomes comparable to the de Broglie wavelengths (indeed, for metallized interconnect regions on the order of 100 Å, de Broglie waveguide modes should dominate the current flow). By using projection calculus methods on the full system-density matrix equations, it has been shown that the interaction of a particular device with its host VLSI system may be classified into coherent (time-reversible) and incoherent (dissipative, irreversible) components. The coherent effects include state renormalization and size quantization arising from the short-range interaction with the regular part of the finite device boundaries; these effects are enhanced by very high, inhomogeneous, controlling electric fields within the device volume. There is also a coherent long-range device interaction with the environment, which stems from effects of device replication and gives rise, for example, to super-lattice phenomena and the consequent overriding of the bulk intradevice carrier dynamics. The latter have already been observed in intercalated structures prepared by molecular-beam epitaxy [102]. The incoherent processes include surface–interface roughness scattering, surface phonons and plasmons, long-range ($\lesssim 1$ μm) device electron, insulator phonon scattering, and phonon-mediated electron–electron interdevice scattering.

The joint action of many subsystems so as to produce structure and functioning on the full-system scale are well known in physical science and are characteristic of nonlinear systems [103]. In recent years, considerable attention has been devoted to so-called dissipative structures [104]: a class of spatially inhomogeneous, ordered structures in which order may be created spontaneously in open systems far from equilibrium and which obey specific nonlinear kinetic (transport) laws. For such nonequilibrium structures, stability is not self-sustaining but is maintained by a continuous exchange of energy and matter with the surroundings. Biological systems, chemically reacting mixtures under open-system conditions, and Benard cell phenomena provide well-known examples.

Comparably complex signal-processing VLSI systems may also support synergetic phenomena as the feature size and complexity approach

that of natural systems. Analogies with many-body phase-transition and synergetic theory lead us to expect strong qualitative differences in stability between sequential and concurrent (array) processing signal systems. The argument is simple; we know from model studies that one-dimensional systems cannot support phase transitions; but, as soon as two- and three-dimensional cross-interactions are introduced, systems may condense or "lock into" ordered macrostructures and make transitions between them. Concurrency has a strong equivalency to multidimensionality, which is, in turn, a prerequisite for cooperative phenomena that are stable against local fluctuations. All the necessary features for synergetic phenomena in VLSI structures appear to exist when feature sizes approach a few hundred angströms, such as in multidevice slaving interactions, nonlinear slaving of intradevice variables by self-consistent control fields, input-current open-system operation.

Of course, device–environment interactions are not unknown even in present scale systems. For example, gated logic systems and programmable logic arrays have device–environment control exercised via the interconnect matrix. A true device–device interaction also appears in large MOS memory chips (near-cell interference in READ–WRITE situations), which is currently treated as a reliability problem rather than an effect on which to capitalize. Of course, capacitative coupling between devices is also very familiar. The true device–environment coupling envisaged here, however, is only possible on almost atomically small scales for which superlattice effects and dissipative coupling become possible.

The main theoretical tools to explore synergetic VLSI already exist: nonlinear quantum transport theory, synergetic theory, renormalization group theory, etc. However, applications will require (a) knowledge of the intended function and skeletal VLSI structure, (b) characterization of the relevant interdevice coupling, and (c) selection of the appropriate control fields and currents. These do not as yet exist and must be the target for future research. We have, however, sketched the outline for such a theory using many-body compaction techniques [105]. Such sophistication is not necessary to understand the qualitative role of these synergetic effects. We can illustrate the basic principles by a simple circuit-theoretic analogy utilizing component–connection-type approaches. First, we examine a special case of isolated devices. Then, we introduce a connection function to describe the system in terms of the devices and show how the properties of the connection function can alter the system dynamics. As the individual device dynamics and connections will be nonlinear, we expect that, although the equations used here are linear, the general nonlinear results will admit of synergetic responses for the system. Thus. Thus, for example, a regular, replicated device structure in the environ-

ment is expected to give a superlattice modulation of the individual device dynamics. Now, "superlattice" in the normal sense is usually used to refer to multiple thin layers of different materials in which the layers are thinner than the electron wavelength so that the atomic potential variation between layers introduces minigaps, which represent a renormalization of the energy spectrum of the electrons within a single material. Here, however, we emphasize the device–device or device–environment coupling, which can be an essentially time-reversible interaction [59], to renormalize the functional behavior of a single device within the array and to change the system dynamics of the array itself. In particular, if the spatial extent of interconnected regions becomes comparable to the range of possible device–device interactions, a second-order device–device correlation or interaction arises, from which the coherent long-range component admits to possible functional superlattice phenomena. In these cases, in very small devices, the bulk intradevice dynamics may be of secondary importance to the system dynamics taken as a whole.

In particular, though, it is apparent that the system equations are functionally similar to those utilized for nonlinear structures, [104], and for such nonequilibrium structures, stability is not self-sustaining but is maintained by a continuous exchange of energy with the surroundings. Comparably complex signal-processing VLSI systems may also support synergetic phenomena as the feature size and complexity approach that of natural systems. Analogies with solid-state and synergetic theory therefore lead us to expect qualitative differences in stability between sequential and concurrent (array) signal-processing systems.

We examine first the state equations for an isolated integrator with input/output conditioning and then examine the applicability of the example. The state equations are then, for the ith device,

$$\dot{u}_i = a_i u_i + b_i y_i, \tag{119a}$$

$$z_i = c_i u_i, \tag{119b}$$

where u_i is the state variable and y_i and z_i the input and output variables, respectively. For an ensemble of N devices, these become

$$\dot{\mathbf{U}} = \mathbf{AU} + \mathbf{BY}, \tag{120a}$$

$$\mathbf{Z} = \mathbf{CU}, \tag{120b}$$

where **A**, **B**, **C** are square diagonal matrices and **U**, **Y**, **Z** column matrices. Solving for the transfer function gives (in the Laplace transform domain with relaxed initial state)

$$\mathbf{Z} = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{BY}. \tag{121}$$

So far, we have considered that each device was isolated from the others. If we describe the system input and output as G and H, respectively, we can describe the connection matrix F through the following [106]:

$$\mathbf{Y} = \mathbf{FZ} + \mathbf{LG}, \tag{122a}$$

$$\mathbf{H} = \mathbf{MZ}. \tag{122b}$$

The connection matrix F describes generally how the input of a particular device is related to the outputs of other devices. We can combine Eqs. (121) and (122) to yield the system transfer function as

$$\mathbf{H} = \mathbf{MCB}(s\mathbf{I} - \mathbf{A} - \mathbf{FCB})^{-1}\,\mathbf{LG}, \tag{123}$$

where we have used the fact that A and, hence, $(s\mathbf{I} - \mathbf{A})^{-1}$ are diagonal.

Equation (123) for the system-transfer function is a special case of the connection-function theory of systems [106] applicable to the integrator. Although we have used this special case, the approach is far more general and is applicable to arbitrary circuits. Furthermore, even though we have assumed an analog signal approach by employing the Laplace transformation for the time variation, the technique is extendible to the class of digital circuits known as linear sequential circuits through the description of system dynamics in an abstract extension field [107]. However, this simple case is adequate to illustrate the major points just discussed.

The quantity $(s\mathbf{I} - \mathbf{A} - \mathbf{FCB})^{-1} = \mathbf{S}^{-1}$ plays the conceptual role of a resolvent for the system and the zeros of det(S) define the various modes of operation. Since B and C are diagonal, any deviation of the system response from that defined by A must arise through the structure of F. For example, if we consider that the system is logically connected, i.e., y_i is connected only to $z_i, j < i$, then F has elements only in the lower triangle below the main diagonal. Since A is diagonal, F does not modify the modes determined by A, i.e., $\det(\mathbf{S}) = \det(s\mathbf{I} - \mathbf{A})$. Only when F has entries across the main diagonal does this change. For example, if $y_i = z_{i-1}$, then F has entries along the diagonal just below the main diagonal. If, however, the last stage is fed back to the first, an entry appears in the upper right corner of F and one new mode is generated, the collective ring-oscillator mode.

In general, the connection function F can be divided into two parts, $\mathbf{F}_1$ and $\mathbf{F}_2$, where $\mathbf{F}_1$ is the portion of F that represents the desired metallizations, i.e., the designed architectural circuit yielding

$$\mathbf{S}_1 = s\mathbf{I} - \mathbf{A} - \mathbf{F}_1\mathbf{CB}. \tag{124}$$

Then, $\mathbf{F}_2$ represents the parasitic interactions (the parasitic device–device couplings) that arise from the line-to-line coupling capacitance, for ex-

ample. Thus, a new resolvent S_2,

$$S_2 = S_1 - F_2CB, \tag{125}$$

arises with a new set of eigenmodes given by $\det(S_2)$. Thus, the structure of the system is altered in the presence of F_2. As F_2 depends on the states of U (voltages, for example) as well as the inputs G, it is entirely conceivable that the system is now strongly nonlinear. In large-scale systems, where sizes are more than 1 μm in scale, F_2 may reasonably be assumed to be negligible. In future VLSI and ULSI systems of submicron dimensions, this is no longer the case, and the presence of F_2 will have to be accounted for in design.

It is clear from the preceding discussion that the structure of the connection function is instrumental in determining the collective modes of system operation. Moreover, the connections need not be the deliberately wired interactions but must include the parametric device–device interactions that can occur in arrays of small devices. Such interactions could occur, for example, from capacitative coupling, charge spill-over, or potential barrier lowering, such as is present in subthreshold currents due to drain-induced barrier lowering [108]. Finally, if the connection matrix F is functionally dependent on the state variables or control signals, i.e., $F = F(U,G)$, then a nonlinear interaction is possible that can lead to synergetic restructuring of the system function. It can be stated that the general results presented in this simple example have been known for some time. However, the importance of device–device correlation–interaction in dense arrays lies in the role this essentially parasitic long-range interaction can play in restructuring F and hence the system dynamics. Essentially, F can be split into design and parasitic fractions. This is the principal point of the present discussion; changes in F due to device–device interactions can lead to a restructuring of F and, therefore, to a restructuring of the entire system dynamics. Since these interactions are principally expected to be nonlinear, synergetic effects can be expected in the system dynamics.

V. DIFFUSION

A. Introduction

One of the most fundamental parameters required for modeling semiconductor devices is the diffusion coefficient $D(F,\omega)$, where F is the electric field and ω the frequency. Not only is the diffusion coefficient neces-

sary for evaluating operating and high-frequency characteristics, it provides a fundamental characterization of velocity fluctuations in the system and their contribution to noise in the device [111,112]. If diffusion is relatively well understood for low fields, this situation does not carry over to the case of high electric fields [112]. The general case for high-field transport in semiconductors differs in that relaxation of the velocity fluctuations is to a nonequilibrium steady state [113–115], and the process is nonlinear [116,117].

Diffusion is a general result of the Brownian motion of the carriers, and early work centered on the calculation of the mean square displacement of the carriers. This led to the Einstein relation for a free particle [118]:

$$D = \lim_{t\to\infty} \langle(\Delta x)^2\rangle/2t. \tag{126}$$

This form has been utilized considerably in studies of Brownian motion in many systems, as is apparent from the many review articles and original papers [56; 60; 111, p. 415; 113; 119–125]. On the other hand, it has also been suggested that the mean square displacement is related to D through the derivative

$$D = \tfrac{1}{2}\, d/dt\langle(\Delta x)^2\rangle, \tag{127}$$

and this form has also found widespread use [114,126,127]. In the steady state, however, in both equilibrium and nonequilibrium situations, $\langle(\Delta x)^2\rangle \simeq t$ for long times. Thus, both Eq. (126) and (127) give the same result for D [128].[7] In this regime, we also overlook the long-time tails that are observed in some hydrodynamic systems [127,128]. For short times, however, differences can arise between Eqs. (126) and (127). We shall return to this later.

B. Diffusion Formalism

We shall begin with a brief review of the formal theory of generalized diffusion and mobility. For the present purposes, we shall neglect the influence of magnetic fields and assume that the electric field $F(x,t)$ is a slowly varying function of position when compared to the spatial extant of the de Broglie wavelength or to the radius of the collision sphere [88]. Transport in a medium that is macroscopically homogeneous with respect to scattering centers is then described by a kinetic transport equation of the form

$$\left[\frac{\partial}{\partial t} + \mathbf{v}\cdot\frac{\partial}{\partial \mathbf{x}} + \mathcal{F}\cdot\frac{\partial}{\partial \mathbf{v}}\right] f(\mathbf{v},\mathbf{x},t) = -\hat{C}f(\mathbf{v},\mathbf{x},t), \tag{128}$$

[7] This is not a totally general result, since non-Gaussian, non-Markovian diffusion is shown in some disordered systems

where $\mathcal{F} = e\mathbf{F}/m^*$. For Boltzmann–Bloch transport, f is the usual distribution function defined over velocity–position phase space (we are assuming a simple, parabolic, isotropic band so that v is interchangeable with $\hbar k/m^*$, although the Monte Carlo calculations of the next section assume the proper nonparabolic band structure) and $\hat{C}$ is the collision operator, which is independent of space, time, and electric field but is velocity dependent. If the field F is a rapidly varying field of very short wavelength, the driving term $\mathcal{F} \cdot \partial f/\partial \mathbf{v}$ requires modification [88, p. 126; 96].

The first integral of Eq. (128) yields the particle-continuity equation

$$\frac{\partial n(x,t)}{\partial t} = -\nabla \cdot \left[\int d^3v \, \mathbf{v} f(\mathbf{v},\mathbf{x},t) \right], \tag{129}$$

where

$$n(x,t) = \int d^3v \, f(\mathbf{v},\mathbf{x},t). \tag{130}$$

Equation (129) may be given the *form* of the traditional phenomenological semiconductor equation by employing the formal causal solution to (128) using the collision operator to form an integrating factor. The result is

$$f(\mathbf{v},\mathbf{x},t) = f_{\text{tr}} - \int_0^t d\tau \exp\left[-\int_\tau^t d\tau' \, \hat{C}(\tau') \right] \times \left[\mathbf{v} \cdot \nabla + \mathcal{F} \cdot \frac{\partial}{\partial \mathbf{v}} \right] f(\mathbf{v},\mathbf{x},\tau), \tag{131}$$

where the transient term is

$$f_{\text{tr}} = \exp\left[-\int_0^t d\tau \, \hat{C}(\tau) \right] f(\mathbf{v},\mathbf{x},0). \tag{132}$$

The transient term will be ignored for now, but we should pointed out that it will have an important role in transport on short time scales, such as for ballistic transport or in very short channel devices. Inserting Eq. (131) into (129) gives, for the classical case,

$$\partial n(\mathbf{x},t)/\partial t = -\nabla \cdot \mathbf{J}(\mathbf{x},t), \tag{133}$$

$$J_\mu(\mathbf{x},t) = -\partial[n(\mathbf{x},t)D_{\mu\nu}(\mathbf{x},t)]/\partial x_\nu - n(\mathbf{x},t)\mu_{\mu\nu}(\mathbf{x},t)F_\nu(\mathbf{x},t), \tag{134}$$

where the generalized diffusion and mobility terms are defined by

$$D_{\mu\nu}(\mathbf{x},t) = \frac{1}{n} \int_0^t d\tau \int d^3v \, v_\mu \exp[-(t-\tau)\hat{C}] v_\nu f(\mathbf{v},\mathbf{x},\tau), \tag{135}$$

$$\mu_{\mu\nu}(\mathbf{x},t) = \frac{e}{m^* n} \int_0^t d\tau \int d^3v \, v_\mu \exp[-(t-\tau)\hat{C}] \left(-\frac{\partial f}{\partial v_\nu} \right). \tag{136}$$

Although these results superficially appear reminiscent of classical diffusion and Doob's theorem [130], Eqs. (131)–(136) provide a rigorous comprehensive formulation of the semiconductor equations for the description of diffusion and device modeling. It should be noted that, in general, the generalized mobility and diffusion do not admit of an Einstein relation [117], except for systems close to thermal equilibrium. Because of the complexity of the collision operator, or, more properly, the resolvent $(\hat{C} + s)^{-1}$ (where s is the Laplace transform variable), it transpires that expressions (135) and (136) are not trivial to evaluate, *even if* the exact solution of $f(\mathbf{v},\mathbf{x},t)$ is known. As with other nonlinear systems, expansions can be made in order to approximate low-order corrections to the Fokker–Planck equation [116]. The quantum case is similar but involves nonlocal diffusion and mobility densities. It will be discussed elsewhere.

Equations (135) and (136) are formally of the correlation-function-type and constitute a more proper definition of the diffusion coefficient itself [60,137]. For longitudinal diffusion, it may then be shown that

$$D(t) = \int_0^t d\tau\, \Phi'(\tau), \tag{137}$$

where $\Phi'(\tau)$ is the reduced velocity-autocorrelation function. On the other hand, we can multiply Eq. (133) by x^2 and integrate over all space to obtain an expression for $D(t)$. This relation is then conveniently compared with Eq. (137) in the Laplace domain to yield

$$\Phi'(s) = s \int_{-\infty}^{\infty} dv\, v[\hat{C} + s]^{-1} v f(v,s)/N, \tag{138}$$

where $f(v,t)$ is the velocity-distribution function (normalized to N) that satisfies

$$\left[\frac{\partial}{\partial t} + \mathcal{F}\frac{\partial}{\partial v}\right] f(v,t) = -\hat{C} f(v,t), \tag{139}$$

and v the reduced velocity. Equation (138) relates $\Phi'(s)$ to $sD(s)$, the longitudinal diffusion coefficient as defined by Eq. (135). The appearance of the resolvent $(\hat{C} - s)^{-1}$ is noteworthy.

C. Correlation Functions for Hot Electrons

The fluctuation response is, in general, complicated because of the many physical processes involved, but the velocity fluctuation can be considered as having two main contributions: $v' = v(t) - \langle v \rangle = v'' + u'$. The first of these, u', is the velocity fluctuation arising from a fluctuation in carrier energy: $u' = u(E + \Delta E) - u(E)$ and the second, v'', arises from

velocity fluctuations about u' [117,132]. These various factors can be observed by studying, not the diffusion coefficient itself, but rather the velocity autocorrelation function $\phi'(t)$, which is the inverse Fourier-cosine transform of $D(F,\omega)$. If we define $\phi'(t)$ as

$$\phi'(t) = \langle [v(t + t_0) - \langle v \rangle][v(t_0) - \langle v \rangle] \rangle, \tag{140}$$

then it is found that, for high electric fields, $\phi'(t)$ decreases initially as an exponential, becomes negative, passes through a minimum, and relaxes finally to zero [117]. This process is basically related to the fact that, in general, energy relaxation is slower than momentum relaxation; and this behavior can be expected to occur via the same processes that lead, for example, to velocity overshoot [133]. Such general behavior was observed in the recent work of Fauquemberque *et al.* [112] but was not adequately explained. We should remark here that such behavior for $\phi'(t)$ is also found generally in hydrodynamic systems [129]. The detailed behavior of $\phi'(t)$ assumes more than academic interest as semiconductor devices begin to assume submicron dimensions. In Si, for example, the time duration of $\phi'(t)$ can be of the order of 1 psec, so as channel lengths drop below, say, 0.1 μm, correlated electron motion and enhanced noise in the devices can be expected to occur.

In general, the diffusion coefficient $D(\omega,F)$ depends on the velocity fluctuations in the electron system and is related to the noise spectral density $S_v(\omega)$ associated with these fluctuations. These are related as (for longitudinal diffusion)

$$\begin{aligned} D(\omega,F) = \frac{S_v(\omega)}{4} &= \int_0^\infty \phi'(\tau) \cos(\omega\tau)\, d\tau \\ &= \int_0^\infty \langle [v(\tau + t) - \langle v \rangle][v(t) - \langle v \rangle] \rangle \cos(\omega\tau)\, d\tau, \end{aligned} \tag{141}$$

where all velocities are understood to be longitudinal. The principal difficulty in calculating transport parameters, particularly $\phi'(t)$, in these systems lies in the complicated energy dependence of the many scattering processes. In the past few years, however, ensemble Monte Carlo techniques have been developed that can be used to calculate these transport parameters with high resolution. As developed by Lebowhol and Price [134] and subsequently used by Ferry and Barker [53,135], the ensemble Monte Carlo technique is a hybrid method in which an ensemble of electrons is adopted. This ensemble is composed of N electrons, with variables $\{R_i\}$, $i = 1,2, \ldots, N$, where the set $R_i = \{k_i, x_i, \ldots\}$ includes all necessary descriptors of each electron's state. At each time step, all R_i are calculated by a Monte Carlo process, and the set $[R_i]$ is treated as an ensemble evolving in time. The ensemble Monte Carlo method has advan-

tages over the normal Monte Carlo technique in that an ensemble-distribution function exists and evolves with $\{R_i\}$. Variables such as velocity or position are calculated from an ensemble average over $\{R_i\}$ at each time step and the variance is controlled by a sufficiently large value for N. One should be aware, however, of the vagaries of stochastic simulations on a computer, and we have used a variety of fields, number of electrons, and seeds for the random number generators without affecting these results. In the calculations reported here, a value of $N = 2500$ was used. This value is sufficiently large to give high-resolution results for the transient dynamic response of the electrons to a high electric field [55,135], for example. Thus the method is capable of yielding good results for the transport characteristics.

The ensemble Monte Carlo method was used to calculate the correlation function in Si for the total velocity, $\phi(t) = \langle v(t + t_0)v(t_0)\rangle = \phi'(t) + \langle v\rangle^2$ (all calculations shown in the figures have $\phi(t)$ normalized to $\langle v^2\rangle$).[8] The ensemble of electrons was initialized as a Maxwellian at the lattice temperature 300 K and was assumed to reside at $x = 0$ at $t = 0$. A homogeneous electric field was applied at $t = 0$, and the ensemble allowed to evolve in time. After a reasonable period of time, the ensemble was in pseudoequilibrium with the field and had a steady drift velocity.[9] After this pseudoequilibrium was achieved, the longitudinal velocity autocorrelation function $\phi(t) = \langle v(t_0 + t)v(t_0)\rangle$ was calculated for several initial times t_0. The stationarity of the system was, therefore, verified as well, and the averaging process was carried out over the ensemble as well as over various initial times. In Fig. 47 is shown the variation of $\phi(t)$ as a function of time t, for several values of the electric field. The initial fall of $\phi(t)$ is primarily due to momentum relaxation, with the local minimum and subsequent rise due to energy relaxation as suggested by Price [117]. The error bars indicate the spread of data points from the calculations and averaging procedures.

The lowest field in Fig. 47, 10 kV/cm, lies below the knee of the velocity–field curve and does not really correspond to hot electrons. In this case, there is no hint of a negative-going portion [where $\phi(t) << \langle v\rangle^2$, the steady-state result]. There is evidence, however, of a tailing behavior

[8] The use of $\phi(t)$, rather than the more normal $\phi'(t)$, was adopted as this allows log–log or semilog plots to be used without zero-crossing complications. The two are, of course, identical for equilibrium cases, but the second equality follows if the process is at least wide-sense stationary. The latter is not a foregone conclusion in nonlinear processes [3] and must be checked in each case. As discussed in the text, this was done and calculations using either $\phi'(t)$ or $\phi(t)$ were found to agree.

[9] It is found, for example, that at 25 kV/cm, the transient and overshoot velocity effects have decayed in less than 0.3 psec.

away from the initial exponential. This behavior is well developed in Fig. 47b, for 25 kV/cm. It is clear from Fig. 47b that the initial decay, the momentum-relaxation portion, deviates substantially from an exponential for times greater than about 0.1 psec, and it is evident that although $\phi(t)$ decays initially as an exponential, it deviates noticeably from this behavior at long times and begins to decay as $t^{-3/2}$. This behavior differs from that reported by Fauquemberque *et al.* [112] but appears to be intrinsic to the momentum-relaxation process [136]. At still higher electric fields, this tailing behavior is washed out because of the much faster energy-relaxation process.

The initial exponential decay portion is significant. The time constant of this portion of the decay of $\phi(t)$ is closely related to and slightly larger than the momentum-relaxation time τ_m associated with the chordal mobil-

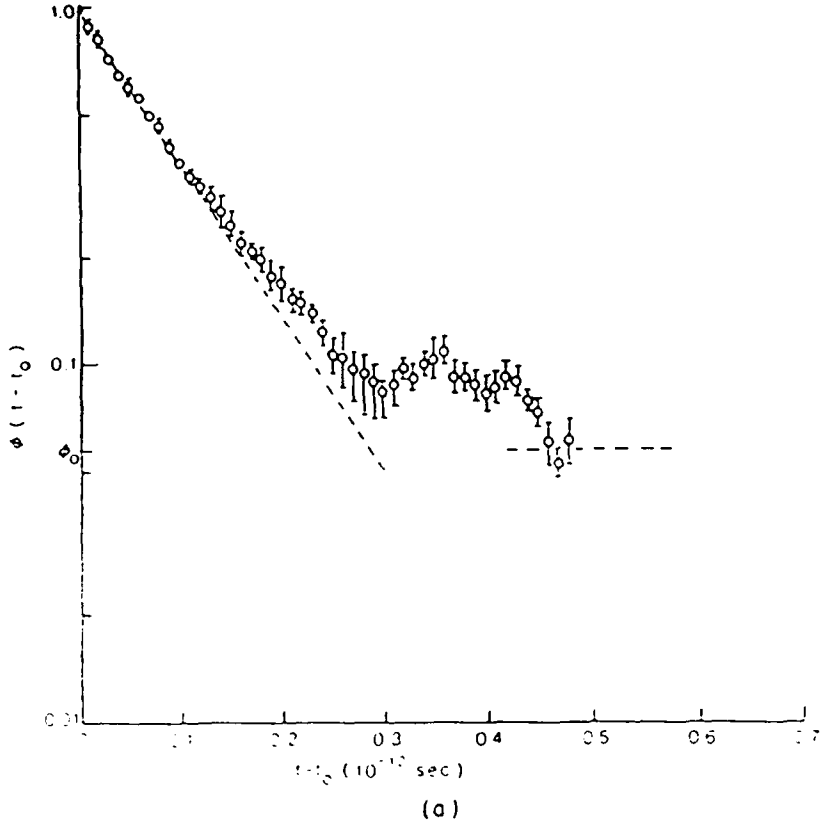


Fig. 47a. The total velocity-correlation function, $\phi(t - t_0) = \langle v(t)v(t_0)\rangle$, as a function of $t - t_0$ for Si at 300 K. (a) field at 10 kV/cm. The curves are normalized to $\langle v^2 \rangle$ and ϕ_∞ is $\langle v - v_0 \rangle^2$, the final value. The individual curves are discussed in the text.

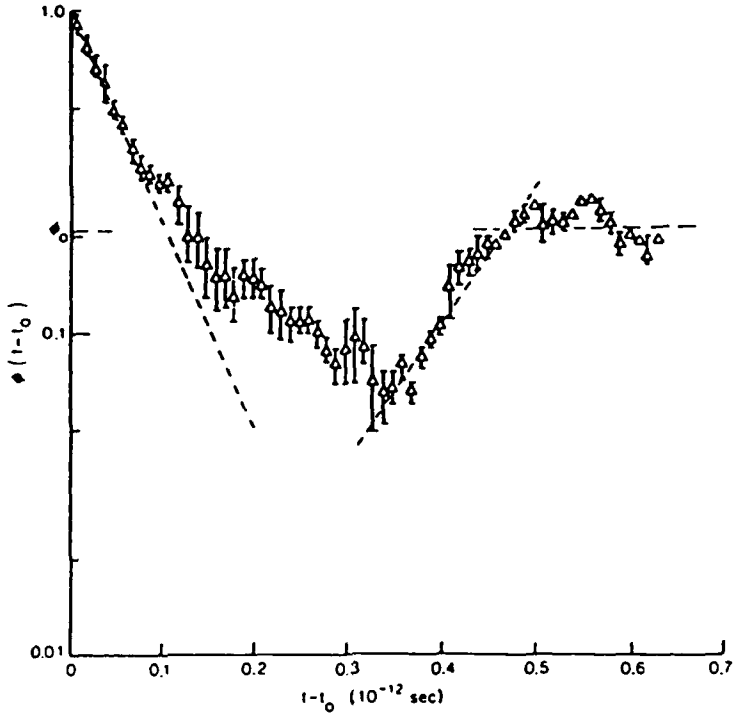


Fig. 47b. Field at 25 kV/cm.

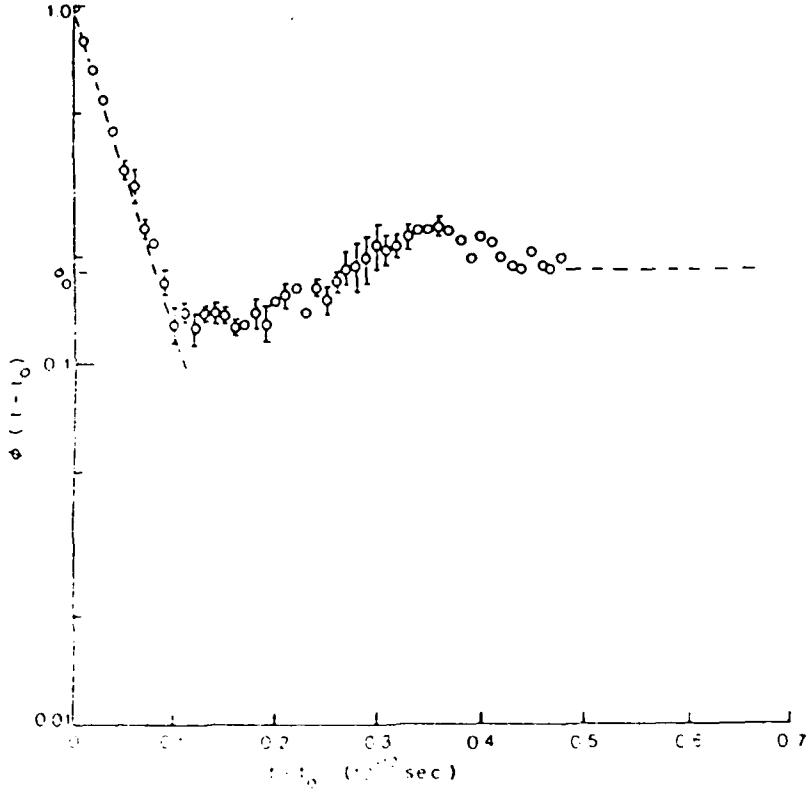


Fig. 47c. Field at 50 kV/cm

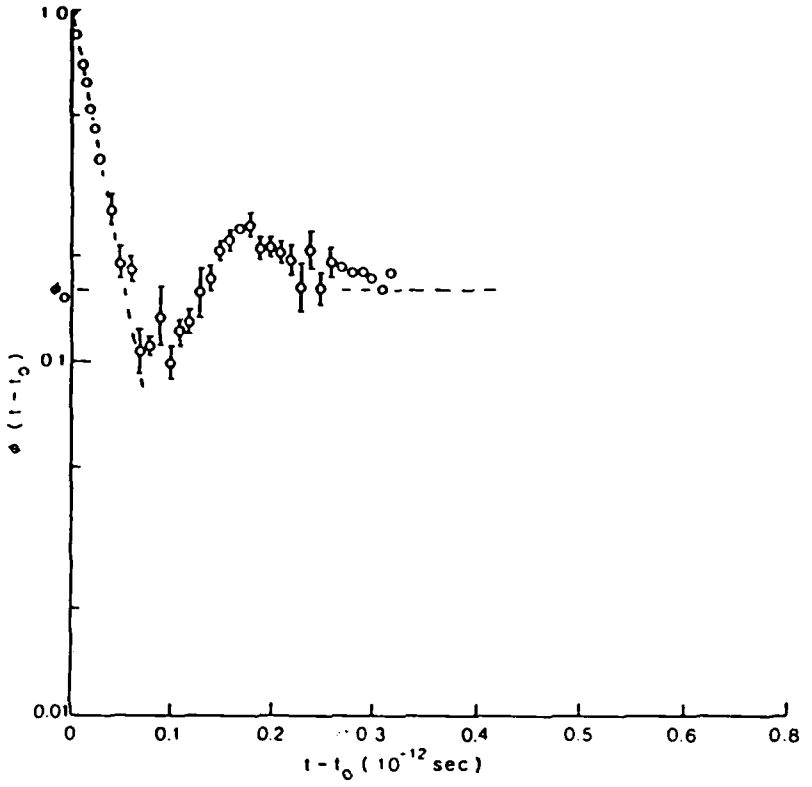


Fig. 47d. Field at 100 kV/cm.

ity $\mu = v_d/F$, rather than the differential mobility dv_d/dE (here we define the effective or average $\tau_m = m^*\mu/e$). The latter quantity has been suggested as the appropriate quantity for longitudinal diffusion [111]. At 25 kV/cm, the velocity is becoming very nearly saturated, so that the differential mobility is more than an order of magnitude smaller than the chordal mobility. This difference is readily distinguished from the data in Fig. 47. The decay of $\phi(t) \simeq \exp(-t/\tau_0)$ and at 25 kV/cm, for example, is best fit with a τ_0 of 7×10^{-14} sec, while $\tau_m \simeq 5.4 \times 10^{-14}$ sec. The results of the decay constant of the exponential portion of $\phi(t)$ being slightly larger than τ_m appears to be a general result, as it was checked at several other values of electric field. Van Kampen [116] has suggested such a difference would occur as a general result of nonlinear relaxation. If $v(t)$ decays as $\exp(-t/\tau_m)$, he suggests that a fully nonlinear treatment of noise would have the correlation function decay with a characteristic time $\tau_0 = \tau_m/(1-\epsilon)$, where $\epsilon \simeq 1.5\langle v\rangle^2/\langle v^2\rangle$. This gives $\tau_0 \simeq 6.8 \times 10^{-14}$ sec at 25 kV/cm, for the τ_m just given, which is within the accuracy of the

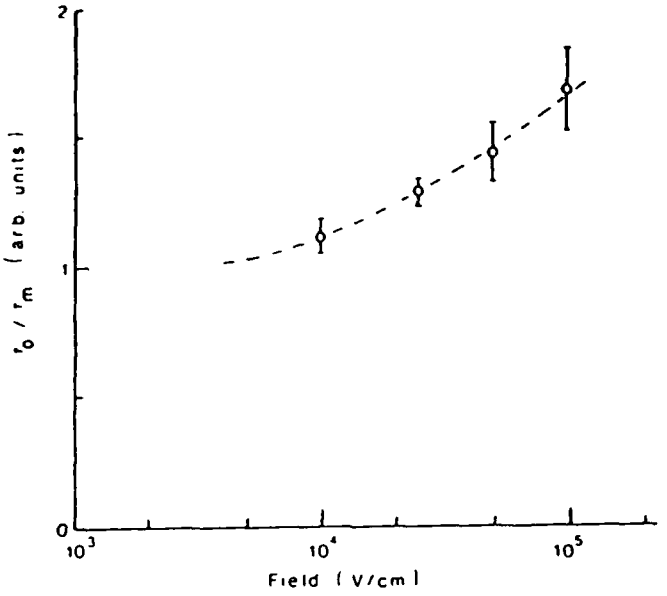


Fig. 48. The deviation of τ_0 from τ_m for Si at 300 K. The velocity-autocorrelation function $\phi(t)$ decreases initially as $\exp(-t/\tau_0)$. Although this initial fall corresponds to momentum relaxation, $\tau_0 > \tau_m$.

present calculations. However, this correction does not show the explicit field dependence of τ_0/τ_m. Both $\langle v \rangle^2$ and $\langle v^2 \rangle$ can be expected to increase (almost quadratically) with F until saturation. Then $\langle v \rangle^2$ should become independent of F. This would imply a reduction of E. It is observed, however, that E increases with F. In Fig. 48, we plot the variation of τ_0/τ_m with electric field. It is observed that τ_0/τ_m increases with the field, although not quite linearly.

In Fig. 49, the Laplace transform and Fourier cosine transform of $\phi'(t)$ are shown for a field of 25 kV/cm. Contrary to linear transport, these functions are not simple, monotonically decreasing functions for large ω and s. Rather, they exhibit peaks at high frequency. The origin of these peaks lies in the enhanced high-frequency conductivity [88, p. 126] in regions where the energy-relaxation process can no longer follow the ac field. Thus, these peaks have their origin in the same processes that lead to velocity overshoot. The oscillations in $S_v(\omega)$ at high frequency appear to be related to the oscillations at long time on $\phi'(t)$. While these oscillations may not be real, their presence and the shape of $S_v(\omega)$ has also been observed by Grondin in GaAs [137]. From this figure, it is apparent that enhanced noise will appear in Si devices at frequencies above $\sim 10^{11}$ Hz and that correlated carrier motion can be expected for times on the order of 1 psec.

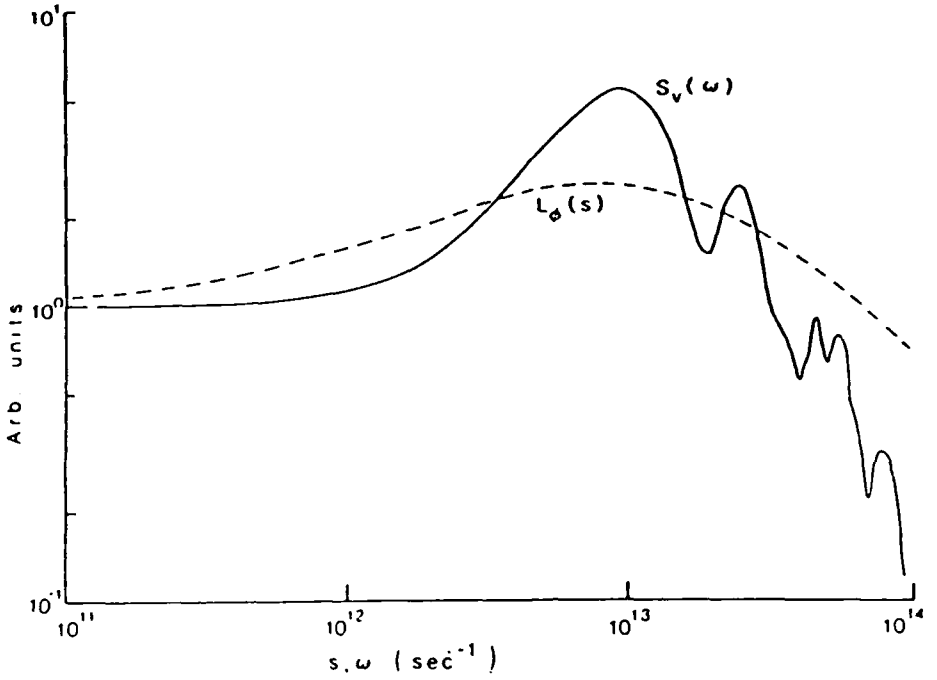


Fig. 49. The Laplace transform $L_\phi(s)$ and Fourier cosine transform $S_v(\omega)$ (noise spectral density), of $\phi'(t) = \langle [v(t + t_0) - \langle v \rangle][v(t_0) - \langle v \rangle] \rangle$ for Si at 300 K, 25 kV/cm.

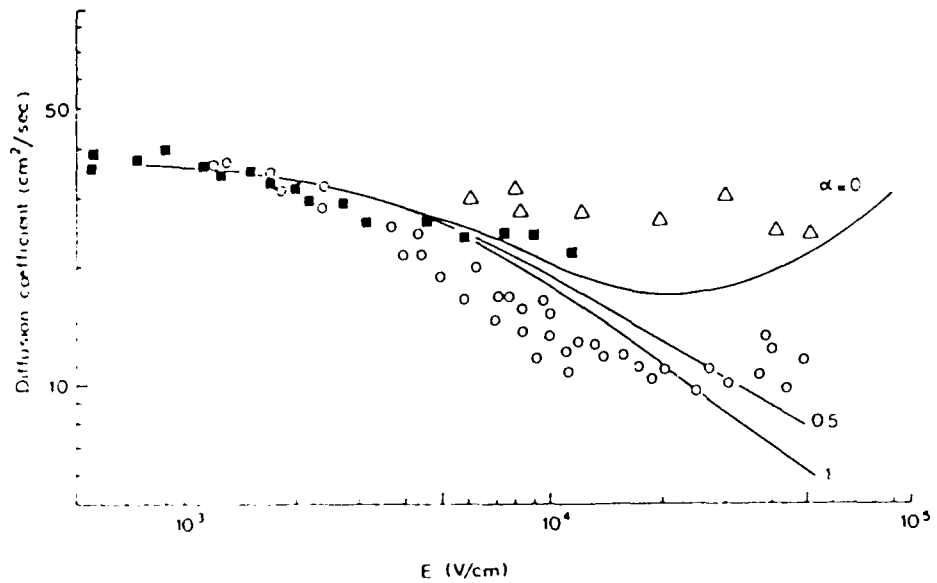
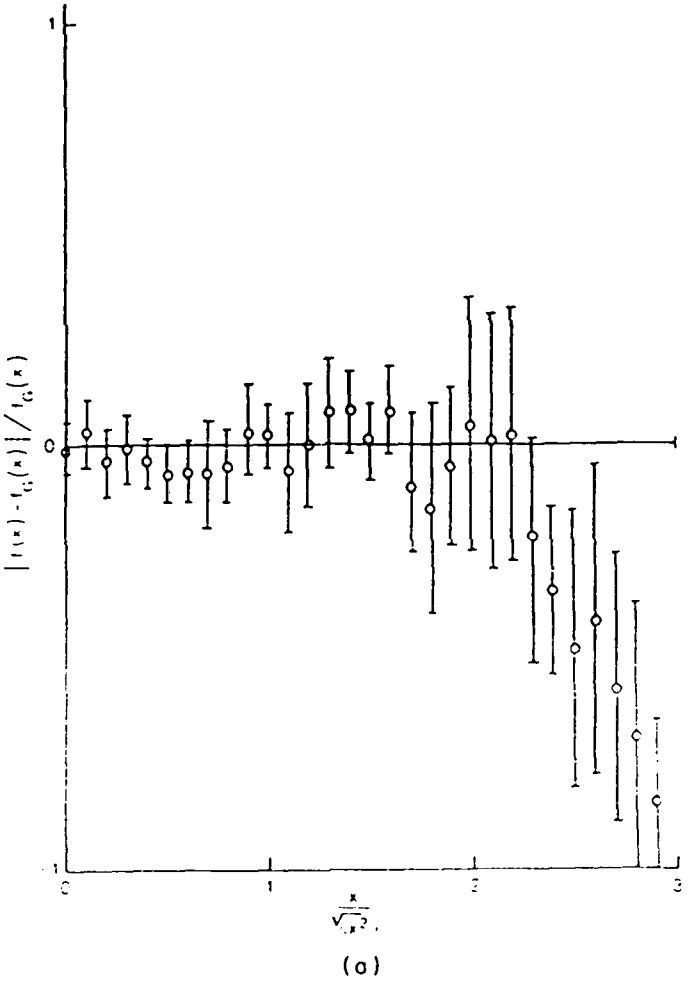


Fig. 50. Comparison of theoretical and experimental values of D for Si at 300 K. (After Ferry [135])

Approaches such as this can be used to calculate the diffusion coefficient as well, through Eq. (141) for $\omega \rightarrow 0$. This has been done theoretically and compared with experiments by the group in Modena [138]. The results are shown in Fig. 50. The curve with $\alpha = 0.5$ is the preferred nonparabolicity-corrected curve for Si. The agreement with experiment is good, but not great. Part of the difference is in the assumption that the packet of electrons diffuses as a Gaussian, which is not a valid assumption.

The spatial distribution shows strikingly non-Gaussian behavior. In Fig. 51a, we illustrate $n(x)$ normalized to a Gaussian with the same value of $[((\Delta x)^2)]^{1/2}$ (i.e., same σ values) for 10 kV/cm. When compared to the



(a)

Fig. 51. The spatial distribution function $n(x)$ normalized to a Gaussian for (a) 10 kV/cm and (b) 100 kV/cm for Si at 300 K. The non-Gaussian nature is discussed in the text.

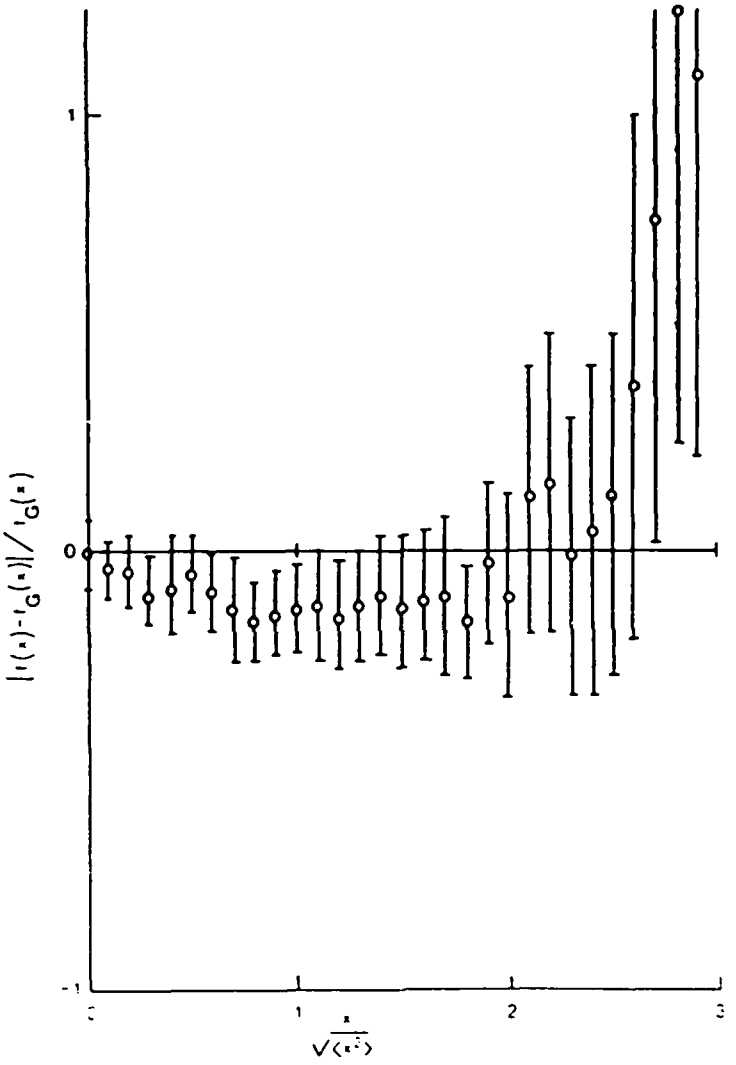


Fig. 51b.

Gaussian, we find $n(x)$ truncated in the tail region. This probably arises from the fact that the electrons in the tail have much higher energy and get hot faster. Once into the velocity-saturation region ($F \approx 50$ kV/cm), a Gaussian again appears; but at higher fields, there are indications that the tailing is reversed. In Fig. 51b, we illustrate this by showing $n(x)$ for 100 kV/cm. At this field, we are approaching breakdown, and there are a significant number of "lucky" electrons [77] in the tail of $n(x)$. These varia-

tions agree with the trends seen in the data of Fig. 50 and may be the cause for the lack of agreement.

D. Transient Diffusion

In recent years, much interest has centered on the transient dynamic response of electrons, especially as it impacts carrier transport through small spatial regions in which the electric field is high. In particular, in the pinch-off region of the channel in a field-effect transistor, the carriers move by drift and diffusion in a very high electric field. Considerable interest has centered on the velocity response, especially that of the overshoot velocity [3,38,39]. The transient response in these high-field conditions is significant in that carriers may completely transit the region prior to obtaining a steady-state high-field distribution. Thus, the transient velocity can be more significant than the steady-state velocity. Considerably less attention has been focused on the transient diffusion that also occurs under these conditions. The lack of attention paid to transient diffusion is easily understood when it is recognized that diffusion is actually a process depending on velocity correlation [139], and the relationship between diffusion and drift, as expressed by the Einstein relation, is a steady-state (stationary distribution) relation [122]. The problem is complicated by the fact that the random-walk equations governing diffusion do not reduce to normal Fick's law behavior on time scales comparable to relaxation processes [122], a result of the general non-Markovian nature of transport on these time scales [62]. The purpose of this section will be to highlight some of these problems and to illustrate them with results calculated by the ensemble Monte Carlo technique.

Diffusion is related to the spatial spreading of an ensemble of carriers with time, as the ensemble responds to both applied drift forces and random forces, such as are generated by collisions. In general, the diffusion coefficient is related to the ensemble position distribution through Eq. (126). In the case of a transient dynamic response, however, the problem is more complicated. A Fokker–Plank equation can be generated whose solution is the transition probability for a particle at x_0,t_0 to transition to x,t. On the short-time scale over which relaxation processes occur, this equation does not reduce to the normal diffusion equation [130]. The problem lies in the fact that the Langevin equation, from which the former equation is obtained, is second order in position. While it remains Markovian in phase space, its projection onto real space does not, except for long times when the ensemble is stationary. Thus Eq. (126) must be corrected for the nonlocal (in time) behavior. Then [122,130],

$$\langle(\Delta x)^2\rangle = (2\langle v^2\rangle/\gamma)\{t - (1/\gamma)[1 - e^{-\gamma t}]\}. \qquad (142)$$

where $1/\gamma$ is the relaxation time and $\langle v^2 \rangle$ the mean squared velocity about the mean.[10] From a fluctuation–dissipation theorem, we generally define $D = \langle v^2 \rangle / \gamma$. If we use Eq. (126), we find

$$D = \frac{\langle (\Delta x)^2 \rangle}{2t \{1 - [1 - e^{-\gamma t}]/\gamma t\}}. \tag{143}$$

If, however, we use Eq. (127), then

$$D = \frac{d\langle (\Delta x)^2 \rangle / dt}{2(1 - e^{-\gamma t})}. \tag{144}$$

The difference in these two is, of course, in the denominator. The denominator of (143) goes as t^2 for small t and that of (144) goes as t plus whatever variation comes from the derivative term. We cannot say which is correct, but if $\langle (\Delta x)^2 \rangle$ varied as t^2 for small t, both would give a constant D. In fact, this does not occur for the general nonlinear relaxation to a nonequilibrium steady state.

In a stationary distribution, the diffusion coefficient can be related to a velocity-correlation function from which the mobility is also derived. This leads to the classical Einstein relation, which, for a nondegenerate semiconductor, is

$$D = \mu k_B T_e / e, \tag{145}$$

where μ and T_e are, respectively, the chordal mobility and electron temperature at a particular electric field. In general, the Einstein relation does not hold for hot electrons, and it is of interest to know how far this result differs from Eqs. (126) or (143) in the case of the transient dynamic response.

We have carried out calculations for the longitudinal diffusion by the ensemble Monte Carlo process. In this case, an ensemble of 10^4 electrons is subjected to transport via a Monte Carlo technique. At each time step, the transport parameters are obtained by an ensemble average. The time evolution of these ensemble averages yields the time evolution of the transport parameters. This technique has previously been shown to yield excellent agreement for the transport coefficients and to agree well with parameterized distribution approaches, which explicitly define selected coefficients. In the present calculations, the carrier ensemble was assumed to have been injected in GaAs at $x = 0$ at $t = 0$, i.e., as a δ-function ensemble. The spread of the distribution and its drift under an applied homogeneous electric field of 25 kV/cm was calculated. The kinetic temperature ($T_L = 300$ K) and drift velocity were also calculated

[illegible] speaking, Eq. (142) is valid only for linear relaxation to equilibrium. Under certain [illegible] however, it is not a bad approximation for quasi-linear relaxation [illegible] Unfortunately, there exists no good theory for [illegible] state.

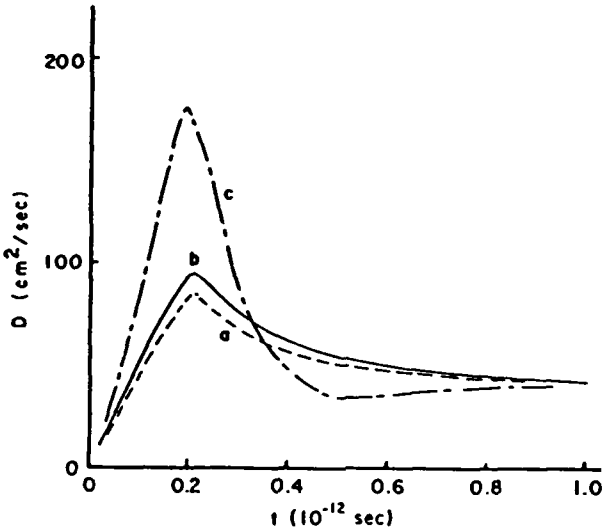


Fig. 52. Diffusion coefficients calculated from Eqs. (126), (143), and (145) (curves a–c, respectively) for an ensemble of carriers injected into GaAs at $x = 0$, $t = 0$, $T = 300$ K, at 25 kV/cm.

during the calculation. Gallium arsenide was chosen because of the complications arising from intervalley transfer and negative differential conductivity. In such material, the transient dynamic response is dominated by the differential repopulation between nonequivalent valleys [72,140]. The diffusion coefficient was calculated by each of Eqs. (126), (143), and (145), and the results are shown in Fig. 52. In using (145), the mobility μ is taken as v_d/F and is not a differential mobility, a definition in keeping with the correlation-function approach. If a derivative approach [Eq. (127)] were used, the results would be closer to those of (145) on the rising portion of the response ($t < 0.2$ psec). However, the derivative approach would give a negative diffusion during the falling portion of the curve, a result not at all in keeping with (145), although understandable on physical terms. In this region, the carrier ensemble appears to actually be contracting as the faster, more energetic carriers transfer to the heavy mass satellite valleys. Within the accuracy of the Monte Carlo method, all three approaches converge as the ensemble approaches a stationary distribution in phase space.

APPENDIX A. DERIVATION OF THE BALANCE EQUATIONS

The balance equations given in Section III are obtained by taking the moments of the Boltzmann transport equation, assuming that the distribu-

tion function in each valley is a drifted Maxwellian. Thus

$$f_j(\mathbf{k}) = C_j \exp[-\hbar^2(\mathbf{k} - \mathbf{k}_{dj})^2/2m_j k_B T_{ej}], \tag{A-1}$$

where $\hbar k_{dj} = m_j v_{dj}$ is the average drift momentum of the electron gas, C a constant for normalization purposes, and T_{ej} the electron temperature, with the subscript j referring to the jth valley. From Eq. (A-1), it follows that $f_j(E) \approx f_{j0}(E) + f_{j1}(E)$, with

$$f_{j0}(E) = C_j \exp(-E/k_B T_{ej}) \tag{A-2}$$

and

$$f_{j1}(E) = v(m_c v_{dj}/k_B T_{ej}) f_{j0}(E). \tag{A-3}$$

The actual moments themselves are complicated by the fact that the semiconductor conduction band is describable by a multivalley structure, which often comprises nonequivalent valley sets, and, hence, is a coupled system. We must also account for the repopulation effects that can occur.

The Boltzmann transport equation is given as

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f + e\mathbf{F} \cdot \nabla_{\mathbf{p}} f = \sum_{\mathbf{p}'} \{W(\mathbf{p},\mathbf{p}')f(\mathbf{p}') - W(\mathbf{p}',\mathbf{p})f(\mathbf{p})\}. \tag{A-4}$$

In Eq. (A-4), $W(\mathbf{p},\mathbf{p}')$ is the scattering rate from state $\mathbf{p}'$ to state $\mathbf{p}$. In the following, we shall ignore the inhomogeneity term $\mathbf{v} \cdot \nabla f$, although it can readily be incorporated. We define the average of $\phi(\mathbf{p})$ in the ith valley as

$$\langle \phi(\mathbf{p}_i) \rangle_i = \frac{1}{n_i} \int \phi(\mathbf{p}) f_i(\mathbf{p})\, d\mathbf{p}. \tag{A-5}$$

To begin, we multiply Eq. (A-4) by an arbitrary function of $\phi(\mathbf{p}_i)$ assuming $\phi(\mathbf{p}_i)$ and $f_i(\mathbf{p}_i)$ represent the ith valley (or set of equivalent valleys in this case). Then, integrating over $\mathbf{p}_i$ yields

$$\begin{aligned} &\frac{\partial}{\partial t}(n_i \langle \phi \rangle_i) - n_i e\mathbf{F} \cdot \langle \nabla_{\mathbf{p}_i} \phi \rangle_i \\ &\quad = \sum_{\mathbf{p}_i} \left\{ \sum_{\nu} [W(\mathbf{p}_i,\mathbf{p}_i')f(\mathbf{p}_i') - W(\mathbf{p}_i',\mathbf{p}_i)f(\mathbf{p}_i)] \right. \\ &\quad \left. + \sum_{\nu} [W(\mathbf{p}_i,\mathbf{p}_j')f(\mathbf{p}_j') - W(\mathbf{p}_j',\mathbf{p}_i)f(\mathbf{p}_i)] \right\} \phi(\mathbf{p}_i). \end{aligned} \tag{A-6}$$

Here, we have separated the intravalley (with equivalent intervalley) and nonequivalent intervalley contributions to the scattering processes. When the initial and final states lie in the same valley (or an equivalent valley), they are describable by the same local distribution function. With this [illegible] the intravalley [illegible] on the right-hand side of Eq. (A-6) [illegible]

tion over $\mathbf{p}_i'$) can be treated by a simple change of variables and this term becomes

$$n_i\langle\Gamma_\phi(\mathbf{p}_i)\rangle_i, \tag{A-7}$$

where

$$\Gamma_\phi(\mathbf{p}_i) = \sum_{\mathbf{p}'} W(\mathbf{p}_i',\mathbf{p}_i)[\phi(\mathbf{p}_i') - \phi(\mathbf{p}_i)]. \tag{A-8}$$

If $\phi = C$, this term vanishes and, as expected, makes no contribution to a density-balance equation.

The nonequivalent intervalley terms can be written as

$$\begin{aligned} n_i\left\langle\left.\frac{d\phi}{dt}\right|_{\text{coll}}\right\rangle &= \sum_{\mathbf{p}_i} \phi(\mathbf{p}_i)\left\{\sum_{\mathbf{p}_j'} W(\mathbf{p}_i,\mathbf{p}_j')f(\mathbf{p}_j') - W(\mathbf{p}_j',\mathbf{p}_i)f(\mathbf{p}_i)\right\} \\ &= \sum_{\mathbf{p}_j'} f(\mathbf{p}_j')\phi(\xi_j)\sum_{\mathbf{p}_i} W(\mathbf{p}_i,\mathbf{p}_j') \\ &\quad - \sum_{\mathbf{p}_i} f(\mathbf{p}_i)\phi(\mathbf{p}_i)\sum_{\mathbf{p}_j'} W(\mathbf{p}_j',\mathbf{p}_i), \end{aligned} \tag{A-9}$$

where we have used the energy-conserving δ function inherent in $W(\mathbf{p}_i,\mathbf{p}_j')$ to define the renormalized momentum ξ_j through the relation

$$\epsilon_j' = \epsilon_i \pm \hbar\omega_0 = \xi_j^2/2m_j^* \pm \hbar\omega_0, \tag{A-10}$$

so that ξ_j is a function of $\mathbf{p}_j'$. Introducing the scattering rate as

$$\Gamma(\mathbf{p}) = \sum_{\mathbf{p}'} W(\mathbf{p}',\mathbf{p}), \tag{A-11}$$

Eq. (A-8) becomes

$$n_i\left\langle\left.\frac{d\phi}{dt}\right|_{\text{coll}}\right\rangle = n_j\langle\phi(\xi_j)\Gamma(\mathbf{p}_j')\rangle_j - n_i\langle\phi(\mathbf{p}_i)\Gamma(\mathbf{p}_i)\rangle_i. \tag{A-12}$$

By inserting the details of the various scattering processes, the individual moment equations can be readily set up. However, because of the multiplicity of scatterers, the individual equations are quite complicated. We shall not delve deeper into their structure here.

APPENDIX B. THE WIGNER DISTRIBUTION FUNCTION

Classical transport physics is based on the concept of a probability distribution function, which is defined over the phase space of the position and momenta of all the particles concerned. The time rate of change of this distribution function is governed by the Liouville equation, the clas-

sical expectation value of any physical observable, which is a function of position and momentum, is obtained by integrating the product of the observable and the distribution function over all of phase space.

In the quantum formulation of transport physics, the concept of a phase-space distribution function is not possible inasmuch as the noncommutation of the position and momentum operators (the Heisenberg uncertainty principle) precludes the precise specification of a point in phase space. However, within the matrix formulation of quantum mechanics, it is possible to construct a "probability" density matrix, which is often interpreted as the analog of the classical distribution function in phase space. The time rate of change of this probability density matrix is governed by the quantum analog of the Liouville equation. Moreover, the expectation value of a physical observable is obtained by taking the trace of the matrix, which is the product of the probability density matrix and the matrix that corresponds to the physical observable.

There is yet another approach to the formulation of quantum transport based on the construction of the Wigner distribution function [100]. As we shall show, this distribution function has no simple interpretation in the sense of probability theory but, in lieu of its special properties, can be used directly for calculating expectation values of observables in a manner quite analogous to that of classical theory, i.e., by integrating the product of the observable and the Wigner distribution function over all phase space.

In this section we shall review the salient features of the Wigner distribution function. Although the Wigner function is generally defined in terms of all the generalized coordinates and momenta of the system in question as

$$P_W(x_1 \cdots x_n, p_1 \cdots p_n) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dy_1 \cdots dy_n \, \psi^* \left(x_1 + \frac{y_1}{2}, \ldots, x_n + \frac{y_n}{2}\right) \times \psi \left(x_1 - \frac{y_1}{2}, \ldots, x_n - \frac{y_n}{2}\right) \exp\left[\frac{i(p_1 y_1 + \cdots + p_n y_n)}{\hbar}\right]. \tag{B-1}$$

we shall discuss the properties of the Wigner function in terms of a single coordinate and momentum. In this case, we let

$$P_W(x,p) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dy \, \psi^* \left(x + \frac{y}{2}\right) \psi \left(x - \frac{y}{2}\right) \exp\left(\frac{ipy}{\hbar}\right). \tag{B-2}$$

where $\psi(x)$ refers to the state of the system in the coordinate representation.

The distribution function of Eq. (B-2) has interesting properties in that

the integration of this function over all momenta leads to the probability density in real space; conversely, the integration of this function over all coordinates leads to the probability density in momentum space. In mathematical terms,

$$\int_{-\infty}^{\infty} P_W(x,p)\,dp = \psi^*(x)\psi(x) \tag{B-3a}$$

and

$$\int_{-\infty}^{\infty} P_W(x,p)\,dx = \phi^*(p)\phi(p), \tag{B-3b}$$

where

$$\phi(p) = (2\pi\hbar)^{1/2} \int_{-\infty}^{\infty} \exp\left(\frac{-ipx}{\hbar}\right) \psi(x)\,dx.$$

It follows immediately from Eq. (B-3) that, for an observable $\hat{F}(\hat{x},\hat{p})$, which is either a function of momentum operator alone, of position operator alone, or of any additive combination therein, the expectation value of the observable is given by

$$\langle F \rangle = \iint \hat{F} P_W(x,p)\,dx\,dp, \tag{B-4}$$

which is analogous to the classical expression for the average value. Herein lies the interesting aspect of the Wigner distribution function: the result of Eq. (B-4) suggests that it is possible to transfer many of the results of classical transport theory into quantum transport theory by simply replacing the classical distribution function by the Wigner distribution function. However, unlike the density matrix, the Wigner distribution function itself cannot be viewed as the quantum analog of the classical distribution function because it is generally nonpositive definite and nonunique [$P_W(x,p)$ of Eq. (B-2) is not the only bilinear expression in ψ that satisfies Eq. (B-3)].

Further resemblance of the Wigner distribution function to the classical distribution function is apparent by examining the equation of time evolution for $P_W(x,p)$. Upon assuming that $\psi(x)$ in Eq. (B-2) satisfies the Schrödinger equation for a system with Hamiltonian $H = (p^2/2m) + V(x)$, it can be readily shown that $P_W(x,p)$ satisfies the equation

$$\frac{\partial P_W}{\partial t} + \frac{p}{m}\frac{\partial P_W}{\partial x} + \theta \cdot P_W = 0, \tag{B-5}$$

where

$$\theta \cdot P_W = -\frac{2}{\hbar} \sum_{n=0}^{\infty} (-1)^n \frac{(\hbar/2)^{2n+1}}{(2n+1)!} \frac{\partial^{2n+1} V(x)}{\partial x^{2n+1}} \frac{\partial^{2n+1} P_W(x,p)}{\partial p^{2n+1}}. \tag{B-6a}$$

Alternately, $\theta \cdot P_w$ can be expressed as

$$\theta \cdot P_w = -\frac{2}{\hbar}\left[\sin\frac{\hbar}{2}\left\{\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right\}\right] V(x)P_w(x,p), \tag{B-6b}$$

where it is understood that the position gradient operates only on the potential energy $V(x)$. It is evident that in the limit $h \to 0$, $\theta \cdot P_w$ in Eqs. (B-6) becomes

$$\theta \cdot P_w = -\frac{\partial V}{\partial x}\frac{\partial P_w}{\partial p}, \tag{B-7}$$

so that Eq. (B-5) reduces to the classical continuity equation; it is also clear that the dominant quantum correction to $\theta \cdot P_w$ is of order h^2, with this term also having dependences on the third derivative of potential energy with respect to position and the third derivative of the Wigner distribution function with respect to momentum.

The Wigner distribution function is derivable [109] from the Fourier inversion of the expectation value with respect to state $\psi(x)$ of the operator $\exp[i(\tau\hat{p} + \theta\hat{x})]$ (here, x and p satisfy the commutation relation $[\hat{x},\hat{p}] = i\hbar$). As such,

$$P_w(x,p) = \frac{1}{4\pi^2}\int\int C_w(\tau,\theta)\exp[-i(\tau p + \theta x)]\,d\tau\,d\theta, \tag{B-8a}$$

where

$$C_w(\tau,\theta) = \int \psi^*(x)\exp[i(\tau\hat{p} + \theta\hat{x})]\,\psi(x)\,dx \tag{B-8b}$$

and the interval of integration is $[-\infty,\infty]$ unless otherwise specified. In order to show that the right side of Eq. (B-8a) is indeed the Wigner distribution function as defined in Eq. (B-2), note from the Baker–Hausdorff theorem [110] that $\exp[i(\tau\hat{p} + \theta\hat{x})$ can be rewritten as

$$\exp[i(\tau\hat{p} + \theta\hat{x})] = \exp(\tfrac{1}{2}i\tau\hat{p})\exp(i\theta\hat{x})\exp(\tfrac{1}{2}i\tau\hat{p}) \tag{B-9}$$

in which case $C_w(\tau,\theta)$ of Eq. (B-8b) becomes

$$C_w(\tau,\theta) = \int_{-\infty}^{\infty}\exp[-\tfrac{1}{2}i\tau\hat{p}]\psi(x)\exp(i\theta\hat{x})\exp[\tfrac{1}{2}i\tau\hat{p}]\psi(x)\,dx. \tag{B-10}$$

which further reduces to

$$C_w(\tau,\theta) = \int_{-\infty}^{\infty}\psi^*(x - \tfrac{1}{2}\tau\hbar)e^{i\theta x}\psi(x + \tfrac{1}{2}\tau\hbar)\,dx. \tag{B-11}$$

Then, by inserting $C_w(\tau,\theta)$ of Eq. (B-11) into the right-hand side of Eq.

(B-8a), integrating over the variable θ by using the relation

$$\int_{-\infty}^{\infty} \exp i\theta(x' - x'')\, d\theta = 2\pi\delta(x' - x''),$$

and letting $\tau = y/\hbar$, the desired result is obtained.

The method just outlined for arriving at the Wigner distribution function is based on the notion of a characteristic function. The characteristic function of an observable $\hat{A}$ with respect to state $|\psi\rangle$ (here, the Dirac notation is utilized for purposes of generality) is defined as

$$C_A(\xi) = \langle\psi|\exp(i\xi\hat{A})|\psi\rangle, \tag{B-12}$$

where ξ is a real parameter. Assuming $\hat{A}$ to possess an eigenvalue spectrum given $\hat{A}|A'\rangle = A'|A'\rangle$, $C_A(\xi)$ can be evaluated in the A' representation as

$$C_A(\xi) = \int dA' \int dA'' \langle\psi|A'\rangle\langle A'|\exp(i\xi\hat{A})|A''\rangle\langle A''|\psi\rangle. \tag{B-13}$$

Since $\langle A'|\exp(i\xi\hat{A}|A''\rangle = \exp(i\xi A')\,\delta(A' - A'')$ in the A' representation, $C_A(\xi)$ in Eq. (B-13) reduces to

$$C_A(\xi) = \int dA' \exp(i\xi A')|\psi_{A'}|^2, \tag{B-14}$$

where $|\psi_{A'}|^2 = |\langle\psi\rangle|^2 \equiv P(A')$, the probability distribution function for measuring A' in state $|\psi\rangle$. Hence, the characteristic function for $\hat{A}$ is the Fourier transform of the probability distribution function $P(A')$. Subsequent inversion of Eq. (B-14) leads to

$$P(A') = \frac{1}{2\pi}\int C_A(\xi)\exp(-i\xi A')\, d\xi. \tag{B-15}$$

The Wigner distribution function was derived by taking the Fourier transform of the characteristic function for $\exp[i(\tau\hat{p} + \theta\hat{x})]$. In view of the connection between the probability distribution function and the characteristic function for a given observable, this approach seems to be a natural way of obtaining a distribution function for momentum and position. Unfortunately, the noncommutative nature of the two observables destroys the convenient probability interpretation of the characteristic function implicit in Eq. (B-15).

In order to demonstrate this point, assume the characteristic function of two noncommuting observables, $\hat{A}$ and $\hat{B}$, to be

$$C_{AB}(\xi_1, \xi_2) = \langle\psi|\exp[i(\xi_1\hat{A} + \xi_2\hat{B})]|\psi\rangle. \tag{B-16}$$

Observables $\hat{A}$ and $\hat{B}$ are assumed to have eigenvalue spectra

$$\hat{A}|A'\rangle = A'|A'\rangle, \tag{B-17a}$$

$$\hat{B}|B'\rangle = B'|B'\rangle \tag{B-17b}$$

and are chosen so that $[A,[A,B]] = [B,[A,B]] = 0$. This assumption is imposed so that the identity

$$\exp[i(\xi_1\hat{A} + \xi_2\hat{B})] = \exp(i\xi_1\hat{A})\exp(i\xi_2\hat{B})\exp(-\tfrac{1}{2}\xi_1\xi_2[\hat{A},\hat{B}]) \tag{B-18}$$

may be used.

Inserting Eq. (B-18) into Eq. (B-16), obtaining the matrix elements of $\exp(i\xi_1\hat{A})$ in the A' representation and $\exp(i\xi_2\hat{B})$ in the B' representation, and assuming $[\hat{A},\hat{B}]$ is a C number, we obtain

$$C_{AB}(\xi_1,\xi_2) = \exp(-\tfrac{1}{2}\xi_1\xi_2[\hat{A},\hat{B}]) \int dA' \int dB' \exp[i(\xi_1A' + \xi_2B')] \times \langle\psi|A'\rangle\langle A'|B'\rangle\langle B'|\psi\rangle. \tag{B-19}$$

We define $F(A',B')$, a generalized Wigner distribution function, to be

$$F(A',B') = \langle\psi|A'\rangle\langle A'|B'\rangle\langle B'|\psi\rangle \tag{B-20}$$

so that

$$F(A'B') = \frac{1}{(2\pi)^2}\int d\xi_1 \int d\xi_2 \exp(\tfrac{1}{2}\xi_1\xi_2[\hat{A},\hat{B}])C_{AB}(\xi_1,\xi_2) \times \exp[-i(\xi_1A' + \xi_2B')]. \tag{B-21}$$

It is evident from Eqs. (B-20)–(B-21) that

$$\int F(A',B')\,dA' \equiv |\langle B'|\psi\rangle|^2 = \frac{1}{2\pi}\int d\xi_2\, C_{AB}(0,\xi_2)\exp(-i\xi_2B') \tag{B-22a}$$

and

$$\int F(A',B')\,dB' \equiv |\langle A'|\psi\rangle|^2 = \frac{1}{2\pi}\int d\xi_1\, C_{AB}(\xi_1,0)\exp(-i\xi_1A'). \tag{B-22b}$$

Thus, Eq. (B-21) establishes the relationship between the characteristic function for two arbitrary noncommuting observables and a generalized Wigner distribution function. The generalized distribution function has the essential properties of the conventional Wigner function in that an integration of the generalized function over the eigenvalue spectrum of one observable leads to the probability density in the canonically conjugate observable [Eqs. (B-22)]. However, there is no simple probability interpretation of $F(A',B')$ in Eq. (B-21) because of the necessary overlap between the states of the noncommuting observables. If $\hat{A}$ and $\hat{B}$ are made to commute so that $|A'\rangle$ and $|B'\rangle$ have a common set of eigenvectors, then $F(A',B')$ reduces to the classical distribution function for A and B

ACKNOWLEDGMENTS

Portions of the work reported were supported by the Office of Naval Research, the Army Research Office, and the North Atlantic Treaty Organization, to which we are grateful.

REFERENCES

1. R. A. Smith, "Semiconductors." Cambridge Univ. Press, London and New York, 1964.
2. P. N. Butcher, *Rep. Prog. Phys.* **30**, 97 (1967).
3. J. G. Ruch, *IEEE Trans. Electron Devices* **ED-19**, 652 (1972).
4. P. N. Butcher and C. J. Hearn, *Electron. Lett.* **4**, 459 (1968).
5. H. D. Rees, *IBM J. Res. Dev.* **13**, 537 (1969).
6. M. Muller, Private communications (1980).
7. F. Stern, *Crit. Rev. Solid State Sci.* **4**, 499 (1974).
8a. D. L. Scharfetter and H. K. Gummel, *IEEE Trans. Electron Devices* **ED-16**, 64 (1969).
8b. B. C. DeLoach, Jr. and D. L. Scharfetter, *IEEE Trans. Electron Devices* **ED-17** (1970).
8c. M. P. Shaw, H. L. Grubin, and P. R. Solomon, "The Gunn–Hilsum Effect." Academic Press, New York, 1979.
8d. J. J. Barnes and R. J. Lomax, *IEEE Trans. Electron Devices* **ED-24**, 1084 (1977).
8e. L. J. Turgeon and D. H. Navon, *IEEE Trans. Electron. Devices* **ED-25**, 837 (1978).
8f. T. Toyabe, K. Yamaguchi, S. Asai, and M. S. Mock, *IEEE Trans. Electron. Devices* **ED-25**, 825 (1978).
8g. K. Yamaguchi, S. Asai, and H. Kodera, *IEEE Trans. Electron Devices* ED-23, 1283 (1976).
8h. H. L. Grubin and T. M. McHugh, *Solid State Electron.* **21**, 69 (1978).
9. E. M. Conwell, "High Field Transport in Semiconductors." Academic Press, New York, 1967.
10. J. B. Gunn, *Solid State Commun.* **1**, 88 (1963).
11. H. Kroemer, *Proc. IEEE* **52**, 1736 (1966).
12. C. Hilsum, *Proc. IRE* **50**, 185 (1962).
13. R. Stratton, *IEEE Trans. Electron Devices* **ED-19**, 1288 (1972).
14. P. S. Cheung, and C. J. Hearn, *J. Phys. C* **5**, 1563 (1972).
15. H. Kroemer, *IEEE Spectrum* **5**, 47 (1968).
16. D. E. McCumber and A. G. Chynoweth, *IEEE Trans. Electron Devices* ED-13, 4 (1966).
17. H. Kroemer, *IEEE Trans. Electron Devices* **ED-13**, 27 (1966).
18. D. J. Colliver, K. W. Gray, D. J. Jones, H. D. Rees, G. Gibbons, and P. M. White, *Proc. Int. Symp. GaAs Related Compounds*, 4th Institute of Physics, London, 1973.
19. P. M. White and G. Gibbons, *Electron. Lett.* **8**, 166 (1972); K. W. Gray, J. E. Pattison, H. D. Rees, B. A. Prew, R. C. Clarke, and L. D. Irving, *Proc. Biennial Cornell Univ. Electron. Eng. Conf.* (1975); H. L. Grubin, *J. Vacuum Sci. Technol.* 13, 786 (1976).
20. E. M. Conwell, *IEEE Trans. Electron Devices* **ED-17**, 262 (1970).
21. K. W. Boer and G. Dohler, *Phys. Rev.* **186**, 793 (1969).
22. P. A. Maksym and C. J. Hearn, *Solid State Electron.* **21**, 1531 (1978).
23. G. S. Kino and P. N. Robson, *Proc. IEEE* **56**, 2056 (1968).

24. M. Reiser, *IEEE Trans. Electron Devices* ED-20, 35 (1973).
25. J. Lindmayer and C. Y. Wrigley, "Fundamentals of Semiconductor Devices," van Nostrand-Reinhold, Princeton, New Jersey, 1965.
26. H. L. Grubin and T. M. McHugh, *Proc. Biennial Cornell Elec. Eng. Conf., 6th, Cornell Univ., Ithaca, New York* p. 409 (1977).
27. H. M. Macksey, D. W. Shaw, and W. R. Wisseman, *Electron. Lett.* 12, 192 (1976).
28. H. L. Grubin, D. K. Ferry, and K. R. Gleason, *Solid State Electron.* 23, 157 (1980).
29. S. Selberherr, A. Schütz, and H. W. Pötzl, *IEEE Trans. Electron Devices* ED-27, 1540 (1980).
30. L. Landau and A. Kompanejez, *Phys. Z. Sowjetunion* 6, 163 (1934).
31. B. Davydov, *Phys. Z. Sowjetunion* 9, 433 (1936); 12, 269 (1937).
32. G. Bauer, *in* "Springer Tracts in Modern Physics 74" (G. Höhler, ed.). Springer-Verlag, Berlin and New York, 1974.
33. P. Price, *IBM J. Res. Dev.* 14, 12 (1970).
34. C. J. Hearn, *Proc. Phys. Soc. London* 86, 881 (1965); also *in* "Physics of Nonlinear Transport in Semiconductors" (D. K. Ferry, J. R. Barker, and C. Jacoboni, eds.), p. 153. Plenum Press, New York, 1980.
35. H. Fröhlich, *Proc. R. Soc. London Ser. A* 188, 521 (1947).
36. H. Fröhlich and B. V. Paranjape, *Proc. Phys. Soc. London* B69, 21 (1956).
37. P. Price, *Solid State Electron.* 21, 9 (1977).
38. T. J. Maloney and J. Frey, *J. Appl. Phys.* 48, 781 (1977).
39. S. Kratzer and J. Frey, *J. Appl. Phys.* 49, 4064 (1978).
40. P. Das and D. K. Ferry, *Solid State Electron.* 19, 851 (1976).
41. R. Bosch and H. Thim, *IEEE Trans. Electron Devices* ED-21, 16 (1974).
42. M. S. Shur, *Electron. Lett.* 12, 615 (1976).
43. B. Carnez, A. Cappy, A. Kaszynski, E. Constant, and G. Salmer, *J. Appl. Phys.* 51, 784 (1980).
44. M. A. Littlejohn, J. R. Hauser, and T. H. Glisson, *J. Appl. Phys.* 48, 4587 (1978).
45. H. F. Budd, *J. Phys. Soc. Jpn. Suppl.* 21, 420 (1966).
46. R. G. Chambers, *Proc. Phys. Soc. London* A65, 458 (1952).
47. H. D. Rees, *Solid State Commun.* 26A, 416 (1968).
48. H. D. Rees, *J. Phys. C* 5, 64 (1972).
49. T. Kurosawa, *J. Phys. Soc. Jpn. Suppl.* 21, 424 (1966).
50. A. D. Boardman, W. Fawcett, and H. D. Rees, *Solid State Commun.* 6, 305 (1968).
51. R. C. Curby and D. K. Ferry, *Phys. Status Solidi* (a) 15, 319 (1973).
52. D. K. Ferry, *Phys. Rev. B* 18, 7033 (1978).
53. M. A. Littlejohn, J. R. Hauser, and T. Glisson, *Appl. Phys. Lett.* 26, 625 (1975).
54. D. K. Ferry, *Phys. Rev. B* 12, 2361 (1975).
55. D. K. Ferry and J. R. Barker, *Phys. Status Solidi* (in press).
56. R. J. Zwanzig, *J. Chem. Phys.* 40, 2527 (1964).
57. M. S. Green, *J. Chem. Phys.* 20, 1281 (1952); 22, 398 (1954).
58. C. Canali, G. Ottaviani, and A. Alberigi-Quaranta, *J. Phys. Chem. Solids* 32, 1707 (1971); D. K. Ferry and J. R. Barker (to be published).
59. D. K. Ferry and J. R. Barker, *Solid State Electron.* 23, 545 (1980).
60. R. Kubo, *J. Phys. Soc. Jpn.* 12, 570 (1957).
61. H. Mori, *Phys. Rev.* 112, 1829 (1958).
62. R. Zwanzig, *Phys. Rev.* 124, 983 (1961).
63. J. R. Barker, *J. Phys. C Solid State Phys.* 6, 2663 (1973).
64. K. K. Thornber, *Solid State Electron.* 21, 259 (1978).
65. J. R. Barker, *Solid State Electron.* 21, 267 (1978).

66. B. Hoeneisen and C. A. Mead, *Solid State Electron.* **15**, 819 (1972).
67. D. K. Ferry and J. R. Barker, *J. Phys. Chem. Solids* (in press).
68. K. Blotekjar, *IEEE Trans. Electron Devices* **ED-17**, 38 (1970).
69. K. Blotekjar and E. B. Lunde, *Solid State Phys.* **35**, 581 (1969).
70. P. S. Cheung and C. J. Hearn, *Electron. Lett.* **8**, 79 (1972).
71. H. Kroemer, *Solid State Electron.* **21**, 60 (1978).
72. H. L. Grubin, D. K. Ferry, J. R. Barker, M. A. Littlejohn, T. H. Gilson and J. R. Hauser, *Proc. Biennial Electron. Eng. Conf., 7th, Cornell Univ., Ithaca, New York* (1979).
73. I. B. Bott and W. Fawcett, *Adv. Microwaves* **3**, 223 (1968).
74. H. L. Grubin, D. K. Ferry, and J. R. Barker, *Proc. IDEM* p. 394, IEEE Press, New York, 1979.
75. P. A. Rolland, E. Constant, G. Salmer, and R. Rauquemberque, *Electron. Lett.* **15**, 374 (1979).
76. Y. Wang, *Int. J. Electron.* **47**, 49 (1979).
77. W. Shockley, *Solid State Electron.* **2**, 35 (1961).
78. K. Hess, Private communication (1980).
79. H. F. Ivey, *Adv. Electron. Electron Phys.* **6**, 138 (1954).
80. M. S. Shur and L. F. Eastman, *IEEE Trans. Electron Devices* **ED-26**, 1977 (1979).
81. R. Papondar, "Electrical Phenomena in Gases." American Elsevier, New York, 1965.
82. D. K. Ferry, J. R. Barker, and H. L. Grubin, *IEEE Electron Devices Lett.* **EDL-1**, 209 (1980).
83. W. Frensly, *IEEE Electron Devices Lett.* **EDL-1**, 137 (1980).
84. I. K. Yanson, *Proc. Int. Conf. Low Temp. Phys., 14th* Vol. 4, p. 230. North Holland Publ., Amsterdam, 1975.
85. L. Reggiani, "Physics of Nonlinear Transport in Semiconductors" (D. K. Ferry, J. R. Barker, and C. Jacoboni eds.), pp. 243–254. Plenum Press, New York, 1980.
86. J. R. Barker and D. K. Ferry, *Solid State Electron.* **23**, 531 (1980).
87. W. Kohn and J. M. Luttinger, *Phys. Rev.* **108**, 590 (1957); **109**, 1892 (1958).
88. J. R. Barker, *in* "Physics of Nonlinear Transport in Semiconductors" (D. K. Ferry, J. R. Barker, and N. C. Jacaboni, eds.). Plenum Press, New York, 1980.
89. G. V. Chester, *Rep. Prog. Phys.* **26**, 411 (1963).
90. R. Kubo, *Rep. Prog. Phys.* **29**, 225 (1966).
91. J. M. Luttinger, "Mathematical Methods in Solid State and Superfluid Physics." Oliver and Boyd, London, 1968.
92. R. Peierls, *in* "Lecture Notes in Physics 31: Transport Phenomena." (J. Ehlers, K. Hepp, and H. A. Weidenmuller, eds.). Springer-Verlag, Berlin and New York, 1974.
93. E. Cohen and W. Thirrig, The Boltzmann Equation, *Acta Phys. Aust. Suppl. 10* (1973).
94. E. T. Jaynes, *Proc. Delaware Seminar Foundations Phys.* (M. Bunge, ed.). Springer-Verlag, Berlin and New York, 1967.
95. A. Ron, *Nuovo Cimento* **34**, 1511 (1964).
96. P. J. Price, *IBM J. Res. Dev.* **10**, 395 (1966).
97. S. Fujita, "Non-Equilibrium Statistical Mechanics." Saunders, Philadelphia, Pennsylvania, 1966.
98. D. K. Ferry and J. R. Barker, *Solid State Commun.* **30**, 361 (1979).
99. J. R. Barker, *in* "Handbook of Semiconductors" (W. Paul, ed.). North-Holland Publ., Amsterdam, 1979.
100. E. P. Wigner, *Phys. Rev.* **40**, 749 (1932).

101. C. H. Scott and E. J. Moore, *Physica* **62**, 312 (1972).
102. G. A. Sai-Halasz, *Proc. Int. Conf. Phys. Semicond., 14th, Edinburgh, Scotland* p. 21. Intitute of Physics, London, 1978.
103. H. Haken, "Synergetics." Springer-Verlag, Berlin and New York, 1978.
104. J. S. Turner, *in* "Lecture Notes in Physics 28: Statistical Physics," p. 248. Springer-Verlag, Berlin and New York, 1974.
105. J. R. Barker and D. K. Ferry, *Proc. Int. Conf. Cybernet. Soc.* 79CH 1424-1 SMC, p. 762. IEEE Press, New York, 1979.
106. M. N. Ransom and R. Saeks, The connection function—Theory and applications, *Int. J. Circuit Theory Appl.* **3**, 5–21 (1975).
107. S. H. Sangani, R. Saeks, and S. R. Liberty, A spectral theoretic approach to faulty analysis in linear sequential circuits, *J. Franklin Inst.* **302**, 239–258 (1976).
108. R. R. Troutman: VLSI limitations from drain-induced barrier lowering, *IEEE Trans. Electron Devices* ED-26, 461–469 (1979).
109. J. E. Moyal, *Proc. Cambridge Philos. Soc.* **45**, 99 (1949).
110. A. Messiah, "Quantum Mechanics," Vol. 1, p. 442. Wiley (Interscience), New York, 1961.
111. J. P. Nougier, *in* "Physics of Nonlinear Transport in Semiconductors" (D. K. Ferry, J. R. Barker, and C. Jacoboni, eds.). Plenum Press, New York, 1980.
112. R. Fauquemberque, J. Zimmermann, A. Kaszynski, E. Constant, and G. Micoondes, *J. Appl. Phys.* **51**, 1065 (1980).
113. M. Lax, *Rev. Mod. Phys.* **32**, 25 (1960).
114. M. Lax, *Rev. Mod. Phys.* **38**, 359 (1966).
115. M. Lax, *Rev. Mod. Phys.* **38**, 541 (1966).
116. N. G. van Kampen, *in* "Fluctuation Phenomena in Solids" (R. E. Burgess, ed.), pp. 139–179. Academic Press, New York, 1965.
117. P. Price, *IBM J. Res. Dev.* **3**, 191 (1959); *in* "Fluctuation Phenomena in Solids" (R. E. Burgess, ed.), pp. 355–380. Academic Press, New York, 1965.
118. A. Einstein, *Ann. Phys.* **17**, 549 (1905).
119. G. E. Uhlenbecr and L. S. Orstein, *Phys. Rev.* **36**, 823 (1930).
120. E. Montroll, *in* "Lectures in Theoretical Physics" (W. E. Brittin, B. W. Downs, and J. Downs, eds.), p. 221. Wiley (Interscience), New York, 1961.
121. A. Widom, *Phys. Rev. A* **3**, 1394 (1971).
122. R. Kubo, *in* "Lecture Notes in Physics 31: Transport Phenomena" (J. Ehlers, K. Hepp, and H. A. Weidenmuller, eds.), p. 75. Springer-Verlag, Berlin and New York, 1974.
123. J. W. Dufty and J. A. McLennan, *Phys. Rev. A* **9**, 1266 (1974).
124. Y. Pomeau and P. Resibois, *Phys. Rep.* **19**, 63 (1975).
125. J. P. Nougier, *in* "Noise in Physical Systems" (D. Wolf, ed.), p. 72. Springer-Verlag, Berlin and New York, 1978.
126. M. Kac, *Am. Math. Mon.* **54**, 295 (1947).
127. W. E. Alley and B. J. Alder, *Phys. Rev. Lett.* **43**, 653 (1979).
128. H. Scher and E. W. Montroll, *Phys. Rev. B* **12**, 2455 (1975).
129. R. Zwanzig and M. Bixon, *Phys. Rev. A* **2**, 2005 (1970).
130. J. L. Doob, *Ann. Math.* **43**, 351 (1942).
131. H. Mori, *Prog. Theor. Phys.* **33**, 423 (1965).
132. V. L. Gurevich, *Sov. Phys. JETP* **16**, 1252 (1963).
133. D. K. Ferry, *in* "Handbook of Semiconductors" (W. Paul, ed.), Vol. 1. North-Holland Publ., Amsterdam (in press).
134. P. A. Lebwohl and P. J. Price, *Solid State Commun.* **9**, 1221 (1970).

135. D. K. Ferry, *Phys. Lett.* **78A**, 375 (1980).
136. D. K. Ferry, (to be published).
137. R. Grondin, Private communication (1980).
138. C. Canali, C. Jacoboni, G. Ottaviani, and A. Alberigi-Quaranta, *Appl. Phys. Lett.* **27**, 278 (1975).
139. R. Kubo, *Rep. Prog. Phys.* **29**, 255 (1966).
140. P. Das, D. K. Ferry, and A. H. Barr, *Surface Sci.* **73**, 147–155 (1978).

PHYSICS AND MODELING CONSIDERATIONS FOR VLSI DEVICES

H.L. Grubin & J.P. Kreskovsky

Scientific Research Associates, Inc.
Glastonbury, Connecticut 06033

G.J. Iafrate

U.S. Army Electronics Technology & Devices Laboratory
Fort Monmouth, New Jersey 07703

D.K. Ferry

Colorado State University
Fort Collins, Colorado 80521

R.F. Greene

Naval Research Laboratory
Washington, D.C. 20375

ABSTRACT

Recent results concerning spatial and temporal transport in submicron devices identify significant aspects of the role of boundary conditions, scaling for suggesting new materials, and structural device changes. These results are discussed as a means of achieving high speed and high frequency devices.

INTRODUCTION

There are several recent results concerning spatial and temporal transport in submicron devices that are likely to have an impact on the design of future high frequency sources. These results, which emerge from Monte Carlo, and momentum moment equation solutions to the "Wigner-Boltzmann" quantum transport equation[1]

$$\frac{\partial f_w}{\partial t} + \frac{p}{m}\frac{\partial f_w}{\partial x} - \frac{2}{h}\left\{\sin\frac{h}{2}\left(\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)\right\}V(x)f_w(x,p) = \left(\frac{\partial f_w}{\partial t}\right)\text{coll}, \quad (1)$$

and the Boltzmann transport equation (BTE)

$$\frac{\partial f}{\partial t} + \frac{p}{m}\frac{\partial f}{\partial x} + \dot{p}\frac{\partial f}{\partial p} = \left(\frac{\partial f}{\partial t}\right)\text{coll}, \quad (2)$$

identify crucial aspects of boundary conditions, the role of scaling in choosing suitable materials, and the significance of alterations in otherwise simple device structures for achieving high speeds. These topics are reviewed below. (It is noted: In equation (1) the position gradient in the brackets operates only on the potential energy, f_w is a single coordinate and momentum distribution function

$$f_w(x,p) = \frac{1}{2\pi h}\int dy\Psi^*\left(x + \frac{y}{2}\right)\Psi\left(x - \frac{y}{2}\right)e^{ipy/h} \quad (3)$$

and $\Psi(x)$ represents the state of the system in the coordinate representation).

SCATTERING MODIFICATIONS TO BALLISTIC TRANSPORT

Ballistic transport implies carrier transport unimpeded by interactions (electrostatic, or otherwise) with other carriers, or with scattering events. The extent to which scattering centers are sensed by transiting electrons is therefore of significance. The first set of calculation shown, Figure 1 (Ref. 2), represents scattering events in GaAs for a collection of electrons entering a uniform field region with an initial energy of approximately 0.30ev. Intervalley Γ-L energy separation for this Monte Carlo calculation is 0.33ev. It is seen that, with the exception of very high fields, approximately 50% of the carriers are unscattered over the first 500°A.

The velocity versus distance curves for this calculation are displayed in Figure 2 (Ref 2), and it is seen that high speeds over useful distances can be achieved even in regions where a high number of scattering events has occured. The optimum conditions for this appear to require moderate fields, generally near the threshold field for electron transfer, and moderate injection energies. The dependence on the latter is displayed in Figure 3 (Ref.2), where the lowest achievable velocities over a distance of 1500°A are for entry electrons with zero initial velocity. It is also seen that an optimum

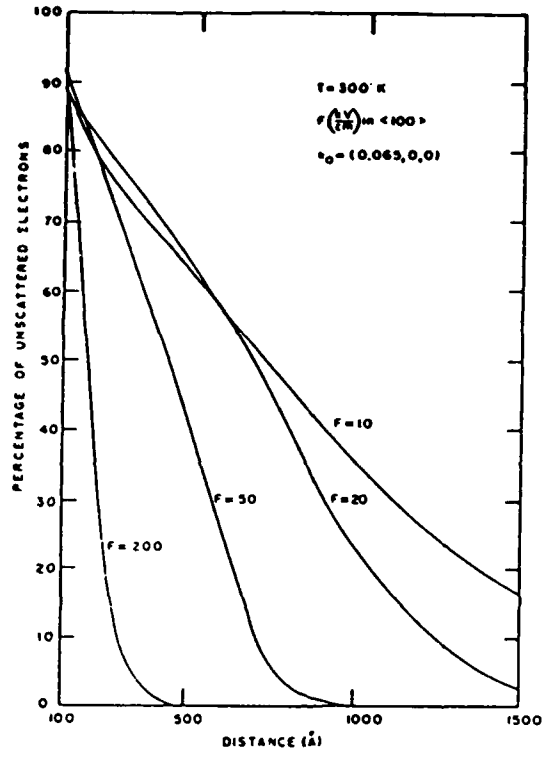


Figure 1. Percentage of unscattered electrons versus distance. Entrance electrons have a finite energy. From reference 2, with permission.

entry energy exists; for energies near the Γ-L separation, scattering quickly reduces the net drift velocity. The optimum conditions identified by these calculations imply that voltage control near the entrance boundary must be near 50mv.

The examples of Figures 1 through 3 are for uniform fields, and carriers subject to sudden changes in field. Studies in which spatial gradients accompany nonuniform fields are more recent, several of which are considered at this Workshop. One aspect is considered below.

SPATIAL AND TEMPORAL TRANSIENTS

A relatively direct way of handling spatial and temporal transients is through moments of the transport equations (1) and (2). Little has been done with the moments to the Wigner - Boltzmann equation and the following deals exclusively with BTE moments. The moments of the BTE form an infinite hierarchy of moment equations. Each equation introduces a higher order moment not defined by the

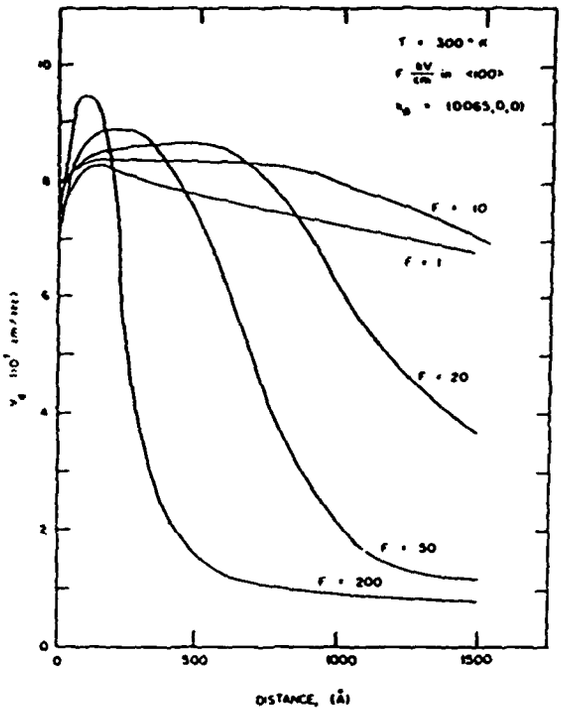


Figure 2. Velocity versus distance for electrons with a finite entrance energy subject to a sudden change in electric field. From reference 2, with permission.

given set of balance equations. The most common form of the balance equations is obtained by assuming a distribution function of the form

$$f = f_{oi} = \frac{n_i}{T_i^{3/2}} \exp - \{\frac{m(u-v)^2}{2 k_o T_i}\} \tag{4}$$

for each species. More generally (Ref. 4)

$$f = (1+A_k\frac{\partial}{\partial u_k} + A_{kl}\frac{\partial^2}{\partial u_k \partial u_l} + A_{klm}\frac{\partial^3}{\partial u_k \partial u_l \partial u_m} + \ldots) f_o \tag{5}$$

where it is insisted that the first term in the expansion gives the correct local values of density, velocity and energy. The coefficients in the above expansion are inversely proportional to density and are model dependent. The first correction to the local equilibrium balance equations, the so-called hydrodynamic approximation yields

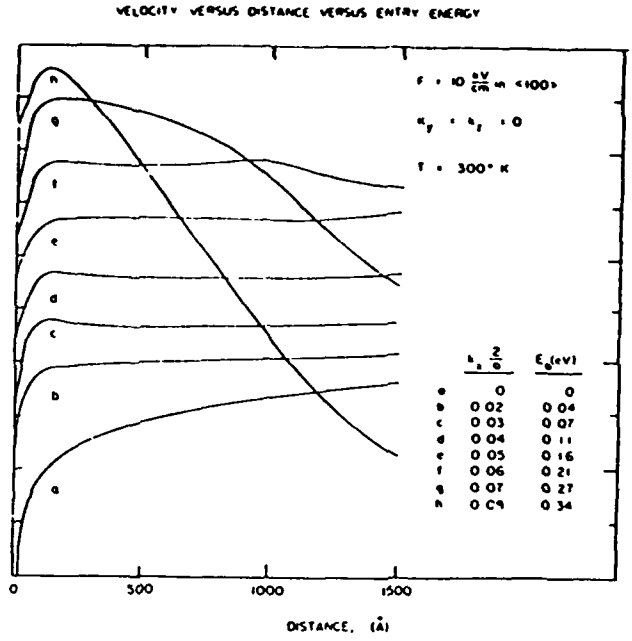


Figure 3. Velocity versus distance for electrons with varying entrance energies subject to a sudden change in electric field. From reference 2, with permission.

$$\frac{\partial n_i}{\partial t} + \nabla \cdot (\vec{v}_i n) = \left(\frac{\partial n_i}{\partial t}\right)_{coll} \tag{6}$$

$$\frac{\partial \vec{P}_i}{\partial t} + \nabla \cdot (\vec{v}_i \vec{P}_i) = -en_i\vec{F} - \text{grad } n_i k_o T_i - \boxed{\text{div } \overleftrightarrow{\sigma}} + \left(\frac{\partial P_i}{\partial t}\right)_{coll} \tag{7}$$

$$\frac{\partial}{\partial t} W_i + \nabla \cdot \vec{v}_i W_i = -en_i \vec{F}.\vec{v}_i - \nabla \cdot v_i n_i k T_i - \boxed{\text{div } \overleftrightarrow{\sigma}^i \, xv_i} \tag{8}$$

$$+ \boxed{\text{div } (\kappa \text{ grad } T)} + \left(\frac{\partial W_i}{\partial t}\right)_{coll}$$

where $P_i = n_i m_i \vec{v}_i$, $W_i = n_i \left\{\frac{1}{2} m_i v_i^2 + \frac{3}{2} n_i k_i T_i\right\}$, $\overleftrightarrow{\sigma}$ is a stress tensor arising from nonuniform velocity distributions, and κ is the thermal conductivity. The stress term is dissipative in that when a nonuniform velocity distribution is impressed on an electron stream there will be reactive forces tending to smooth them out. Detailed information regarding suitable values for these terms are not available, and they are regarded as phenomenological entries; they and their modification by the collision terms are to be studied.

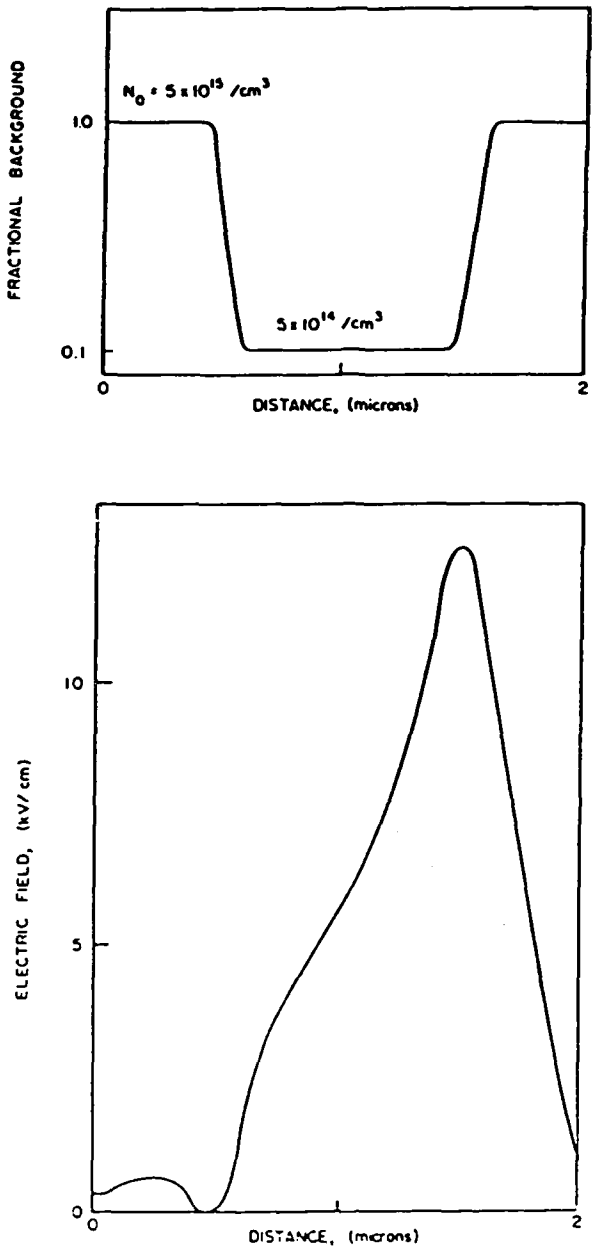


Figure 4. a. Doping distribution for an N^+-N^--N^+ advice. b. Self consistent electric field for a bias of 1 volt. (From Ref. 4, with permission).

The boxed terms above represent contributions arising from the nonspherical nature of the distribution function. The underlined terms are ignored in the drift and diffusion approximation. Figure 4 shows the result of a solution to the balance equations for two level transfer for a two-terminal device whose structure is receiving considerable attention as an illustration of velocity overshoot. The structure considered here consists of a low doped region L<1μm sandwiched between two regions of higher doping concentration. The structure is subject to a bias of 1 volt. Four interfaces are involved but most attention is focused on the two interior faces where significant gradients in field occur, all associated with differences between the background and mobile carrier density. The carrier concentration is displayed in figure 4c, for the case of injecting cathode contacts (Ref. 4). It is seen that (a) significant charge injection is present in the low doped region, and (b) electron transfer occurs near the downstream section of the low doped region and is significant in the downstream N^+ region. It is noted that the field dependence of carrier density in the Γ valley, $N_\Gamma(F)$ is not the same as that associated with uniform field steady state values. Spatial gradients result in a spatial lag.

Also shown in figure 4c is the drift velocity of carriers in the Γ-valley, and it is apparent that the carriers acquire speeds considerably in excess of the steady state peak velocity of electrons in GaAs. These velocities are not, however, much different in value than those associated with Γ valley electrons under uniform field conditions. The velocity shown in figure 4c is not, however the same as the carrier velocity computed under uniform field conditions. The velocities computed here include temperature as well as momentum gradients, and are properly defined as the current flux/carrier density. Thus, increased velocity here is often associated with decreased carrier density, through current continuity. The key to all device design is that high velocities must be accompanied by only marginal transfer out of the high mobility section of the conduction band. This will require device lengths shorter than the active region of the device shown in this figure. (Modifications to this statement arising from variations in material parameters are considered in Ref. 4).

The calculations of figure 4 show the situation where a high carrier velocity occurs near the downstream edge of the low doped region. Often high velocities are required near the upstream boundary as indicated by figures 1 through 3. Figure 5 (Ref. 6) shows a spatially dependent result for carriers subjected to a prespecified value of electric field. The calculation is for silicon and the highest carrier velocities occur near the 0.4μm boundary. An accompanying plot of carrier density (Ref. 6, figure 9) shows the lowest level of carrier density in the region of highest velocity.

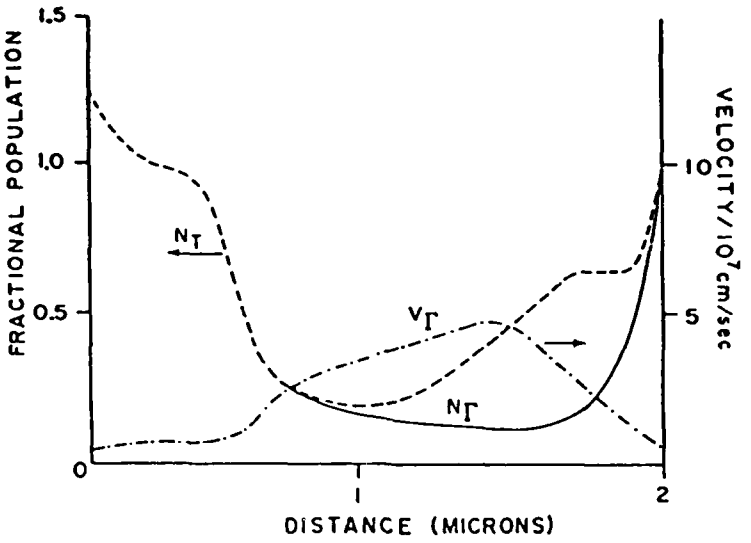


Figure 4C. Fractional population and central value velocity. (From Ref. 4., with permission.)

Coupling the above results to the need for moderate fields for high speed operation a useful multiterminal device very likely will require the presence of a local moderate field region that "kicks" electrons into high speed regions. The simplest high field injecting region could come from local charge depletion or "notches" (Fig. 6). Note that in Figure 6 the gate is treated generically. It could be a metal Schottky contact, a p^+ AlGaAs layer, etc.

SCALING

In the absence of, or in concert with structural variations in devices it is also necessary to examine changes in material parameters. The change in material parameters should have as its goal high mobility and high characteristic velocities. Thornber (Ref. 8) provided a general set of guidelines for choosing material parameters through an alteration of the scattering rates, one of which is duscussed below.

The collision term in the Boltzmann transport equation is

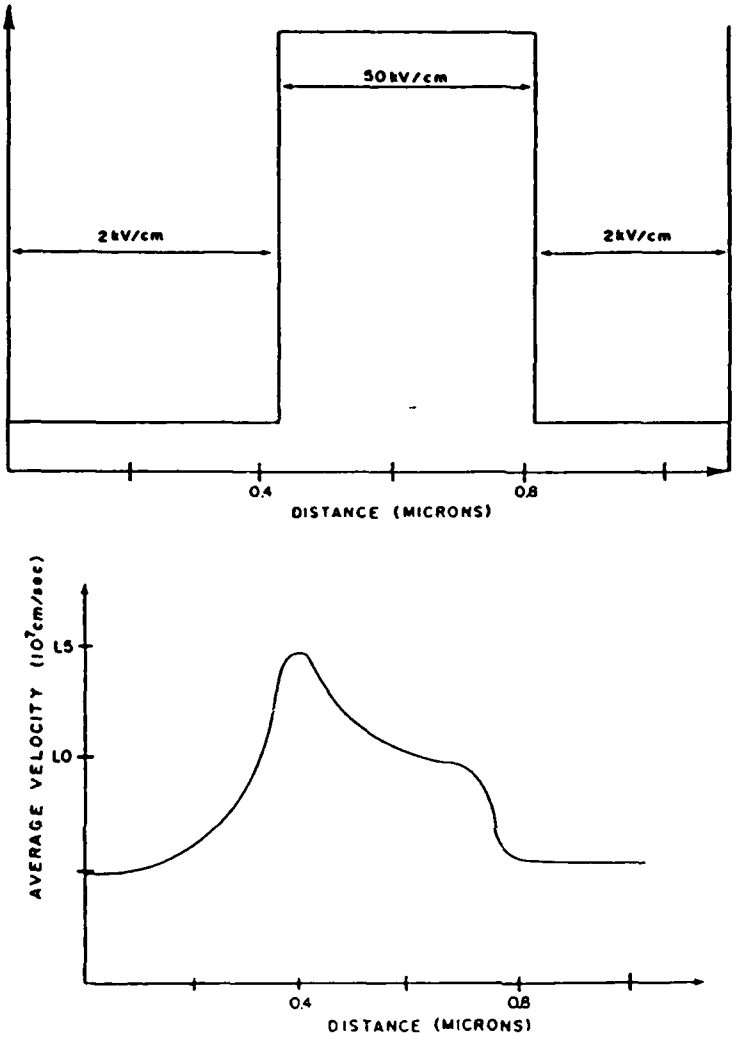


Figure 5. Variation of carrier velocity in silicon for electrons subject to the illustrated field profile. From Ref. 6 with permission

$$\left(\frac{\partial f}{\partial t}\right)_{coll} = \int d^3p' \{f(p')W(p',p) - f(p)W(p,p')\} \qquad (9)$$

where $W(p',p)$ is the total scattering rate from $\vec{p}$ to $\vec{p}'$. Thornber suggested several scaling changes in the BTE, to alter the drift velocity.

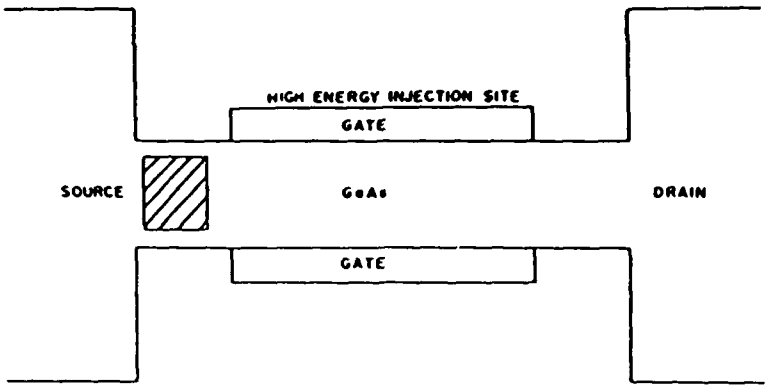


Figure 6. Schematic of a high energy injection region in a gated submicron structure.

$$v = \int d^3p f \vec{p}/m \int d^3p f. \tag{10}$$

The one illustrated below involves uniformly altering the scattering rate by a constant Γ, and results in the following relation

$$v_\Gamma(x,t,F) = v(\Gamma x, \Gamma t, F/\Gamma) \tag{11}$$

Where the right-hand side is the unscaled velocity. In this scaling mobility is altered but saturated drift velocity is not.

A dramatic consequence of the effects of scaling is illustrated below, but it is necessary to note that the possibility of relevant material scaling over a broad range of electric field values is remote. Rather, it is more likely to be appliable over a restricted range of field values. This is illustrated for GaAs. For GaAs, over a field range of approximately 3 → 15 kv/cm, intervalley Γ-L coupling is a dominant scattering mechanism. Figure 7 shows the velocity field characteristic for Γ-L-X ordering with three different values of the Γ-L deformation potential. All these valleys are taken as parabolic with N_o=0, while all other material parameters are those of Littlejohn, et al. (Ref.9). (We note that similar values for material constants yield differences between Monte Carlo and balance equation calculations.) In Figure 7 the fields at points 1, 2 and 3 occur at values 5.6 kv/cm, 8.2 kv/cm and 12 kv/cm, respectively. The dominant Γ valley total momentum scattering rates are 4.78×10^{12}/sec., 7.32×10^{12}/sec., and 10.0×10^{12}/sec., respectively. The ratios F/Γ are 1.17×10^{-12}kv-s/cm, 1.12×10^{-12}kv-s/cm and 1.20×10^{-12}kv-s/cm, respectively in general agreement with the rule of Equation 11.

Figure 8 illustrates the consequences of uniform scaling on velocity overshoot. The solid curve displays overshoot for a gallium arsenide element subject to a field of 27kv/cm (Ref. 10). The dashed

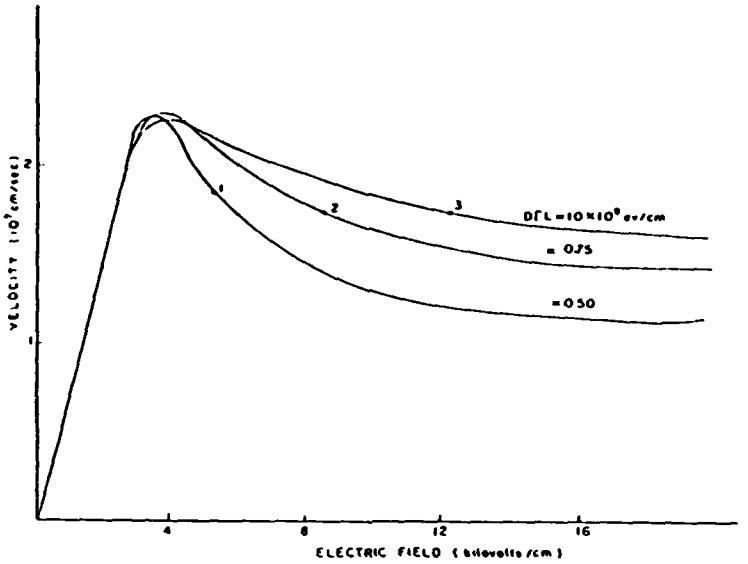


Figure 7. Steady state velocity versus field for Γ-L-X ordering as obtained from balance equations. Different curves are for different values of the deformation potential coupling constant for intervalley Γ-L scattering.

(dotted) curve is a sketch of a scaled curve for Γ=2(Γ=0.5) and a field of 54kv/cm (13.5kv/cm). For Γ=2, high overshoot velocities occur over a short period of time and require high bias levels. By reducing Γ to values below 1.0 high overshoot velocities are retained for a longer period of time and at lower values of field than for GaAs. This result is highly significant insofar as it suggests higher transit velocities for extended temporal scales, and indicates a direction of material selection for high frequency sources.

GENERAL CONCLUSION

It has been known since 1969 that the compound semiconductors, GaAs in particular, are theoretically capable of providing high frequency, near 100GHZ oscillations. Overshoot has also been known since that time (Ref. 11). The difficulty in attaining high frequency three-terminal and sometimes two-terminal operation has over the past decade been attributed to inadequate contact regions (Ref. 12), material preparation, etc. Many of the high frequency problems were thought to be reduced by going to small dimensions, where in addition to shorter transit lengths the benefits of overshoot would emerge. While it is still too early to state how these benefits can be implemented in practice, it is clear that special contacts or injection regions

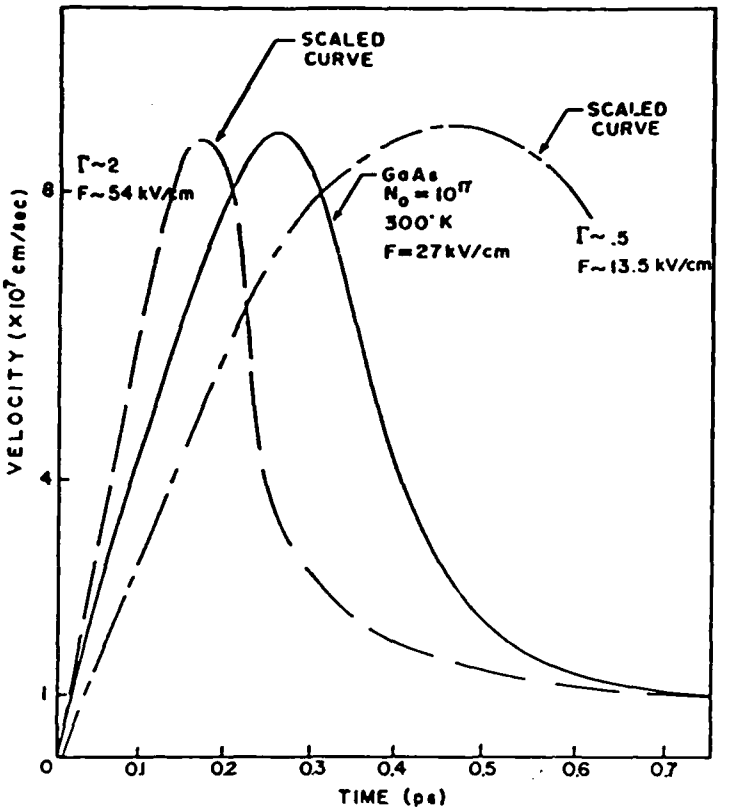


Figure 8. Transient velocity characteristics for scaled and unscaled scattering rates. Solid curve is from reference 10.

are required. It appears, therefore, that the 1969 questions as to why devices are not operating closer to theoretical limits is still valid today. The dual solution approaches of the last decade also appear equally valid today, where the pursuit of novel device structures emphasizing the boundary role is continuing, while simultaneously new material directions are sought.

ACKNOWLEDGEMENT

This work was supported by the Office of Navel Research, and the Army Research Office, to whom we are grateful.

REFERENCES

1. G.J. Iafrate, H.L. Grubin and D.K. Ferry, J. de Physique C7, 307 (1981).
2. G.J. Iafrate, R. Malik, K. Hess and J. Tang. To be published.
3. See, e.g., A. Sommerfeld, Thermodynamics and Statistical Mechanics. Sec. 43. Adacemic Press, NY (1956).
4. H.L. Grubin and J.P. Kreskovsky, Surface Science 39 (1983).
5. H.L. Grubin, G.J. Iafrate and D.K. Ferry, J. de Physique C7, 201 (1981).
6. E. Constant and B. Boittiaux, J. de Physique C7,73 (1981).
7. R.K. Cook and J. Frey, IEEE Trans. Electron Devices, Ed-29 (1982).
8. K. Thornber, J. Appl. Phys. 51, 2127 (1980).
9. M.A. Littlejohn, J.R. Hauser and T.H. Glisson, J. Appl. Phys. 48, 4587 (1978).
10. H.L. Grubin and D.K. Ferry, J. Vac. Sci. & Tech. 19, 540 (1981).
11. P.N. Butcher and C.J. Hearn, Electron. Lett. 4, 459 (1968).
12. M.P. Shaw, H.L. Grubin and P.R. Solomon. The Gunn-Hilsum Effect, Academic Press, NY 1979.

Reprint 10

CONSIDERATIONS ON THE FINITE, VERY-SMALL SEMICONDUCTOR DEVICE AND SUPERLATTICE ARRAYS

David K. Ferry, Robert O. Grondin, and
Robert K. Reich
Colorado State University, Fort Collins, CO

Harold L. Grubin
Scientific Research Associates, Glastonbury, CT

Gerald J. Iafrate
Electronics Technology and Device Laboratory
Fort Monmouth, NJ

ABSTRACT

The size of expected future very-small devices will result in their being strongly coupled to the environment in which they are located. In this paper, we examine several of the limitations on the device physics that can be expected to arise in these structures, including the role of the environment and its effect on ballistic transport. In addition, we look at submicron arrays of devices in a lateral surface superlattice (LSSL) and examine transport in such an LSSL through a Wigner function and Monte Carlo approach.

INTRODUCTION

Many people have examined the limit to which semiconductor devices can be scaled downward[1-3]. While small devices in the range 0.1 - 0.3 μm gate length have been made[4], problems such as interconnections, electromigration and thermomigration of metallization, and power density within the device strongly affect the packing density and device size that can be achieved[5]. Yet, no one has evaluated the operation and performance of very-small semiconductor devices, i.e. those that can be conceived in the sub-0.1 μm size range, and the interactions within arrays of such devices. One reason for this is that the transport within such a device cannot be treated in isolation. Because of the size of such a very-small device, it is coupled strongly

to the environment in which it is located. The basic transport equations cannot be separated from their casual boundary conditions and both of these factors must be modified to account for the influence of the device environment[6]. In previous work, some of us have presented a formalism to address this[4,7], but this formalism remains untried due to its inherent complexity. In the present work, we wish to examine qualitatively several of the limitations on the device physics that can be expected to arise in the very-small semiconductor device, particularly with respect to the finiteness of source and sink regions, contact regions and barriers, and the influence of these latter quantities on the ballistic transport that may exist in the device[8,9]. In particular, we examine these effects through the use of discrete area-preserving phase-space maps which display the characteristics of transport in a generic device. In addition, we treat the cooperative interactions that can arise in densely packed arrays of devices.

BALLISTIC TRANSPORT

Potential barriers within the device can play a significant role in the quantum ballistic transport of carriers through it. Such barriers are found in very-small devices, for example, to confine carriers to the active region[3], and are an intimate part of devices such as the planar-doped barrier transistor[10], tunneling barrier devices[11], real-space transfer devices[12], or superlattice avalanche photodiodes[13]. When a barrier is present at the contact region, care must be taken to adequately handle turning-point reflection of the electrons from this barrier. Even when the electron has sufficient energy to pass over the barrier, there is a well-known quantum mechanical reflection at the barrier interface. If the potential barrier is smooth, i.e. introduces a transition over many wavelengths, the reflection and wave function matching can be handled by well-known approximation techniques such as the WKB approximation[14], in which the potential barriers represent turning points for a near-classical path. In the very-small device, however, the barriers are expected to be sharp on the scale of the electron wavelength and care must be exercised in matching wavefunctions and determining reflections. The reflection problem is further complicated in real-space transfer devices[12] due to the different band structure on either side of the barrier. Here, additional terms arise due to the spatial variation of the effective mass.

If the potential barriers are slowly varying on the scale of the wave packet, the trajectories are largely those of the classical motion. Even if this is not the case, as we expect for the very-small device, nearly semiclassical trajectories can be expected if the variation of the action is limited to a few low-order derivatives[15]. In semi-classical systems, the phase space of the classical motion forms a natural framework in which to examine problems such as these.

While classical one-dimensional transport appears to be basically simple, there exists recent work that suggests that even this simple problem contains a number of unexpected subtleties[16]. In this paper, we begin to investigate the role of barriers in submicron devices through the use of such finite, area-preserving phase-space mappings of classical dynamics. The use of such mappings reveals a variety of complicated structure[16,17], and has had recent success in explaining the cause of excess noise in Josephson junction parametric amplifiers[18,19]. The results that we obtain indicate that if these mappings are applicable to the VLSI scale, then present concepts of submicron transport may require substantial generalization.

The types of subleties to which we are referring are best illustrated by Fig. 1. There, we are using an area-preserving mapping of particle position (horizontal) and momentum (vertical) within a device active region bounded by two Gaussiam potentials V(x) (overlaid in the figure). In addition, an electric field has been applied.

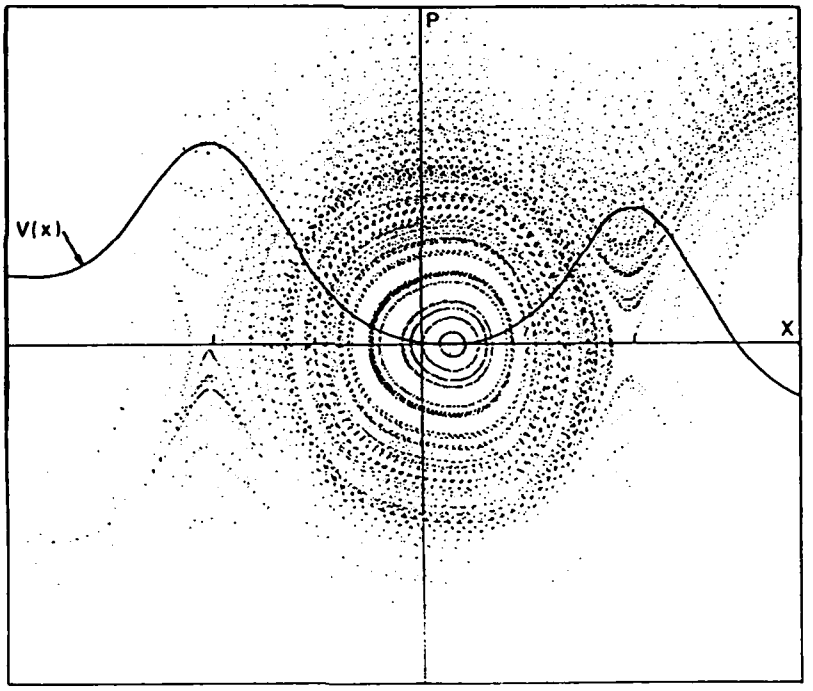


Figure 1. The phase-spacing mapping of an initial uniform distribution of electrons in a generic very-small semiconductor device. The contact regions are represented by a twin-gaussian potential V(x). The horizontal direction is position while the vertical direction is momentum.

The canonical mappings are area-preserving since we are considering a collisionless (conservative) system and looking at the ballistic transport. The classical differential equations of motion are

$$\dot{p} = - \frac{\partial V(x)}{\partial x} \ , \qquad \dot{x} = + p/m \ , \tag{1}$$

and the discreet classical area-preserving maps are

$$p_{n+1} = p_n - T \left. \frac{\partial V}{\partial x} \right|_{x_{n+1}} \ , \qquad x_{n+1} = x_n + T\, p_n/m \ . \tag{2}$$

In Fig. 1, the potential has been scaled so that the two potentials have a weak overlap. For small values of the total energy, resonant orbits occupy the central part of the figure. These orbits are in the region of the classically integrable motion. For larger values of the energy, however, the orbits are such that the particles are swept out of the well by the field. An energy dissipating collision can drop a high energy particle into the resonant region, thus trapping it within the structure. We expect these particles to contribute a diffusive component of current. Large angle scattering, however, can move a particle to a back-flowing orbit, which effectively causes a reflection of particles from the device input.

A number of interesting factors arise. The potential is generic, in the sense that it is similar in form not only to the devices mentioned above, but also to the sinusoidal force term in the Josephson junction devices[18,19]. In the latter case, interaction between devices can lead to a parametric pumping of the potential which yields period-doubling bifurcations and chaotic behavior. The system (2) is not truly generic though, as the set (x,p) are not proper action-angle variables. We have examined the behavior through potential pumping, with dissipation, through replacing the first equation of (2) with

$$p_{n+1} = p_n - T \frac{\partial V_o^{(x_{n+1})}}{\partial x} - T\gamma p_n + TF_o \sin(\Omega t_n) \ , \tag{3}$$

where V_o is the set of Gaussian potentials. The factor γ is an effective damping factor. In Fig. 2, we plot the (F_o,Ω) plane results. The curve is a separatrix below which a stable device results. Above the curve, the device is unstable. No period-doubling bifurcations within a single device are found, contrary to[18].

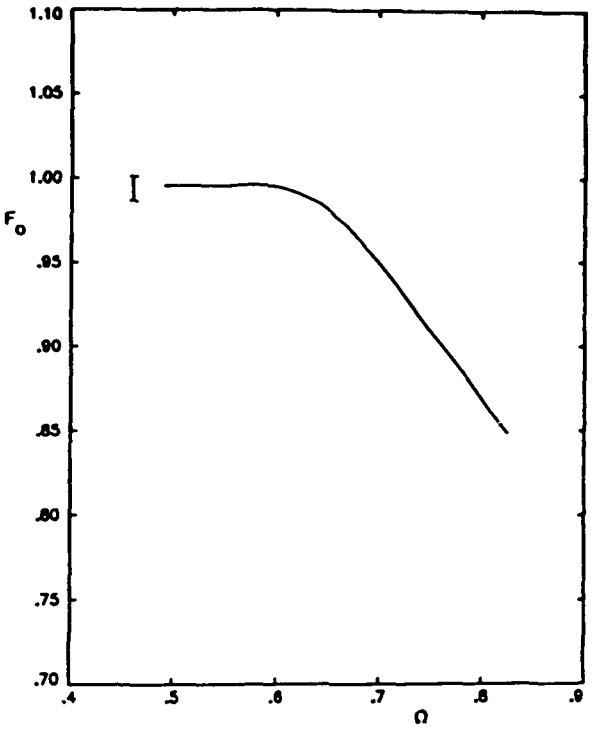


Figure 2. The separatrix in (F_o,Ω) space for parametric pumping of the device of Fig. 1. For values of (F_o,Ω) above the curve, electrons are ejected from the device, while for values below the curve the device is stable.

DEVICE ARRAYS

When device sizes begin to shrink toward the 0.1μm or less region, the line-to-line capacitance in dense device arrays begins to dominate the total node capacitance[4]. This parasitic capacitance leads to a direct device-device interaction outside of the normal circuit or architectural design. In conventional descriptions of LSI circuits, each device is assumed to behave in the same manner within the total system as it does when it is isolated. In the dense arrays discussed here, this will no longer be the case.

The possible device-device coupling mechanisms are numerous and include such effects as the capacitive coupling mentioned above, but also include such effects as wave function penetration or tunneling and charge spill-over. Formally, however, one may describe these effects on system and device behavior by assuming the simplest form of coupling. Arrays of devices, interacting in this manner, form a

lateral surface superlattice[20,21]. Lateral superlattices, in which the superstructure lies in a surface or heterostructure layer, offer considerable advantages for obtaining superlattice effects in planar technology. While a surface MOS structure is formally similar to an array of CCD devices[20,22], superlattices can also be fabricated through the use of electron and ion beam lithography and selective area epitaxial growth[21]. If the coupling is capacitive, then the limitation to a spacing less than the de Broglie wavelength is removed[4]. We have examined transport in such lateral surface superlattices (LSSL) through a Wigner function approach. Before proceeding, however, it is worth noting that the circuit theory view of LSSLs[4] is generically that of cellular automata[23]. Many of the image processing applications proposed for LSSLs[24] arise from the "games" aspect of cellular automata[25].

Superlattice structures give rise to sinusoidal energy minibands with relatively narrow widths. The shape of such bands results in interesting electrical transport properties. Lateral surface superlattices have cosinusoidal bands in two dimensions, with the third dimension quantized with discrete energy levels.

We have calculated the transport properties for such LSSLs from a Wigner function formulation. A complete integral equation is obtained for the Wigner function and must be solved to obtain the transport coefficients[26,27]. However, the form of the solutions can be found by taking the average velocity and energy from the first and second moments of an equivalent Wigner representation of the transport equation[28]. A constant electric field is applied in the plane of the sinusoidal bands. For simplicity, a constant relaxation time τ can be used. Solving these two equations simultaneously leads to the velocity and energy as functions of the field[29]

$$\langle v_z \rangle = (\varepsilon + \langle E_\perp \rangle - \langle E_o \rangle) \frac{(2\pi/\hbar L)^2 eF_z \tau}{1 + (2\pi e F_z \tau/\hbar L)^2} \quad , \tag{4}$$

$$\langle E \rangle = \varepsilon + \langle E_\perp \rangle + \frac{\langle E_o \rangle - \varepsilon - \langle E \rangle}{1 + (2\pi e F_z \tau/\hbar L)^2} \quad , \tag{5}$$

where $L = 2\pi/D$, D is the LSSL spacing, ε is one-half the band-width, and $\langle E_o \rangle$ and $\langle E_\perp \rangle$ are the equilibrium and transverse energies, respectively.

The velocity as a function of field has the same basic analytical expression as that obtained by Lebwohl and Tsu[30] except for the energy factor in front of the expression for the field. The difference is in effect caused by the different equilibrium distribution chosen. In the latter paper, the authors assumed the initial equilibrium dis-

tribution is very close to a zero temperature Fermi-Dirac function, while here the distribution is a finite temperature Wigner distribution and includes the physics of the energy band shape. Note that the velocity as a function of the electric field shows the negative differential mobility predicted earlier and also exhibits the general shape expected of the velocity-field curve.

To more exactly illustrate the mobility and negative differential conductivity, we have carried out a Monte Carlo calculation. The scattering processes are calculated for the model of Iafrate et al.[21], and thus are for the system of GaAs/GaAlAs. The scattering rates for acoustic and polar optical phonons have been obtained using a two dimensional density-of-states for cosinusoidal energy bands. A Van Hove singularity occurring in the density-of-states produces a singularity in the scattering rate, which was removed by including the self-energy corrections due to the phonons in the vicinity of the singularity. The widely-spaced discrete energy levels in the third dimension allows scattering and transport in that direction to be neglected.

In this surface superlattice, as in others, the conduction band splits into subbands. Here the lowest energy subband was nearly flat. Therefore, transport dominantly occurs in the next higher minibands. The satellite valleys and next subband are at energies of 0.2eV and 0.3eV, respectively, above the subband considered. Their contribution to the transport of electrons is insignificant since there are no intermediate energies through which the electrons can scatter to aid population of upper bands.

The overall transport properties of this system are calculated by an ensemble Monte Carlo technique. The results of the simulation are the velocity-field curves and are shown in Fig. 3. The lower curve results for a field applied along one of the (10) basis vectors of the square lattice array of cylinders while in the top curve the field is applied along a (11) direction. At low fields, both curves show a linear region as expected for most structures. At approximately 8-10 kV/cm the curves begin to bend over to the peak near 13 kV/cm. As the field is further increased, the velocity begins to decrease and for this model continues to decrease to zero as the field tends to infinity.

In summary, transport and scattering in a generic surface superlattice structure exhibits a negative differential mobility arising from Bloch oscillations[31]. This surface superlattice negative differential mobility is expected to be useful at much higher frequencies than that due to the conventional Gunn effect. Alternatively, related instabilities may be an ultimate limit on very large scale integration[3,4,22].

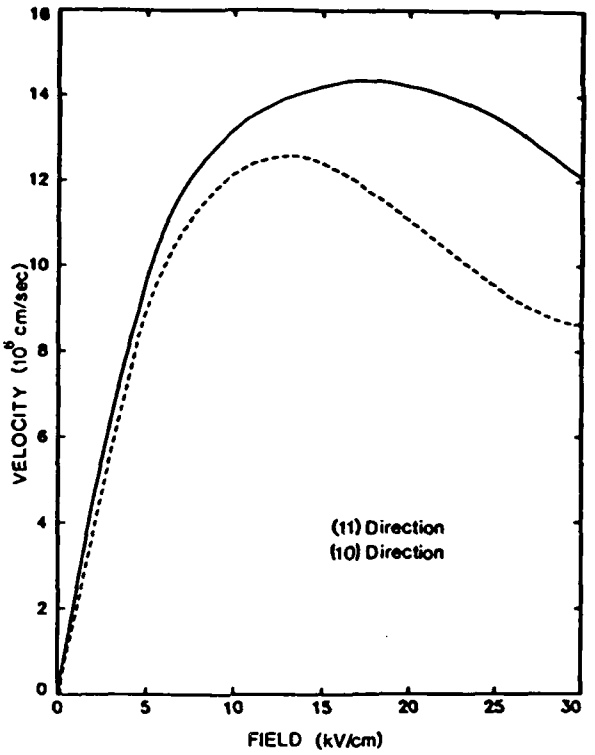


Figure 3. The drift velocity (as a function of electric field) for a LSSL at 300 K. The GaAs/GaAlAs model of ref. 21 has been used.

ACKNOWLEDGEMENT

This work was supported by the Office of Naval Research, the Army Research Office, and the North Atlantic Treaty Organization.

REFERENCES

1. B. Hoeneisen and C. A. Mead, Sol. State Electronics, 15, 819, 891 (1972).
2. J. T. Wallmark, in Inst. of Physics Conf. Series, 25, 133 (1974).
3. J. R. Barker and D. K. Ferry, Sol. State Electronics, 23, 519, 531 (1980).
4. These are reviewed in D. K. Ferry, Adv. in Electronics and Electron Physics, 58, 312 (1982).
5. R. W. Keyes, IEEE Trans. Electron Dev., ED-26, 271 (1979), and references contained therein.

6. M. P. Shaw, H. L. Grubin, and P. R. Solomon, "The Gunn Hilsum Effect", (Academic Press, New York, 1979).
7. J. R. Barker and D. K. Ferry, in "Proc. IEEE Conf. on Cybernetics, Science, and Society", 1979, pp. 762-5 (IEEE Press CH1424-1, 1979).
8. J. R. Barker, D. K. Ferry, and H. L. Grubin, IEEE Electron Dev. Letters, EDL-1, 209, (1980), and references contained therein.
9. D. K. Ferry, J. Zimmermann, and P. Lugli, IEEE Electron Dev. Letters, EDL-2, 228, (1981). See also the papers in IEEE Trans. Electron Devices, August (1981).
10. R. J. Malik, M. A. Hollis, L. F. Eastman, D. W. Woodard, C. E. C. Wood, and T. R. AuCoin, in Proc. Eighth Biennial Cornell Electr. Engr. Conf., Ithaca, N.Y., (1981).
11. P. A. Blakely, J. R. East, and G. I. Haddad, "Tunneling Transistors for Millimeter-Wave Operation", submitted for publication.
12. M. Keever, H. Shichijo, K. Hess, S. Banerjee, L. Witkowski, H. Morkoc, and B. G. Streetman, Appl. Phys. Letters, 38, 36, (1981).
13. F. Capasso, W. T. Tsang, A. L. Hutchinson, and G.F. Williams, Applied Physics Letters, 40, 38, (1982).
14. See, e.g., L. I. Schiff, "Quantum Mechanics", Third Ed., (McGraw-Hill, New York, 1968), pp. 269-79.
15. L. S. Schulman, "Techniques and Applications of Path Integration", (Wiley-Interscience, New York, 1981).
16. See, e.g., "Topics in Nonlinear Dynamics", Ed. by S. Jorna, A.I.P. Conf. Proc. 46 (American Institute of Physics, New York, 1978).
17. V. I. Arnold and A. Avez, "Ergodic Problems of Classical Mechanics", (Benjamin, New York, 1968).
18. B. A. Huberman, Applied Physics Letters, 37, 750, (1980).
19. N. F. Pedersen and A. Davidson, Applied Physics Letters, 39, 830, (1981).
20. R. T. Bate, Bull. Am. Phys. Soc. 22, 407 (1977).
21. G. J. Iafrate, D. K. Ferry, and R. K. Reich, Surf. Sci. 113, 485 (1982).
22. D. K. Ferry, Phys. Stat. Sol. (b) 106, 63 (1981).
23. A. W. Burks, "Essays on Cellular Automata" (Univ. Illinois Press, Urbana, 1970).
24. See J. R. Barker, these proceedings.
25. M. Gardner, Scientific American 224, 112 (Feb. 1971).
26. I. B. Levinson, Sov. Phys. JETP 30, 362 (1970).
27. D. K. Ferry, H. L. Grubin, and G. J. Iafrate, "Nonlinear Transport in Semiconductors", (Academic Press, in preparation).
28. G. J. Iafrate, H. L. Grubin, and D. K. Ferry, J. Physique 42, (Suppl. 10) C7-307 (1981).
29. R. K. Reich and D. K. Ferry, Phys. Letters A, in press.
30. P. A. Lebwohl and R. Tsu, J. Appl. Phys. 41, 2664 (1970).
31. R. K. Reich, R. O. Grondin, D. K. Ferry, and G. J. Iafrate, Phys. Letters A, in press.

Reprint 11

BOUNDARY SCATTERING EFFECTS IN HIGH FIELD TRANSPORT IN SUBMICRON STRUCTURES

R.F. Greene - U.S. Naval Research Laboratory
Washington, D.C. 20375

H.L. Grubin - Scientific Research Associates, Inc.
Glastonbury, Conn. 06033

G.J. Iafrate - U.S. Army Electronics Technology and Device
Lab., Ft. Monmouth, New Jersey 07703

ABSTRACT

The effects of boundary scattering in submicron structure transport at high electric field strengths are investigated in the context of the balance equations which arise from the Boltzmann transport equation. Integrated balance equations are presented into which surface dissipation can be inserted. The evaluation of the bulk and surface dissipation terms in the balance equations requires knowledge of the distribution function. We calculate this by introducing scattering boundary conditions appropriate for hot electrons and apply this to an analytic path-integral solution of the high field Boltzmann transport equation. The latter treats lattice and electron-electron scattering approximately. We show that surface scattering strongly perturbs and distorts the distribution function with corresponding effects on the transport process. The consequences of this for submicron device technology are discussed briefly.

INTRODUCTION

There are important strengths in both of the main approaches[1] to the theoretical study of high field transport effects in semiconductors: direct Monte Carlo simulation, and the use of the balance equations arising from averages over the Boltzmann transport equation (BTE). There is some merit in continuing the latter approach, e.g. because of what appears to be its ability to easily

treat transient space charge effects, etc., where self-consistency is important. To extend this approach to include boundary scattering effects is the purpose of this paper. Inclusion of surface dissipation effects is handled by first writing the balance equations as integrals over the width of the semiconductor structure. We find it useful to introduce a generalization of the Fuchs boundary condition to include hot electron effects on the distribution function which underlays the balance equations. With this boundary condition we find an explicit solution of the nonlinear BTE in the presence of lattice and electron-electron scattering for the simple case of a uniform semiconductor film. The distribution function near a scattering surface is shown to have a marked velocity gradient which is important for evaluation of the scattering terms in the balance equations. The significance of surface scattering effects for device technology is briefly discussed.

BALANCE EQUATIONS

The well-known balance equations[2,3] for energy, momentum, and particle density are

$$\frac{d}{dt}\langle\phi\rangle + \frac{\partial}{\partial \underline{r}}\cdot\langle\phi\underline{v}\rangle - \langle e\underline{F}\cdot\frac{\partial\phi}{\partial \underline{p}}\rangle = \left\langle\int d^3p' W(\underline{p}',\underline{p})\left[\phi(\underline{p}')-\phi(\underline{p})\right]\right\rangle \quad (1)$$

where W is the bulk scattering rate, ϕ stands for energy, momentum component or unity, and $\langle\ \rangle$ means the momentum space average weighted by the distribution function f. We can include surface scattering effects W_s after integrating (1) across the semiconductor structure. For a uniform film in the region $(-d/2 \leq z \leq +d/2)$ in a uniform electric field $\underline{F}$ we then get

$$\frac{d}{dt}\int dz\,\langle\phi\rangle - \int dz\,\langle e\underline{F}\cdot\frac{\partial\phi}{\partial \underline{p}}\rangle = \int dz\left\langle\int d^3p' W(\underline{p}',\underline{p})\left[\phi(\underline{p}') - \phi(\underline{p})\right]\right\rangle + \left\langle\int d^3p' W_s(\underline{p}',\underline{p})\left[\phi(\underline{p}') - \phi(\underline{p})\right]\right\rangle \quad (2)$$

Evaluation of the dissipation terms on the RHS of (2) depends on the form of f, which we must investigate from the BTE.

SCATTERING BOUNDARY CONDITIONS IN HIGH FIELD TRANSPORT

A perfectly ordered interface with phonon coupling switched off would provide[4] specular reflection and no size effect (in spherical bands), i.e. one would have $f(\underline{v}^+) = f(\underline{v}^-)$, where $\underline{v}^-$ and $\underline{v}^+$ are incident and specularly reflected velocities at a boundary. But, with surface disorder and phonon coupling, both energy and momentum can be surface-dissipated: this will affect both device transit time and noise level. To handle the wide range of surface scattering

conditions that may be encountered in submicron device technology we propose the use of the following boundary condition on f:

$$f(\underline{v}^+) = Pf(\underline{v}^-) + (1 - P)f^* \tag{3}$$

in which P is a Fuchs type of specularity parameter[4] which indicates roughly the fraction of incident electrons which are specularly reflected, and (1 - P) indicates the fraction scattered into a Maxwellian distribution f*, characterized by some temperature T* and quasi-Fermi level E_F^*. This boundary condition differs from the low field case in that T^* and E_F^* can differ from the thermodynamic equilibrium values because the electrons incident on the boundary can likewise have a highly nonequilibrium distribution. We will, in fact, assume that surface scattering mechanisms are inefficient in relaxing the energy, so that we can impose the two conditions:

$$\int d^3v^+ \; mv^2 \; f(\underline{v}^+) = \int d^3v^- \; mv^2 \; f(\underline{v}^-) \tag{4}$$

$$\int d^3v^+ \; v_z^+ \; f(\underline{v}^+) = -\int d^3v^- \; v_z^- \; f(\underline{v}^-) \tag{5}$$

where the latter condition conserves particle number. These two conditions are sufficient to determine T^* and E_F^*. (Note that the boundaries are planes z = 0 and z = d).

HIGH FIELD BTE SOLUTION WITH SCATTERING BOUNDARIES

We will be able to apply the foregoing scattering boundary condition to the solution of the high field BTE in which lattice and (e,e) scattering are handled following the bulk relaxation time approximation method of Keyes[5]. The BTE is then

$$\frac{\partial f}{\partial t} + \dot{\underline{r}} \cdot \frac{\partial f}{\partial \underline{r}} - e\underline{F} \cdot \frac{\partial f}{\partial \underline{p}} = \frac{\hat{f} - f}{\tau_{ee}} + \frac{f_e - f}{\tau_L} \tag{6}$$

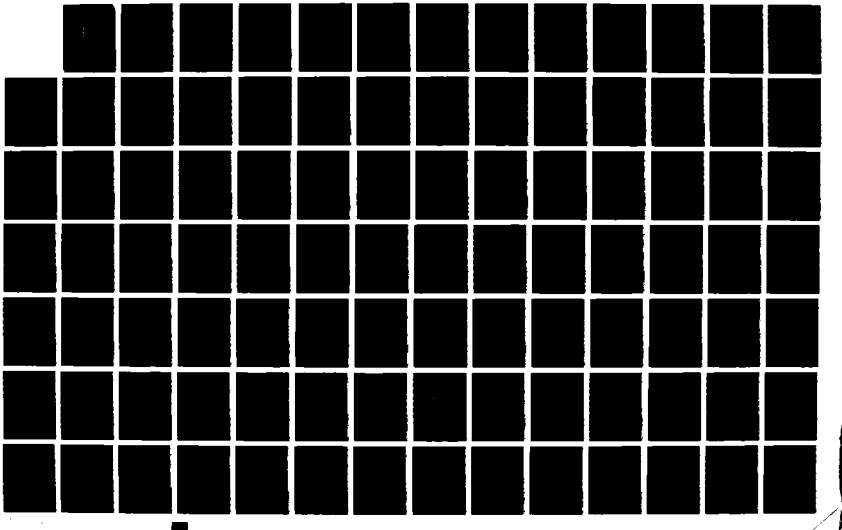
Here

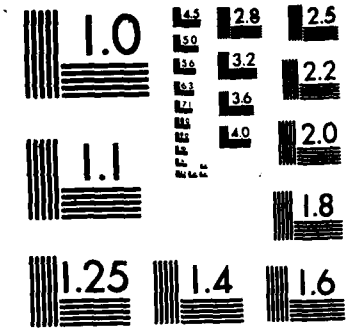
$$\hat{f} = \exp \frac{1}{kT} (\hat{E}_F - E - m\underline{v} \cdot \hat{\underline{v}}) \tag{7}$$

is the drifted Maxwellian with constant parameters $\hat{v}$, $\hat{T}$, and $\hat{E}_F$ determined from the balance equations. In this approximation $1/\tau_L$ is the phonon and defect scattering rate into the equilibrium distribution f_e, and $1/\tau_{ee}$ is the (e,e) scattering rate into the drifted Maxwellian $\hat{f}$. Although this is a crude approximation to the full collision integral BTE, the solutions will provide good insight into the distortion of the distribution function by surface scattering and a good enough value of f to use in the dissipation terms of the balance equations.

AD-A193 300 STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL NEAR-MICRON AND SUB-.. (U) SCIENTIFIC RESEARCH ASSOCIATES INC GLASTONBURY CT H L GRUBIN ET AL. 3/6

UNCLASSIFIED 15 FEB 88 SRA-R88-920010-F N00014-81-C-0452 F/G 9/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

It is useful to set

$$f_1 = f - \hat{f} \tag{8}$$

so that (6) takes the form

$$\frac{\partial f_1}{\partial t} + \dot{\underline{r}} \cdot \frac{\partial f_1}{\partial \underline{r}} - e\underline{F} \cdot \frac{\partial f_1}{\partial \underline{p}} + \frac{f_1}{T} = e\underline{F} \cdot \frac{\partial \hat{f}}{\partial \underline{p}} + \frac{f_e - \hat{f}}{T_L} \tag{9}$$

where

$$1/T = 1/T_L + 1/T_{ee} \tag{10}$$

The corresponding form of the scattering boundary condition is

$$f_1(\underline{v}^+) = Pf_1(\underline{v}^-) + (1 - P)(f^* - f) \tag{11}$$

In contrast to the low field case, there are no safe linearizations to be made in (9). Nevertheless, an analytic solution of (8, 9 & 10) can be obtained satisfying the boundary condition (11) (with (4) and (5)) by means of the path integral method[6,7,4] in which one makes a convective-derivative transformation of (9) to a coordinate system moving with the unscattered electron. In the present high field case the motion includes both the equilibrium Hamiltonian $H_o = mv^2/2$ but also the nonequilibrium Hamiltonian $H_1 = -e\phi$, where $\underline{F} = -\partial\phi/\partial\underline{r}$ is the large driving field. In that trajectory coordinate system the BTE takes the form (x is the direction along $\underline{F}$)

$$\left.\frac{Df_1}{Dt}\right)_{H_o+H_1} + \frac{f_1}{T} = \left\{\left[\frac{eF}{k\hat{T}}(\hat{v} - v_x) - \frac{1}{T_L}\right]\hat{f} + \frac{fe}{T_L}\right\} \tag{12}$$

which may be integrated immediately to give

$$f_1(t) = f_{10}e^{-\int_o^t dt'/T'} + e^{-\int_o^t dt'/T'}\int_o^t dt'' \, e^{\int_o^{t''} dt'/T'} \cdot \left\{\left[\frac{eF}{k\hat{T}}(\hat{v} - v_x) - \frac{1}{T_L}\right]\hat{f} + \frac{fe}{T_L}\right\} \tag{13}$$

Here t now denotes position along the trajectory, and we can take t = o as the point where a trajectory emerges from a boundary. Since the trajectory is not that of equilibrium, v_x varies with t according to

$$v_x = v_{xo} - eFt/m \tag{14}$$

and there are corresponding variations in T, f_e, and $\hat{f}$. To get some idea of the resulting form of f(t) we replaced T_L and T_{ee} by constants and obtained

$$f_1(t) = (f_{10} + f_2(t)) \exp(-t/T) \tag{15}$$

where

$$f_2(t) = \hat{f}_o\left[e^{-\hat{\omega}^2(t^2-2t\hat{t}_1)}-1\right] + \frac{\sqrt{\pi}\,\hat{f}_o}{2\hat{\omega}T_{ee}}\, e^{\hat{\omega}^2\hat{t}_1^2}\left[\mathrm{erf}(\hat{\omega}(t-\hat{t}_1)) + \mathrm{erf}(\hat{\omega}\hat{t}_1)\right] + \frac{\sqrt{\pi}f_{eo}}{2\omega T_L}\, e^{\omega^2 t_1^2}\left[\mathrm{erf}(\omega(t-t_1)) + \mathrm{erf}(\omega t_1)\right] \tag{16}$$

Here f_{10}, $\hat{f}_o$, f_{eo}, etc. refer to t = o, and

$$\hat{\omega}^2 = \frac{e^2F^2}{2mk\hat{T}}\ ;\quad \omega^2 = \frac{e^2F^2}{2mkT} \tag{17}$$

$$\hat{t}_1 = \frac{1}{2\hat{\omega}^2}\left[\frac{eF}{k\hat{T}}(v_{xo}-\hat{v}) + \frac{1}{T_L}\right]\ ;\quad t_1 = \frac{1}{2\omega^2}\left[\frac{eF}{kT}v_{xo} + \frac{1}{T_L}\right] \tag{18}$$

With this solution we can follow f_1 from t = o across the film to $t_d = d/T|v_z|$, reflect it by means of (11), integrate back again to z = +d and use (11) again. This gives us the explicit form of the boundary values $f_1(\underline{v}^+)$ and $f_1(\underline{v}^-)$;

$$f_1^-(1 + gP) = gf_{2d} + g(1 - P)(f_o^* - \hat{f}_o) \tag{19}$$

$$f_1^+(1 + gP) = gPf_{2d} + (1 - P)(1 + 2gP)(f_o^* - \hat{f}_o) \tag{20}$$

where $g = \exp -d/T|v_z|$, and $f_{2d} = f_2(t = d/T|v_z|)$. We can then evaluate T^* and E_F^* by substitution of the boundary values (19,20) into the energy and particle conservation conditions (4) and (5). The details of this will be given elsewhere, but for the case $d/T<v_z>>1$, i.e. where the film is thick enough compared to the total mean free path to approximately decouple the two boundary effects, one gets the two conditions

$$\langle v^2 f_o^* \rangle_+ = \langle v^2 \hat{f} \rangle_- \tag{21}$$

$$\langle v_z f_o^* \rangle_+ = \langle v_z \hat{f} \rangle_- \tag{22}$$

where the averaging is done only over the v_z^+ directions. This gives roughly,

$$kT^* \sim k\hat{T} + m\hat{v}^2/2, \text{ and } E_F^*/kT^* \sim \hat{E}_F/k\hat{T} \tag{23}$$

Now T^* and E_F^* can be determined in terms of $\hat{v}$, $\hat{T}$, and $\hat{E}_F$, and the approximate form of f used in the dissipation terms of the balance equations (2).

It is interesting to sketch out the contours of f close to the scattering boundary at z = -d/2, shown in Fig. 1 for a film wider than the mean free path. The incident distribution is close to the drifted Maxwellian $\hat{f}$, while the reflected distribution contains both a drifted Maxwellian part and the "thermalized" component f^*. In this figure we have also shown a typical trajectory integrated over, in (12).

The asymmetry of the distribution is thus quite strong in a submicron semiconductor film with scattering boundaries of the type considered, and such large angle scattering may be expected from interface disorder[4]. There are then two coincident electron streams of different velocity, and one may expect off-diagonal or viscous effects as well as noise arising from turbulence. Such effects should be investigated in the future since they may have major scientific and technological significance.

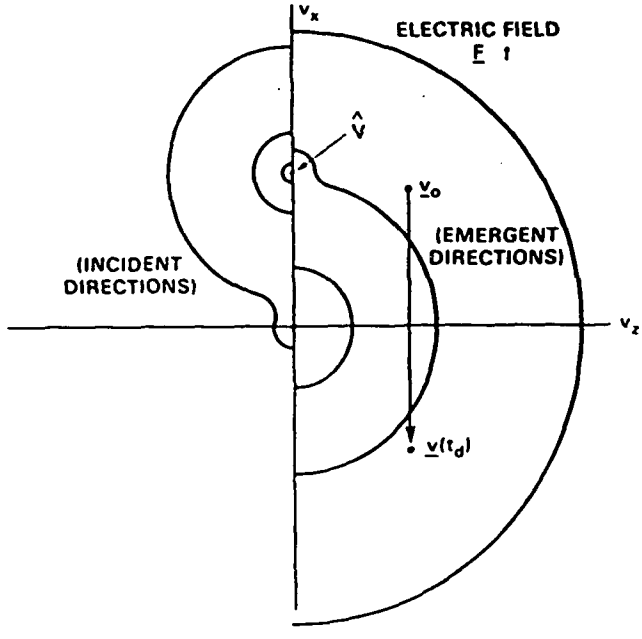


Figure 1. Contours of the electron distribution function near a scattering boundary in a high electric field $\underline{E}$. A typical trajectory is shown, with initial (emergent) velocity $\underline{v}_o$ terminating at $\underline{v}(t_d)$ at the intersection with the other boundary: $t_d = d/T|v_z|$.

CONCLUSION

Boundary scattering effects have a significant effect on electrical transport in submicron structures, particularly at high electric field strengths. This may be a significant consideration in the choice of material systems and structures for submicron devices. Ordered heteroepitaxial interfaces are likely to present significantly less transport limitations than disordered interfaces, such as Si/SiO_2.

These boundary scattering effects may be treated via integrated balance equations, where the dissipation terms are evaluated using an approximate solution of the high field BTE together with suitably generalized boundary conditions which can handle hot electrons. The effect of surface scattering is to introduce an extremely sharp assymmetry in the electron distribution the effect of which on turbulence and noise should be investigated.

REFERENCES

1. "Physics of Nonlinear Transport in Semiconductors", editors D. K. Ferry, J. R. Barker, and C. Jacoboni. (NATO Advanced Study Institute Series, vol. 52)(Plenum Press, N.Y., 1980).
2. P. J. Price, IBM Journal of Res. & Dev. 14, 12 (1970).
3. H. L. Grubin, D. K. Ferry, G. J. Iafrate, and J. R. Barker, in Ch. 6 of "VLSI Electronics", editor: N.G. Einspruch (Academic Press, N.Y., 1981).
4. R. F. Greene, in Ch. 2 of "Solid State Surface Science", vol. I editor: M. Green, (M. Dekker, N.Y., 1969).
5. R. W. Keyes, J. Phys. Chem. Solids 6, 1 (1958).
6. R. G. Chambers, Proc. Roy. Soc. Lond. A202, 378 (1950).
7. H. Budd, Phys. Rev. 127, 4 (1962).

Reprint 12

Surface Science 132 (1983) 594–622
North-Holland Publishing Company

THE ROLE OF BOUNDARIES ON HIGH SPEED COMPOUND SEMICONDUCTOR DEVICES

H.L. GRUBIN and J.P. KRESKOVSKY

Scientific Research Associates, Inc., P.O. Box 498, Glastonbury, Connecticut 06033, USA

Received 2 September 1982; accepted for publication 11 November 1982

The role of boundaries and interfaces on the electrical characteristics of long and submicron scale compound semiconductor devices is discussed.

1. Introduction

Theoretical and experimental studies of compound semiconductor devices over the past decade have demonstrated that conditions at the device boundaries are the most important determinant of the operating characteristics of the device. Studies in long two-terminal gallium arsenide devices [1], indium phosphide [2], germanium [3], and cadmium sulphide [4] have demonstrated that conditions at or near the contacts control the current–voltage relations and the electrical characteristics of any resulting instabilities. For two-terminal devices it is often possible to correlate the pre- and post-threshold device behavior. For three-terminal devices, most interface studies have focused on the role of the layer just under the principle region of electron transport. In the case of gallium arsenide field effect transistors, heterostructure interfaces have been incorporated with the object, e.g., of confining carriers to the active region in the case of a heavily doped active region, [5,6], or of providing a sea of carriers in a nominally undoped region.

In near and submicron scale devices, the relative importance of the boundaries increases as these regions occupy a sizeable fraction of the device active region length, and the up- and downstream boundaries begin to communicate with each other. Controlling these boundaries is likely to be the most significant task of device physics studies in the immediate future. The motivation for this outlook is based upon the fact that transport on a near and submicron scale involves nonequilibrium effects on a picosecond time frame, and its effectiveness is based on transport by high mobility carriers. Thus, questions such as "How many carriers are injected into the Γ valley of gallium arsenide, and with what energy and velocity?" enter the picture prominantly.

0039-6028/83/0000-0000/$03.00 © 1983 North-Holland

The purpose of this paper is to highlight the effects of boundaries on compound semiconductor devices, particularly gallium arsenide. In doing so, the discussion is separated into two parts: (1) the role of boundaries on long, low frequency (< 20 GHz) gallium arsenide devices, and (2) the role of the boundary on high frequency near and submicron scale devices.

For long devices in which a rich body of experimental work exists, a review of boundary effects is given in terms of a "pinned" cathode field and "pinned" cathode current model. The basis for these models is the notion that contact boundaries may be *phenomenologically* represented as either tunneling or thermionic emission dominated regions, with a varying barrier height. The description of long devices assumes that all picosecond scale transients have occurred, and that all band structure population statistics are adquately represented by steady state conditions.

For the shorter devices it is the short time transients and the spatial and temporal transients within the bands that are calculated. The description requires solutions of the Boltzmann transport equation and resolution on the scale of a fraction of a bulk mean free path is needed. These problems are discussed in section 3 where a review of such phenomena as velocity overshoot is given. The ability to attain contact and boundary effects permitting the realization of the high overshoot speeds in devices is the core on which the studies of section 3 are based.

All of the theoretical studies presented arise from solutions of differential equations. For long devices,the "drift and diffusion" equations are solved. For short devices, moments of the Boltzmann transport equation are solved. In both cases true contacts are represented as boundary conditions to the differential equations. Thus, the studies illustrate the *effects* of the contacts and/or interfaces on device behavior. The results of these studies, when successful, tend to highlight what is unknown about device material parameters. For the case of long devices, it is the cathode boundary condition that is unknown. Here the sensitivity of the results to varying the cathode field identifies the significance of the contacts on device behavior [1]. Similar sensitivity studies are discussed in connection with solutions of the Boltzmann transport equation, where a range of parameters is chosen to identify the conditions for "injecting" and "blocking" contacts. A sensitivity analysis of the electronic contribution to the thermal conductivity is also included as its effects are dramatic.

2. Boundary conditions to negative differential conductivity devices

In this section a brief review of the influence of boundaries on the behavior of NDC devices is given. Fig. 1 displays typical boundary-dependent data from three different gallium arsenide two-terminal devices [7]. The lower portion of

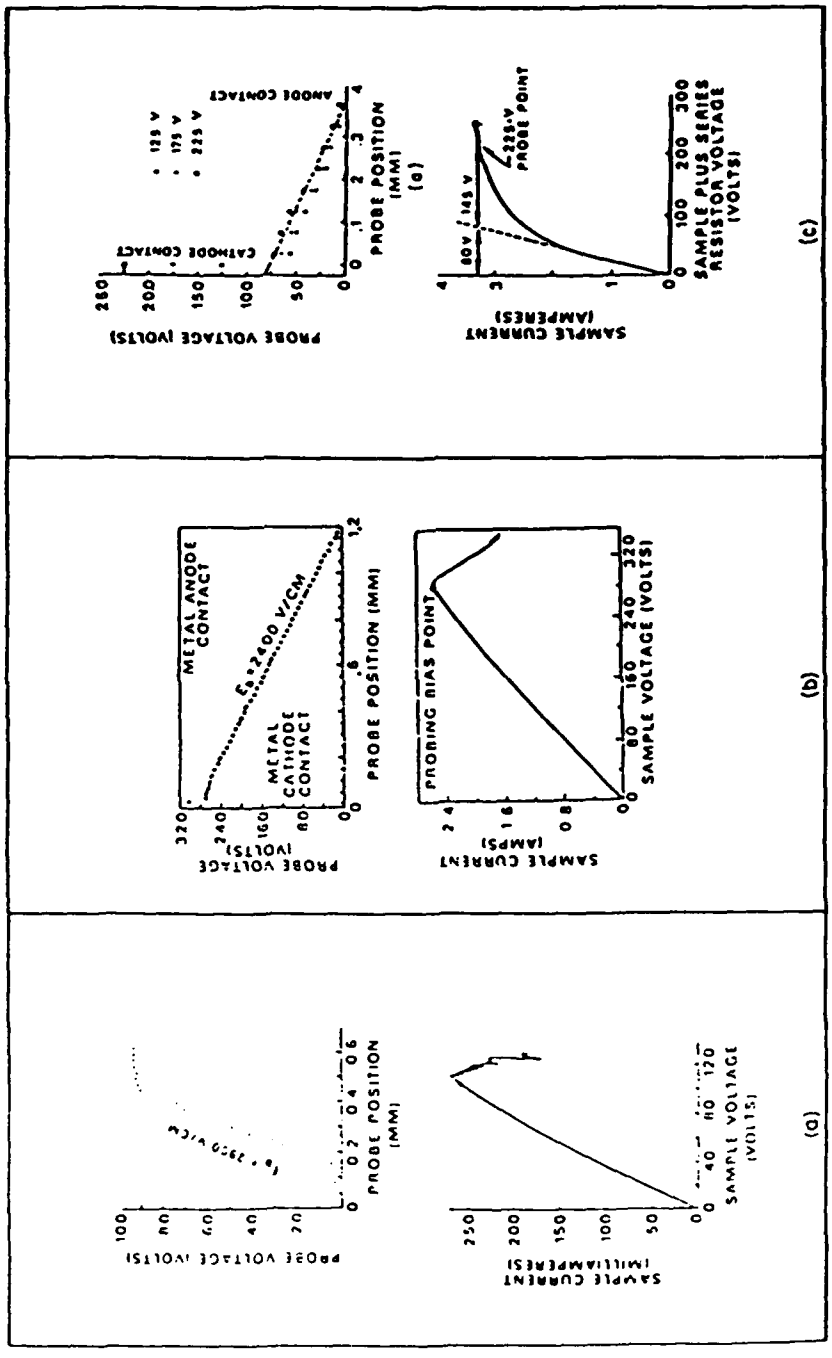


Fig. 1. (a) Current versus voltage for a low cathode field GaAs sample with a conductivity mobility of 5900 $cm^2/V \cdot s$. Sample was sculptured into an "H" shape to remove active region from boundaries. Also shown is the probed voltage versus distance. Probing bias is indicated by closed circle. (b) As in (a) but for an intermediate cathode field. Sample was "parallelepiped" shaped without any deliberate shaping. (c) As in (b) but for a high cathode field sample. From ref. [7] with permission.

each diagram displays current versus voltage characteristics, while the upper portion shows voltage versus distance at one bias point. Fig. 1a shows measurements for a device in which the metal contacts are far removed from the active region of the device. The current–voltage relation is relatively linear until a point where current oscillations occur. The field profile just prior to the oscillation is relatively uniform within the active region of the device, and is near zero at the ends of the active region. Fig. 1b represents a set of measurements in which the metal contact abutted the active region of the device. The current–voltage characteristics remained linear to threshold which again was manifested by a current oscillation. Notably different here is the lower average field prior to the instability and the enhanced voltage drop at the cathode. Fig. 1c displays results for another device with a metal contact abutting the active region. For this case there is a sublinear current voltage characteristic and no instability. The probed voltage versus distance shows a large voltage drop at the vicinity of the cathode.

The electrical characteristics associated with fig. 1 have been described as representative of "ohmic" (fig. 1a), "slightly blocking" (fig. 1b), and "strongly blocking" (fig. 1c) contacts. One of the earliest models employed for explaining these results assumed a "pinned" value of cathode electric field [1]. Other models in which the cathode conductivity [8] or doping [9] have also been suggested with varying degrees of success. The "pinned" cathode field model developed, partially as a consequence of the way the governing equations describing current instabilities was written. Here the one-dimensional differential equation for total current,

$$J(T) = q\left(NV - D\frac{\partial N}{\partial X}\right) + \epsilon\frac{\partial E}{\partial T}, \tag{1}$$

was rewritten, using Poisson's equation, as

$$J(T) = qN_0V + \epsilon\left(V\frac{\partial E}{\partial X} - D\frac{\partial^2 E}{\partial X^2} + \frac{\partial E}{\partial T}\right). \tag{2}$$

This is a second order partial differential equation requiring two boundary conditions on $E(X, T)$ and one initial condition. In the above $J(T)$ represents current density, N is carrier density, V and D are field dependent velocity and diffusion.

Solutions to eq. (2) have been used successfully to simulate device results similar to those of fig. 1. For gallium arsenide in which the threshold field for negative differential mobility is approximately 3.2 kV/cm = E_{TH}, qualitatively similar results occur for solutions with cathode fields falling into any one of the following three groups: $0 \leq E(X = 0, T) \leq E_{TH}$, $E_{TH} \leq E(X = 0, T) < 4E_{TH}$, $E(X = 0, T) > 4E_{TH}$. The simulations with pinned fields falling in either group 1, 2 or 3 yield electrical characteristics similar to those of figs. 1a, 1b and 1c, respectively. The crucial feature of this model is that the cathode field is

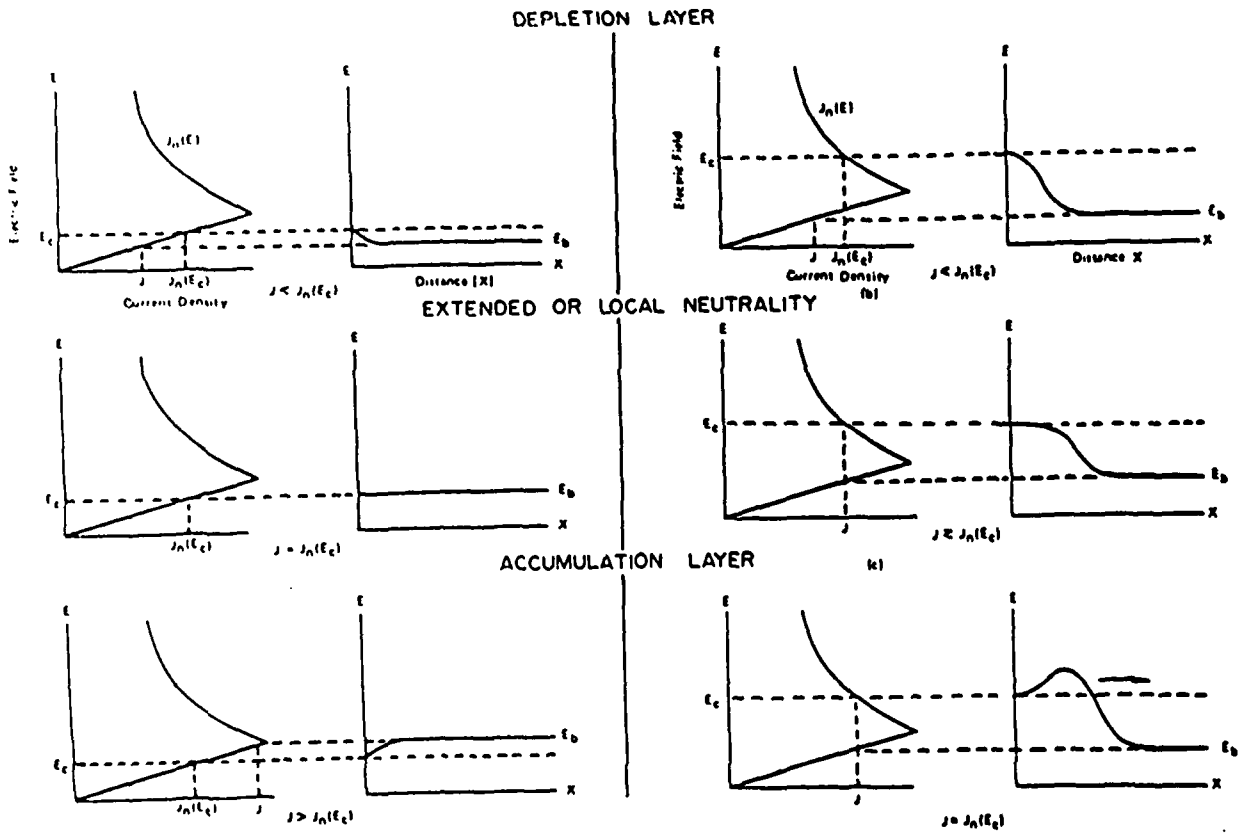


Fig. 2. (a) Electric field versus neutral current density, and electric field versus distance for a pinned cathode field, E_c, within the "ohmic" portion of the velocity field relation. (b) As in (a) but for a pinned cathode field within the NDM region. Curves show trends for cathode depletion, charge neutral, and cathode accumulation. From ref. [10], with permission.

pinned, necessitating that any instabilities in current occur at a critical value of current density. The field profiles associated with cathode fields in the range $0 \leq E(X=0, T) \leq E_{TH}$, and $E_{TH} \leq E(X=0, T) < 4\,E_{TH}$ are sketched in figs. 2a and 2b respectively. For reference, a velocity field curve with velocity scaled to current as $qN_0V(E) = J_n(E)$, and with a region of negative differential mobility is also included. Fig. 2 is understood as follows: The second column of each section shows the electric field versus distance profiles. $E(X)$ begins with a value E_c at the cathode and extends downstream to a value E_b. By current continuity, the current everywhere within the device is given by $J = qNV(E_b)$. For $J < J_n(E_c)$, a region of charge depletion forms near the cathode for E_c below and within the region of negative differential mobility (NDM). Increasing the current until $J \equiv J_n(E_c)$ introduces charge neutrality everywhere for $E_c < E_{TH}$. However, because E is a double valued function of V, for E_c within the NDM region, approximate charge neutrality exists near the cathode for $J \leq J_n(E_c)$, and for regions sufficiently far downstream from the cathode. Charge neutrality breaks down between these two regions. Finally, for $J > J_n(E_c)$ an accumulation layer forms near the cathode. For fig. 2a, the accumulation layer is stable until the bulk field exceeds E_{TH}. For fig. 2b, the accumulation layer, followed downstream by the depletion layer, is often unstable and leads to cathode originated instabilities [10].

The situation corresponding to fig. 1c is often represented by very high cathode fields. The field profiles are those appropriate to a wide region of charge depletion near the cathode. The profiles are electrically stable.

The characteristic feature of these nonuniform field profiles is that their structure is significantly affected by the field being a double valued function of velocity. The pinning of the cathode field is not necessarily common, however, to all semiconductor devices. For example, it was also applied to InP devices, where it worked for a significant number of cases. However, a broad class of device behavior could not be accounted for through its use [2]. These devices showed anomalously high efficiency and significantly low DC current levels. Spontaneous Gunn type oscillations did not occur. Rather, device operation required a tuned circuit. The details of the oscillation were thought to depend critically on the cathode boundary condition, which in this case was taken as a fixed cathode conduction current [10,11].

The distinction between "pinned" cathode field and "pinned" cathode conduction current is placed in perspective in fig. 3 and in the following equation

$$J(T) = J_c(E_c) - \epsilon \frac{\partial E_c}{\partial T}. \tag{3}$$

Eq. (3) is the equation for total current through the boundary to the device. $J_c(E_c)$ represents the current–field relation at the cathode [12] which may be expected to differ from that of the semiconductor device. Two such types of

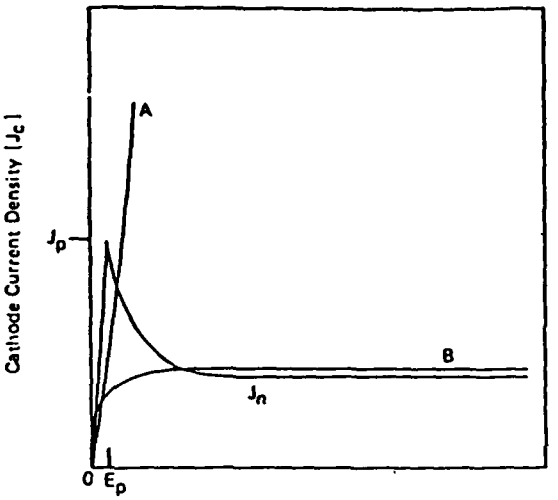


Fig. 3. Cathode conduction current curve for an approximately "pinned" cathode field, curve A; and pinned cathode conduction current (at high field fields), curve B. Curves are obtained from eq. (4). Also shown, for reference, is the neutral current versus field curve for GaAs. From ref. [10], with permission.

curves are represented by curve A and curve B of fig. 3. Curve A is closely related to the pinned cathode field model while curve B is associated with the pinned cathode current model. The similarity in "form" of curves A and B to moderate barrier height tunneling and thermionic emission dominated contacts is deliberate [10], and the equation used to arrive at these curves is shown below

$$J_c(E_c) = -J_R\left\{\exp\left(-\frac{qE_cL_c}{nk\tau}\right) - \exp\left[-\left(\frac{1}{n}-1\right)\frac{qE_cL_c}{k\tau}\right]\right\}. \tag{4}$$

which was adapted from studies on the unalloyed metal/semiconductor contact [13]. Its use here presumes a similar description. For the unalloyed contact, n is the ideality factor and describes the contact as dominated by thermionic emission ($n \approx 1$) or by tunneling ($n \gg 1$). J_R is the reverse current flux and may be related to the barrier height phenomenologically through the Richardson equation [10].

Detecting a particular contact effect on a device is a difficult procedure. For long devices current voltage characteristics as represented by fig. 1 are often signatures of a contact classification. For short devices proximity effects introduce an additional complication and current–voltage measurements are less valuable. One type of measurement which may serve to provide information about the boundary is a noise measurement.

Here the situation to envision is that if the field is pinned within the

negative differential mobility region, increasing the bias will result in an increase in the length of the NDM region. Any fluctuation will sustain enhanced amplification and the noise will increase. If increasing the bias resulted in carrier injection into the device, the field at the cathode is likely to decrease with increasing bias and the noise is expected to decrease. While these results should be folded in with the field dependence of velocity and diffusion, a simple analytical noise calculation assuming a three piece linear approximation to represent GaAs has been performed [14].

In this calculation, the "impedance field method" [15] is applied to calculating noise due to thermal velocity fluctuations amplified within the device. The mean squared noise voltage per unit band width [15] is computed,

$$\left\langle \frac{\delta V_N^2}{\Delta f} \right\rangle = 4q^2 \int_{vol} |\nabla Z|^2 \, ND \, \mathrm{d(vol)}, \tag{5}$$

where ∇Z is the impedance field vector [15]. The calculation is performed for a 10 μm long element with a doping of $10^{15}/\mathrm{cm}^3$. The element sustains the field profile of fig. 2b where it is seen that the NDM region increases with increasing bias. The calculations, which are discussed in detail elsewhere, [14] are expressed in terms of the noise figure [16]:

$$\mathrm{NF} = 1 + \left\langle \frac{\delta V_N^2}{\Delta f} \right\rangle \frac{1}{4k_0 T |R|}, \tag{6}$$

where R is the real part of the device impedance. The results of the calculation are displayed in fig. 4, where the noise figure is sketched as a function of bias current and transit angle. The results appear as a signature of the effects of the cathode boundary. First, at low values of transit angle $\theta = \omega T(\Delta)$, where $T(\Delta)$ is the transit time across the negative differential mobility region, the noise figure increases with increasing bias. This corresponds to an increase in the length of the negative differential mobility region and enhances amplification of any fluctuation originating there. More interesting structure is present at higher frequencies and higher bias where the noise figure increases and then shows a singularity. On the other side of the singularity there is a "U"-shaped region ending again at a singularity. The strong increase in noise figure represents the approach of $|R| \Rightarrow 0$. Here, at low frequencies, the real part of the impedance is positive, and becomes negative at frequencies somewhere between $\pi \leqq \theta \leqq 2\pi$. In going from positive to negative values it passes through zero, hence the singularity. The frequency range for small signal negative resistance increases with increasing bias [10] reflecting the broadening of the negative differential mobility regions – a broad "U"-shaped region appears. Both the increasing noise figure at low frequencies, and the "U"-shaped region at high frequencies are characteristics of an increasing depletion layer width. Note that increasing the bias still further will result in an electrical instability [10].

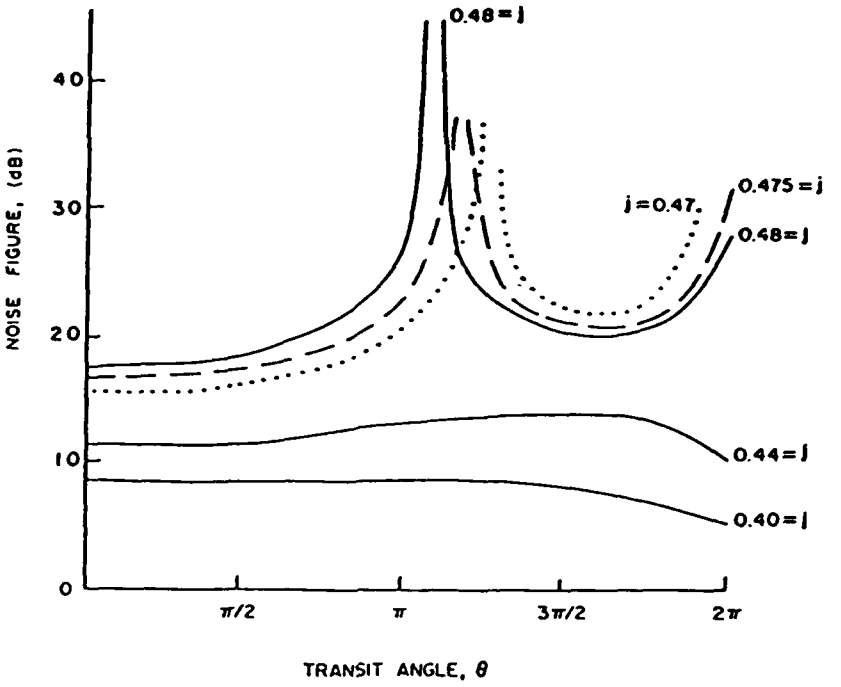


Fig. 4. Noise figure versus transit angle and normalized bias current $j = J_b/N_0eV_p$, for a 10 μm long element with a depletion layer profile. From ref. [14], with permission.

3. Electron transport in near and submicron devices: the role of the boundaries

The discussion of the above sections dealt with devices whose lengths were typically 10 μm long or longer. Transport for these devices is generally discussed though use of the drift and diffusion equation (eq. (1)). For near and submicron length devices electron transfer is generally not complete until a substantial fraction of the device has been traversed. Consequently the drift and diffusion description is inadequate and solutions to the Boltzmann transport equation are required.

For two level transfer in GaAs the steady state velocity field curve for carriers in the central and satellite valleys are shown in fig. 5, where the net velocity V is

$$V = N_1V_1 + N_2V_2. \tag{7}$$

Here the subscript 1 denotes transport in the central valley (Γ valley for GaAs) and the subscript 2 denotes transport in the satellite valley. Most device design is concerned with controlling the time spent by the numbers of carriers in

in valleys 1 or 2; and circuits, interfaces or contacts are sought which will allow for this control. Solutions to the Boltzmann transport equation which provide the required nonequilibrium transport behavior are obtained by a number of

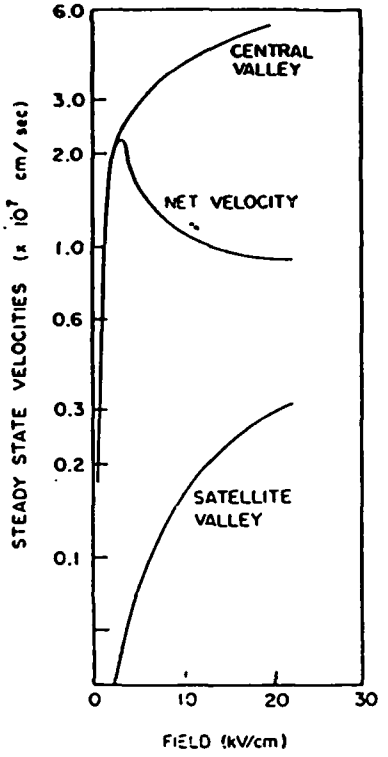


Fig. 5. Steady state velocity versus field curves for Γ and satellite valleys for gallium arsenide. From ref. [22], with permission.

different techniques, the most familiar of which is the Monte Carlo method [17]. Other methods include the Rees iterative technique [18] and balance equation solutions [19], which are considered below.

The balance equations discussed below are the first three moments of the Boltzmann transport equation. They form a subset of the infinite hierarchy of moment equations, and as such do not form a closed set of equations. Each equation introduces a higher order moment not defined by the given set of equations. The most common form of the balance equations is obtained by assuming a distribution function of the form

$$f = f_{0_i} \propto \frac{N_i}{\tau^{3/2}} \exp\left[- \frac{m(U - U_i)^2}{2k_B T} \right], \tag{8}$$

for each species i. More generally [20]

$$f_i = \left(1 + A_k \frac{\partial}{\partial U_k} + A_{kl} \frac{\partial^2}{\partial U_k \partial U_l} + A_{klm} \frac{\partial^3}{\partial U_k \partial U_l \partial U_m} + \dots\right) f_{0_i}. \tag{9}$$

where it is insisted that the first term in the expansion gives the correct local values of density, velocity and energy. The coefficients in the above expansion are inversely proportional to density and are model dependent. The first correction to the local equilibrium balance equtions, the so-called hydrodynamic approximation yields

$$\frac{\partial N}{\partial T} + \nabla \cdot (V_i N_i) = \left(\frac{\partial N_i}{\partial T}\right)_c, \tag{10}$$

$$\underline{\frac{\partial P_i}{\partial t}} + \nabla \cdot (V_i P_i) = -eN_i F - \text{grad } N_i k_0 \tau_i - \boxed{\text{div } \sigma^i} + \left(\frac{\partial P_i}{\partial T}\right)_{\text{coll}}, \tag{11}$$

$$\frac{\partial}{\partial t} W_i + \nabla \cdot V_i W_i = -eN_i F \cdot V_i - \nabla \cdot V_i N_i k \tau_i - \boxed{\text{div } \sigma^i \times V_i} + \boxed{\text{div}(\kappa \text{ grad } \tau_i)} + \left(\frac{\partial W_i}{\partial T}\right)_{\text{coll}}, \tag{12}$$

where τ_i is an electron temperature,

$$P_i = N_i m_i V_i, \quad W_i = N_i\left[\tfrac{1}{2} m_i V_i^2 + \tfrac{3}{2} N_i k \tau_i\right],$$

σ is a stress tensor arising from nonuniform velocity distributions, and κ is the thermal conductivity. The stress term is dissipative in that when a nonuniform velocity distribution is impressed on an electron stream, there will be reactive forces tending to smooth them out. At this point, however, these terms are regarded as phenomenological entries.

The boxed terms above represent contributions from the nonspherical nature of the distribution function. These contributions have not been included in evaluating the collision integrals, i.e., $(\dots)_{\text{coll}}$ but will be discussed in a future paper. The collision integrals are discussed in ref. [21]. The underlined and boxed terms are ignored in the drift and diffusion approximation.

Under uniform field conditions the mean response of carriers to a sudden change in electric field is shown in fig. 6. The high peak velocity is a consequence of carriers being retained in the Γ valleys for time duration upwards of 0.2 ps before undergoing transfer to the upper valleys. The peak velocity in fig. 6 is extremely high. It represents an upper bound on the carrier velocity that may be expected at this field and provides the motivation for designing boundaries that permit achievement of these values.

The peak velocity, however, is sensitive to bias rise time and hints at problems to be addressed in designing appropriate device boundary conditions. See fig. 7. Furthermore, because a finite time is required for carriers to

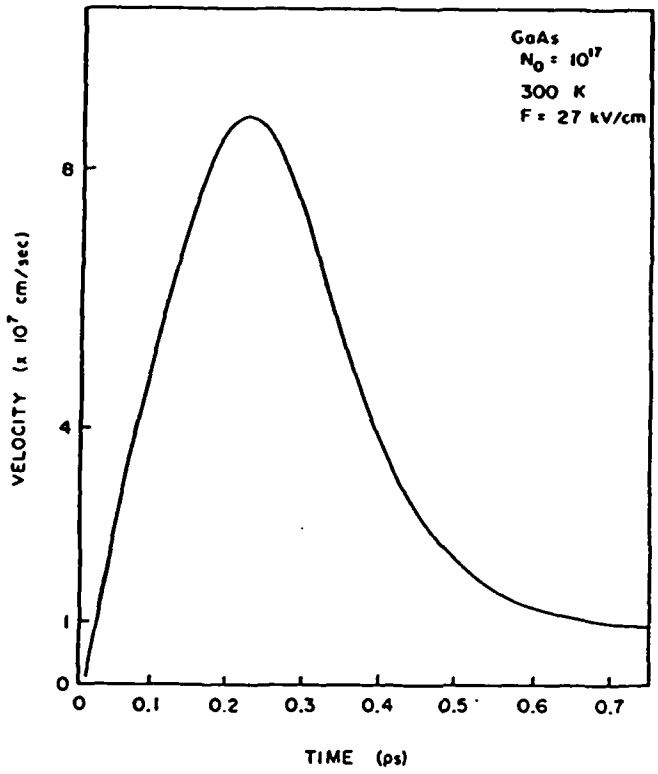


Fig. 6. Transient velocity versus time, for electrons subject to a sudden change in field of magnitude 27 kV/cm. From ref. [22], with permission.

transfer into subsidiary portions of the conduction band, on a near and submicron scale, the velocity field characteristics are expected to be length dependent. See fig. 8. Thus, near and submicron devices will categorically be sensitive to both boundary effects and device length, neither of which will be independent of the other.

Figs. 5 to 8 are signatures of high speed submicron transport. What are the consequences for device behavior? Can a suitable set of contact or interface conditions be found to achieve high speeds associated with overshoot? While a detailed study on a scale similar to that for long devices is not yet available, some results are known. We discuss these below for a collection of 2 μm long devices, essentially undoped, $N_0 = 5 \times 10^{15}/\text{cm}^3$, each subjected to a bias of 2 V. These one-dimensional studies are in steady state, i.e., $\partial/\partial T = 0$, and are subjected to the following cathode boundary conditions:

$$\frac{\partial}{\partial X} \log N_1 = -A, \tag{13}$$

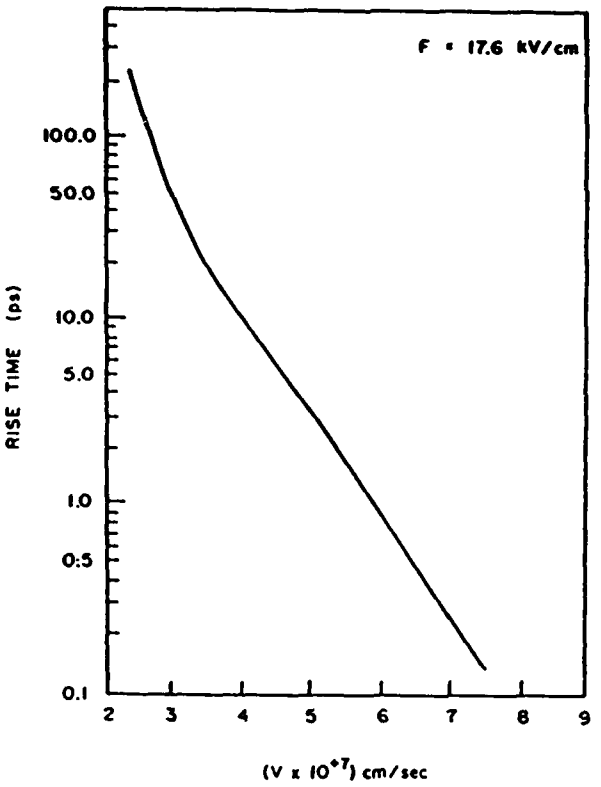


Fig. 7. Peak carrier velocity versus bias rise time for electrons subject to a final field of 17.6 kV/cm.

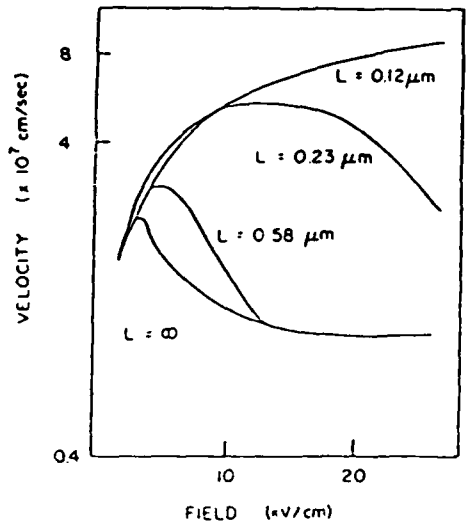


Fig. 8. Exit velocity of carriers at the terminus of a device of specified length, and indicated value of electric field. Carriers enter with zero velocity and are subject to a sudden change in field. From ref. [23], with permission.

$$V_1 = -\mu_c F. \tag{14}$$

$$\tau_1 = B. \tag{15}$$

$$\partial N_2/\partial X = \partial V_2/\partial X = \partial \tau_2/\partial X = 0. \tag{16}$$

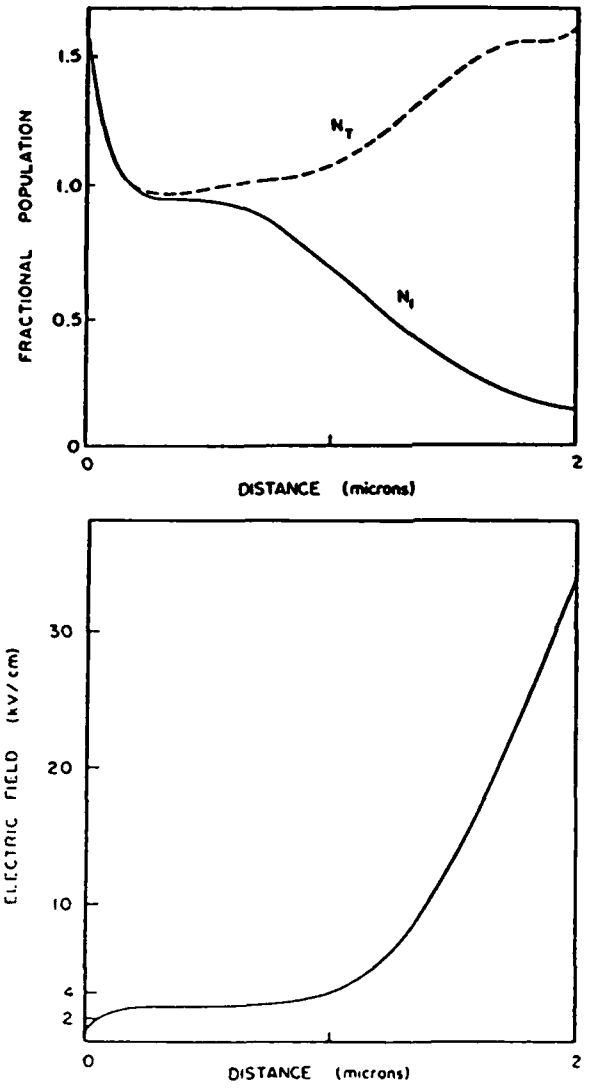


Fig. 9. (a) Fractional population of carriers, N/N_0 in a 2 μm long device with a uniform nominal doping of 5×10^{15}/cm^3, subject to a bias of 2 V. N_T is the total carrier density. N_1 is the carrier density in the central valley. (b) Electric field versus distance. For this calculation $\mu = 12{,}000$ cm^2/V s, $A = 0.2$ and $B = 300$.

For A positive (negative) local charge accumulation (depletion) occurs at the cathode boundary. B is generally greater than, or equal to, 300 K and is a measure of the mean thermal energy of the Γ-valley carriers. As in section 2, the results are placed into three categories, "ohmic", "slightly" and "strongly" blocking contacts. The "ohmic" results are shown in fig. 9.

For "ohmic" boundaries an appropriate set of constants, with ref. [19], are: $A = 0.2$, $\mu_c = 12{,}000$ cm^2/V·s, and $\tau_1 = 300$. For this case carriers enter the device with speeds greater than that associated with the central valley mobility. Consequently there is an accumulation of carriers at the boundaries, resulting in low values of cathode field. The field starts off at nearly 1 kV/cm and approximately 2500 Å must be traversed before significant transfer occurs. Increased transfer results in a lowering of the mean carrier velocity, necessitating an increase in mobile charge as the anode is approached. The average velocity across this device obtained from the relation

$$V_{av} = J/qN_0 \tag{17}$$

is $V_{av} = 1.78 \times 10^7$ cm/s, with $N_0 = 5 \times 10^{15}$/cm^3. The advantage of overshoot is not fulfilled by this contact device length configuration.

Lowering the cathode mobility to a value below that associated with the Γ valley results in a more rapid dispersal of carriers, and cathode adjacent charge depletion, associated with slightly blocking contacts, occurs. This is seen in fig. 10 for $A = -0.11$, $\mu_c = 6000$ cm^2/V·s, and $\tau_1 = 300$ K. It is noticed that the cathode field for this case is approximately 4 kV/cm, which is higher than that associated with the "ohmic" contact condition of fig. There are, however, important similarities between figs. 9 and 10. In both cases the carriers adjacent to the cathode are, for all practical cases, Γ-valley electrons. Very little transfer, which is determined by carrier energy (temperature), has occurred. In addition, sufficiently downstream from the cathode the carriers appear to be ignoring the cathode condition and are dominated by the downstream voltage drop, whose spatial distribution is about the same for both. The average velocity for this case is $V_{av} = 1.75 \times 10^7$ cm/s, slightly below that of fig. 9.

A significant change occurs when the mean energy of the Γ-valley entering carriers is elevated. For the parameters of fig. 10, but with $\tau_1 = 1200$ K, a substantial amount of transfer occurs at the cathode, resulting in a lowering of the current through the device. The cathode field is approximately 7 kV/cm, higher than that associated with fig. 10, and the downstream field is lower (see fig. 11). The average velocity of the carriers in this case is lowered to 1.28×10^7 cm/s, even though the central valley carriers are traveling at higher speeds (see fig. 5).

The presence of moderately high cathode fields is attractive if a sufficient number of carriers can be retained in the central valley where they can sustain high transit velocities. While this case is discussed in more detail below, the simple ruse of injecting excess carriers into a device with the contact conditions

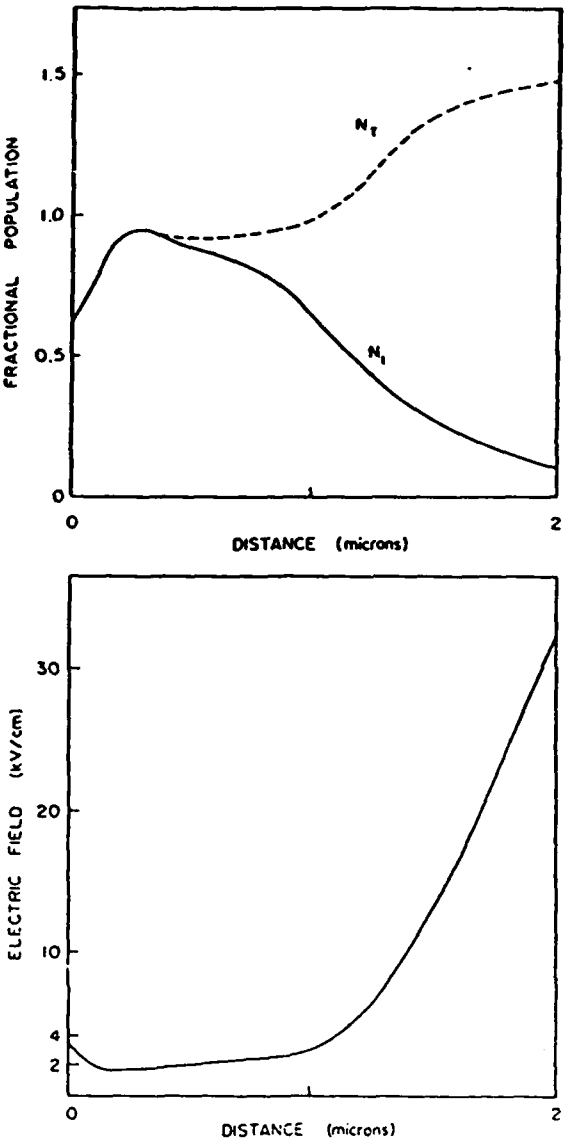


Fig. 10. As in fig. 9, but with $\mu_c = 6000$ cm^2/V·s, $A = -0.11$ and $B = 300$ K.

of fig. 11 does not always yield the sought after current levels. This is illustrated in fig. 12, where now $A = 0.2$. Here, the excess charge serves to lower the cathode field, which does not change the downstream characteristics in any significant way. The average velocity for this case is virtually unchanged

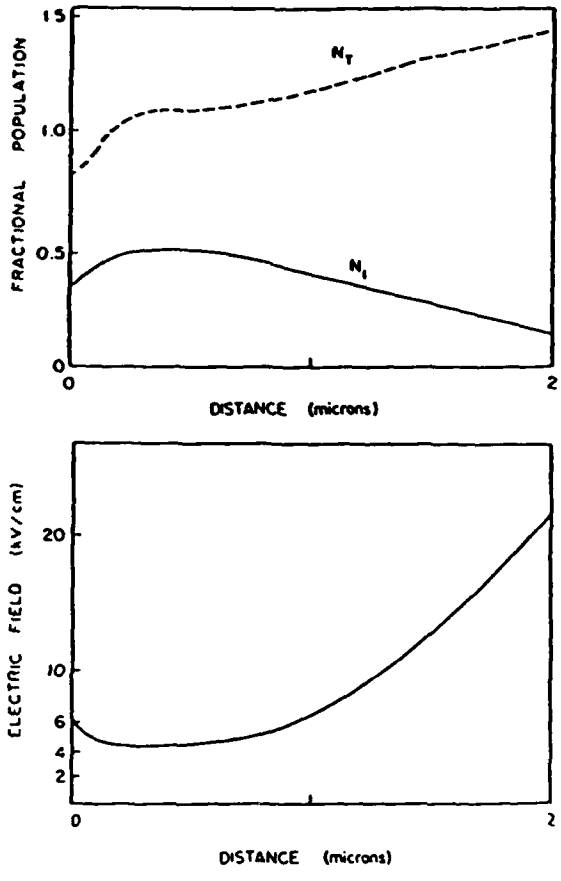


Fig. 11. As in fig. 9, but with $\mu_c = 6000\ cm^2/V \cdot s$, $A = -0.11$ and $B = 1200$ K.

when compared to fig. 11. In this case $V_{sc} = 1.30 \times 10^7$ cm/s.

The situation of a strongly blocking contact is illustrated in fig. 13 for $A = 9.0$, $\mu = 1000\ cm^2/V \cdot s$, and $T = 3000$ K. For this case virtually all carriers are swept away from the cathode, with the satellite valley carriers accounting for most of the transport in the transition region. The cathode field is extremely high, approaching 60 kV/cm. The average velocity is approximately zero.

The above results focus attention on the role of the upstream boundary in the distribution of charge near the interface. The results are more general than those obtained using the steady state field dependent velocity relation, especially in identifying the fact that electron transfer, *even in the case of partially blocking contacts*, may not occur until some point downstream from the

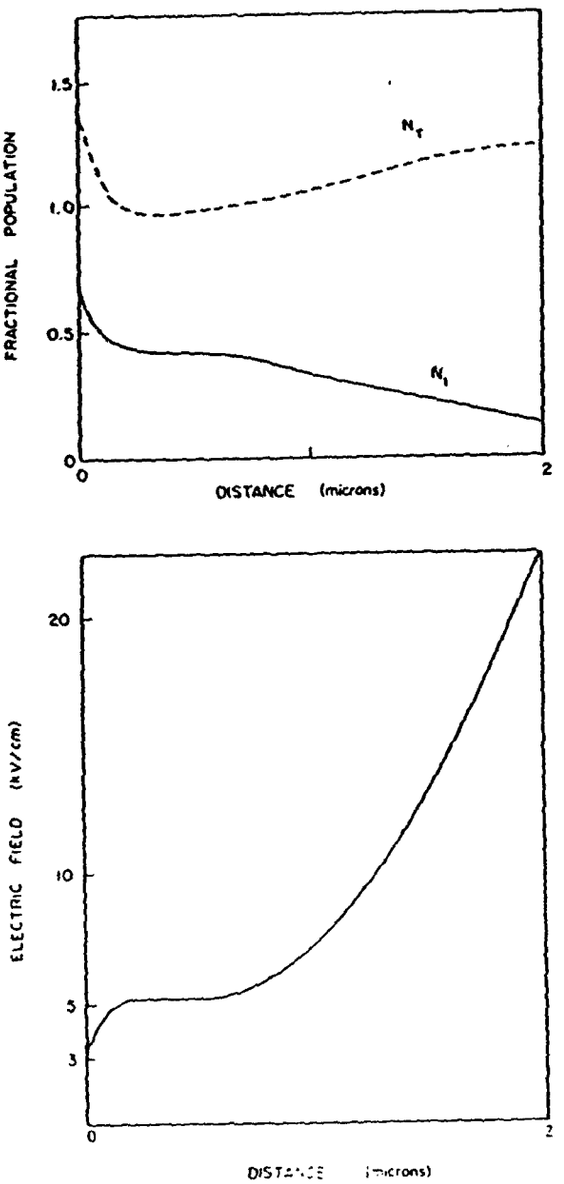


Fig. 12. As in fig. 8 but with μ_0 = [illegible] cm^2/V-s, $A = 0.2$ and $B = 1200$ K

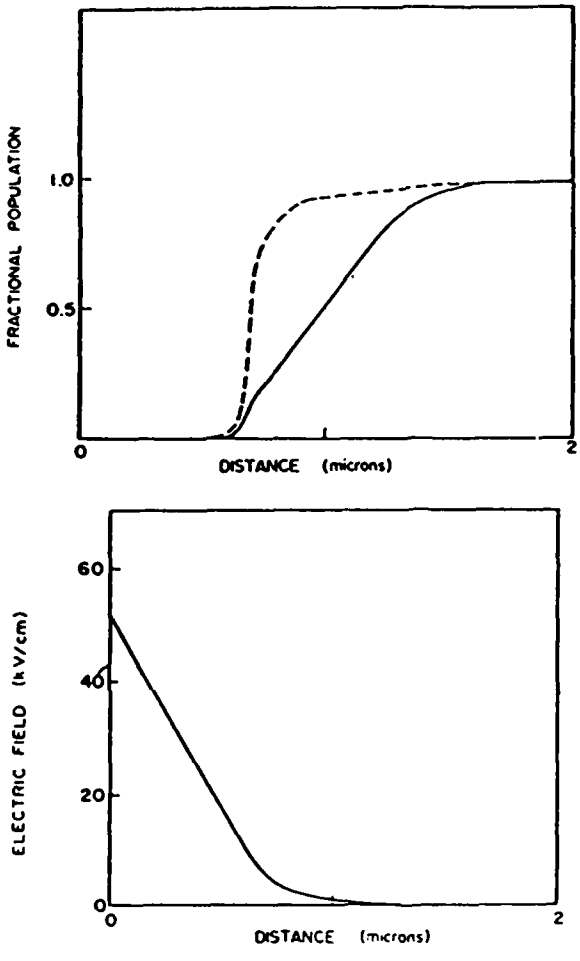


Fig. 13. As in fig. 9, but with $\mu_c = 1000\ cm^2/V \cdot s$, $A = -9.0$ and $B = 3000$ K.

cathode contact. The implication of this is that the detailed role of the contact in controlling current instabilities may need generalization. On another matter, it is apparent that the advantages of velocity overshoot are absent from the 2 μm long calculation. The reason for this is direct: the carriers have traversed a path of sufficient length to exceed the threshold energy for electron transfer. In this case electron transfer results in velocity saturation and currents appropriate to the steady state parameters. Thus, while a blocking contact may be expected to yield high entrance velocities for one species of carriers, electron transfer prevents a large fraction of these carriers from enjoying the speed advantage. The message here is that a forgiving contact is useful only when

coupled to an appropriate device design. Within the framework of the displaced Maxwellian, improvements in device structure are synonymous with values of electron temperature that are below that required for electron transfer.

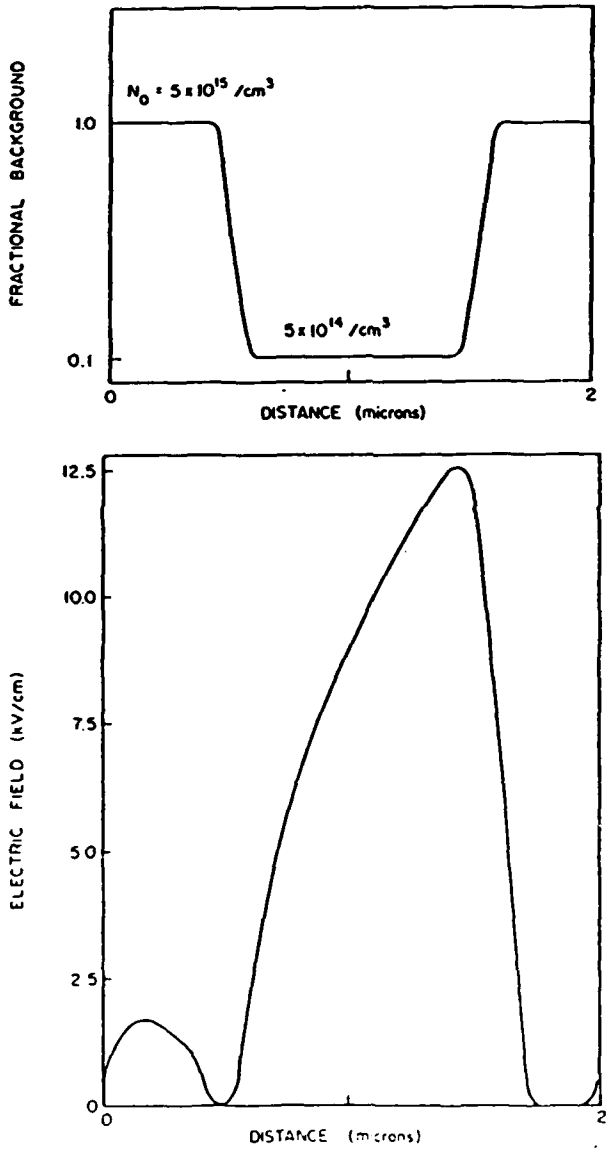


Fig. 14. (a) Doping distribution for N-N$^-$-N device. Bias is 1 V, and conditions at the cathode boundary are as in fig. 9. (b) Electric field distribution.

One way of reducing the electron temperature is simply to go to shorter device structures, and recent results for gallium arsenide at room temperature suggest that device structures not greater than 0.25 μm may be required. A structure that has recently been discussed within the framework of near and submicron devices is the N-N$^-$-N structure shown in fig. 14. This structure has four interfaces to contend with. The most significant aspect of this structure is that it introduces an abrupt change in field at the N-N$^-$ interface, which tends to emphasize overshoot contributions. The detailed input to this calculation is as follows. The N region is doped to 5×10^{15}/cm^3 with the N$^-$ region an order of magnitude less in doping. For this device the cathode ohmic conditions of fig. 9 were imposed, and an average field of 5 kV/cm was imposed. We note that over a distance of approximately 1.0 μm there is very little electron transfer. Current is carried mainly by Γ-valley electrons whose carrier velocity peaks near 6×10^7 cm/s, showing substantial overshoot. The electron temperature for this calculation is shown in fig. 15, and is reasonably low. V_{av} for this case is 1.3×10^7 cm/s.

The results of this calculation are very encouraging. However, they are not solely a consequence of structural device changes. The retention of Γ-valley electrons in the fig. 15 calculation is a consequence of using an electronic thermal conductivity appropriate to 5×10^{15}/cm^3 through the Wiedemann-Franz ratio. A reduction in electronic thermal conductivitiy results in steeper gradients in electron temperature. Fig. 16 shows results for a value of thermal conductivity in which substantial transfer occurs within 0.5 μm of entry into the N$^-$ region. The average current has dropped to 6.5×10^6 cm/s. The first set of results which virtually eliminate transfer must be regarded as an upper bound on current, whereas the second set is a lower bound. Actual device results are likely to fall between the two. One important measurement likely to provide significant information is whether a 1 μm, 10^{14}/cm^3 element will show small signal gain.

Returning to the boundary conditions, there are several points to be made. First, there is a clear indication as to the procedures necessary for generating an entire set of "ohmic" and blocking contacts. It is not likely, however, that the boundary prescriptions of figs. 9 to 13 are unique; other sets of conditions can be envisioned to yield similar results under DC conditions. Distinguishing between different boundary condition effects will come from time dependent studies. The question then is: "How model dependent are the results?"

The governing equations are derived starting from the condition of a displaced Maxwellian, which assumes strong electron-electron interaction. Boundary scattering is likely to substantially alter this interaction - near the boundary [25]. Results that appear to be model independent are those associated with the entrance velocity. If central valley carriers enter with speeds greater (less) than that dictated by equilibrium band structure considerations, carrier accumulation (depletion) will occur.

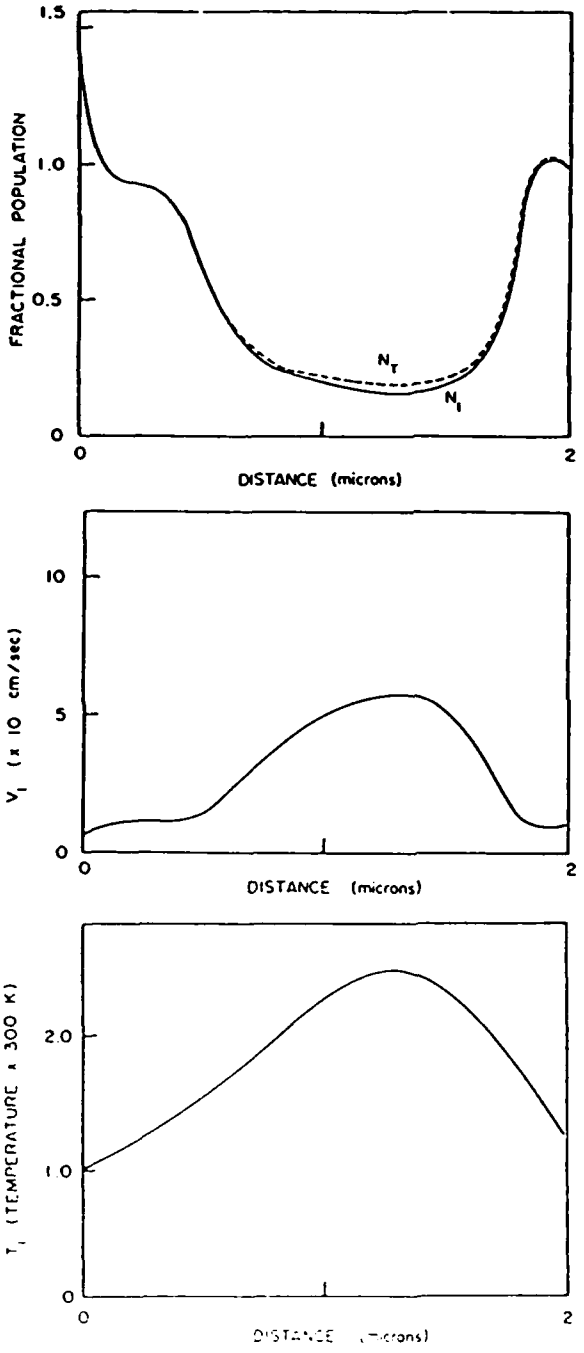


Fig. 15. (a) Fractional population (b) Central valley velocity (c) Electron temperature, for calculation of fig. 14.

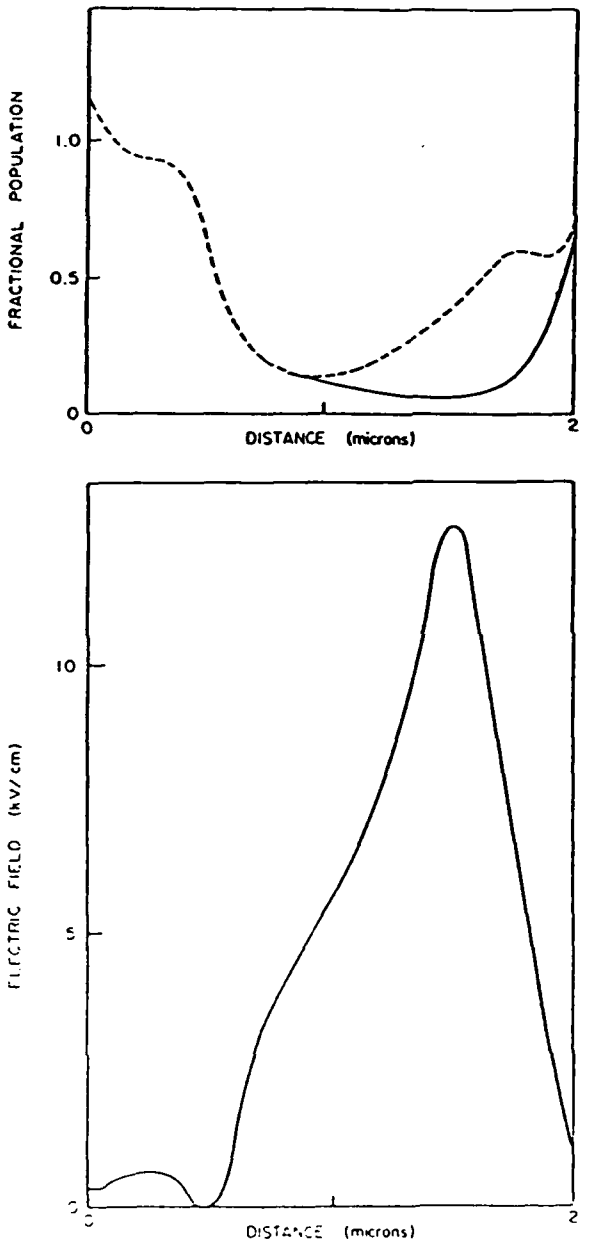


Fig. 15. (a) Fractional population (b) Electric field distribution, for parameters of fig. 14 with a reduced thermal conductivity.

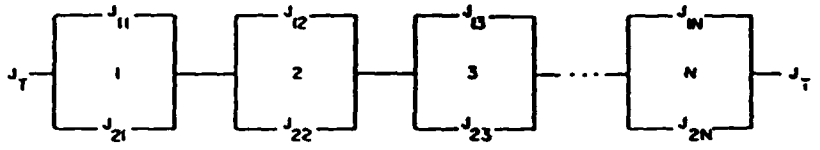


Fig. 17. Representation of spatial dependence of current flow through two level device as a serial chain of parallel elements. The first subscript in each element represents current through the central valley 1 or satellite valley 2.

Another point worth raising, particularly with regard to the absence of overshoot in figs. 9 to 13, is the fact that for at least half of these devices the space distributions were in steady state equilibrium with results similar to those of the drift and diffusion equations. While going to shorter device lengths does not necessarily prevent this from occurring [26], the length dependence has not been explored.

The results of figs. 9 to 13 also suggest that transport in a multivalleyed system may be controlled by one of the valleys. In the above examples, control is by the central valley; satellite valley boundary conditions are relatively benign. To see the reasons for this it is necessary to turn to fig. 17, which represents current flow through the nonlinear elements as a linear chain of parallel resistors. In this figure

$$J_{1_i} = N_{1_i} V_{1_i}, \quad i = 1, 2, \ldots, N,$$

$$J_{2_i} = N_{2_i} V_{2_i}, \quad i = 1, 2, \ldots, N.$$

The voltage across element i is $F_i \Delta X$. Thus if, e.g., the Γ valley in element 1 sustains a net carrier density below the background, it will have a higher resistance and voltage drop than an element with a density closer to background. This, of course, is consistent with the results of fig. 11 and suggests that the portion of the conduction band with boundary conditions most strongly departing from the uniform field conditions will be the dominant boundary condition.

The contact boundary conditions discussed in figs. 9 to 13 address only a small part of the problem. It is certainly unrealistic to assume that undesirable departures in doping will be absent. A simulation of a nominally $10^{17}/\text{cm}^3$ doped device, with a 10% decrease in doping over a distance of 1000 Å, is shown in fig. 18. The distortion in electric field, for the set of boundary conditions listed in the caption, is such as to prevent any real overshoot from occurring.

The above results which reflect the influence of space charge on transport in devices should be compared to uniform field calculations to indicate the goals that perhaps should be taken for some device structures. Fig. 19 displays velocity versus distance curves for electrons subjected to a sudden value of

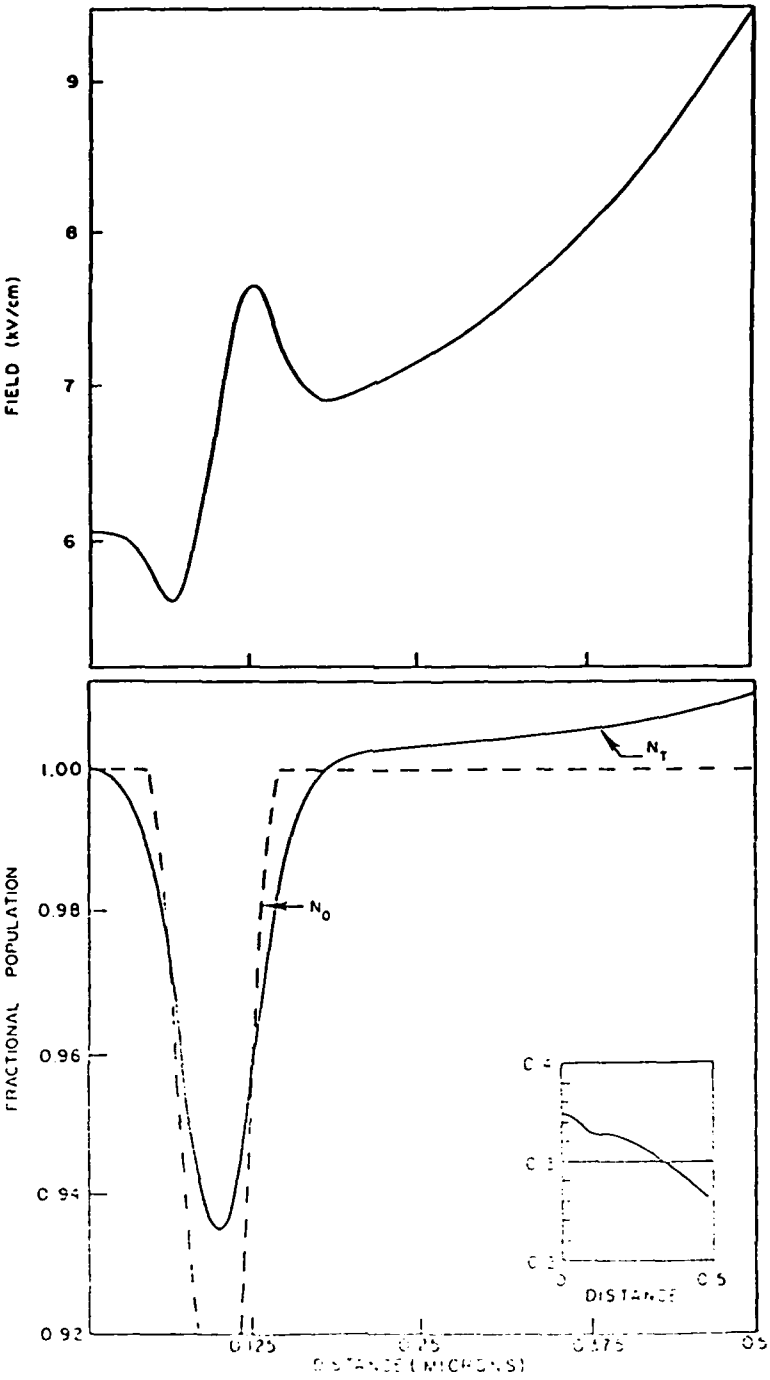


Fig. 18. (a) Electric field versus distance. (b) Carrier density versus distance for a "notched" device. Shown in the inset is the central valley carrier distribution. From ref. [25] with permission.

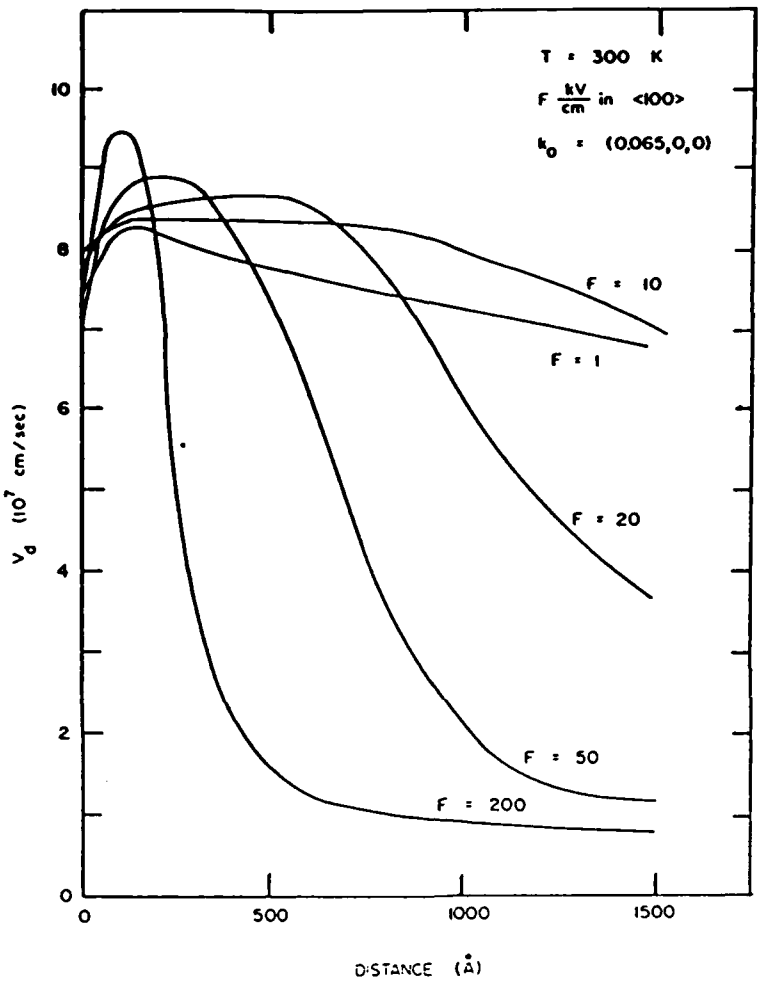


Fig. 19. Velocity versus distance for electrons with a finite entrance energy subjected to a sudden change in electric field. From ref. [27], with permission.

electric field. In this case the energy of the entering carriers is just below that required for electron transfer. It is seen that moderate values of electric field are necessary to sustain high transit velocities. In fig. 20 the dependence of transit velocity on entrance energy is calculated. It is seen that there is a window for high transit velocities, and it is this velocity level that is sought.

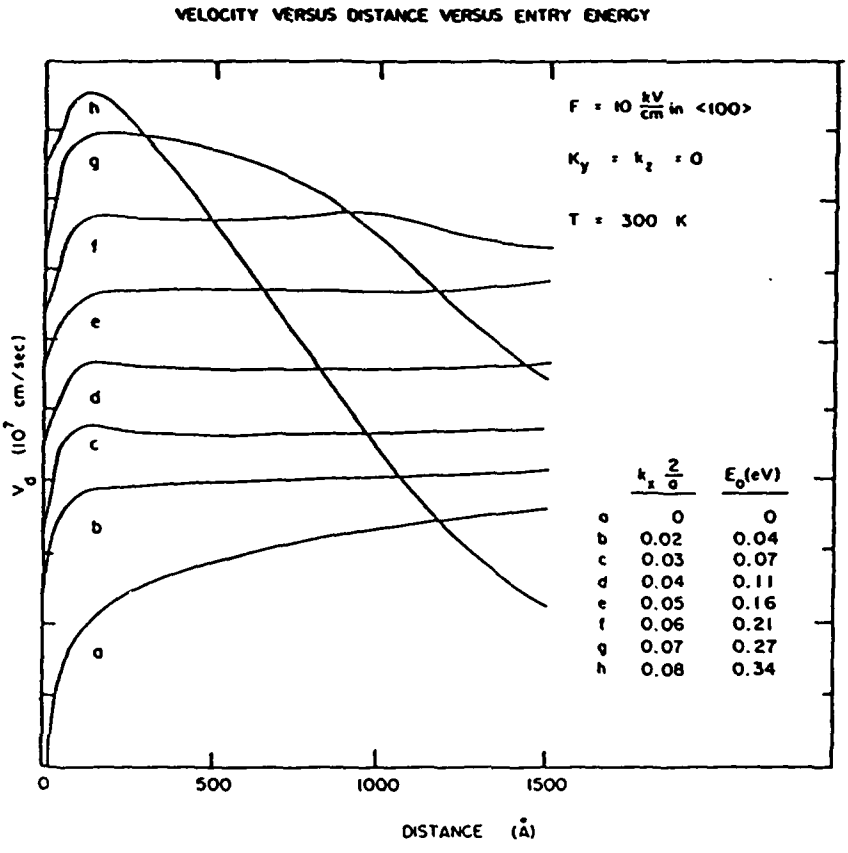


Fig. 20. Velocity versus distance for electrons with varying entrance energies, subject to a sudden change in electric field. From ref. [27], with permission.

4. Summary

The experimental situation is such that, with the exception of long compound semiconductor devices, there are very few data on the role of boundaries and contacts to submicron devices. The reason for the paucity of data lies in the fact that most submicron devices are three terminal device designs and the third terminal tends to mask the role of the contact boundaries. This is extremely unfortunate since it is likely that two terminal device measurements will indicate what can be achieved in controlling the entrance dynamics of the carriers. To date, most two terminal device measurements on simple device structures have concentrated on the role of transport *within* the device, and raise the question of whether "ballistic" motion is possible [28]. Based on the

history of vacuum tube dynamics [29] it should be recalled that, if transport is ballistic, the electrical characteristics will be controlled by the contacts.

The situtation in submicron devices is further complicated by communication between the up- and down-stream contacts. Thus it may be expected that the influence of a blocking contact on the electrical characteristics of long and submicron devices will be different. For submicron devices simple current voltage measurements may be rendered useless as a diagnostic tool. This is certainly not the case in long devices.

The role of numerical simulations in these boundary and device studies has been to act as surrogates for measurements that are not feasible. In one case, obtaining cathode boundary fields from measurements was not possible. Thus for long devices the sensitivitiy of the numerical results to numerical changes in the boundary conditions, when coupled to experiments, provided the key to the role of contacts on device behavior [1]. For submicron devices, the difficulties of direct correlation of experiment with specific transport phenomena are apparent and simulation through parametric studies will provide a key to the role of boundaries. But the description of transport on a submicron scale is still inadequate and the descriptive role of boundaries is correspondingly weak. For example, most space charge dependent problems still treat the background as a "jellium" distribution. The discrete nature of impurities is ignored, as are structural variations in the contacts. The extent to which this affects such measurements as current-versus-voltage is yet to be determined. Notwithstanding these uncertainties, a considerable amount of information can be obtained by extrapolation from the ideal cases which can provide bounds on the limits of transport through both the boundary and active region of the device.

Acknowledgments

The author acknowledges critical discussions with M.P. Shaw, G.J. Iafrate and D.K. Ferry. This work was supported by the Office of Naval Research and the Army Research Office.

References

[1] M.P. Shaw, P.R. Solomon and H.L. Grubin, IBM J. Res. Develop. 13 (1969) 587.

[2] D.J. Colliver, K.W. Gray, D.J. Jones, H.D. Rees, G. Gibbons and P.M. White in: Proc. 4th Intern. Symp. on GaAs and Related Compounds, Boulder, CO, 1972, Inst. Phys. Conf. Ser. 17 (Inst. Phys., London, 1973) p. 286.

[3] W.G. Gunn and D.K. Ferry, Appl. Phys. Letters 11 (1967) 337; J. Appl. Phys. 42 (1971) 2602.

[4] K.W. Boer, H.J. Hansch and V. Kummel, Z. Physik 155 (1969) 170.
K.W. Boer and G. Dohler, Phys. Rev. 186 (1969) 793.
see also M.P. Shaw, P.R. Solomon and H.L. Grubin, Solid State Commun. 7 (1969) 1619.

[5] See, e.g., D. Boccon-Gibodi, J.P. André, P. Baudet and J.P. Hallais, IEEE Trans. Electron Devices ED-27 (1980) 1141.
[6] D. Delagebeaudeuf and N.T. Linh, IEEE Trans. Electron Devices ED-28 (1981) 790.
[7] P.R. Solomon, M.P. Shaw, H.L. Grubin and R.D. Kaul, IEEE Trans. Electron Devices ED-22 (1975) 127.
[8] W.S.C. Gurney, Electron. Letters 7 (1971) 711;
see also K.W. Böer and G. Dohler, Phys. Rev. 186 (1969) 793.
[9] E.M.Conwell, IEEE Trans. Electron Devices ED-17 (1970) 262.
[10] M.P. Shaw, P.R. Solomon and H.L. Grubin, The Gunn-Hilsum Effect (Academic Press, New York, 1979).
[11] H.D. Rees, in: Metal-Semiconductor Contacts, Inst. Phys. Conf. Ser. 22 (Inst. Phys., London, 1974).
[12] H. Kroemer, IEEE Trans. Electron Devices ED-15 (1968) 819.
[13] V.L. Rideout, Solid State Electron. 18 (1975) 541.
[14] H.L. Grubin, R.F. Greene and G.J. Iafrate, to be published.
[15] W. Shockley, J.A. Copeland and R.P. James, Quantum Theory of Atoms, Molecules and the Solid State (Academic Press, New York, 1966).
[16] H.W. Thim, Electron. Letters 7 (1971) 106.
[17] T. Kurosawa, J. Phys. Soc. Japan Suppl. 21 (1966) 424.
[18] H.D. Rees, Solid State Commun. 26A (1968) 416; J. Phys. C5 (1972) 64.
[19] R. Bosch and H. Thim, IEEE Trans. Electron Devices ED-21 (1974) 16.
[20] See, e.g., A. Sommerfeld, Thermodynamics and Statistical Mechanics (Academic Press, New York, 1956) section 43.
[21] H.L. Grubin, D.K. Ferry, G.J. Iafrate and J.R. Barker, in: VLSI Electronics, Vol. 3 (Academic Press, New York, 1982) p. 198.
[22] H.L. Grubin and D.K. Ferry, J. Vacuum Sci. Technol. 19 (1981) 540.
[23] H.L. Grubin, D.K. Ferry and J.R. Barker, in: Proc. 1979 IEDM (1979) p. 394.
[24] B. Faugemberg, M. Pernisek and E. Constant, in: Proc. 1982 Workshop on the Physics of Submicron Structures, to be published.
[25] R.F. Greene, H.L. Grubin and G.J. Iafrate, in: Proc. 1982 Workshop on the Physics of Submicron Structures, to be published.
[26] H.L. Grubin, G.J. Iafrate and D.K. Ferry, in: Proc. 1981 IEDM (1981) p. 633.
[27] G.J. Iafrate, R. Malek, K. Hess and J. Tang, to be published.
[28] M.S. Shur and L.F. Eastman, IEEE Trans. Electron Devices ED-26 (1979) 1977.
[29] See, e.g., H.F. Ivey, Advan. Electron. Electron. Phys. 6 (1959) 138.

Transient Transport in Semiconductors and Submicron Devices

D. K. FERRY
Center for Solid State Electronics Research
Arizona State University
Tempe, Arizona

H. L. GRUBIN
Scientific Research Associates, Inc.
Glastonbury, Connecticut

G. J. IAFRATE
U.S. Army Electronics Technology and Devices Laboratory
Form Monmouth, New Jersey

I. Introduction 413
II. Overshoot Velocity Effects 415
III. Experiments on High-Speed and Submicron-Length Devices 422
IV. Moment-Balance Equation 431
V. The Correlation Functions 441
A. Self-Energy Corrections 442
B. The Intracollisional Field Effect 444
References 446

I. Introduction

The thrust of integrated electronics in recent years has been toward the realm of very-large-scale integration (VLSI). This has led to a steady increase in the number of devices on a single chip and to a reduction in the typical dimension of a single device. While research metal-oxide-semiconductor field-effect transistors (MOSFETs) have been made with channel lengths in the 0.1–0.25-μm range (Elliott *et al.*, 1979; Hunter *et al.*, 1980), simple experimental structures have been made in the 0.01–0.1-μm range (Broers *et al.*, 1978). These new developments and advances into

submicron semiconductor devices have been hampered by our lack of detailed understanding of nonequilibrium semiconductor transport on scales intermediate to the true atomic scale (< 10 Å) and the bulk solid-state macroscopic scale ($> 1\ \mu m = 10^4$ Å). The transport in this regime is complicated by the high electric fields and the resultant hot-carrier phenomena. For example, 1 V across a 0.1-μm channel produces an average electric field of 10^5 V/cm, enough to produce hot-electron effects in any semiconductor.

As early as 1963 hot-electron effects were suggested to be important in MOSFETs (Grosvalet *et al.*, 1963), and the role of velocity saturation was considered shortly thereafter (Trofimenkoff, 1965). Subsequently, the role of hot carriers in metal–oxide–semiconductor (MOS) and metal–semiconductor (MES) devices has been investigated and reviewed. Many of the problems that arise in transport in submicron devices are due to the very fast time scales inherent in these small devices. For example, an electron traveling at 10^7 cm/sec can cross a 0.1-μm channel in 10^{-12} sec, which is a time on the scale of the appropriate relaxation times for momentum, energy, and charge. On this time scale, the electrons encountering a high-field region of such dimension, as, for example, the pinch-off region in a MOSFET, do not have adequate time to establish any sort of equilibrium distribution, a point made rather pointedly earlier by Ruch (1972) and Maloney and Frey (1977). Additional complications arise from the fact that the collision duration is no longer negligible on this time scale and strongly affects the transport dynamics (Ferry and Barker, 1980a).

Many semiconductors exhibit velocity overshoot, in which the velocity can reach levels well above the steady-state value in transient response to high electric fields. Such effects have not generally been felt to be important in logic devices because the contribution to the delay time from the transit time is somewhat smaller than the contribution due to the interconnect capacitance. Yet, it is not generally appreciated that such overshoot effects have an indirect contribution by increasing the apparent saturation velocity in small structures. Thus it is not the direct contribution to a reduction in transit time, but the indirect contribution to the increased effective saturation velocity, that makes overshoot effects important in very small semiconductor devices. Clearly, the emphasis in studies of device scaling has been on questions other than whether or not our current understanding of device physics is adequate to handle devices on a smaller spatial scale.

In the modeling of semiconductor devices, the major physical effects are dominantly tied up with the manner in which the charge fluctuations and current response are coupled to the local electric field, formally related through the continuity equation. To determine the current response accurately, one must solve an appropriate transport equation, and it is in

these transport equations that many of the major modifications arising from these short time scales occur.

In the following sections we examine, in some detail, the considerations of overshoot that arise in transient transport. In the next section we first discuss the intuitive view of overshoot and the velocity autocorrelation that illustrates overshoot. We then turn to a discussion of the experiments that have been carried out in attempt to measure overshoot. Finally, we undertake a detailed theoretical treatment of the quantum transport equations to illustrate the proper balance equations.

II. Overshoot Velocity Effects

In an otherwise perfect crystal, the mobile electrons have a well-defined energy distribution, with some carrier energies very near the threshold for phonon emission due to the need for detailed balance to exist. As a result, even for small applied fields electron–phonon interactions will occur, and ballistic transport in the microscopic sense is not observable. Instead ballistic transport in the sense of the mean is conceptually possible where on a short time scale the mean velocity is approximately given by

$$\langle V \rangle = qFt/m^*. \tag{1}$$

Equation (1) implies an ensemble average in which electron–electron interactions are sufficiently rapid for the carriers to be in thermal equilibrium with each other. Further, for GaAs, t in Eq. (1) is generally less than the relaxation time for the Fröhlich interaction that is, $t \ll 0.3$ psec.

When we examine transport on a time scale longer than that for which Eq. (1) is approximately valid, dramatic nonequilibrium size-dependent effects occur. The phenomena have been referred to as velocity overshoot for reasons that will be apparent. We concentrate on gallium arsenide, which for simplicity we treat as a two-valley semiconductor.

At low values of electric field and at room temperature almost all of the carriers are in the central valley. In the absence of ionized impurities, detailed balance dominates and the mobility, a "mean" quantity, is determined by intravalley longitudinal-optic-(LO) phonon scattering. At high values of electric field intervalley phonon scattering becomes the dominant scattering mechanism, and there is an increased level of spontaneous phonon emission. For the central valley, velocity saturation occurs, but at values of velocity higher than the steady-state velocity. The general situation is represented by steady-state curves for the mean carrier velocity in the central and satellite valleys. Due to differences in the effective mass, the carrier velocity is highest for central valley electrons.

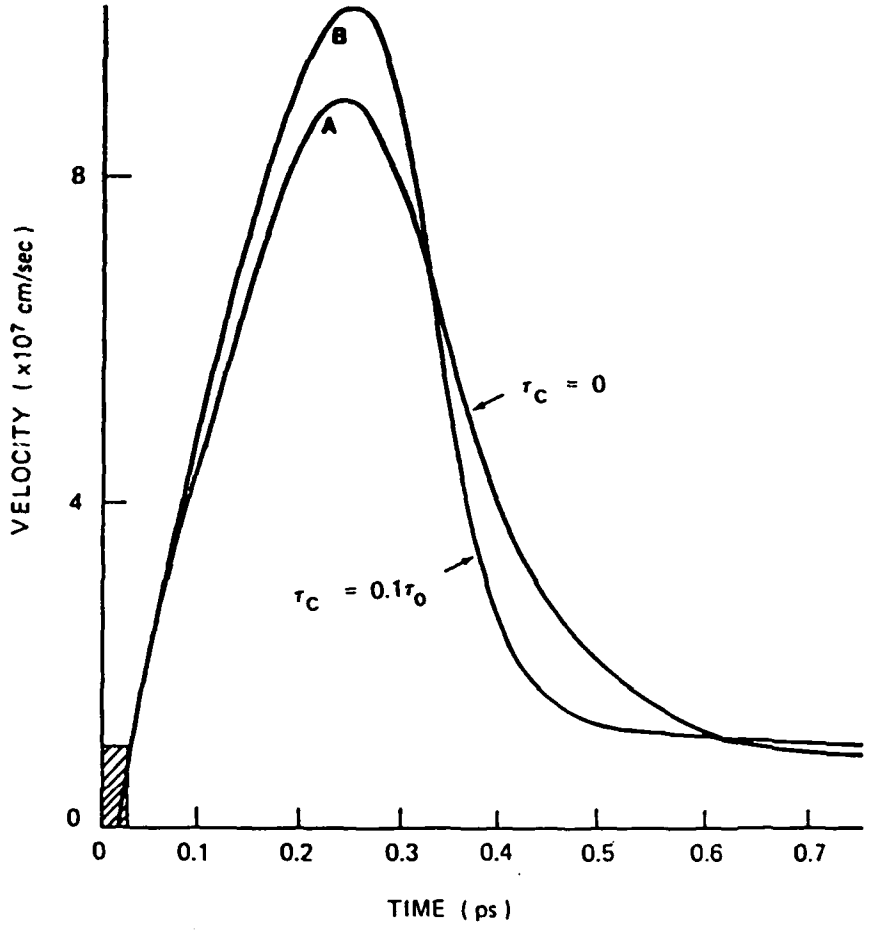


Fig. 1. Mean transient carrier velocity versus time for zero and finite collision duration, for GaAs, at 300 K. Results are obtained from the first three moments of the quantum kinetic equation. $N_0 = 10^{17}$; $F = 27$ kV/cm.

Once we recognize that a finite amount of time is necessary for carriers to transfer from the central to satellite valleys, we see the immediate possibilities of high transient velocities. Typical transient curves are shown in Fig. 1: one result is for zero collision duration and a second for finite collision duration. The crosshatched region, where the expired time is less than the LO-phonon relaxation time, is referred to as the mean ballistic regime.

The calculation of Fig. 1 can be translated into a velocity-versus-distance transient. Then repeating the calculation at different values of field, we obtain the results in Fig. 2 (Grubin and Ferry, 1981). It is clear that as the device dimensions are reduced, the apparent saturated drift velocity increases while the presence of negative differential mobility is eliminated.

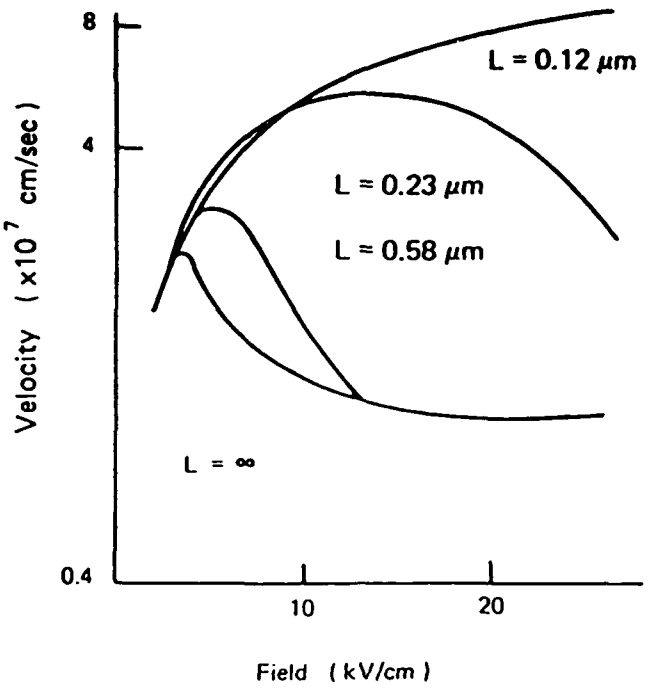


Fig. 2. Field dependence of the effective drift velocity (L/t) for GzAs. The parameter is the transit distance.

This particular result is highly significant insofar as it suggests the possibility of engineering a size-dependent device, which in one case offers negative differential mobility, and in another velocity saturation. Similar effects are also observed in Si, where the overshoot arises from differences in the relaxation time. If the energy relaxation time τ_E is greater than the momentum relaxation time τ_M, then the velocity first rises to a peak and then decays to a steady value as the scattering rate increases due to the increase in the average energy.

A conduction electron in a semiconductor is not a free electron. Within the effective mass approximation, the conduction electron is a quasi-particle whose effective mass describes an averaged (and renormalized) interaction with the atoms and bound electrons. It should be apparent that the correlation between electrons is important. This is more evidenced when it is recognized that a kinetic equation, such as the Boltzmann transport equation, for the average single-particle distribution function is an approximation to the full many-body problem of large numbers of conduction electrons (Bogoliubuv, 1946; Kreuzer, 1981). Although it is possible to project such single-particle equations, these must be cast so that a description of the correlations between the electrons—the correlation functions—is recoverable. Indeed, it was just this view that led Kubo (1957,

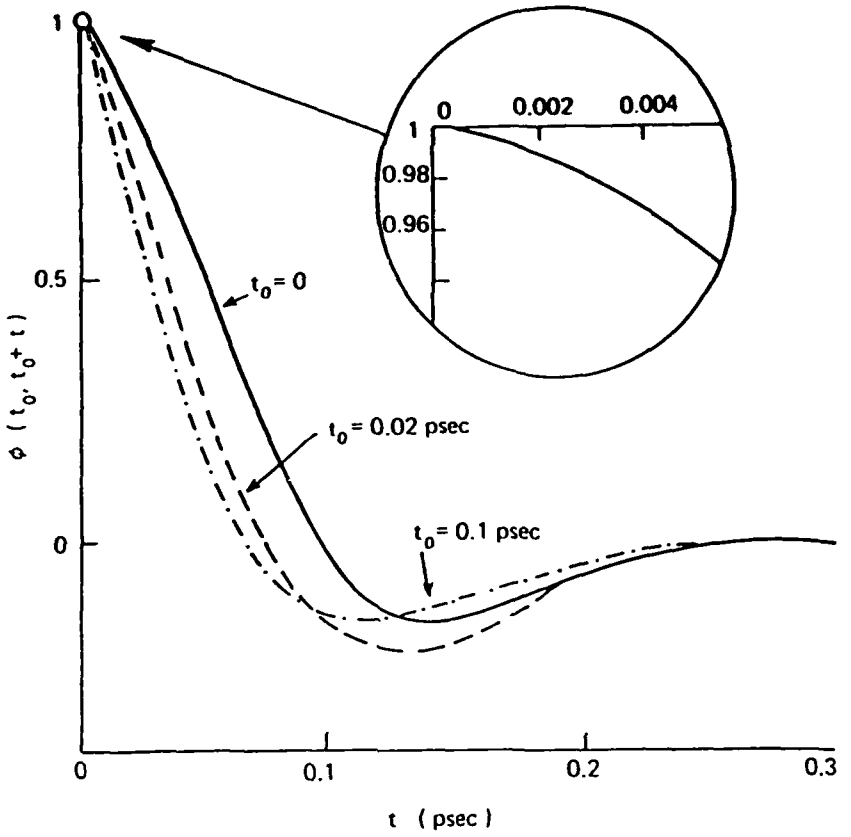


Fig. 3. Normalized velocity autocorrelation function $\phi_v(t - t_0, t_0)$ for electrons in Si for three initial times t_0. From the inset, it is apprarent that ϕ_v is constant for only a very short time.

1974) to formalize transport theory entirely in terms of the correlation functions. We propose to show how proper treatment of the initial transient response limits the range of ballistic transport.

In a semiconductor subject to an applied electric field, the carriers respond to this field, as well as to a random force, which leads to velocity fluctuations, as well as to the concept of temperature (thermal fluctuations). The response to the applied field (applied at $t = 0$) can be written as

$$v_d(t) = (eF/m^*) \int_0^t \phi_v(t', 0)\, dt'. \tag{2}$$

where $\phi_v(t', 0)$ is the nonstationary, two-time velocity autocorrelation function whose amplitude is normalized to unity at $t' = 0$. Although Kubo obtained Eq. (2) for the equilibrium situation, its validity has also been established for the nonequilibrium, high-field case (Zimmermann *et al.*, 1981). A linear increase in the velocity, as in Eq. (1), can only occur so long

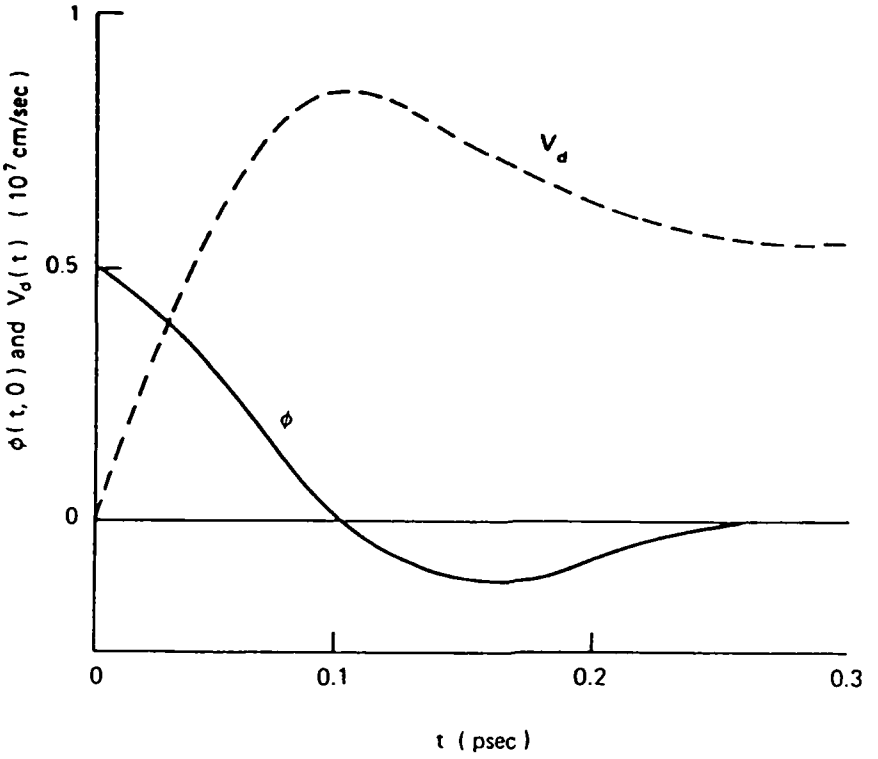


Fig. 4. The typical evolution of $\phi_v(t - t_0, t_0)$ and the transient drift velocity illustrating the significance of Eq. (2). The velocity peak occurs at the point where ϕ_v passes through zero, and the subsequent decline in velocity is related to the negative excursion of ϕ_v.

as ϕ_v is constant in time. To demonstrate this period, we have calculated the transient dynamic response and velocity autocorrelation function for silicon, using an ensemble Monte Carlo approach used previously (Ferry and Barker, 1981). Figure 3 shows the initial decay of $\phi_v(t', t_0)(t' > t_0)$ for three different values of t_0 including $t_0 = 0$. What is evident from this figure is that the time duration over which ϕ_v is constant, and for which the velocity rises in a ballistic manner, is exceedingly short, perhaps only 0.004 psec (in Si), even though the mean free time ($\sim$0.16 psec at $t_0 = 0$) is much longer. The decay of ϕ_v represents a decay in the correlation of the velocity fluctuations, as well as an increase in the dissipation because of collisions. This is due to the fact that collisions start to break up the correlation as soon as they occur, but the relaxation time is a hydrodynamic average over the ensemble, rather than a value for actual collision.

We also note from Eq. (2) that the presence of overshoot in the transient velocity is intimately connected with a negative portion of $\phi_v(t, t_0)$. This is illustrated more clearly in Fig. 4. Here the velocity and ϕ_v are calculated for electrons in Si at 300 K for an electric field of 50 kV/cm applied at $t = 0$

Doob's (1942) theorem states that for a stochastic process that is stationary, Gaussian, and Markovian, the autocorrelation function of the process is a pure exponential. Applying this to our system when the steady state is reached in the presence of the electric field and where the autocorrelation function of the velocity fluctuations is not an exponential, we infer that the system is neither Gaussian nor Markovian. The fact that the system is not Gaussian has long been recognized, since in the presence of a field in the hot-carrier regime, the velocity distribution departs strongly from a Gaussian distribution. The same is true in real space: If initially the system has a δ-function spatial distribution and then evolves freely in response to the external field, diffusion (i.e., velocity fluctuations) causes the packet to spread around its mean position, and the departure from a Gaussian is best seen by calculating the third and fourth cumulants of the distribution. As is well known, these cumulants vanish for a Gaussian but differ significantly from zero in the hot-electron case (Ferry and Barker, 1981; Lugli *et al.*, 1981). Clearly, the nonlinear, nonequilibrium transport has driven us away from classical statistical mechanics.

An equation such as Eq. (2) is readily obtainable from a retarded Langevin equation (Zimmermann *et al.*, 1981). Such an approach is appropriate since it yields an equation for $v_d(t) = \langle v \rangle$ that we shall derive in a later section. Indeed, there we shall calculate balance equations, correct even on the short time scale, that may be expressed as

$$m^* \frac{dv_d}{dt} = qE - m^* \int_0^t dt' \, X_v(t') v_d(t - t'), \tag{3a}$$

$$\frac{d\bar{E}}{dt} = qFv_d(t)[1 - \phi_v(t, 0)] - \int_0^t dt' \, [\bar{E}(t - t') - E_0] X_e(t'), \tag{3b}$$

where X_v and X_e are decay functions related to the velocity and energy correlation functions. Indeed, the time integrals of X_v and X_e give the relaxation rates $1/\tau_m$ and $1/\tau_e$ in the limit as $t \to \infty$. In this regard, Eq. (3a) is particularly consistent with the correlation function approach of Eq. (2), as it is an ensemble average of a generalized, retarded Langevin equation. Except for the details of the correlation function and the convolution in the relaxation terms, Eq. (3a) differs little from earlier forms. Such is not the case for the energy equation, Eq. (4), because of the presence of the memory function in the driving term.

In studying the transient regime, one considers generally a population of electrons at equilibrium with the crystal lattice (i.e., described by a Maxwell–Boltzmann distribution), which, in our case, are suddenly subjected to an external homogeneous electric field. Then the k distribution of carriers evolves from the equilibrium one to a far-from-equilibrium, steady-state distribution imposed in the presence of the electric field. Equations

such as Eq. (3) are useful in this regard. Moreover, the presence of the non-Markovian convolution integrals is a result of the nonlinear evolution of the system, in which the scattering rates themselves are evolving in time. This is evident in the nonstationary character of $X_\alpha(t')$, which is related to $\phi_\alpha(t_0, t - t_0)(t' = t - t_0)$ in Figs. 3 and 4. The essential feature of this function is that it exhibits a negative part that exists in the range of time during which the ensemble drift velocity experiences overshoot, and this negative part persists even when the system has reached its nonequilibrium steady state. Such behavior characterizes a system that evolves with two kinds of relaxation, namely momentum and energy, characterized by very different time scales for each in semiconductors. It follows that, in this case, the autocorrelation function departs strongly from an exponential, except when the electric field is sufficiently small.

We now want to examine the frequency dependence of the small-signal conductivity, in which stationarity is reached. Using Eq. (3) in the frequency domain, we obtain the frequency dependence of the mobility

$$\mu(\omega) = (e/m)[i\omega + X(i\omega)]^{-1}, \tag{4}$$

and $X(i\omega)$ contains all the details of the frequency dependence. The complex structure of $X(i\omega)$, including the effects of energy and momentum relaxation in the system, generally leads to a real mobility, which, as frequency increases, at first increases and then decreases after experiencing a maximum (most often in the far-infrared domain). Indeed, as has been established by Kubo (1974), the generalized susceptibility of the system (the mobility, in the present case) is proportional to the Fourier transform of the velocity fluctuation correlation function [at least in the steady state, but Eq. (3a) is valid in this case as well], and Eq. (3a) allows this result to be recovered once it is recognized that the correlation function has an important negative part, as shown in Fig. 3. When the carrier dynamics are modeled with equations like Eq. (3a), which are non-Markovian in nature, we can recover the essential results known from previous studies on hot-carrier effects, namely, velocity overshoot, the negative part of the velocity fluctuation correlation function, and the proper spectrum of the mobility as a function of frequency.

An approach similar to that of Eq. (4) was previously proposed for application to the calculation of the dynamic conductivity of simple metals (Götze and Wölfle, 1972) but is certainly older (Zwanzig, 1960). It was introduced by Götze and Wölfle in order to explain the high-frequency behavior of the conductivity, and the connection to the classical Drude theory is possible only in the low-frequency domain, or long time limit, of the system. More recently a similar approach has been used for semiconducting surface inversion layers (Ting *et al.*, 1976).

III. Experiment on High-Speed and Submicron-Length Devices

Velocity overshoot is a transient nonequilibrium effect. It occurs on time scales of a picosecond or less and almost certainly represents the response of a distribution of carriers to *sudden* changes in field. Both requirements may preclude manifestation in today's commercially available devices, although some new design structures should overcome these difficulties (Grubin *et al.*, 1982). Independent of design, the time scales of the phenomenon place unrealistic demands on most electrical instrumentation and have resulted in all-optical measurements, as discussed in the following.

In Fig. 5 the physical time scales associated with momenta and energy relaxation are compared to existing and projected optical instrumentation. The scattering rates are those appropriate to gallium arsenide and represent interaction within different portions of the conduction band, at different values of field. The initial overshoot phenomenon occurs at the lower end of the scale, where the energy scattering rates may be several orders of magnitude less than those for momenta. The first attempt (Shank *et al.*, 1981) to measure this transient response involved the "pump-and-probe" technique summarized later in this section. The method was applied to three different bias-dependent overshoot situations. Since the efficacy of the

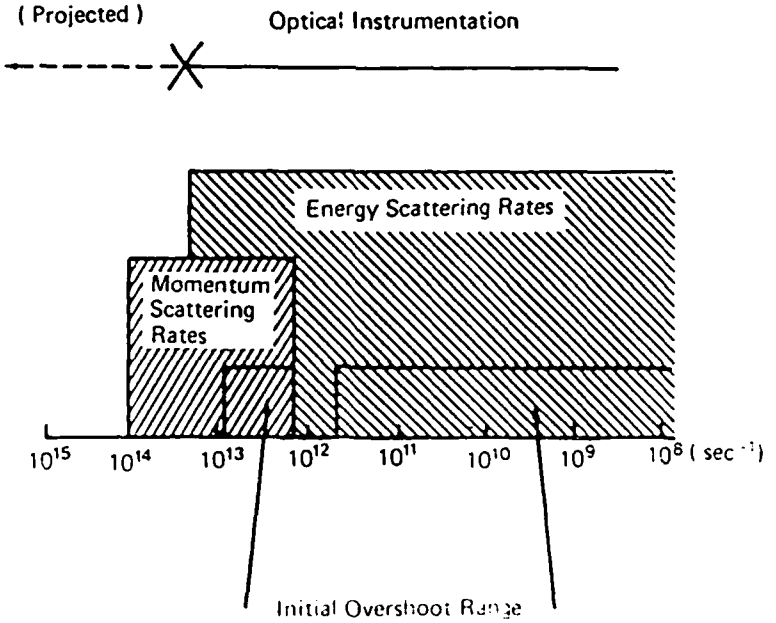


Fig. 5. Comparison of momentum and energy relaxation times and optical instrumentation time scales. Lower blocks denote initial overshoot range.

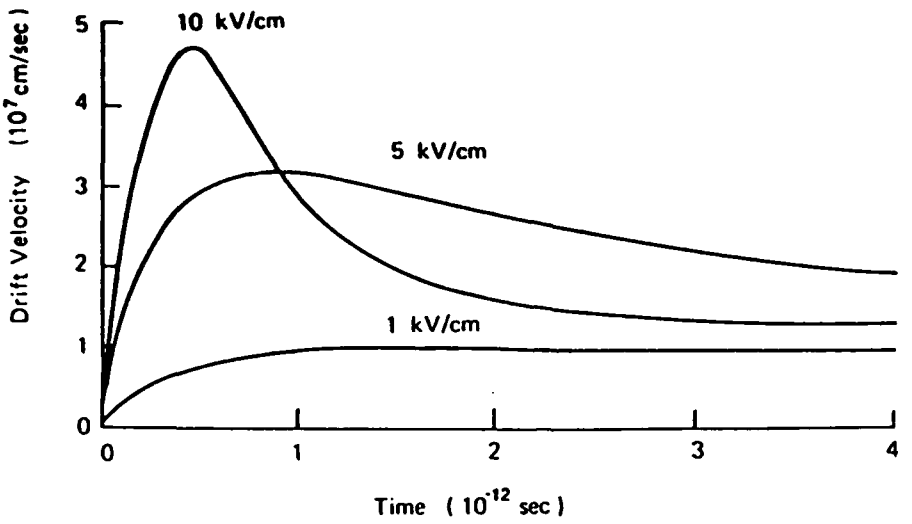


Fig. 6. Time dependence of electron drift velocity, at three values of electric field, for GaAs. (From Ruch, 1972.)

technique is dependent on the degree to which bias-dependent results are correlated with overshoot as presently understood, some discussion of the theoretical predictions is in order.

The phenomenon of overshoot was discussed in an earlier section and illustrated in Figs. 1–3. Quantitative differences occur at different bias levels, as any number of publications have shown. For example (Ruch, 1972), in Fig. 6 calculations at 1, 5, and 10 kV/cm show progressively higher peak velocities and shorter time intervals before the mean carrier velocity attains its steady-state value. The attainment of progressively higher velocities is due to Γ valley carriers acquiring high speeds before transfer. The realization of shorter velocity relaxation is related to two features: (1) Carriers are heated more rapidly, to higher electron temperatures, at higher bias levels; this, in turn, leads to (2) more rapid scattering rates. It should be noted, however, that the presence of shorter velocity relaxation does not imply steady-state equilibrium at 1 psec. Energy relaxation generally requires larger time intervals than velocity relaxation.

The pump-and-probe method used to examine overshoot involves photoexcitation by picosecond lasers. As illustrated in Fig. 7, a beam from a passively mode-locked dye laser is split into two beams; one part excites the GaAs layer with a short pulse, while the second optical pulse is spectrally broadened and used to probe the absorption spectrum at delayed times.

The experiments were performed at 77 K on the structure shown in Fig. 8, with pump and probe through the etched window. A variable optical delay line provides the timing between pumping and probing. A voltage is applied

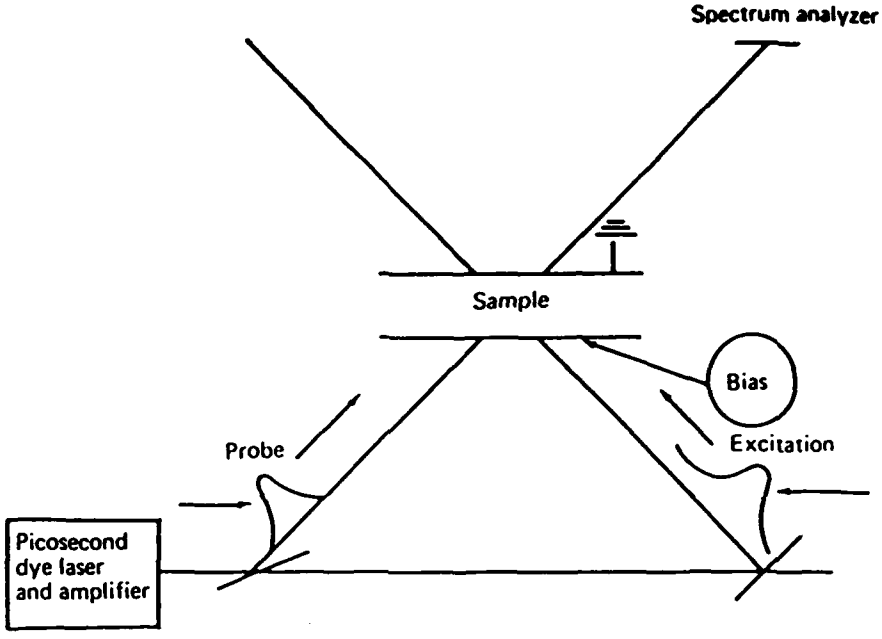


Fig. 7. Schematic of experiment for determining overshoot.

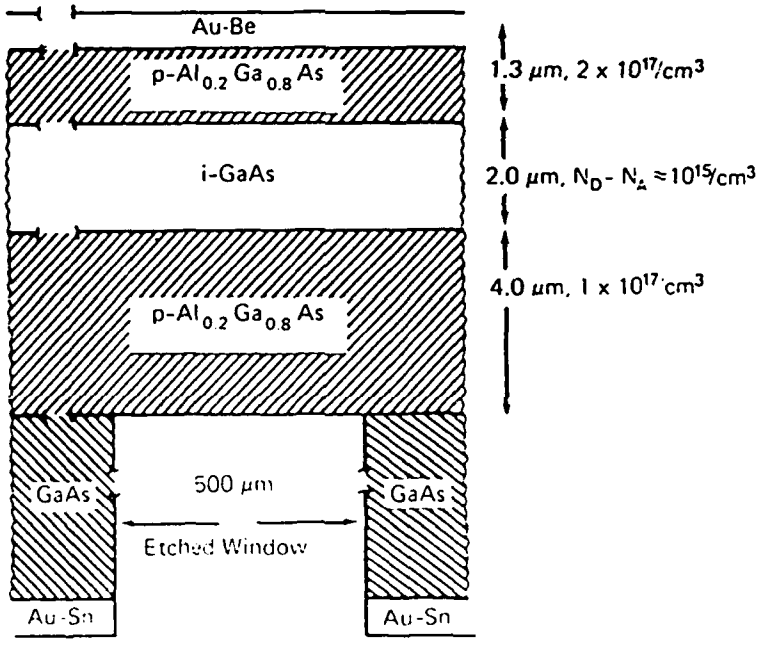


Fig. 8. Sketch of sample used in overshoot experiment.

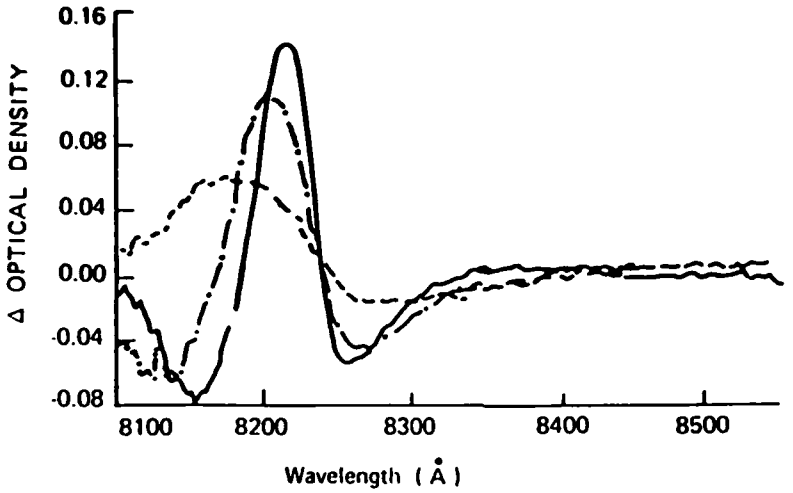


Fig. 9. Induced optical density change in GaAs at $t = 20$ psec following excitation by a 0·5 psec pulse at 8030°A. ——, 14 kV/cm; —·—, 22 kVcm; – – –, 55 kV cm. From Shank *et al.*, 1981.)

across the sandwich structure and modifies the optical absorption, as described by Franz (1958) and Keldysh (1958). The optical absorption is further modified by carriers optically injected near the band edge. The optically injected electrons and holes drift to opposite ends of the sample, thereby altering the net field across the device. The small space-charge-induced field perturbs the Franz–Keldysh effect, and the temporal evaluation to steady state is monitored.

The experiments measure the optical absorption spectra as a function of relative delay time. The differential optical absorption spectra are then obtained by subtracting the spectra before and after a specified relative delay time. The differential spectra after a 20-psec delay are shown in Fig. 9 for three applied voltages. Note the dependence of the period of oscillation on the applied field. The amplitude of the absorbance change is obtained by adding the area under the positive and negative portions of the differential absorption spectra, and this is shown as a function of time in Fig. 10. The points in Fig. 10 are data, and the solid line represents a fit through the equation

$$\begin{aligned}\Delta\alpha(t)\,\Delta\alpha(\infty) = \{d[1-\exp(-\alpha_p d)]\}^{-1} \\ \times\left(\left[\int v_e\,dt - \int v_h\,dt\exp(-\alpha_p d)\right]\right. \\ \left.+\,\alpha_p^{-1}\left\{1-\exp\left(-\alpha_p\int v_h\,dt\right)-\exp(-\alpha_p d)\right.\right. \\ \left.\left.\times\left[1-\exp\left(-\alpha_p\int v_e\,dt\right)\right]\right\}\right).\end{aligned} \tag{5a}$$

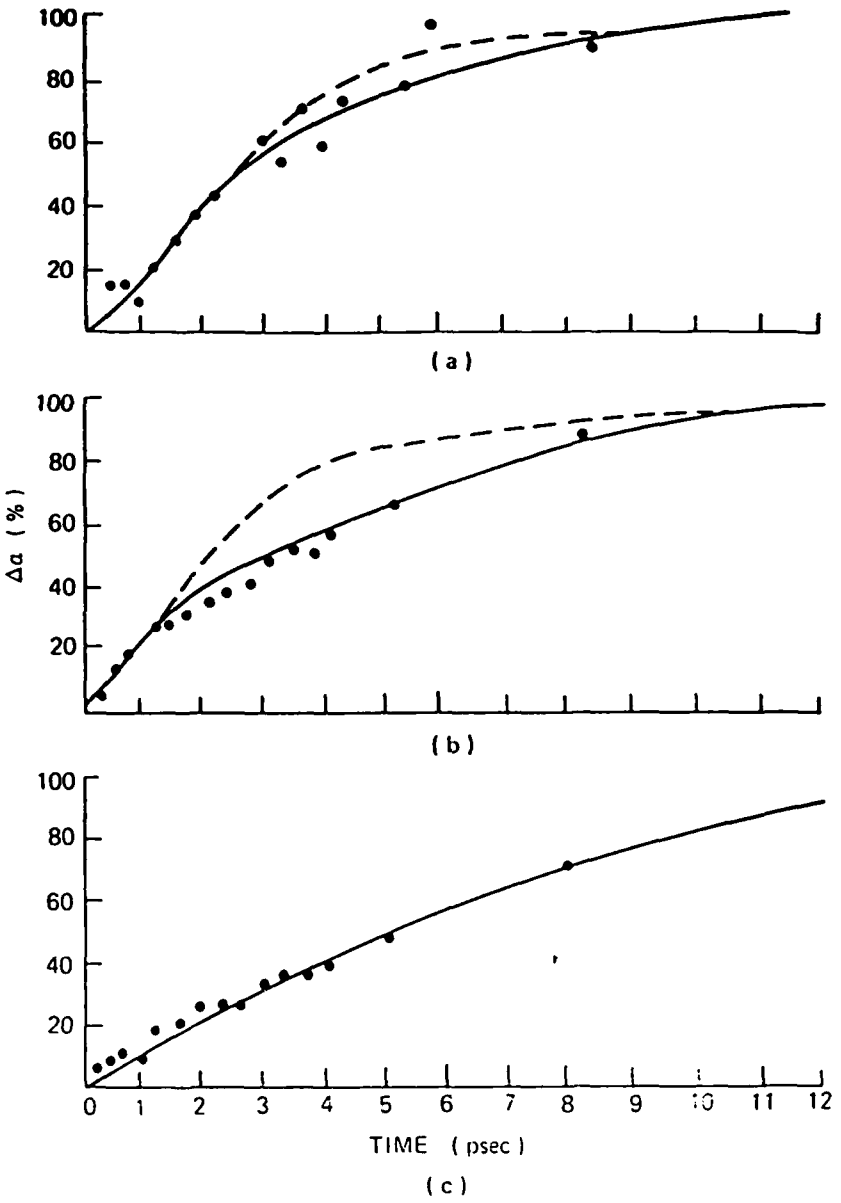


Fig. 10. Plot of induced absorbance as a function of time. (a) 14 kV cm, (b) 22 kV/cm; (c) 55 kV cm. (From Shank *et al.*, 1981.)

where α_p is the pump absorption constant and d the width of the region over which the field is applied. The velocities in Eq. (5a) are of course time dependent, and although Shank *et al.* (1981) did not feel it necessary, within the resolution of their data, to incorporate the full time-dependent velocity overshoot calculation, it was necessary for at least two cases to assume a two-valued electron velocity-time relation. For the 14-kV cm result the

dashed curve represents a fit with a single value of electron velocity chosen to match the data near $t = 0$, where

$$\frac{\partial}{\partial t}\frac{\Delta\alpha(t)}{\Delta\alpha(\infty)} \simeq \frac{v_e + v_h}{d}. \tag{5b}$$

Here v_h is taken as 10^7/cm · sec. There is an apparent decrease in the slope of the absorbance, and a better fit, represented by the solid curve, is obtained from a two-velocity fit. For the 22 kV/cm result, the dashed curve is for a single value of velocity = 4.4×10^7 cm/sec. The solid curve is for a two-valued velocity with $v = 4.4 \times 10^7$ cm/sec for $t < 1.1$ psec and 1.2×10^7 cm/sec for $t > 1.1$ psec. For the 55-kV/cm result a single fit of 1.3×10^7 cm/sec appears adequate.

There are several points worth noting. First, the slopes of the absorbance curves are higher at the lower bias levels suggesting (Shank *et al.*, 1981) the presence of longer overshoot relaxation. Further, the initial overshoot velocity is higher for the 22-kV/cm measurement than the 14-kV/cm measurement, a result also consistent with calculation. The interpretation is less direct with the 55-kV/cm measurement, where the initial slope is significantly below that of the lower field measurement. Shank *et al.* (1981) argue that velocity relaxation occurs during a shorter time interval, as the results of Figure 6 suggests, and that the measurements are not sensitive to this time duration. However, because of the increased value of the high-field overshoot velocity (near 10^8 cm/sec), some variation in the slope of absorbance is expected. At 55 kV/cm, the influence of the AlGaAs layers, particularly of the boundaries, and of the finite hole relaxation must be considered.

As illustrated in Figs. 1 and 2, the velocity overshoot and nonequilibrium calculation can be translated into velocity versus field and distance. For 0.12 μm it was seen that the velocity exhibited saturation without negative differential mobility. Device length in excess of 2 μm showed the full steady-state velocity–field relation. The first published attempts at demonstrating this length dependence was due to Lavel *et al.* (1980). For these transport measurements the device structure consists of a microstrip line deposited on a lightly doped GaAs expitaxial layer (Fig. 11). The upper strip has a gap, on each side of which are heavily doped contact regions. A voltage is applied across this gap. The gap is then uniformly illuminated, and excess carriers are assumed to be injected uniformly across the gap so that the electric field across the gap is simply $E_x(x, t) = V(t)/L$, where $V(t)$ is the gap voltage. The photoinduced current is thought to arise from a thin layer of high conductivity near the surface and to be a direct measure of the carrier velocity. Measurements of the photoinduced current have been made on devices of different gap length. The results of measurement for a gap length

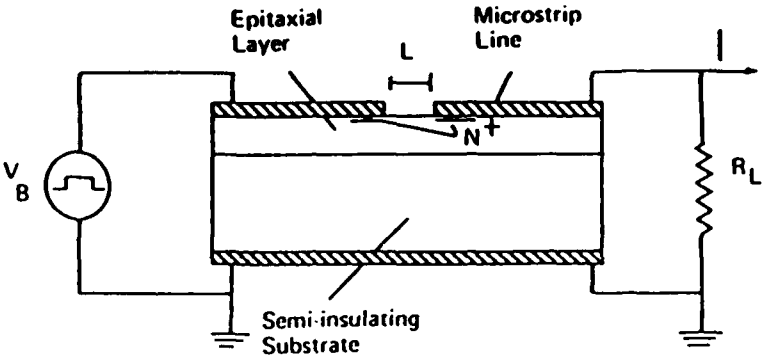


Fig. 11. Structure of a system for measuring velocity overshoot. $L = 385\,\mu m$. (From Laval *et al.*, 1980.)

of $385\,\mu m$ are shown in Fig. 12. The results show a peak value near the steady-state value of electron transfer. They also show the unusual feature of a ratio of peak to minimum current in excess of 3:1. This suggests that nonuniformity in the carrier density, and hence in the electron field, is present, arising either from contact, nonuniform surface depletion, or band structure contributions. Measurements on the 385-μm gap device have also been performed under varying incident light wavelengths and intensity with relatively little alteration in results. It had been suspected that under sufficiently uniform field conditions a sufficient number of carriers could have been placed in the satellite valleys with a substantial reduction in photocurrent. This proved not to be the case, although shorter gap-length devices did show current reduction. The results for shorter devices are

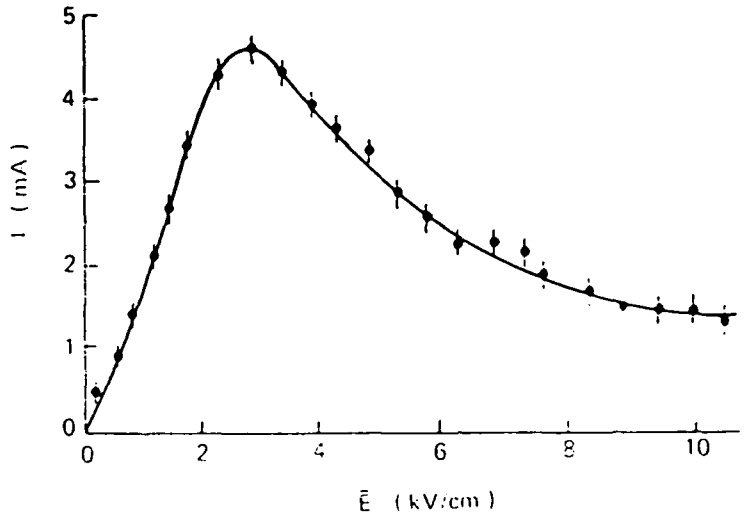


Fig. 12. Photoconduction current versus average electric field (From Laval *et al.*, 1980.)

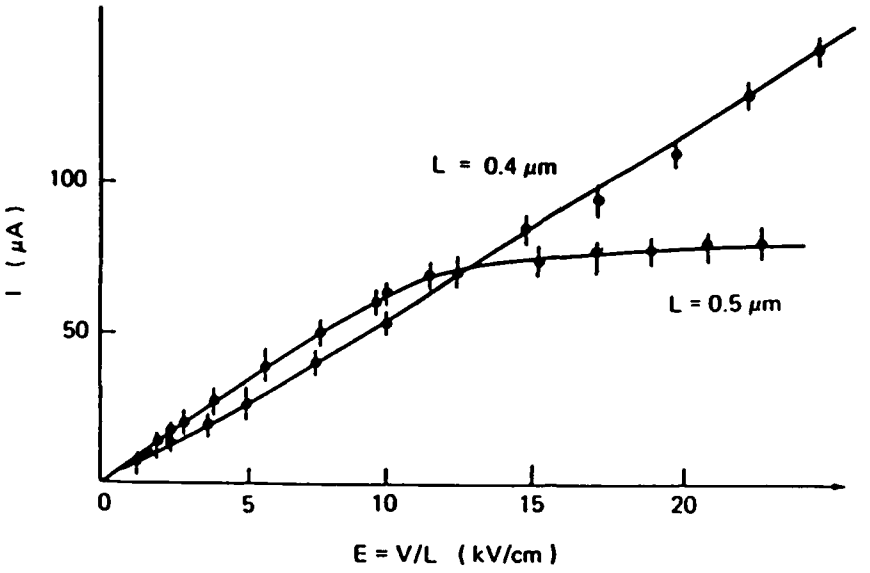


Fig. 13. Photoconduction current versus average electric field for short gap devices and 1.55 eV photon energy (λ = 8000 Å). (From Laval *et al.*, 1980.)

shown in Fig. 13. The 0.4-μm gap device displays no saturation; instead it shows an almost linear current voltage characteristic. The longer gap device shows saturation and perhaps electron transfer. We note the higher mobility at low fields for the 0.5-μm device. When the photoenergy is increased, there is a marked reduction in photoinduced current due to injection of carriers into the heavy mass valley, as shown in Fig. 14.

The two groups of measurement just discussed concern the measurement of velocity transients and the effects thereof. All-optical device measurements have also been recently developed. These measurements, developed by Auston *et al.* (1980) and Smith *et al.* (1981), although "all optical," are analogous to the pulse generator and sampling gate of conventional systems. The technique is illustrated in Fig. 15.

In this measurement technique a train of picosecond pulses is split into two beams. One beam passing to the generator photoconductor causes its electrical resistance to drop, allowing an electrical pulse to pass toward the device. A second pulse passes to a sampler photoconductor, where the signal from the device provides the necessary bias. By varying the relative arrival time of the optical pulses to the generating and sampling photoconductor through time delay of the latter, the arrival time of the propagating electrical pulse is sampled as a function of time delay. One measurement of interest was obtained when the device tested was a GaAs FET. Here the gate of the device received a signal from the generator photoconductor, while the signal to the sampler photoconductor originated at the drain contact. The

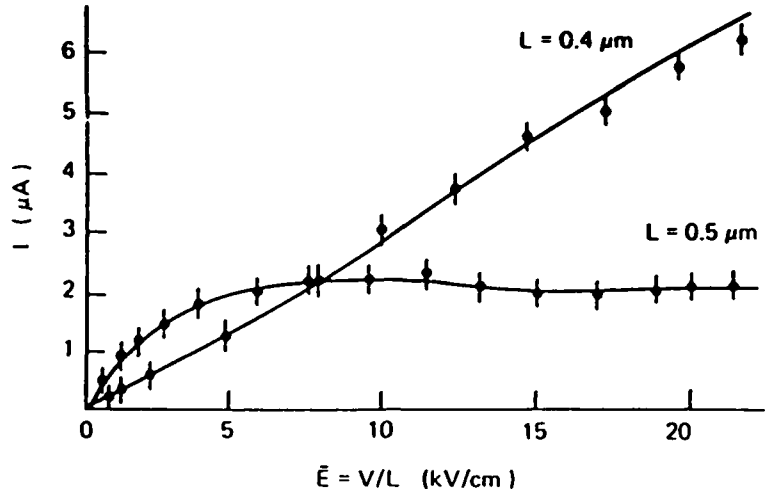


Fig. 14. Photoconduction current versus average electric field for short-gap devices and 1.7 eV photon energy (λ = 7300 Å). (From Laval *et al.*, 1980.)

measured response as a function of time delay is shown in Fig. 16. The device was a 0.6-μm gate length GaAs FET and was biased to provide maximum gain. The dashed curve in Fig. 16 is the measured impulse response with a 50-Ω coaxial cable replacing the FET. The device response displayed two characteristic time components: a fast component with a 40-psec FWHM, followed by a lengthy falling edge. The device response was obtained by subtracting the circuit response from the sum of squares, resulting in a response time of approximately 23 psec.

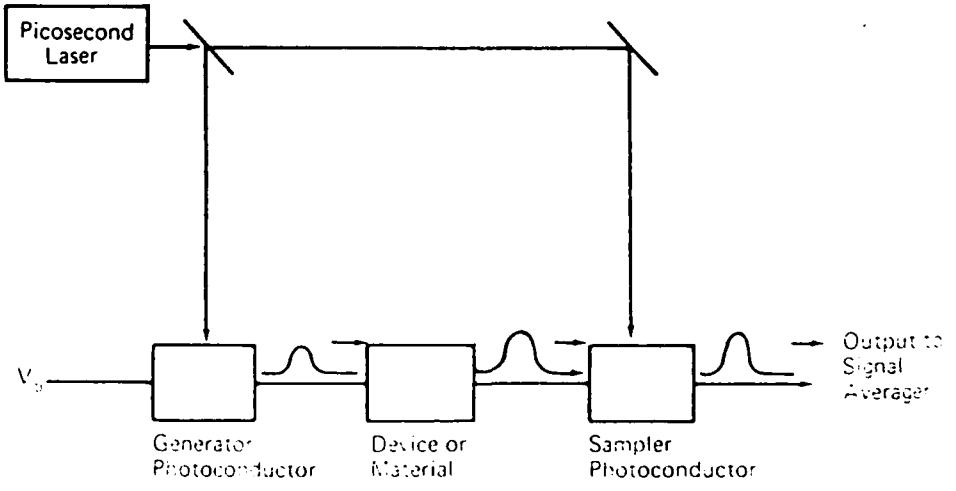


Fig. 15. Schematic of picosecond optical device measurement

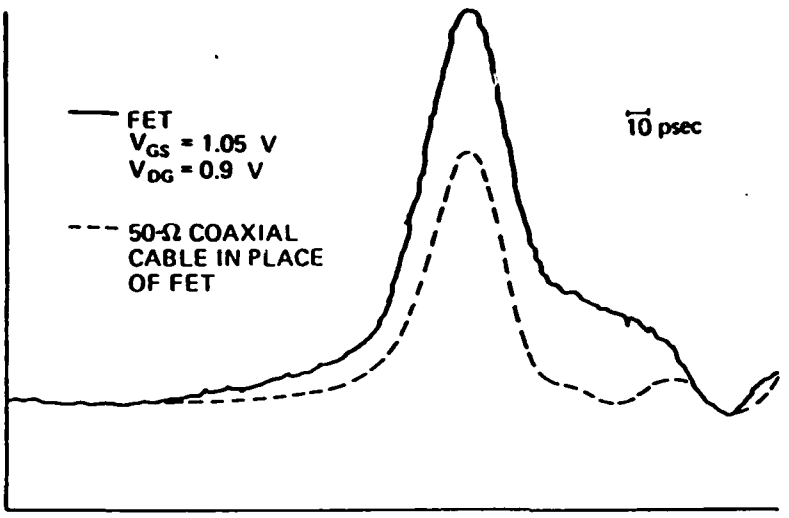


Fig. 16. Electronic response of device for the picosecond optical device setup of Fig. 15. The solid line is for the FET, $V_{GS} = -1.05$ V, $V_{DS} = +0.9$ V; the dashed line is for a 50-Ω coaxial cable transmission line in place of the FET. (From Smith *et al.*, 1981.)

The three groups of experiments just described demonstrate a successful merging of picosecond optics and electronics as a means of measuring high-speed phenomena. It is anticipated that the optical–electronic approach will likely lead to the development of new very-high-speed microelectronic devices.

IV. Moment–Balance Equations

Essentially all investigations of hot-electron transport in semiconductors are based on a one-electron transport equation, usually the Boltzmann transport equation (BTE). Indeed, the overriding theoretical concern in such high-field transport is primarily the solution of the transport equation to ascertain the form that the nonequilibrium distribution function takes in the presence of the electric field. However, for transport purposes, this is not an end product, since integrals must be carried out over the distribution function in order to evaluate the transport coefficients. In applications to semiconductor devices, however, the full solution of the BTE is usually too complicated to be determined at each spatial point within the device, and transport equations for relevant observables, such as energy and momentum, are preferred (Grubin *et al.*, 1982). Such transport equations are obtained by taking moments of the kinetic equation, and these often relate directly to the normal hydrodynamic semiconductor equations. In general,

the complicated nature of the kinetic equation precludes solving it analytically, and the existence of the various moment equations is based upon a number of assumptions, the most common of which is that the distribution function can be represented as a displaced Maxwellian.

In small semiconductor devices, the time scales are such that the use of the most common kinetic equation, the BTE, must be questioned (Barker and Ferry, 1980). Traditional semiclassical approaches, such as that of the BTE, assume that the response of the carriers to any applied force occurs simultaneously with the applied force, even though the system may undergo subsequent relaxation to a non-equilibrium steady state. On the short time scale of interest though, a truly causal theory introduces memory effects that lead to convolution integrals in the transport coefficients (Zwanzig, 1961), so that the resultant kinetic equation is not of the Markovian type. For the steady state, this results in a collision operator that depends upon the frequency of the driving field. Indeed, it is the non-Hermitian properties of the collision operator that have allowed Prigogine (1980) to postulate a formal transformation theory to couple dynamics and thermodynamics.

Previously, we have discussed the development of balance equations from a modified BTE (Ferry and Barker, 1980b), based upon a full quantum kinetic equation. Direct and general methods for the derivation of such quantum kinetic equations for the one-electron density matrix, for electrons in phonon and other arbitrary fields of force, have existed for some time (Kreuzer, 1981). Such treatments avoid the random-phase approximation and yield kinetic equations that indeed are non-Markovian. In addition, effects due to interference between the driving fields and the scatterers are obtained, leading to collisional shielding of the driving field, as well as to an intracollisional field effect, which is a disturbance of the scattering by the driving field. By the use of such a quantum kinetic equation for the distribution function, the moment equations can be obtained in a straightforward manner and are modified from their more normal BTE-derived counterparts. These equations show memory effects, as well as additional retardation due to the nonzero collision duration. However, the basis of these equations is still subject to rather stringent approximations. Indeed, a unique form must be assumed for the distribution function: it is specified as a retarded, displaced Maxwellian of rather special form. It remains unclear to what extent the form assumed for the distribution function affects the resulting moment equations and whether the actual form assumed is a proper ansatz.

The concerns over the detailed form of the moment equations can be removed by deriving these equations directly from the quantum transport equations. The exact solutions of the Liouville equations describe the time evolution of a statistical ensemble at any time interval after an external field

is applied. If the rate of intercarrier scattering is high, then after a short time interval τ_1, smaller than any appropriate time scale of interest, the evolution of the nonequilibrium density matrix must be independent of the initial distribution, and there should be a reduction in the number of parameters necessary to describe the nonequilibrium response of the system (Bogoluibov, 1946). It is therefore possible to assume a nonequilibrium statistical operator that is smoothed in its microfluctuations and from the very beginning describes the slow evolution of the system for time intervals that are larger than τ_1 (Zubarev, 1961). This approach is essentially similar to that of McLennan (1961) and to that suggested more recently by Procaccia *et al.* (1979). It has also been utilized by Kalashnikov (1970) directly for hot-electron problems in semiconductors. By utilizing such an approach, both the relevant moment equations and the form of the distribution function itself are obtained directly prior to the extension to the semiclassical transport properties. In this chapter, we review this approach. Both the nonequilibrium density matrix and the moment equations are developed.

When one introduces a retarded kinetic equation to describe the transport on a short time scale, this retarded equation is a significant deviation from the usual assumption of a simultaneous response to driving forces, and is a consequence of extending the concept of a causal response to the short time scale. Causal behavior is usually associated with ignoring a large percentage of the individual dynamic variables. Indeed, in the derivation of the quantum kinetic equation, the general density matrix is projected onto its "relevant", diagonal part, with the terms involved in the "irrelevant", off-diagonal part thus ignored. The density function can be transformed via a proper correspondence principle into a phase-space function (Moyal, 1949), and then a semiclassical limit can be taken. Such an approach is not totally required, however, since in obtaining a generalized kinetic equation, no truly quantum effects are included. Consequently, the Hamiltonian could as easily be a classical one, and the resulting equation would be the same (Zwanzig, 1960). Thus, the density function evolves in the $2N$ dimensional phase space (N particles) of position and momentum.

We want to consider a subspace of phase space, however, which is composed of what would be constants of the motion, if the system were conservative (which is not a requirement). The constants of the motion are $(2N-1)$-dimensional tori in phase space (i.e., for one dimension, phase space has two dimensions and the surface E = constant is a closed curve). Such parameters can be the average energy, the average momentum, etc. We therefore define a set of parameters P_m (Zwanzig, 1961)

$$P = \{P_m\} = \{E, \mathbf{P}, \ldots\}, \tag{6}$$

and to each of these corresponds a measurable average

$$Q_m \equiv \langle P_m \rangle. \tag{7}$$

For short-hand notation, we write $P = \{P_1, P_2, \ldots, P_N\}$. A parameterized distribution function can be defined as $g(P, t)$ which is the probability of transition from the initial values $g(P, 0)$ to the set of parameters P_m at time t. Specifically, the average in Eq. (7) is defined by the first moments as

$$Q_m = \int P_m g(P, t)\, dP. \tag{8}$$

The individual Q_m satisfy transport equations. We want to develop these equations themselves and shall do so for the quantum transport case, where achievement is easier. In part, we must first work from the equation of motion for $g(P, t)$ instead.

In particular, there are two situations in which the dynamic systems of interest will show the necessary property of weak instability and evolve into statistical behavior, which shows up on the macroscopic level as irreversibility (Prigogine, 1981). One is when the system is mixing, and the second is when there is no true invariant of the motion other than the energy. In these cases the dynamic system has undergone strong chaos, and individual trajectories no longer exist (Arnold and Avez, 1968). Our time scale τ_1 is taken to be the time for this weak instability to fully form, and we presume that this arises owing to carrier–carrier interactions. Then, for time $t > \tau_1$, correlations with lifetimes less than τ_1 can be ignored and the state of the system may be described by a reduced set of macroscopic observables $Q_m(t)$ which are the average values, over the nonequilibrium ensemble, of the set of dynamic variables P_m. These dynamic variables, and their conjugate forces $f_m(t)$, may then be used to define a nonequilibrium statistical ensemble. The approach is based upon Poincare's theorem on integral invariants generalized to quantum systems (Zubarev, 1974). Thus, the system density matrix should be constructable from the principal invariants of the system. Additionally, an auxiliary (local) equilibrium density matrix ρ_l is defined as a quasi-equilibrium quantity. This latter function serves as an idealized initial condition for the system after the randomization, characterized by τ_1, has occurred, and from which the system evolves under dynamic laws governed by its Hamiltonian. This means that at all times $t > \tau_1$, the nonequilibrium density matrix, which must satisfy the Liouville–Neumann equation, is a functional of ρ_l.

It now seems reasonable to use the nonequilibrium statistical operator, which is smoothed in the microfluctuations and, from the beginning time $t > \tau_1$, describes the slow, irreversible evolution of the system, to obtain the

balance equations necessary. We therefore consider the nonequilibrium distribution of conduction electrons, in a semiconductor in a strong electric field $\mathbf{F}(\mathbf{x})$, which are interacting with the lattice. The Hamiltonian of the system can be written as

$$H = H_e + H_F + H_L + H_{eL}, \tag{9}$$

where H_e and H_L are the Hamiltonians of the free carriers (in the effective mass approximation) and the lattice, H_{eL} describes the interaction between the carriers and the lattice, and

$$H_F = -\int \mathbf{F}(\mathbf{x}) \cdot d\mathbf{x} \tag{10}$$

is the potential due to the external field (we ignore the dynamic screening of this field).

Within a time τ_1, the energy and momentum that the carriers obtain from the field are redistributed among themselves due to carrier–carrier interactions. Thus, the distribution function describes electrons that are in equilibrium with each other, but not with the lattice, and, on a time scale greater than τ_1, the system may be characterized by the average values of the set of operators $P_m(\mathbf{x})$, defined by

$$P_m = \{H_e, \mathbf{P}_e, N, \mathbf{P}_L, H_L + H_{eL}\}, \tag{11}$$

where $\mathbf{P}_e$ and $\mathbf{P}_L$ are operators of total momentum of the electrons and the lattice, respectively, and N is the carrier number operator. The term H_e contains the carrier Hamiltonian, the coherent part of any interaction with the environment, and the carrier–carrier interaction, although we only explicitly retain the first of these in a renormalized effective mass treatment. However, it is assumed that this term is diagonal. Although we have included the terms for the lattice, it will be assumed that the lattice remains in equilibrium. The various P_m satisfy the equations of motion as

$$dP_m/dt = (i/\hbar)[H, P_m] = (i/\hbar)\hat{H} P_m. \tag{12}$$

where the caret over H signifies a superoperator. Then, for a homogeneous system (we may easily extend this to an inhomogeneous system),

$$\dot{H}_e = e\mathbf{F} \cdot \mathbf{P}_e / m + \dot{H}_{eL}, \tag{13a}$$

$$\dot{\mathbf{P}}_e = e\mathbf{F} - \dot{P}_{eL}, \tag{13b}$$

$$\dot{N} = 0, \tag{13c}$$

$$\dot{\mathbf{P}}_L = \dot{\mathbf{P}}_{eL}, \tag{13d}$$

$$\dot{H}_L - \dot{H}_{eL} = -\dot{H}_{eL}. \tag{13e}$$

where

$$\dot{P}_{m(L)} = (i/\hbar)\hat{H}_{eL}P_m. \tag{14}$$

We now introduce a set of time-dependent parameters $f_m(t)$ (which are c numbers) constructed to be thermodynamic conjugates of the Q_m as follows:

$$f_m(t) = \{-\beta_e, -\beta_e v_d, -\beta_e(\mu - mv_d^2), 0, -\beta\}, \tag{15}$$

where β_e is the inverse carrier temperature ($= 1/k_B T_e$), v_d the drift velocity, and μ the chemical potential (or Fermi energy), and each of the f_m is a homogeneous function of t. Here f_m for P_L is zero, since the lattice can have no average momentum, and $\beta = 1/k_B T$, where T is the lattice temperature.

The asymptotic, time-smoothed quantities $B_m(g)$ may be defined from the P_m and the f_m as ($t > 0$) (Zubarev, 1974)

$$B_m(t) = s\int_{-\infty}^{0} dt' e^{st'} f_m(t + t')P_m(t'), \tag{16}$$

where

$$P_m(t') = e^{-i\hat{H}t'/\hbar}P_m(0). \tag{17}$$

The quantities B_m satisfy the Liouville equation in the limits $s \to 0$ and are the quasi-invariant portions of the products $P_m f_m(t)$ with respect to the evolution under the Hamiltonian H. Thus, they are integrals of motion in this limit, and any total statistical operator constructed from these quantities will also be an integral of Liouville's equation. The operation of taking the quasi-invariant part is equivalent to taking the causal response, a treatment used to assure that retarded solutions are obtained (Gell-Mann and Goldberger, 1953). To assure that f_m and $Q_m = \langle P_m \rangle^t$ (the t superscript denotes a time-varying average, i.e., an ensemble average evaluated at t) are thermodynamically conjugate, we require

$$\langle P_m \rangle^t = \langle P_m \rangle_t^t, \tag{18}$$

where $\langle \cdot \rangle^t$ denotes the average taken over the nonequilibrium density operator $\rho(t)$ and $\langle \cdot \rangle_t^t$ indicates an average taken on the quasi-equilibrium density operator $\rho_t(t)$, both of which are defined shortly.

We can now write the nonequilibrium density operator in the form

$$\rho(t) = \exp\{-\Phi - \sum_m B_m(t)\}. \tag{19}$$

where

$$\Phi = \ln \mathrm{Tr} \exp\{-\sum_m B_m(t)\}. \tag{20}$$

The choice of Eq. (19) for $\rho(t)$ assures that this quantity reduces to the generalized Gibbsian grand ensemble (or grand canonical distribution) in the thermal equilibrium state and results in a positive entropy production in the equilibrium and the nonequilibrium states. Indeed, by defining the entropy operator as

$$S(t,0) = \Phi_0 + \sum_m P_m f_m(t), \tag{21}$$

with

$$\Phi_0 = \ln \mathrm{Tr} \exp\{-\sum_m P_m f_m(t)\}, \tag{22}$$

we can combine Eqs. (16) and (19) as (Kalashnikov, 1970)

$$\rho(t) = \exp\{-S(t,0) - \int_{-\infty}^{0} dt' e^{\varepsilon t'} \dot{S}(t+t',t')\}, \tag{23}$$

after an integration by parts and the extension of Eq. (21) to

$$\dot{S}(t+t',t') = \sum_m \{\dot{P}_m(t') f_m(t+t') + [P_m(t) - \langle P_m \rangle^t] \dot{f}_m(t+t')\}. \tag{24}$$

Clearly then, we may define the quasi-equilibrium density operator as that local operator for which $\dot{S} = 0$. Thus, we are led to

$$\rho_l(t) = \exp\{-S(t,0)\} \tag{25}$$

for the quasi-equilibrium operator. Here, the entropy $S(t,t')$ is a true two-time function since it represents that quantity in a nonstationary environment.

The total density matrix of the real system, $\rho(t)$, deviates from the quasi-equilibrium due to the microscopic relaxation processes in $\dot{S}$ that occur in a small, coarse-grained region of the system. Yet, the quasi-equilibrium operator $\rho_l(t)$ is determined from the actual $\rho(t)$ by the fact that the state parameters P_m involved in $\rho_l(t)$ are determined by the requirements of Eq. (18). The difference between $\rho(t)$ and $\rho_l(t)$ relaxes toward the latter due to the scattering processes and can be treated within a perturbation expansion, much as in the interaction picture.

In the expansion of $\dot{S}$ in Eq. (24), we can replace $\dot{P}_m$ by the set of Eqs. (13) and $\dot{f}_m$ from the expansion

$$\frac{df_m}{dt} = \mathbf{P}_c \cdot \nabla f_m / m + \sum_l \frac{\delta f_m}{\delta \langle P_l \rangle} \langle \dot{P}_l \rangle^t$$

$$= \mathbf{P}_c \cdot \nabla f_m / m + \frac{i}{\hbar} \sum_l \frac{\delta f_m}{\delta \langle P_l \rangle} \langle [H, P_l] \rangle^t. \tag{26}$$

where the δ indicates a functional derivative. In general, Eq. (13) tells us that the $\dot{P}_l$ can be expanded as

$$\dot{P}_l = \sum_v \alpha_{lv} P_v + \dot{P}_{l(L)}. \tag{27}$$

We further note that

$$\sum_{lv} f_l \alpha_{lv} \langle P_v \rangle = \frac{i}{\hbar} \left\langle \left[H_0, \sum_l f_l P_l \right] \right\rangle = \frac{i}{\hbar} \left\langle \left[H_0, \sum_l f_l P_l \right] \right\rangle_l = 0, \tag{28}$$

where

$$H_0 = H - H_{eL},$$

since $\sum_l \alpha f_l P_l$ commutes with the quasi-equilibrium density ρ_l. Differentiating Eq. (28) with respect of f_l yields

$$\sum \alpha_{lv} \langle P_v \rangle + \sum_{mv} f_m \alpha_{mv} \frac{\langle P_v \rangle}{\partial f_l} = 0. \tag{29}$$

Taking into account the thermodynamic relation (Zubarev, 1974)

$$\frac{\hat{c}\langle P_v \rangle}{\partial f_l} = -\frac{\partial^2 \ln(\Phi_0)}{\partial f_l \partial f_v} = \frac{\partial \langle P_l \rangle}{\partial f_v},$$

we can multiply Eq. (29) by $\partial f_i / \partial \langle P_l \rangle$ and sum over l. As a result we obtain

$$\sum_{lv} \frac{\partial f_i}{\partial \langle P_l \rangle} \alpha_{lv} \langle P_v \rangle + \sum_m f_m \alpha_{mi} = 0, \tag{30}$$

where we have used the relation

$$\sum_{lv} \frac{\partial f_i}{\partial \langle P_l \rangle} \frac{\hat{c}\langle P_v \rangle}{\hat{c} f_l} = \delta_{iv}. \tag{31}$$

By means of Eqs. (30) and (29), we can rewrite Eq. (26) as

$$\frac{df_m}{dt} = \dot{P}_e \cdot \nabla f_m / m - \sum_l \alpha_{lm} f_l + \sum_l \frac{\hat{c} f_m}{\hat{c}\langle P_l \rangle} \dot{P}_{l(L)}. \tag{32}$$

This equation can be regarded as a kinetic equation for the variables f_m and complements the set of kinetic equations for P_m given earlier.

It is obvious that terms of both first and second order in H_{eL} appear in Eq. (24). Therefore, to obtain the balance equations up to second order in the electron–phonon interaction (the first nonvanishing order), we use the following iterative expansion of the density matrix

$$\rho(t) = \left\{ 1 + \int_{-\infty}^{0} dt' e^{\varepsilon t'} \int_{t'}^{t} d\tau e^{-iS(\tau,0)} \hat{S}_e e^{iS(\tau,0)} \right\} \rho_l(t), \tag{33}$$

where S is given by Eq. (24). We obtain Eq. (33) in the following fashion. First, we note that from Eq. (23) we may write

$$\rho(t) = e^{-A-B}, \qquad \rho_I(t) = e^{-A},$$

with

$$A = S(t,0), \qquad B = \int_{\infty} dt'\, \dot{S}(t+t',t').$$

It is convenient to introduce an operator $\kappa(\tau)$ by means of

$$e^{-(A+B)\tau} \equiv \kappa(\tau)e^{-A\tau}, \tag{34}$$

which is equivalent to the operator equation

$$\kappa(\tau) = 1 - \int_0^{\tau} d\xi\, \kappa(\xi)e^{-A\xi}Be^{A\xi}. \tag{35}$$

From this relation, it is recognized that $\kappa(t)$ is the evolution operator in the interaction representation. Finally, with the change of variables $\tau \to 1$, $\xi \to t$, Eqs. (34) and (35) become Eq. (33). We can now use this to evaluate the averages and compute the balance equations.

To begin, we rewrite Eq. (24), using Eq. (27), is a more simple format:

$$\begin{aligned}\dot{S}(t+t',t') &= \sum_m \left\{ \dot{P}_{m(L)}(t')f_m(t+t') + [P_m(t') - \langle P_m\rangle^{t'}] \right. \\ &\qquad \left. \times \left[\frac{1}{m}\mathbf{P}_e \cdot \nabla f_m + \sum_n \frac{\delta f_m}{\delta\langle P_n\rangle}\dot{P}_{n(L)} \right] \right\} \\ &= \sum_m \{\dot{P}_{m(L)}(t')f_m(t+t') + \dot{S}(t,0)\}.\end{aligned} \tag{36}$$

so that

$$\begin{aligned}\int_{-\infty}^{0} dt'\, \dot{S}(t+t',t') &= \int_{-\infty}^{0} dt' \sum_m \dot{P}_{m(L)}f_m(t+t') \\ &\quad + \sum_m [P_m(0) - \langle P_m(0)\rangle^t] f_m(t)\end{aligned} \tag{37}$$

in the $s \to 0$ limit. Taking $f_m(t) = 0$ for $t < 0$ ($\beta_e - \beta = 0$ for $t < 0$), we can carry out an average over Eq. (13) as (after changing the integration variable)

$$\begin{aligned}\langle P_m\rangle &= \sum_n \gamma_{mn}\langle P_n\rangle + \int_0^{t} dt' \sum_n \langle \dot{P}_{m(L)}; \dot{P}_{n(L)}(t')\rangle^t f_n(t-t') \\ &\quad + \sum_n \gamma_{mn}\langle P_n; P_l(0)\rangle^t f_l(t).\end{aligned} \tag{38}$$

where the quantum correlation functions are defined by

$$\langle A; B(t')\rangle^t = \frac{1}{N}\int_0^t dt \langle A[e^{-tS}B(t')e^{tS} - \langle B\rangle_i^t]\rangle' n \tag{39}$$

in the absence of spatial correlations. Here $B(t')$ evolves under the total Hamiltonian according to Eq. (12). For $N\langle E\rangle = \langle H_e\rangle$, we find the energy balance equation to be

$$\begin{aligned}\frac{d\langle E\rangle}{dt} = \frac{e\mathbf{F}}{Nm}\cdot\langle \mathbf{P}_e\rangle &- \int_0^t dt' \langle \dot{H}_{e(L)}; \{\beta_e'(t-t')\dot{H}_{e(L)}(t') \\ &+ \beta_e(t-t')\mathbf{v}_d(t-t')\cdot\mathbf{P}_{e(L)}(t')\}_i^t\rangle' \\ &- \beta_e(t)\frac{e\mathbf{F}}{Nm}\cdot\langle \mathbf{P}_e; \{H_e(0) + \mathbf{v}_d(t)\cdot\mathbf{P}_e(0)\}_i^t\rangle'.\end{aligned} \tag{40}$$

where $\beta_e' = \beta_e - \beta$. The momentum balance equation becomes

$$\begin{aligned}m\frac{d\mathbf{v}_d}{dt} = \frac{1}{N}\langle P_e\rangle = e\mathbf{F} &- \int_0^t dt' \langle \mathbf{P}_{e(L)}; \beta_e'(t-t); \dot{H}_{e(L)}(t') \\ &+ \beta_e(t-t')\mathbf{v}_d(t-t')\cdot\mathbf{P}_{e(L)}(t')\}\rangle' \\ &- \beta_e(t)\frac{e\mathbf{F}}{N}\langle N; \{H_e(0) + \mathbf{v}_d(t)\cdot\mathbf{P}_e(0)\}_i^t\rangle'.\end{aligned} \tag{41}$$

Over a broad interval of the electric field strength, the kinetic energy associated with the electron drift velocity is much smaller than the mean energy of the carriers that is, $mv_d^2 \ll k_B T_e$. In this case, we may neglect the cross terms in the correlation functions such as $\dot{P}_{e(L)}\dot{H}_{e(L)}$ or $\dot{N}H_e$ just as in thermal equilibrium. Then, the correlation functions simplify and the balance equations become

$$\frac{d\langle E\rangle}{dt} = e\mathbf{F}\cdot\mathbf{v}_d - \int_0^t dt' \beta_e'(t-t')\langle \dot{H}_{e(L)}; \dot{H}_{e(L)}(t')\rangle', \tag{42}$$

and

$$m\frac{d\mathbf{v}_d}{dt} = e\mathbf{F} - m\int_0^t dt' \beta_e(t-t')\mathbf{v}_d(t-t')\langle \dot{P}_{e(L)}; \dot{P}_{e(L)}(t')\rangle'. \tag{43}$$

The remaining correlation functions are the collision relaxation memory functions (Zwanzig, 1964), and the form of these equations is similar to that obtained for the drifted Maxwellian approximation, although the details of the scattering functions are somewhat different. The second factor is

negligible for $t \to \infty$ but is significant during the initial transient phase and appears even in careful Monte Carlo calculations, even in the semiclassical limit. That this memory effect should appear in the energy equation follows from the fact that $E = \frac{3}{2}k_B T_e = \frac{3}{2}\beta_e^{-1}$ (when the drift energy can be neglected). Thus, the temperature is, in effect, a measure of the fluctuations and cannot be changed until the correlations of these fluctuations is broken by the field.

The correlation functions are generally written from Eqs. (40) and (41) as

$$J = \left\langle \dot{P}_{m(L)}; \sum_l f_l(t+t')P_{l(L)}(t') \right\rangle$$

$$= \frac{1}{N}\int_0^1 d\tau \left\langle P_{m(L)} \left\{ e^{-\tau S} \sum_l f_l(t+t')\dot{P}_{l(L)}(t')e\tau^S \right\} \right\rangle.$$

Now,

$$\sum_l e^{-\tau S} f_l(t+t')\dot{P}_{l(L)}(t')e^{\tau S} = \frac{i}{\hbar}\sum_l e^{-\tau S}[H_{eL}(t'), f_l(t+t')P_l]e^{\tau S}$$

$$= \frac{i}{\hbar}\frac{d}{d\tau}\{e^{-\tau S}H_{eL}(t')e^{\tau S}\},$$

and

$$J = \frac{i}{N\hbar}\left\langle \dot{P}_{m(L)} \int_0^1 d\tau \frac{d}{d}\{e^{-\tau S}H_{eL}(t')e^{\tau S}\}_j \right\rangle$$

$$= \frac{1}{N\hbar^2}\langle [H_{eL}(t'), [H_{eL}, P_m]] \rangle^t$$

$$= -\frac{1}{N\hbar^2}\mathrm{Tr}\{P_m[H_{eL}, [H_{eL}(t'), \rho(t)]]\}. \tag{44}$$

V. The Correlation Functions

In the last form in Eq. (44), the correlation function has the suggestive form of a decay function for the generalized parameter P_m. In particular, if we rewrite Eq. (44) as

$$J = -(1/N\hbar^2)\mathrm{Tr}\{P_m \hat{K}(t')\rho(t)\}. \tag{45}$$

then $\hat{K}$ can be recognized as a collision superoperator and is formally equivalent to that introduced by Van Hove (1955) and by Prigogine and Resibois (1961). In this regard, we can say that the evolution of $\hat{K}$ is governed by a projection into the space of correlations, and the consequent evolution of ρ lies in a subspace that includes a portion of this space of correlations. The role of $\hat{K}$ is thus related to the building up of correlations from asymptotic averages and sorts out the noise correlations that have a long-time effect and manifest themselves at the macroscopic level (George *et al.*, 1972). Now, as is clear from Eq. (44), $\hat{K}(t)$ is, in general, a function that involves only the interaction $H_{el.}$ except in the resolvent producing the time evolution. Since the resolvent also involves $H_{el.}$ from Eq. (17), we expect that J involves not only relaxation effects, but also self-energy corrections, particularly in high-density situations much as those found in picosecond laser pulse experiments. In addition, we expect that the presence of high dc electric fields will lead to interactions between the field and the scatterers—an intracollisional field effect. These effects are discussed later.

A. Self-Energy Corrections

In the presence of a strong monochromatic source, one, in general, would expect saturation due to state-filling also to be a factor. State-filling of the conduction and valence bands, as opposed to band-filling (Zitter, 1969), arises from monochromatic radiation and results in a δ-function-like spike (in energy) in the distribution function (Hearn *et al.*, 1962). The occurrence of such state-filling depends upon the generation rate and the length of time that the excitation occurs in a specific energy state. Normally, it is to be expected that such effects will occur under picosecond laser pulses, although one would expect the time width of the resulting saturated absorption to be of the same scale as the laser pulse. In Ge, it is found that the state-filling that would normally be expected to occur is restricted by the process of energy-gap narrowing, or band renormalization as it is sometimes called, induced by the electron–hole pairs (Ferry, 1978).

Energy-gap narrowing at high concentrations of electrons and holes arises primarily from two principal mechanisms: self-energy or exchange energy contributions to the bandgap and the free-carrier-induced shifts of the plasmon frequencies. The narrowing of the energy gap due to the interacting nature of the free carriers has been considered by Inkson (1976) using a dynamically screened potential. The introduction of free carriers alters the quasi-Fermi level so that the exchange integral, as well as its contribution to the exchange energy throughout the band, must change. He has estimated

that at low temperature the gap closure is given by

$$\Delta E_g = (2e^2/\pi\varepsilon_0)\{k_f + a[\tfrac{1}{2}\pi - \tan^{-1}(k_f/a)]\}, \tag{46}$$

where $a = \lambda/\varepsilon_0^{1/2}$ and λ is the screening wave vector. Now, Eq. (46) is a zero-temperature approximation and its validity at higher temperatures is restricted. However, k_f reflects primarily an estimate of the number of interacting carriers within the conduction band (or valence band). An estimate of this effect can be found by deriving k_f from the Fermi energy, given by

$$E_F - E_c = k_B T_e \ln(n/N_c), \tag{47}$$

where

$$N_c = \tfrac{1}{4}(2m_c k_B T_e/\pi\hbar^2)^{3/2} \tag{48}$$

is the effective density of states in the conduction band and T_e is the electron temperature.

The shift in the phonon frequencies, induced by the free carriers, also causes a gap narrowing. Brooks (1955) relates the energy-gap variation to the change in the lattice vibration frequencies from ω_i to ω_i' when the electron–hole pair is excited across any particular gap. The gap change arises from a change in chemical potential and can be expressed as

$$\Delta E_g = \sum_i [f(\omega_i', n) - f(\omega_i, 0)], \tag{49}$$

where $f(\omega, n)$ is the standard formula for the free energy of an oscillator of frequency ω, which itself is a function of n (or p). The variation of the energy gap at high densities (high temperatures) for Si was shown by Heine and Van Vechten (1976) to be dominated by the anharmonicity in the transverse-acoustic (TA) phonons. The TA modes depend critically upon the covalent nature of the bonding in the diamond structure. Without this bonding, the TA modes go unstable. The generation of electron–hole pairs removes bond charge, thus destabilizing and softening the TA mode. It is significant that estimates of the peak electron density in picosecond laser experiments at the damage threshold is within a factor of 5–8 of that necessary to completely destablize the TA mode. Heine and Van Vechten have shown that the density dependence of the anharmonic TA mode may be expressed as

$$\omega_{TA}' = \omega_{TA}[1 - f_{cv} n_{e0} \varepsilon_0^* 4(\varepsilon_0^* - \varepsilon_0) N_A], \tag{50}$$

where $f_{cv} = 0.85 \pm 1$ is the bond-charge shift per electron–hole pair, $\varepsilon_0^* = 24$ the dielectric constant of the competing β-tin phase, and N_A the atomic density.

As already mentioned, the strong energy-gap renormalization leads to a gap narrowing that is sufficiently strong so that state-filling probably does not occur. Indeed, it has been suggested by Wautelet and Van Vechten (1981) that under extreme conditions where $T_e \gg T$, anomalously large changes in the bandgap will occur. This will, in turn, lead to a gradient in T due to the large rate of phonon emission that is localized and to gradients in the electrochemical potentials of the excited electrons and holes. As the energy gap is a minimum at the site of the highest density, these gradients oppose the normal diffusion of the excited carriers, since the carriers are essentially self-trapped in a potential well.

B. The Intracollisional Field Effect

At high fields, such as will occur in very small devices, the collision duration is significant and correction terms appear to account for the nonzero time duration of each collision. If the instantaneous collision approximation that normally is used is relaxed, an additional field contribution appears as a differential superoperator term [see, e.g., the discussion in Ferry (1980)] in the collision integrals evaluated in the momentum representation, resulting in an intracollisional field effect (ICFE).

The ICFE can be partially understood by a simple example. In the usual case, the collision occurs instantaneously, so that the carrier enters the collision sphere at a point a and instantaneously exits at a second point b. However, the collision does not occur instantaneously, but requires a nonzero collision duration τ_c. In this case, it can now be accelerated by the field *during the collision*. Thus it exits not at b, but at b' some time $\Delta t = \tau_c$ later. As τ_c becomes comparable to τ, the mean time between collisions, this ICFE will have a significant effect on the transport dynamics, particularly in the transient-response region.

The mathematical details of the ICFE have previously been given by Barker (1980), so we shall not go into these details here. Rather, we shall give some of the supportive evidence for the observability of the effect. In very large fields, such as can occur in SiO_2 near breakdown, the ICFE can indeed be very significant (Ferry, 1979). Two major modifications of the scattering integral occur as a result of this intracollisional process. First, the total energy-conserving δ function is broadened by the presence of the electric field. Second, the threshold energy required for the emission of an optical phonon is modified, which causes a shift (in energy) of the δ function. This latter process is easily understood in physical terms. In the emission of an optical phonon, where the electron is scattered against the

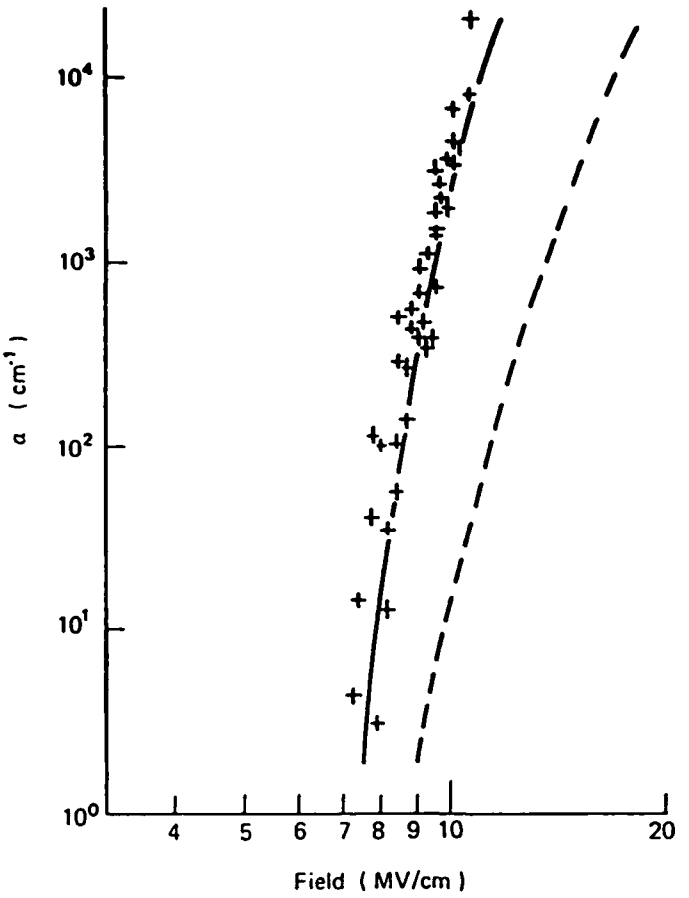


Fig. 17. Ionization rates in SiO_2. The curves are theoretical, and the solid curve includes effects due to the intracollisional field effect. The data are from Solomon and Klein (1975). (After Ferry, 1979.)

electric field, the field will absorb a portion of the electron's energy during the collision, and hence a reduction in energy loss to the lattice will be favored. The opposite effect, an enhancement in energy loss to the lattice occurs for emission along the electric field. These effects can be incorporated into the appropriate scattering integrals in the iterative technique, and this has been carried out. For electric fields above $5\text{–}6 \times 10^6$ V/cm, the broadening and shift of the scattering resonances produce a noticeable effect upon the velocity–field relationship and cause a reduction of the polar runaway field. This reduction in threshold can be observed in the impact ionization rates in SiO_2. In Fig. 17, we show these rates as measured by Solomon and Klein (1975) and as calculated earlier (Ferry, 1979). It is necessary to include effects arising from the ICFE to adequately fit theory to experiment.

References

Arnold, V. I., and Avez, A. (1968). "Ergodic Problems of Classical Mechanics." Benjamin, New York.

Auston, D. H., Johnson, A. M., Smith, P. R., and Bean, J. C. (1980). *Appl. Phys. Lett.* **37**, 371.

Barker, J. R. (1980). *In* "Physics of Nonlinear Transport in Semiconductors" (D. K. Ferry, J. R. Barker, and C. Jacoboni, eds.), pp. 126–152. Plenum, New York.

Barker, J. R., and Ferry, D. K. (1980). *Solid-State Electronic.* **23**, 519–530.

Bogoliubov, N. N. (1946). "Problem Dinam. Teorii. u. Stat. Fiz." Moscow, USSR.

Broers, A. N., Harper, J. M. E., and Molzen, W. W. (1978). *Appl. Phys. Lett.* **33**, 329.

Brooks, H. (1955). *Adv. Electron. Electron Phys.* **7**, 85.

Doob, J. L. (1942). *Ann. Math.* **43**, 785.

Elliott, M. T., Splinter, M. R., Tones, A. B., and Reekstin, J. P. (1979). *IEEE Trans. Electron Devices* **ED-26**, 469.

Ferry, D. K. (1978) *Phys. Rev. B* **18**, 7033.

Ferry, D. K. (1979). *J. Appl. Phys.* **50**, 1422.

Ferry, D. K. (1980). *In* "Physics of Nonlinear Transport in Semiconductors" (D. K. Ferry, J. R. Barker, and C. Jacoboni, eds.), pp. 577–588. Plenum, New York.

Ferry, D. K., and Barker, J. R. (1980). *Solid-State Electron.* **23**, 545–549.

Ferry, D. K., and Barker, J. R. (1980). *J. Phys. Chem. Solids* **41**, 1083–1087.

Ferry, D. K., and Barker, J. R. (1981). *J. Appl. Phys.* **52**, 818–824.

Franz, W. (1958). *Z. Naturforsch* **13**, 484.

Gell-Mann, M., and Goldberger, M. L. (1953). *Phys. Rev.* **91**, 398.

George, C., Prigogine, I., and Rosenfeld, L. (1972). *Nature* **240**, 25.

Götze, W., and Wölfle, P. (1972). *Phys. Rev. B* **6**, 1228.

Grosvalet, J., Motsch, C., and Tribes, R. (1963). *Solid-State Electron.* **6**, 65.

Grubin, H. L., and Ferry, D. K. (1981). *J. Vac. Sci. Tech.* **19**, 540–544.

Grubin, H. L., Ferry, D. K., Iafrate, G. J., and Barker, J. R. (1982). *In* "VLSI Electronics Microstructure Science" (N. G. Einspruch, ed.), pp. 198–301. Academic Press, New York.

Grubin, H. L., Kreskovsky, J. P., Iafrate, G. J., and Ferry, D. K. (1984) *In* "Proceedings of the Workshop on The Physics of Submicron Structures." Plenum, New York.

Hearn, C. J., Landsberg, P. T., and Beattie, A. R. (1962). *In Proc. 6th Int. Conf. Phys. Semicond.* Institute of Physics, London, United Kingdom, p. 857.

Heine, V., and Van Vechten, J. A. (1976). *Phys. Rev. B* **13**, 1622.

Hunter, W., Holloway, T. C., Chatterjee, P. K., and Tasch, A. F., Jr. (1980) *IEEE Electron Dev. Lett.* EDL-2, 4.

Inkson, J. C. (1976) *J. Phys. C* **9**, 117.

Kalashnikov, V. P. (1970) *Physica* **48**, 93.

Keldysh, L. V. (1958) *Sov. Phys. JETP (Engl. Transl.)* **7**, 788.

Kreuzer, H. J. (1981) "Nonequilibrium Thermodynamics and Its Statistical Foundations" Clarendon, Oxford.

Kubo, R. (1957) *J. Phys. Soc. Jpn.* **12**, 570.

Kubo, R. (1974) *In* "Lecture Notes in Physics 31: Transport Phenomena" (G. Kirczenow and J. Marro, eds.), pp. 75–125. Springer-Verlag, Berlin and New York.

Laval, S., Bru, C., Arnodo, C., and Castagne, R. (1980) *Proc. 1980 IEDM*, p. 52.

Lugli, P., Zimmermann, J., and Ferry, D. K. (1981). *J. Phys. Colloq. (Orsay, Fr.)* **42**(Suppl. 10), 103–110.

Maloney, T. J., and Frey, J. (1977). *J. Appl. Phys.* **48**, 781.

McLennan, J. A., Jr. (1961) *Phys. Fluids* **4**, 1319.

Moyal, J. E. (1949). *Proc. Cambridge Philos. Soc.* **45**, 99.
Prigogine, I. (1980). "From Being to Becoming." Freeman, San Francisco.
Prigogine, I. (1981). *In* "Order and Fluctuations in Equilibrium and Nonequilibrium Statistical Mechanics; XVIIth International Solvay Conference on Physics" (G. Nicolis, G. Dewel, and J. S. Turner, eds.), pp. 35–76. Wiley, New York.
Prigogine, I., and Resibois, P. (1961). *Physica* **27**, 629.
Procaccia, I., Ronis, D., Collins, M. A., Ross, J., and Oppenheim, I. (1979). *Phys. Rev. A* **19**, 1290.
Ruch, J. G. (1972). *IEEE Trans. Electron Devices* **ED-19**, 652.
Shank, C. V., Fork, R. L., Greene, B. I., Reinhart, F. K., and Logan, R. A. (1981). *Appl. Phys. Lett.* **38**, 104.
Smith, P. R., Auston, D. H., and Augustyniak, W. M. (1981). *Appl. Phys. Lett.* **39**, 739.
Solomon, P., and Klein, N. (1975). *Solid State Commun.* **17**, 1317.
Ting, C. S., Ying, S. C., and Quinn, J. J. (1976). *Phys. Rev. Lett.* **37**, 215.
Trofimenkoff, F. N. (1965). *Proc. IEEE* **53**, 1765.
Van Hove, L. (1955). *Physica* **21**, 517.
Wautelet, M., and Van Vechten, J. A. (1981). *Phys. Rev. B* **23**, 5551.
Zimmermann, J., Lugli, P., and Ferry, D. K. (1981). *J. Phys. Colloq. (Orsay, Fr.)* **42**(Suppl. 10), 95–101.
Zitter, R. N. (1969). *Appl. Phys. Lett.* **14**, 73.
Zubarev, D. N. (1961). *Dokl. Akad. Nauk, SSSR* **40**, 92. [(1972). *Sov. Phys. Dokl. (Engl. Transl.)* **6**, 776].
Zubarev, D. N. (1974). "Nonequilibrium Statistical Thermodynamics." Consultant's Bureau, New York.
Zwanzig, R. (1960). *J. Chem. Phys.* **33**, 1338.
Zwanzig, R. (1961). *Phys. Rev.* **124**, 983.
Zwanzig, R. (1964). *J. Chem. Phys.* **40**, 2527.

Structural dependent electrical characteristics of submicron gallium arsenide devices

H. L. Grubin and J. P. Kreskovsky

Scientific Research Associates, Inc., Glastonbury, Connecticut 06033

(Received 31 January 1984; accepted 9 May 1984)

Numerical studies of the transient and dc electrical behavior of submicron $N^+N^-N^+$ gallium arsenide structures are discussed. It is shown that the transient results are dominated, during the first fraction of a picosecond, by displacement current contributions. Velocity overshoot is less important. Under dc conditions and high bias levels, submicron effects may be masked by transport within the N^+ regions.

PACS numbers: 85.30.De, 72.80.Ey

I. INTRODUCTION

Recent activities in the area of submicron transport in compound semiconductor structures have tended to focus on the two and three terminal $N^+N^-N^+$ diode.[1-3] The diode has four interfaces along the principal direction of transport—metal/N^+, N^+N^-, N^-N^+, and the downstream N^+/metal interface. In addition, if gate contacts are included, there are the transverse interfaces. It has been argued that the structure is capable of supporting ballistic transport,[4] space charge limited transport,[5] reflections at the downstream N^-N^+ interface, etc. Its dc electrical properties have been measured[6] and time dependent electrical measurements are now possible.[7] Each of these experimental studies has attempted to determine whether nonequilibrium transient transport or velocity overshoot is present in the $N^+N^-N^+$ structure. There has, however, been uncertainty in the interpretation of the experiments, primarily because the presence of injected space charge can also yield current levels in excess of the steady state uniform field values. Thus, when experiment yields higher than expected current levels, are they due to velocity overshoot, space charge injection, or some combination thereof? These questions are addressed in this paper through studies of the dc and transient electrical characteristics of the $N^-N^-N^+$ structure.

The studies involve numerical solutions to the first three space and time dependent moments of the Boltzmann transport equation for the semiconductor gallium arsenide. Each of the following five questions has been examined: (1) How are the electrical properties of the $N^+N^-N^+$ structure affected by the length of the N^- region? (2) Will transient electrical measurements reflect velocity overshoot, or are the measurements a reflection of displacement current contributions, only to be followed by a weaker overshoot contribution? (3) How important is the shape of the N^+N^- interface, *vis a vis* the electrical properties of the structure? (4) What are the consequences of either deliberate or accidental variations in the doping profile? (5) To what extent will the ratio of the N^+/N^- doping level affect the electrical characteristics of the structure? Each of the five questions is addressed in the discussion below, although particular attention is given to the first two topics.

Most of the calculations performed for this paper are for the one dimensional structure of Fig. 1, in which the N^- region is assigned a nominal doping level of $10^{15}/cm^3$ and the N^+ region is at 10^{17} cm^3. There are conceptual problems with this jellium model, particularly on the submicron scale,[8] but they are ignored here. The length of the N^- region is specified at the doping level of $10^{16}/cm^3$ and varied from 0.416 to 0.116 μ. The entire structure is fixed at a length of 1.0 μ. The design of the structure dictates that nonuniform fields and charge densities form within it. Thus, the relevant quantity that relates to experiment is current density $J(t) = -eN(x,t)V(x,t)$, rather than velocity. It is here that one of the more significant aspects of the study appears. A

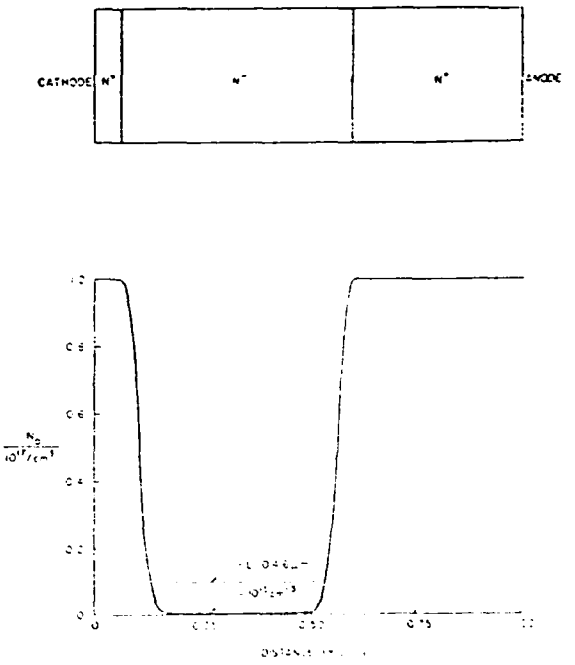


FIG. 1. (a) Sketch of the $N^+N^-N^+$ structure used for the simulations. (b) Profile of the doping level used in the simulations for L_{N^-} = 0.416 μ.

summary of the results follows and is supported by the detailed description of the calculations.

II. BRIEF SUMMARY OF RESULTS

For the time dependent calculations, the current transient shows an ostensible similarity to that commonly regarded as the signature of velocity overshoot. For the latter the mean carrier velocity shows a rapid rise to a peak value in excess of the steady state value. After a time duration of a fraction of a picosecond, the velocity descends toward its equilibrium steady state value. This decrease is associated with differences between the energy and momentum scattering rates and is enhanced through electron transfer. For the $N^+N^-N^+$ structure, there is a similar transient associated with the current density where the peak current is reached within a fraction of a picosecond. Here, however, the calculations show that initial transient is associated primarily with the rearrangement of the *self-consistently calculated* internal electric field. *The initial current transient is thus primarily due to displacement current effects.* Transient overshoot contributions while present, become effective at later times and are significantly below their uniform field predictions.

Steady state, time independent solutions are also obtained. These are the "long-time" solutions following the transient. In determining the time constants involved in the approach toward steady state, it is noted that momentum relaxation often occurs within a picosecond time frame, while energy relaxation takes somewhat longer. For the calculations with $L_{N^-} = 0.416\mu$, steady state in the presence of nonuniform fields occurs after a time-duration of approximately 13 ps. The steady state solutions show the excess carrier velocities discussed by the present authors[1] and others.[2–4] Of importance here are the steady state current voltage characteristics which display the anticipated nonlinear behavior associated with space charge injection and field dependent mobilities.

Because of the intuitive relation between the space charge injection properties of the submicron $N^+N^-N^+$ structure and those associated with Child's law, a power law $J \alpha \psi^\gamma$ was extracted. For the $N^+N^-N^+$ structures the calculations revealed a low voltage power law where $\gamma > 1.5$, as compared to the Child's law relation where $\gamma = 1.5$. For higher bias levels, the power law was considerably below 1.5 and is a consequence of the field dependent mobility of the gamma valley in gallium arsenide and the presence of electron transfer. The extent to which a reduction in the rate of charge injection into the N^- region is contributing to the decreased power law was not determined.

It is important to note that the detailed transient reported in this study is for the $N^+N^-N^+$ structure. It would be incorrect to assume that the transient would be similar to, e.g., a uniformly doped structure with nonuniformities resulting from boundary conditions that reflected, e.g., the presence of the metal contact. Calculations for the latter structure have been completed[9] and while the initial transient reflects both displacement and conduction current contributions, the details of the time transients are dependent in [illegible] on the boundary conditions.

Another important feature concerns the presence of negative differential conductance. It is known from uniform field studies that the long-time asymptotic solutions to the governing equations exhibit the presence of negative differential mobility in the field dependent velocity. For the calculations reported below, in which current density rather than velocity is sought and where the fields are nonuniformly distributed, *there was no evidence of negative differential conductivity.*

III. THE NUMERICAL SIMULATIONS

The nonuniform field calculations were performed for gallium arsenide subject to two-level transfer with the parameters listed in Table I. The restriction to two-level transfer is computational. For one-dimensional problems, seven partial differential equations including Poisson's equations are solved. For three-level transfer, three additional equations are required for one-dimensional constraints. Further, for the range of electric field values considered in this paper, transfer to the X valley is relatively insignificant.

For the parameters of Table I, the transient and steady state calculations were performed for six different values of bias; from 0.25 to 1.5 V, in steps of 0.25 V. The scattering events considered in the calculation are listed in Table I. Ionized impurity scattering was ignored, as it would be insignificant within the N^- region. Its contributions would, however, be of importance within the heavily doped surrounding regimes. Thus, the current levels achieved in this calculation will tend to be somewhat optimistic.

The equations used in the study have been discussed earlier and are

$$\frac{\partial n_i}{\partial t} + \nabla \cdot (\mathbf{v}_i n_i) = \left(\frac{\partial n_i}{\partial t}\right)_{\text{coll}}, \tag{1}$$

$$\frac{\partial \mathbf{P}_i}{\partial t} + \nabla \cdot (\mathbf{v}_i \mathbf{P}_i) = -en_i\mathbf{F} - \text{grad}\, n_i kT_i - \text{div}\, \sigma + \left(\frac{\partial P_i}{\partial t}\right)_{\text{coll}}, \tag{2}$$

$$\frac{\partial}{\partial t} W_i + \nabla \cdot \mathbf{v}_i W_i = -en_i\mathbf{F} \cdot \mathbf{v}_i - \nabla \cdot v_i n_i kT_i - \text{div}\, \sigma \times \mathbf{v}_i - \text{div}\, \mathbf{Q}_i + \left(\frac{\partial W_i}{\partial t}\right)_{\text{coll}} \tag{3}$$

for particle, momentum, and energy balance, respectively. In Eqs. (2) and (3), $P_i = n_i m_i v_i$, $W_i = n_i(1/2 m_i v_i^2 + 3/2 kT_i)$, σ is a stress tensor and Q is of the form to represent a flow of "heat." Strictly, both of these terms arise from the nonspherical nature of the distribution function.[10] They are treated phenomenologically through equivalent macroscopic constitutive relations

$$\sigma_i^{xx} = \eta \frac{\partial v^i}{\partial x}, \tag{4}$$

$$\mathbf{Q}_i = -\kappa\, \text{grad}\, T_i. \tag{5}$$

Justification for the use of Eqs. (4) and (5) is considered elsewhere. It is noted that η is dimensionally a viscous coefficient and κ is dimensionally a thermal conductivity coefficient. The values used [illegible]

TABLE I. Parameters and boundary conditions used in calculation.

Parameters	Γ	L	Common
Number of equivalent valleys	1	4	
Effective mass (m_e)	0.067	0.222	
Γ–L separation (ev)			0.33
Polar Optical Scattering			
Static dielectric constant			12.90
High frequency dielectric constant			10.92
LO phonon (ev)			0.0354
Γ–L Scattering			
Coupling constant (ev/cm)			0.800×10^{9}
Phonon energy (ev)			0.0278
L–L Scattering			
Coupling constant (ev/cm)		2.0×10^{9}	
Phonon energy (ev)		0.0354	
Acoustic Scattering			
Deformation potential (ev)	7.0	9.2	
Nonpolar Scattering (L)			
Coupling constant (ev/cm)		0.300×10^{9}	
Phonon energy (ev)		0.0343	
Miscellaneous			
Viscosity coefficient (Γ)		1.15×10^{-9} g/cm s	
Thermal condition coefficient (Γ)		4.14×10^{-5} J/cm s K	

Boundary Conditions	
$x=0$	$x=L$
$N_\Gamma = 10^{17}/\text{cm}^3$	Second derivatives of all quantities are zero except the potential, which is specified.
N_Γ = equilibrium value	
$\frac{\partial^2 v_\Gamma}{\partial x^2} = v_L = \psi = 0$	
$T_\Gamma = T_L = 300$ K	

In addition, Poisson's equation is solved

$$\nabla^2\psi = \frac{e}{\epsilon}(n_\Gamma - n_L - n_0), \tag{6}$$

where n_Γ, n_L, and n_0 denote the population of the Γ, L, and background levels. With the incorporation of Eqs. (4) and (5), the momentum and energy balance equations become second order in velocity and temperature. Solutions to these equations for the set of boundary conditions listed in Table I provide the total current through the device which is written as

$$J(t) = -e(n_\Gamma v_\Gamma + n_L v_L) + \epsilon\frac{\partial F}{\partial t}. \tag{7}$$

An external circuit is not included in these calculations.

The first set of results is shown in Fig. 2. Figure 2(a) displays the total current flowing through the device following application of a voltage pulse of magnitude 1.0 V. The calculation is performed in two stages. The first involves obtaining a steady state solution at 0.01 V. For the second, using this as an initial condition the bias is raised in one time step, to 1.0 V. Application of the bias in one time step replicates the procedure of most of the uniform field calculations.

As seen in Fig. 2, the current displays an initial peak at approximately 0.15 ps, followed by a drop in current and a subsequent rise toward a steady state value. For uniform field calculations in which the voltage is increased in one time step, there is an initial displacement current whose magnitude is determined entirely by the computational time step. Thereafter, all displacement currents are zero and all transients are particle current transients. (It is noted that with a load line, displacement currents would exist.) The situation with the nonuniform field calculation and displacement current contributions is different. Figure 2(b) displays the total current through the device [it is a repeat of Fig. 2(a)]

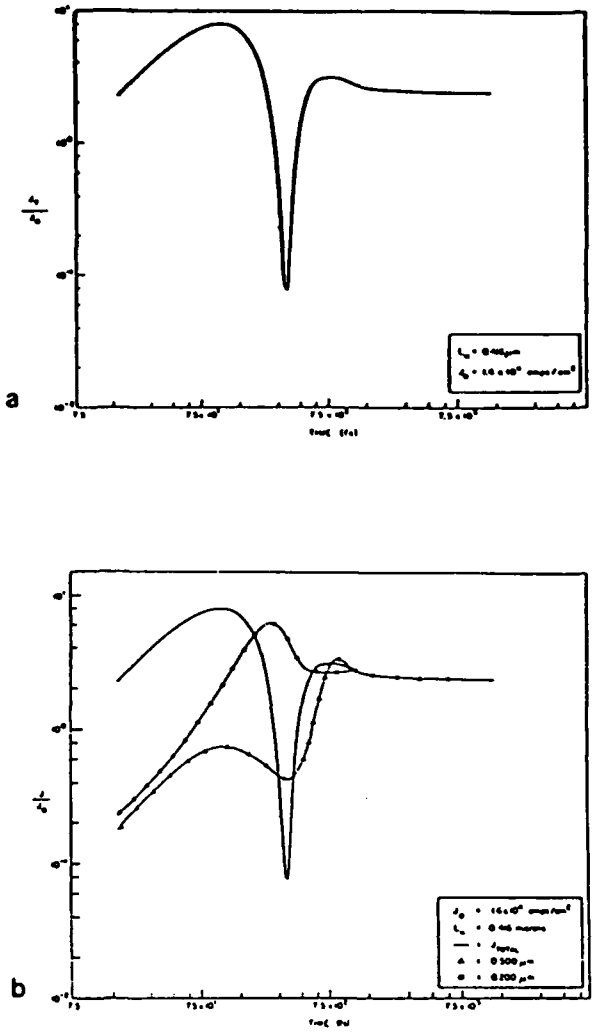


FIG. 2. (a) Total current vs time following application of a bias of 1.0 V to the $N^+N^-N^+$ structure with $L_{N^-} = 0.416\mu$. (b) Total current vs time following application of a bias of 1.0 V to the $N^+N^-N^+$ structure with $L_{N^-} = 0.416\mu$ (—). Particle current at position 0.200μ(-●-). Particle current at position 0.500μ (-▲-). $J_R = 1.6\times10^4$ A/cm^2.

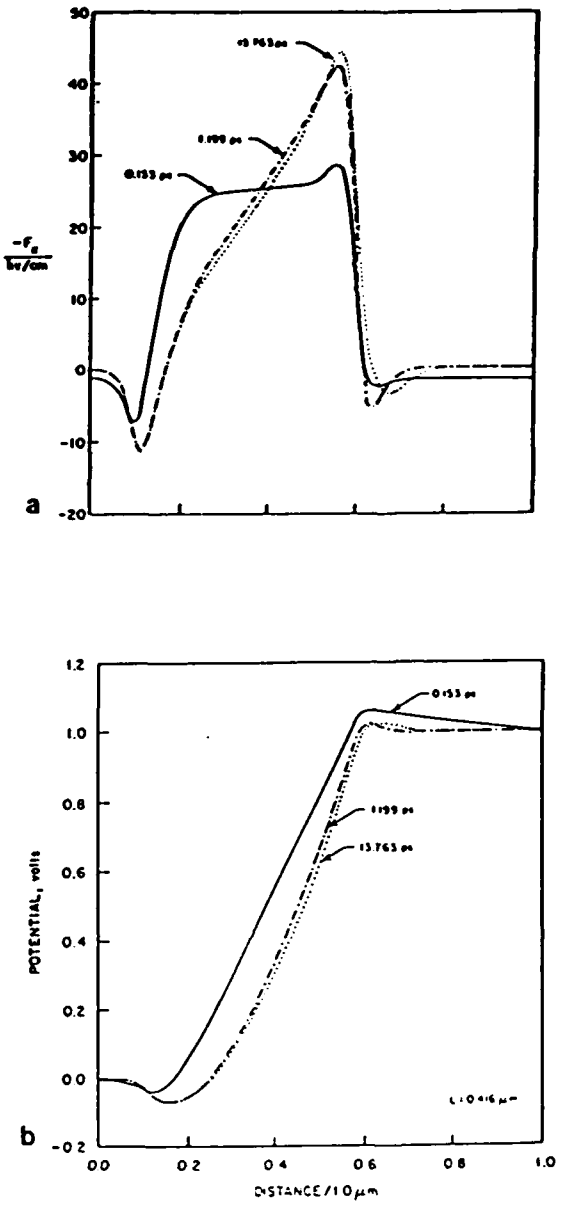


FIG. 3. (a) Electric field and (b) potential vs distance profiles at different instants in time. Calculation is for the $N^+N^-N^+$ structure with $L_{N^-} = 0.416\mu$, $\phi = 1.0$ V.

and the particle current at two different locations within the structure—0.200 and 0.500μ. *The magnitudes of the particle current indicate that the initial transient is strongly influenced by displacement current contributions and that it would be inappropriate to assume that the initial current transient is a measure of velocity overshoot.*

The details of the transient, specifically as it relates to displacement current contributions, are reflected in the time dependence of the electric field and potential profiles (Fig. 3) and the spatially dependent charge density profiles (Fig. 4). It is noted, however, that as in the uniform field calculations, immediately following the voltage step the electric field increases everywhere by the ratio of the applied bias to the length of the structure, in this case 10 kV/cm. This initial increase introduces a displacement current whose magnitude does not correctly represent the physical transient, but rather the impulsive change in the applied potential over a single, small but finite time step. Physically accurate calculations follow the initial time step and are discussed below.

Prior to the application of the step potential a retarding field is formed at the upstream N^+N^- interface limiting further injection of space charge into the N^- region. This retarding field, which at its maximum is positive in sign, is significantly reduced following application of the step potential; carrier injection into the N^- region is thereby resumed. Two events accompany this enhanced injection. *First*, to accommodate the increased charge within the N^- region

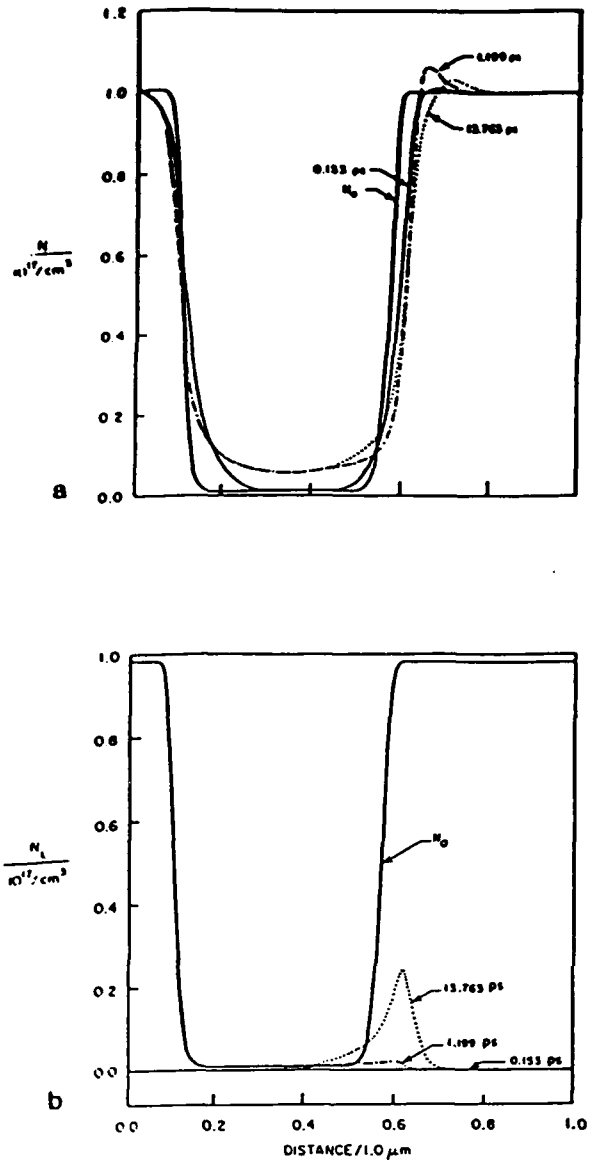


FIG. 4. (a) The same as in Fig. 3, but for the total carrier density and (b) satellite carrier density. N_0 denotes the background doping profile.

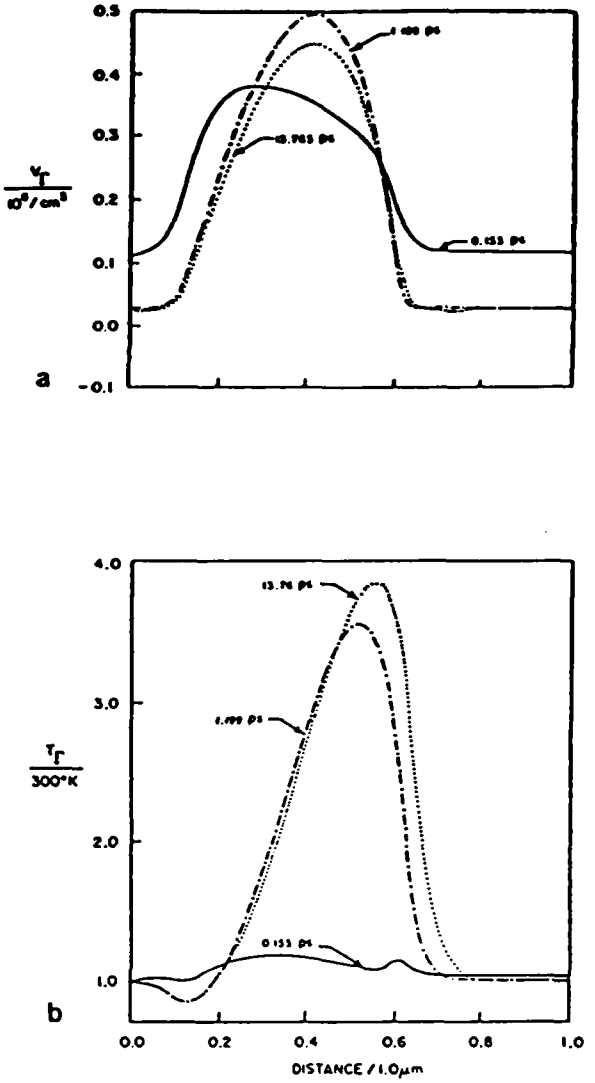


FIG. 5. (a) The same as in Fig. 3, but for the velocity and (b) temperature carriers in the gamma valley.

Gauss's law dictates that the electric field within the region must become increasingly negative. Second, the space charge injection is self-limiting in that as the process of injection proceeds, the retarding field begins to reform and positive field values result at the upstream N^+N^- interface. The process of an increased and positive retarding field accompanied by an increasingly negative field within the N^- region and the constraint of the fixed potential leads to the spatially dependent displacement currents inferred from Fig. 2(b). It is noted that in a steady state a significant amount of injected charge resides at the downstream N^-N^+ interface and results in the presence of a downstream retarding field.

Notwithstanding the displacement current contributions, [illegible] determine the extent to which the carrier velocity can exceed the equilibrium values. For the structure considered herein with $L_{N^-} = 0.416\,\mu$, most of the current is transported by the gamma valley carriers. For this case, the mean carrier velocity thereby exceeds the steady state value. This is seen in Fig. 5, which also includes a plot of the gamma valley temperature. The steady state distribution is qualitatively similar to that reported in a number of different studies.[1-3] In particular, the presence of a local cooling at the N^+N^- interface is noted.[11]

There are several noteworthy features associated with Fig. 5. *First*, there is the progressive movement of the velocity layer toward the downstream N^-N^+ interface. This migration is associated with the spatial and temporal derivatives on the left-hand side of Eq. 2. *Second*, there is a progressive

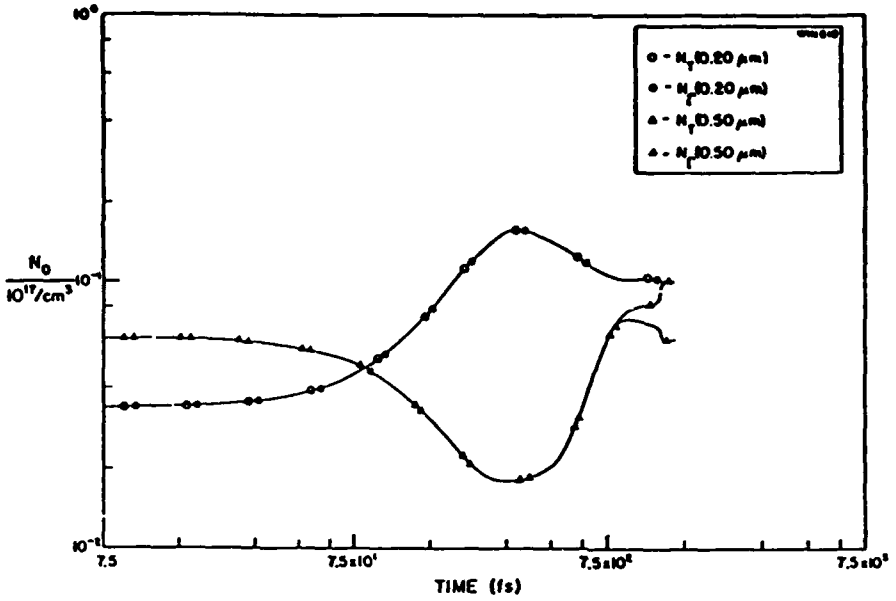


FIG. 6. Time dependent total and gamma valley carrier density at two points, $x = 0.200$ and 0.500 μm within $N^+N^-N^+$ structure with $L_N = 0.416$ μs. $\phi = 1.0$ V.

decrease in the velocity in the N^+ regions as dictated by the decreasing field within these regions. Indeed, the possibility exists for the carriers to sustain a transient separation at the N^+N^- interface, with carriers on the upstream portion of the interface moving toward the cathode and carriers at the downstream portion of the interface moving toward the downstream boundary. This separation is accompanied by compensating displacement current contributions.

The results clearly indicate the presence of velocity overshoot under nonuniform field conditions. Under uniform field conditions, the transient following the peak velocity is dominated by electron transfer. The question of interest thus becomes "Does similar phenomena occur when nonuniform fields are present?"

Figure 6 is a plot of the time evolution of the total carrier density and the gamma valley carrier density at two points within the $N^+N^-N^+$ structure. It is seen that at the upstream portion of the structure very little transfer occurs. Most of the transfer is at the downstream portion of the structure. One necessary conclusion of this study is that the *uniform field calculations bear little resemblance to the transients occurring in the $N^+N^-N^+$ structure.*

The situation as described above depicts the transient solution in detail for the $N^+N^-N^+$ structure of Fig. 1. Of additional interest are transients for structures with smaller N^- regions and for uniformly doped materials with metal contacts or heterostructures abutting the active region. The former two studies have been completed and will be reported at a later time. Of equal significance and discussed next are the dc current–voltage relations and their dependence on the length of the N^- region.

It is expected that the shorter the active region the higher the dc drive current.[12] Most discussions associated with this increased drive current are based on the fact that electron transfer effects are minimized. What is generally ignored is [illegible] arises from the excess charge injected into the N^- region. This point was also made in Ref. 3 where the dependence of current and voltage on N^- region length was also examined. A second point of importance concerns determining which portion of the structure dominates its transport. It may be intuitively expected that for the structure considered it is the N^- region that dominates. This appears to be the case for the above discussion. But one may expect that *for a sufficiently small N^- region, no single region dominates.* In the calculations reported here, the absence of a single dominating region becomes apparent at higher voltage levels and for the case when $L_N = 0.116\mu$. Here, variations in the total charge density tend to screen variations in the doping profile of the structure, and the potential drop across the N^- region is small enough to allow a substantial drop across the downstream N^+ regions. Electron transfer occurs *away* from the N^- region. This, of course, is not unexpected. It is implicit in the design of Gunn oscillators with doping variations assigned the task of domain nucleation sites. The current–voltage characteristics are, therefore, expected to reflect a complex set of electrical phenomena. These are displayed in Fig. 7.

Figure 7 displays a series of current versus voltage curves for $N^+N^-N^+$ structures with the indicated N^- region length. Each curve displays J/J_{ref} vs ψ/ψ_{ref}. J_{ref} is the computed value of current at $\psi_{ref} = 0.25$ V. ψ_{ref} is indicated in the figure caption. As discussed above it is noted that J_{ref} increases as the N^- region decreases in length. At low bias levels, the current–voltage relation appears to follow a power relation that is slightly less than $J/J_{ref} = (\psi/\psi_{ref})^{\gamma}$ with $\gamma = 1.8$. At higher values of bias there is considerable sublinearity in the current–voltage relation due in part to electron transfer to the satellite valleys. Electron transfer is also occurring within the N^- region of the device. Thus the curves while straightforward in appearance, mask [illegible] electron [illegible]

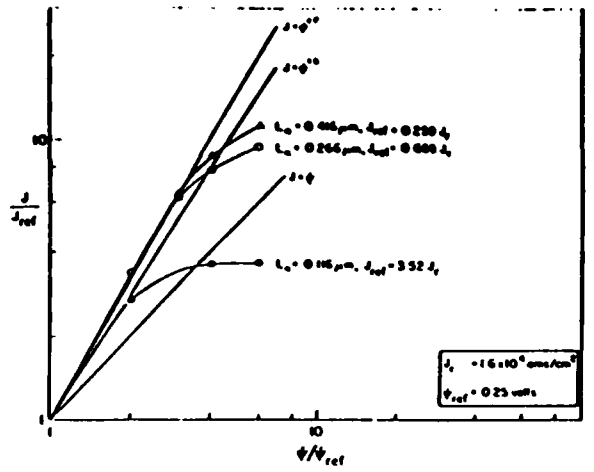


FIG. 7. Steady state current vs voltage for $N^+N^-N^+$ structures of indicated length. $J_r = 1.6 \times 10^4$ A/cm^2, $\phi_r = 1.0$ V.

As indicated above, a considerable amount of electron transfer occurs in the downstream portion of the N^+ region when the N^- region is decreased in length. Indeed, the detailed calculations indicate that the relative amount of *electron transfer increases as the N^- region decreases in length.* At first glance, this result appears to contradict all that has been discussed about transport in submicron devices. But it is not unusual when it is realized that as the N^- region decreases in length a greater fraction of the voltage drop falls across the N^+ regions of the device. It is this latter feature that is responsible for the enhanced transfer. To place this in different terms, the active region length of the device *increases* as the N^- region becomes insignificantly small.

One question of immediate interest is, therefore, how does one retain submicron features while decreasing the length of the N^- region? Heterostructure confinement offers one possibility. A second, which we have not yet explored involves increasing the magnitude of the N^+ region doping, thereby reducing the potential drop across that region. It must also be noted that impurity scattering has been ignored here. This type of scattering is likely to alter the low field results. At high values of electric field impurity scattering is less effective and the results obtained above are likely to be qualitatively correct. Inclusion of the effects of ionized impurity scattering is currently in progress.

A final question of interest concerns the sensitivity of the results to structural changes in the doping profile. A set of calculations were performed for the $N^+N^-N^+$ structure in which the N^- region decreased linearly from 10^{17}/cm^3 to 10^{15}/cm^3 over a distance of 0.271 μ. The initial transient contained significant displacement current contributions and provides further evidence for its importance. In another case, the structure of the N^- region was altered by incorporating a local region of higher donor density within the N^- region. The donor density was increased to 10^{16}/cm^3 over a distance of 0.03 μ and resulted in a current increase of 15% over the unaltered region calculation. A reduction in donor density over the same region to a value of 10^{14}/cm^3 offered no current changes of any significance for the range of bias level studied. The result will be reported in detail at a later time.

IV. CONCLUSIONS

To summarize, a broad range of numerical calculations have been performed for the $N^+N^-N^+$ structure which sustains a highly nonuniform distribution of charge and potential. Both transient and steady state calculations have been performed and discussed. There are two significant observations of this study: (i) displacement currents dominate the initial transients; and (ii) the current voltage characteristics reflect both space charge injection and electron transfer. Electron transfer may dominate for both long and very short N^- regions.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research and DARPA, to whom the authors are grateful.

[1] H. L. Grubin and J. P. Kreskovsky, Surf. Sci. 132, 594 (1983).
[2] R. Fauquemberque, M. Pernisek, and E. Constant, *Proceedings of the 1982 Workshop on The Physics of Submicron Structures* (Plenum, New York, 1984), p. 171.
[3] J. R. East and P. A. Blakey, *Proceedings of the 1982 Workshop on the Physics of Submicron Structures* (Plenum, New York, 1984), p. 287.
[4] M. S. Shur and L. T. Eastman, IEEE Trans. Electron Devices ED-26, 1677 (1979).
[5] J. R. Barker, D. K. Ferry, and H. L. Grubin, IEEE Electron Device Lett. EDL-1, 209 (1980).
[6] R. Zuleeg, IEEE Electron Device Lett. EDL-1, 234 (1980).
[7] C. V. Shank, R. L. Fork and B. I. Greene, Appl. Phys. Lett. 38, 104 (1981). This experiment was performed on GaAs/AlGaAs double heterostructures.
[8] H. L. Grubin and D. K. Ferry, J. Vac. Sci. Technol. 19, 549 (1981).
[9] H. L. Grubin and J. P. Kreskovsky (to be published).
[10] H. L. Grubin and J. P. Kreskovsky (to be published).
[11] R. K. Cook and J. Frey, IEEE Trans. Electron Devices ED-28, 951 (1981).
[12] D. K. Ferry and H. L. Grubin, Microelectron. J. 12, 5 (1981).

Reprint 15

VLSI ELECTRONICS MICROSTRUCTURE SCIENCE, VOL 10

Chapter 6

The Role of Boundary Conditions in Near- and Submicrometer-Length Gallium Arsenide Structures*

H. L. GRUBIN
J. P. KRESKOVSKY

Scientific Research Associates, Inc.
Glastonbury, Connecticut

I.	Introduction	237
II.	Transport through Moment of the Boltzmann Transport Equation	240
III.	Solution of the Governing Equations	245
	A. Uniform Fields	245
	B. Nonuniform Fields and Uniform Doping	249
	C. The Effect of Device Length	254
	D. Steady State and Transient Behavior Injecting Cathode ($L = 1.0\ \mu m$)	255
	E. Steady State and Transient Behavior and Partially Blocking Cathode ($L = 5.00\ \mu m$) Gunn Oscillations	274
	F. Nonuniform Fields and Length Scaling	281
	G. Transients in $n^+-n^--n^+$ Structures and Length Scaling	292
IV.	Conclusions	316
	References	317
Appendix	Dimensionless Equations Used in the Numerical Simulations	317

I. INTRODUCTION

It is now generally accepted that electrical instabilities in bulk III–V semiconductors are controlled by the *details* of the boundary as well as *details* of the interior regions. By boundary we mean the metal–

* This study was supported by ONR and DARPA.

Copyright © 1985 by Academic Press, Inc.
All rights of reproduction in any form reserved.
ISBN 0-12-234110-4

semiconductor interface, the n^+-n^- interface, the semiconductor–vacuum interface, etc. The situation with submicrometer devices is such that, by virtue of the thin interior region, the interface is expected to exercise principle control over transport within the semiconductor and devices constructed thereof.

Transport within any device, particularly with regard to boundaries is three dimensional. The distribution function within the device mirrors scattering events at the boundaries, particle confinement, and a host of detailed surface properties. Difficulties arise simply in describing the role of the boundary theoretically and identifying its influence experimentally. In this volume, there are several chapters dealing with the role of the boundary for transport parallel to the interface. Here, however, the discussion will be confined to transport normal to the interface. Particular emphasis will rest with identifying the role of the boundary in controlling transport in near- and submicrometer-length devices.

In examining the role of the boundary, cognizance is taken of the chapter by Hess *et al.* [1] on the dependence of transport on the energy and velocity distribution of electrons entering a *uniform* field region. In the discussion below, however, emphasis is on spatially dependent transport in which both the space charge and the field distribution within the device are nonuniform. The reason for including nonuniformities in the discussion is that they are consequences of the presence of contacts and/or the existence of nonuniformities in the doping profile. The significance of including them in the study lies in the fact that transient effects in the presence of spatial inhomogenieties are both qualitatively and quantitatively different from those calculated under uniform field conditions. Several examples illustrate these differences. First, under uniform field conditions, long-time steady state velocities show the presence of a dc negative differential conductivity in gallium arsenide arising from electron transfer. Under nonuniform field conditions where current rather than velocity is the relevant quantity, calculations for devices with injecting, partially blocking contacts and highly nonuniform $n^+-n^--n^+$ structures show highly nonlinear current–voltage ($I-V$) relationships. These $I-V$ characterizations do not, however, display negative differential conductivity.* Another point of importance involves the character of the transient. For uniform fields the signature of velocity overshoot lies in an initial high peak velocity followed by electron transfer and a rapid settling toward steady state. Under nonuniform field conditions, the initial transient is dependent upon the structure of the device. For n^+-

* The absence of dc negative differential conductivity (NDC) from the calculations reported later does not imply the universal absence of dc NDC from transferred electron semiconductors. It is possible to envision the *mathematical* possibility of a boundary with a region of NDC, which when coupled to a transferred electron semiconductor, will yield dc NDC.

n^-–n+ regions, the initial transient sustains major position-dependent displacement current contributions. These displacement current contributions arise from the internal rearrangement of electric fields and have the effect of increasing the lapsed time before the field reaches its steady state value. This results in a decreased velocity overshoot transient but not a decreased spatial overshoot, as discussed below.

In another matter, it must be recognized that the role of metal boundaries and/or properly designed heterostructure interfaces is significantly different than the role of the n^+–n or n–n^+ interface on device operation. The key element here, even for transport normal to the interface, may be carrier confinement. A 0.25-μm structure with carriers confined to this region will behave differently than a 1-μm-long n^+–n^-–n^+ element in which the n^- region is only 0.25 μm in length. For the n^+–n^-–n^+ structure at sufficiently high fields enough of the potential can fall across the downstream n^+ region to cause it to maintain high current densities and electron transfer.

A key element in the study of these electron devices lies in the description of the interface and how it is modeled. Here, it may be argued that there are several philosophical approaches that we may take. In one-dimensional descriptions, the metal–semiconductor interface may be treated as a mathematical boundary, with the variables chosen to represent the boundary dictated by the form of differential equations chosen to describe transport within the semiconductor. For example, in the drift and diffusion formulation of transport, the equation for total current is often expressed in terms of a second-order partial differential equation in field. Thus, the boundary condition involves specifying the field at the cathode and anode. In one study [2] the electric field was specified as a time-independent value and the resulting dc current voltage characteristic and time-dependent behavior, when it occurred, was shown to be a sensitive function of the chosen boundary value. More general discussions have included a time-dependent cathode field [3].

Another point of view may tend to ignore the mathematical boundary as an appropriate representation of the interface effect. Instead, at a position far removed from the boundary, an effective field may be introduced to account for the consequences of e.g., a dipole layer, or indeed the dipole layer may be introduced [4]. The region must then be coupled to a set of time-dependent rate equations that account for either thermionic emission or field-assisted tunneling through the generated barrier [5].

Independent of the point of view taken to model the effect of the interface in the presence of an applied field, the carriers will enter the semiconductor with a well-defined distribution of energies that are likely to be significantly different from those far from the interface. A case in point is gallium arsenide where the following question may be asked. When the distribution of carriers, velocity, and energy in the Γ, L, and X valleys are known at the up and

downstream interface, then through solution to the governing interior equations, it may be expected that the current–voltage relation and transient behavior of the structure in principle is predictable; given this, can the obverse side be seen? Namely, can we extract from a given set of electrical measurements on near- and submicrometer structures a family of interfacial characteristics within which material variations lead to predictive device behavior. This approach is clearly iterative and has been attempted. It may also be necessary if we have any hopes of engineering structures for high-speed applications. Indeed, there are already indications that this approach may be successful. The evidence lies in the success of the boundary field models to explain, on one level, the broad range of electrical behavior of gallium arsenide and indium phosphide [6] and the apparent relationship of these boundary field models to the energy, momentum, and carrier distribution of entering electrons.

The preceding discussion expresses the construction and viewpoint of this chapter. Namely, device boundaries and interfaces dictate that transport must reflect their presence. The purpose of this chapter is to illustrate this. The discussion is separated into two distinct parts, with the first part dealing with the equations governing near- and submicrometer transport. The description of transport is through moments of the Boltzmann transport equation. The second part of the discussion deals with boundary- and length-dependent transport. Initially, several uniform field transient calculations are included to introduce the language of transient transport and to form a basis for comparison with the nonuniform field results.

The nonuniform field results are discussed in Section III. Here, two distinct classes of devices are considered. The first consists of a uniformly doped structure in which all space charge nonuniformities arise from variations in the upstream boundary (cathode) conditions. The second device structure is the $n^+-n^--n^+$ structure in which nonuniformities in the space charge arise primarily from the n^+-n^- and n^--n^+ interfaces. Transient calculations with both structures show distinct local displacement current contributions, which will camouflage, in many cases, the presence of transient overshoot.

A brief summary of the basic findings of the study is contained in Section IV.

II. TRANSPORT THROUGH MOMENT OF THE BOLTZMANN TRANSPORT EQUATION

Spatial and temporal transients are determined through solution to a set of coupled equations. These include Poisson's equation

$$\nabla^2\phi = +(e/\epsilon)(n - n_0), \tag{1}$$

where n_0 is a prespecified background concentration and n denotes the free carrier contribution arising from various portions of the conduction band. For the discussion below, only two sections of the conduction band are considered, Γ and L. Thus,

$$n = n_1 + n_2, \tag{2}$$

where n_1 designates the population of the Γ valley and n_2 the population of the L valley.

Poisson's equation is coupled to the first three moments of the Boltzmann transport equation, the first set of which involves continuity. For the Γ valley

$$\frac{\partial n_1}{\partial t} = -\frac{\partial}{\partial x_j}\frac{n_1 \hbar k_1^j}{m_1} - n_1\Gamma_1 + (n - n_1)\Gamma_2, \tag{3}$$

where Γ_1 denotes the rate at which carriers are scattered from the Γ valley to all sections of the L valley and Γ_2 denotes return scattering. It is noted that for parabolic bands, we assume

$$\hbar k_1^j = m_1 V_1^j. \tag{4}$$

An equation similar to Eq. (3) describes transient population changes in the L valley. When the two are combined, a global continuity equation results,

$$\frac{\partial n}{\partial t} = -\frac{\partial}{\partial x_j}\left[n_1\frac{\hbar k_1^j}{m_1} + (n - n_1)\frac{\hbar k_2^j}{m_2}\right]. \tag{5}$$

The quantity

$$n_1(\hbar k_1^j/m_1) + (n - n_1)(\hbar k_2^j/m_2) \equiv C^j \tag{6}$$

is the velocity flux density of the system. It is convenient to relate this term to a mean spatially dependent drift velocity

$$V^j = C^j/n. \tag{7}$$

It is noted that the total current density,

$$J^j = -eC^j + \epsilon(\partial F^j/\partial t), \tag{8}$$

is conserved; i.e.,

$$\partial J^j/\partial x_j = 0. \tag{9}$$

The second pair of moment equations is that of momentum balance. For the Γ-valley carrier

$$\frac{\partial}{\partial t} n_1 \hbar k_1^j = -\frac{\partial}{\partial x_i}\frac{\hbar k_1^i}{m_1} n_1 \hbar k_1^j + e n_1 \frac{\partial \phi}{\partial x_j} - \frac{\partial}{\partial x_i}\psi_1^{ij} - n_1 \hbar k_1^j \Gamma_3. \tag{10}$$

Here, Γ_3 represents the net rate of momentum scattering and ψ_1^{ij} represents

the components of the pressure tensor

$$\psi_1^{ij} = \frac{1}{4\pi^3}\frac{\hbar^2}{m_1}\int (\mathbf{k}-\mathbf{k}_1)_i(\mathbf{k}-\mathbf{k}_1)_j f\, d\mathbf{k}. \tag{11}$$

For the situation in which f represents a displaced Maxwellian

$$\psi_1^{ij} = n_1 k_B T_1 \delta_{ij}, \tag{12}$$

where T_1 is the electron temperature of the Γ-valley carriers. For the situation in which there are nonspherical contributions to f, additional diagonal as well as off-diagonal components of the pressure tensor arise. For the following calculations, the distribution function has been generalized from the displaced Maxwellian

$$f_0 = A\exp[-\hbar^2(k-k_1)^2/2m_1 k_B T_1] \tag{13}$$

into the form (see, for example, Sommerfield [7, Section 43])

$$f = \left(1 + a_i\frac{\partial}{\partial k_i} + a_{ij}\frac{\partial^2}{\partial k_i\,\partial k_j} + a_{ijk}\frac{\partial^3}{\partial k_i\,\partial k_j\,\partial k_k} + \cdots\right) f_0, \tag{14}$$

subject to the conditions

$$\frac{1}{4\pi^3}\int f\, d\mathbf{k} = \frac{1}{4\pi^3}\int f_0\, d\mathbf{k}, \tag{15}$$

$$\frac{1}{4\pi^3}\frac{\hbar^2}{m_1}\int f(k-k_1)^2\, d\mathbf{k} = \frac{1}{4\pi^3}\frac{\hbar^2}{m_1}\int f_0(k-k_1)^2\, d\mathbf{k}. \tag{16}$$

The nonspherical nature of the distribution function suggests the separation

$$\psi_1^{ij} = \psi_{01}^{ij} + \hat{\psi}_1^{ij}, \tag{17}$$

where ψ_{01}^{ij} is given by Eq. (12) and $\hat{\psi}_1^{ij}$ represents the additional contribution. The nonspherical contributions are not calculated from first principles. Instead, the treatments of fluid dynamics are followed with

$$\hat{\psi}_1^{ij} = -\mu_1\left[\frac{\partial V_1^i}{\partial x_j} + \frac{\partial V_1^j}{\partial x_i} - \frac{2}{3}\frac{\partial V_1^k}{\partial x_k}\delta_{ij}\right], \tag{18}$$

where it is noted that

$$\sum_{i=1}^{3}\hat{\psi}_1^{ii} = 0. \tag{19}$$

In one dimension (along x)

$$\hat{\psi}^{ij} = -\frac{4}{3}\mu_1\frac{\partial V_1^x}{\partial x}. \tag{20}$$

In two dimensions, the derivative of the stress tensor is

$$\frac{\partial}{\partial x_i}\hat{\psi}_1^{ij} = \mathbf{i}\left\{(-\mu_1)\left[\frac{4}{3}\frac{\partial^2 V^x}{\partial x^2} + \frac{\partial^2 V^y}{\partial x^2} + \frac{\partial^2 V^x}{\partial x\,\partial y} - \frac{2}{3}\frac{\partial^2 V^y}{\partial x\,\partial y}\right]\right.$$
$$\left. + \mathbf{j}\left\{(-\mu_1)\left[\frac{4}{3}\frac{\partial^2 V^y}{\partial y^2} + \frac{\partial^2 V^x}{\partial y^2} + \frac{\partial^2 V^y}{\partial x\,\partial y} - \frac{2}{3}\frac{\partial^2 V^x}{\partial x\,\partial y}\right]\right\}. \quad (21)$$

In the discussion below, an even simpler version of Eq. (21) is assumed:

$$\frac{\partial}{\partial x_i}\psi_1^{ij} = -\hat{\mu}_i\frac{\partial^2 V_1^j}{\partial x_i^2}, \quad (22)$$

with the constraint Eq. (15) only approximately satisfied. Thus, the relevant equation for momentum balance is

$$\frac{\partial}{\partial t}n_1\hbar k_1^j = -\frac{\partial}{\partial x_i}\frac{\hbar k_1^i}{m}n_1\hbar k_1^j + en_1\frac{\partial\phi}{\partial x_j} - \frac{\partial}{\partial x_j}n_1 k_B T_1$$
$$+ \frac{\hat{\mu}_1\hbar}{m_1}\frac{\partial^2 k_1^j}{\partial x_i^2} - n_1\hbar k_1\Gamma_3. \quad (23)$$

For the L valley, the relevant momentum balance equation is

$$\frac{\partial}{\partial t}(n-n_1)\hbar k_2^j = -\frac{\partial}{\partial x_i}\frac{\hbar k_2^i}{m_2}(n-n_1)\hbar k_2^j + e(n-n_1)\frac{\partial\phi}{\partial x_j}$$
$$- \frac{\partial}{\partial x_j}(n-n_1)k_B T_2 + \hat{\mu}_2\frac{\hbar}{m_2}\frac{\partial^2 k_2^j}{\partial x_i^2}$$
$$- (n-n_1)\hbar k_2^j\Gamma_4. \quad (24)$$

The third and final pair of balance equations is that associated with energy transport. Straightforward application of the moment equations yields

$$\frac{\partial}{\partial t}W_1 = -\frac{\partial}{\partial x_j}\frac{\hbar k_2^j}{m_1}W_1 + n_1 e\frac{\hbar k_1^j}{m_1}\frac{\partial\phi}{\partial x_j} - \frac{\partial}{\partial x_j}\frac{\hbar k_1^i}{m_1}\psi_1^{ij} - \frac{\partial}{\partial x_j}Q_1^j$$
$$- n_1 U_1\Gamma_5 + (n-n_1)U_2\Gamma_6, \quad (25)$$

where

$$U_1 = \tfrac{3}{2}k_B T_1, \quad (26)$$

$$W_2 = n_1\left[\frac{\hbar^2 k_1^2}{2m_1} + U_1\right], \quad (27)$$

and

$$Q_1^j = \frac{\hbar^3}{8\pi^3 m_1^2}\int (k-k_1)_j(k-k_1)_i^2 f\,d\mathbf{k}, \quad (28)$$

where the summation convention over i is assumed. For spherically symmetric distribution functions, Q_1^i is zero. For nonspherical situations it represents a flow of heat and is treated phenomenologically through analogy to Fourier's law

$$Q_1^i = -\kappa_1 \frac{\partial}{\partial x_j} T_1. \tag{29}$$

It is important to note at this point that the use of the relationships given by Eqs. (18) and (29) are not fundamental. Rather, they are expressions of ignorance of the detailed role of the distribution on transport, particularly near the boundaries.

In the analysis that follows, Eq. (25) is not solved. Rather, it is combined with Eqs. (3) and (24) to yield

$$\begin{aligned}\frac{\partial}{\partial t} n_1 U_1 = &-\frac{\partial}{\partial x_j}\frac{\hbar k_1^j}{m_1} n_1 U_1 - \frac{2}{3} n_1 U_1 \frac{\partial}{\partial x_j}\frac{\hbar k_1^j}{m_1} + \kappa_1 \frac{\partial^2}{\partial x_j^2} T_1 \\ &+ \frac{\hbar^2 k_1^2}{2m_1}[2n_1\Gamma_3 - n_1\Gamma_1 + (n-n_1)\Gamma_2] - n_1 U_1 \Gamma_5 \\ &+ (n-n_1) U_2 \Gamma_6. \end{aligned} \tag{30}$$

In Eq. (30), the nonspherical contributions of the stress tensor, [Eq. (10)] are ignored. For the second species of carriers

$$\begin{aligned}\frac{\partial}{\partial t}(n-n_1)U_2 = &-\frac{\partial}{\partial x_j}\frac{\hbar k_2^j}{m_2}(n-n_1)U_2 - \frac{2}{3}(n-n_1)U_2\frac{\partial}{\partial x_j}\frac{\hbar k_2^j}{m_2} \\ &+ \kappa_2 \frac{\partial^2}{\partial x_j^2} T_2 + \frac{\hbar^2 k_2^2}{2m_2}[2(n-n_1)\Gamma_4 + n_1\Gamma_1 - (n-n_1)\Gamma_2] \\ &- (n-n_1)U_2\Gamma_7 + n_1 U_1 \Gamma_8. \end{aligned} \tag{31}$$

Equations (1), (3), (5), (23), (24), (30), and (31) are the equations governing transport in the systems considered in this chapter. The equations are more general than others in that nonspherical contributions to the Boltzmann transport equation (BTE) moments have been included. The scattering integrals $\Gamma_1 - \Gamma_8$ and the form they take have been discussed in the past where these evaluations have been in terms of the displaced Maxwellian only. These integrals have not been generalized to include nonspherical contributions.

The governing equations are expressed in dimensionless form prior to transformation into difference equations. The dimensionless equations are discussed in the appendix to this chapter. Solution of the governing equation requires imposition of boundary conditions. These represent a crucial aspect

of the study and are discussed as they are needed. The band structure parameters used in the study for two-level transfer are also discussed in the appendix.

III. SOLUTION OF THE GOVERNING EQUATIONS

A. Uniform Fields

Calculations for uniform fields are discussed first, as they offer an important starting point for examining transients under nonuniform field conditions. Uniform fields result from assuming a donor level n_0 that is spatially constant to the boundary and specifying that

$$n_x = n_{1x} = V_{1x} = V_{2x} = T_{1x} = T_{2x} = 0 \tag{32}$$

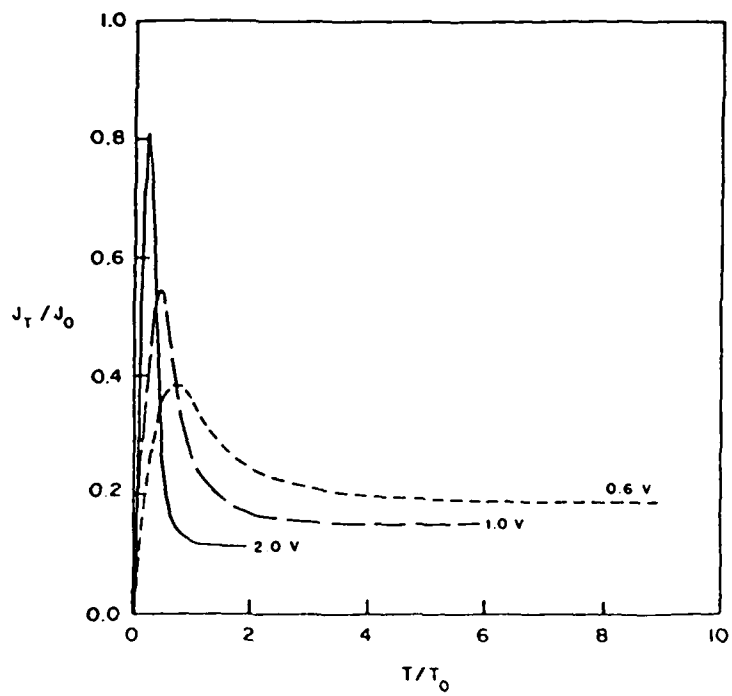


Fig. 1. Magnitude of the current transient [Eq. (8)] following application of a sudden change in bias. Parameters for this calculation are listed in Table A-1. The results of this calculation are qualitatively similar to those obtained in many studies, the first for GaAs being Ruch [8]. The terminus of each calculation reflects the physical time required for steady state. The longest time duration is that associated with the lowest bias level. For this calculation 2.0 V corresponds to an average field of 20 kV/cm; 1.0 V yields 10 kV/cm; etc. $J_0 = 8 \times 10^4$ A/cm^2; $T_0 = 1$ psec.

at both the cathode and anode boundaries. The subscript x in Eq. (32) denotes a first derivative. Figure 1 displays the velocity transient for a 1-μm-long element with a doping level of $5.0 \times 10^{15}/cm^3$. The parameters involved are listed in Table A-I. The length specification is artificial. For each calculation, the bias was raised in one time step from 0.01 V to the value indicated in the figure. We note the high carrier velocity occurring at approximately 0.5 psec and the long-term asymptotic lower steady-state value. Also apparent in the figure is the presence of a region of negative differential mobility. Figure 2 displays the time rate of change of carriers in the gamma valley. Electron transfer is apparent at times following the peak velocity. Figure 3 displays the time dependence of the electron temperature following application of the voltage pulse. The feature to be noted from this figure and Eq. (30) is that for uniform fields any time dependence in T_i is due entirely to scattering events and is thus a measure of when ballistic transport may be ignored. Another point of interest is that under uniform field conditions the population of carriers in either the central or satellite valley is governed by the scattering rates which are in turn governed by the value of the carrier temperature. This will be featured prominently later when contact effects are considered. The time dependence of the Γ-valley velocity is displayed in Fig. 4. (The long-time asymptotic values do not and should not display negative differential mobility. The mean steady state distribution of velocities as well

TABLE I
GaAs Parameters Used in Calculations

Parameter	Γ	L	Common
Number of equivalent valleys	1	4	
Effective mass (m_e)	0.067	0.222	
$\Gamma - L$ separation (eV)			0.33
Polar optical scattering			
Static dielectric constant			12.90
High-frequency dielectric constant			10.92
LO phonon (eV)			0.0354
$\Gamma - L$ scattering			
Coupling constant (eV/cm)			0.800×10^9
Phonon energy (eV)			0.0278
$L - L$ scattering			
Coupling constant (eV/cm)		2.0×10^9	
Phonon energy (eV)		0.0354	
Acoustic scattering			
Deformation potential (eV)	7.0	9.2	
Nonpolar scattering (L)			
Coupling constant (eV/cm)		0.300×10^9	
Phonon energy (eV)		0.0343	

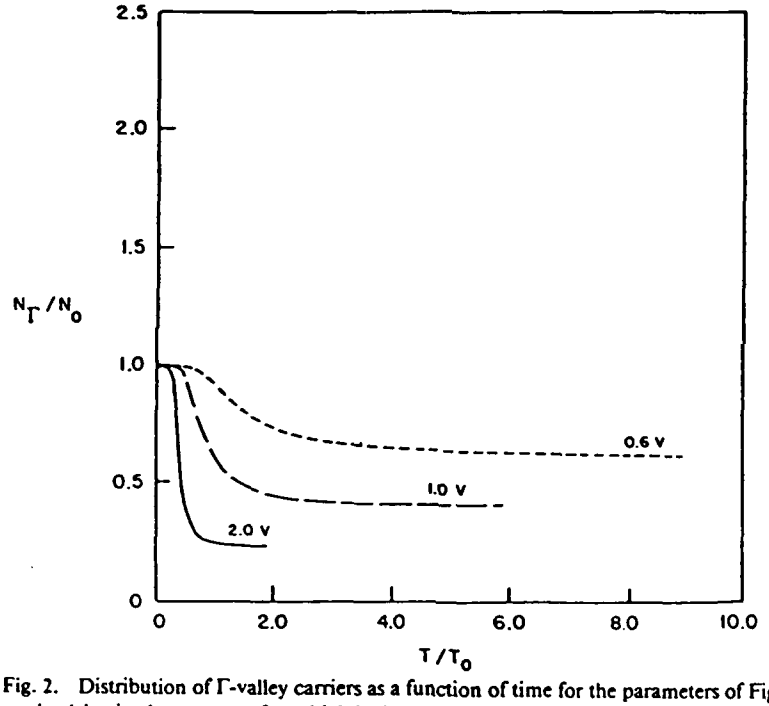


Fig. 2. Distribution of Γ-valley carriers as a function of time for the parameters of Fig. 1. Note the delay in electron transfer, which is shortest for the highest bias level. $N_0 = 5 \times 10^{15}$/cm^3; $T_0 = 1$ psec.

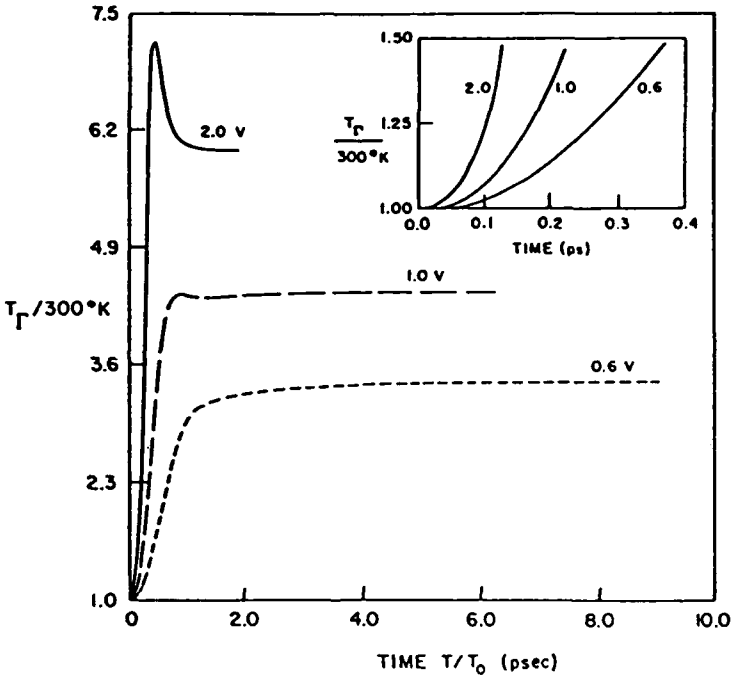


Fig. 3. Transient distribution of temperature following application of a sudden change in bias for the parameters of Fig. 1. The presence of a temperature overshoot is noted, a feature resulting from the enhanced scattering at elevated temperatures. The inset displays the temperature during the first 0.4 psec and demonstrates through application of Eq. (30) the onset of scattering. $T_0 = 1$ psec.

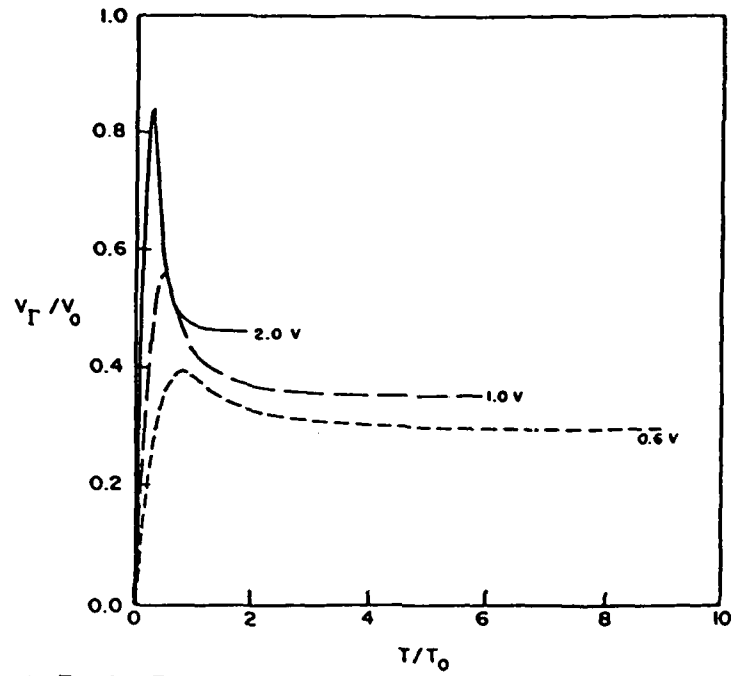


Fig. 4. Transient Γ-valley velocity distribution for the parameters of Fig. 1. The initial velocity peak corresponds closely in value to the peak current transient prior to electron transfer. The decreased velocity represents enhanced scattering at elevated temperatures. $V_0 = 10^8$ cm/sec; $T_0 = 1$ psec.

as that within the Γ valley is shown in Fig. 5. It is noted that nonparabolic effects are not included here.)

B. Nonuniform Fields and Uniform Doping

The origin of nonuniform fields and space charge layers in uniformly doped structures lies in the conditions imposed at the upstream and downstream boundaries. Under conditions in which current is flowing through the structure, the upstream boundary conditions manifest themselves as *cathode boundary current–field relationships*. It is the influence of the cathode boundary that will dominate the following discussion. To develop the concept of boundary controlled transport several qualitative features of the mathematics governing transport are considered.

Under time-independent steady state conditions, the velocity flux density

$$C = n_1(x)V_1(x) + [n(x) - n_1(x)]V_2(x) \equiv n(x)V(x) \tag{33}$$

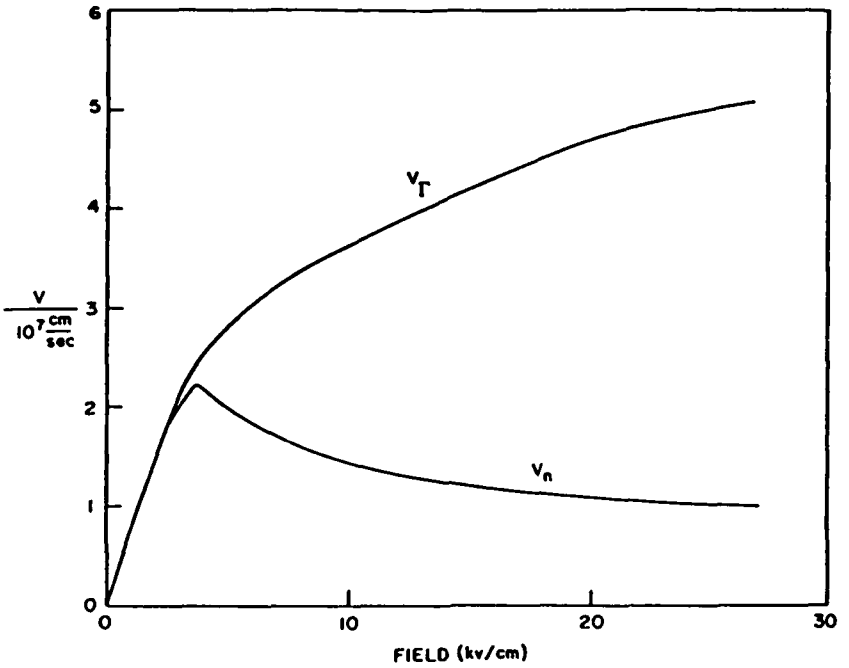


Fig. 5. Steady state field dependent velocity for electrons in the Γ-valley of gallium arsenide for the parameters of Table I. Also, steady state mean field dependent electron velocity V_n. [See Eq. (6).]

is a constant independent of position. Denoting, through the subscript c the carrier density and the mean velocity at the first computed point within the semiconductor, the following exercise is performed:

$$C - n_0 V_c = (n_c - n_0) V_c, \tag{34}$$

where n_0 is the uniform background doping level. For the purpose of specificity V_c is assigned to be a monotonically increasing function of field and to have the form represented by the curve $n_0 V_c$ in Fig. 6. Note that for uniform field conditions, V_c would necessarily be the same as the bulk field dependent velocity and exhibit negative differential mobility. Also included in Fig. 6 is a sketch of one possible variation of $n_c V_c$. The field dependence of n_c is thereby defined implicitly in Fig. 6. It is also noted that $n_0 V_c$ and $n_c V_c$ are chosen to intersect, although there is no reason *a priori* to assume any universality to this property. Figure 6 also includes a schematic of the velocity flux density, $n_0 V_n$, (assuming negative differential mobility) associated with uniform fields and two horizontal lines representing two different values of the current flux density within the device.

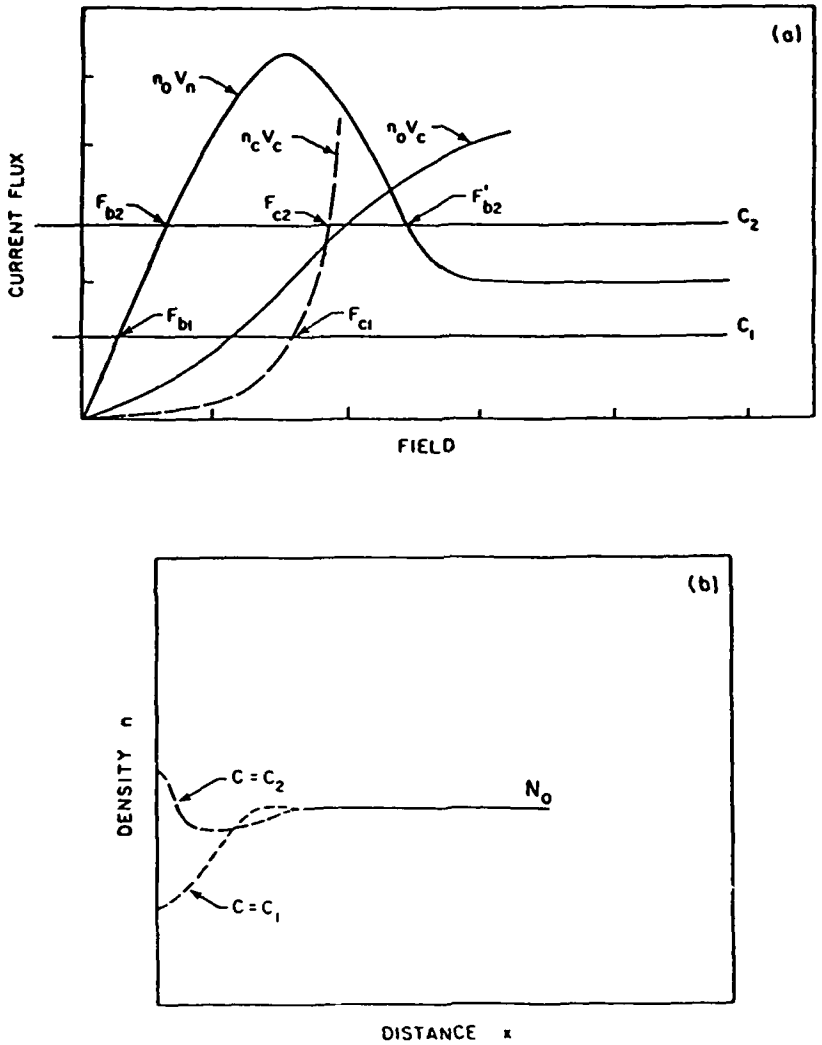


Fig. 6. (a) Schematic representation of a current–field relationship within the interior of the semiconductor n_0V_n and a possible current–field relation at the first computed point following the boundary of the semiconductor n_0V_c. (C_1 and C_2 represent constant current levels in the device.) The cathode field for the low current level case is denoted by F_{c1}. The neutral interior field is represented by F_{b1}. Similar remarks apply to the higher current level. (It is important to note that studies using the drift and diffusion equations indicate that for $\phi_x = 0$ at the boundary, the transition from cathode depletion to cathode accumulation requires all three characteristics, n_0V_c, n_cV_c, and n_0V_n to intersect at the same point [9]. (b) Schematic of possible cathode adjacent depletion and accumulation, followed by broad depletion, for the two bias current levels of Fig. 6a.

Figure 6a takes on significance when the intersection of the line of constant current C and the neutral field characteristic n_0V_n is taken to represent uniform field region values within the interior of the semiconductor; and the intersection of C with the cathode characteristic n_cV_c is taken to represent field values at the boundary of the semiconductor [9].

Consider first the low-current case C_1. Here, the assumed current field relationships are such that for

$$F_c > F_{b1}, \tag{35}$$

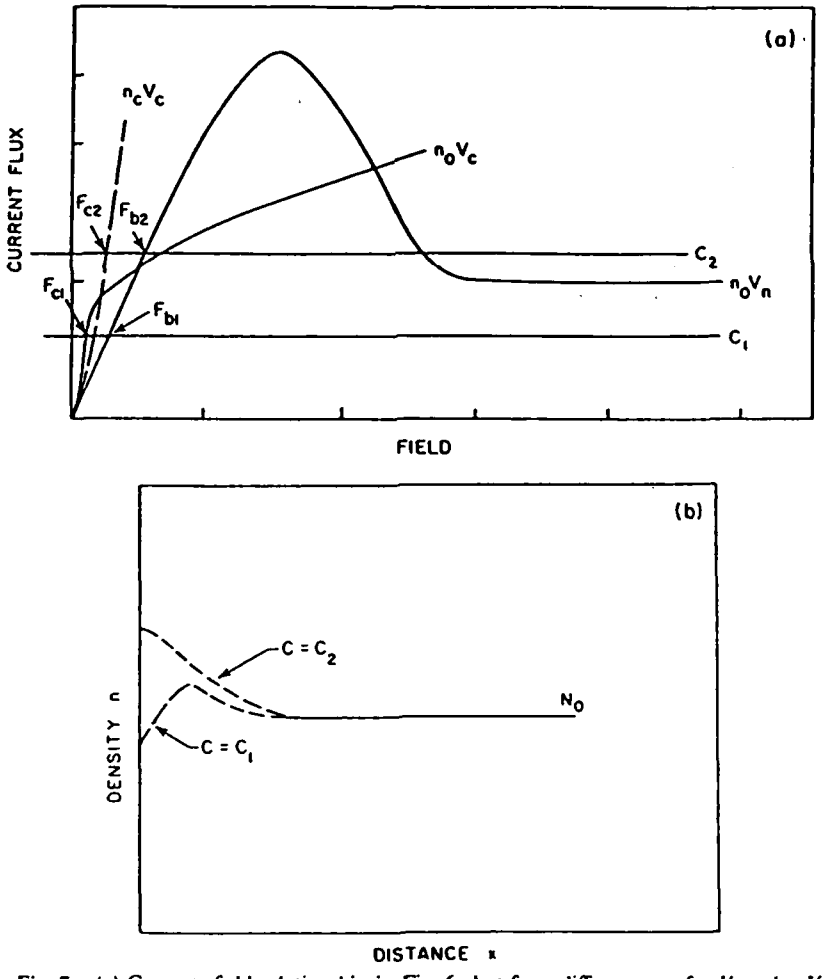


Fig. 7. (a) Current–field relationship in Fig. 6a but for a different set of n_0V_c and n_cV_c curves. (b) Schematic of possible cathode adjacent depletion and accumulation followed by broad accumulation for the two bias levels of Fig. 7a.

$n_0 > n_c$. For a specific distance between the upstream boundary and the interior of the structure, a range of charge depletion forms, as sketched in Fig. 6b.

Consider next the higher-current case C_2. For this situation

$$n_0 V_c(F_c) < n_c V_c(F_c) \tag{36}$$

and a region of local charge accumulation forms at the upstream boundary. Because the field dependence of the mean carrier velocity exhibits a region of negative differential mobility, the downstream interior field is either greater than or less than the cathode field and either a range of charge accumulation forms within the interior of the structure or a range of charge depletion forms within the interior. The latter is illustrated in Fig. 6b. It is important to note that nothing has been said about the stability of these profiles. Indeed, in some cases the profiles are electrically unstable [6].

Consider Fig. 7a with a different set of upstream boundary characteristics. For the low-current case

$$F_c < F_{b1}, \tag{37}$$

and a region of charge accumulation layer forms over a specific distance between the upstream boundary and the interior of the structure. However, at the upstream boundary

$$n_0 V_c(F_c) > n_c V_c(F_c), \tag{38}$$

indicating that a region of local charge depletion forms at the upstream boundary. A sketch of a possible space charge profile is shown in Fig. 7b. For the high-current case, both within the interior and at the upstream boundary regions of charge accumulation form. A sketch of this charge layer is also shown in Fig. 7b.

The preceding discussion indicates that the interplay between the boundary and the interior of the semiconductor is able to introduce a rich variation in the space charge distribution. A situation evoking considerable interest with respect to this interplay is one that may be regarded as a singular solution. This occurs when the current flux density C_2 intersects the neutral characteristic at two points and the curves $n_c V_c$, $n_0 V_c$, and $N_0 V_a$ intersect at the same field value (thus, $V_a = V_c$ and $n_0 = n_c$). The general description and consequences of the approach to this event in long samples, as a precursor for nucleation of high electric field traveling dipole layers has been broadly delineated in a variety of publications [6,9]. The consequence of this in terms of solutions to the BTE moments is discussed in Subsection II.E.

It should be apparent from the preceding discussion that the detailed description of the influence of the boundary requires a description of the field dependence of the mean *entrance* velocity and carrier distributions. In the following calculations in which transport is described through solutions

to moments of the BTE, these field dependencies are constrained by the boundary conditions to the governing equation and are expressed as solutions of

$$C = n_c V_c(F_c) \Rightarrow F_c = g(C). \tag{39}$$

When F_c is a double-valued function of C a regional approach is taken.

C. The Effect of Device Length

Up to this point, nothing has been discussed concerning the influence of device length on the cathode characteristic properties, nor on the transport properties through the structure. There are, however, several points of note. First, in the calculations, the field at the cathode is not specified but is computed self-consistently from the governing equation subject to the constraints of the boundary conditions. For constant n_0 the details of transport from the cathode are dominated by the uniform field neutral velocity characteristic. For uniform fields and long devices, the velocity exhibits negative differential mobility as displayed in Fig. 5. However, as the device decreases in length, the mean carrier velocity for uniform fields is altered as displayed in Fig. 8. (See Grubin *et al.* [10] for a discussion of how the calculation was

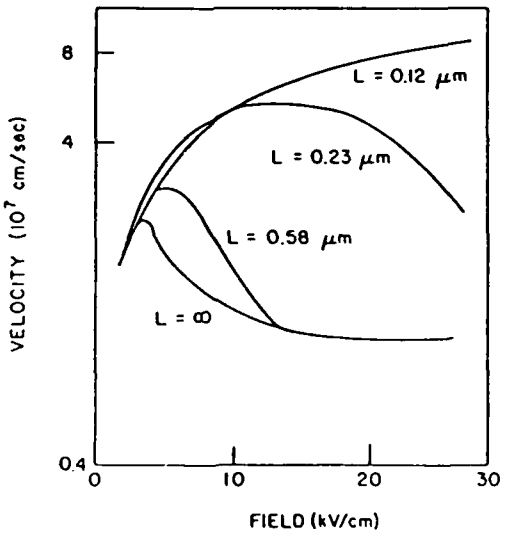


Fig. 8. Velocity versus distance for the uniform field velocity transient. For these calculations, as in Figs. 1–4, velocity transient is in response to a sudden change in electric field. Initial velocity is zero and $L = \int_0 V(t)\,dt$. (From Grubin *et al.* [10].)

performed along with the imposed condition.) Perhaps the most significant phenomena occurring as the semiconductor length is reduced are the progressive decrease in negative differential mobility of GaAs and the increase in the velocity and, hence, current. These effects are illustrated below.

D. Steady State and Transient Behavior Injecting Cathode ($L = 1.0\ \mu m$)

The preceding discussion is independent of the detailed description provided by the governing equations chosen to represent device transport. The governing equations and their associated boundary conditions provide a mechanism by which a set of contact descriptors can be extracted. For example, it is expected that the specific properties of the physical contact or boundary will influence the distribution of carriers within the valence and conduction bands of the semiconductor in the vicinity of the boundary. (One of the earlier studies involving the role of the boundaries on transient transport through solutions of the Boltzmann transport equation was that of

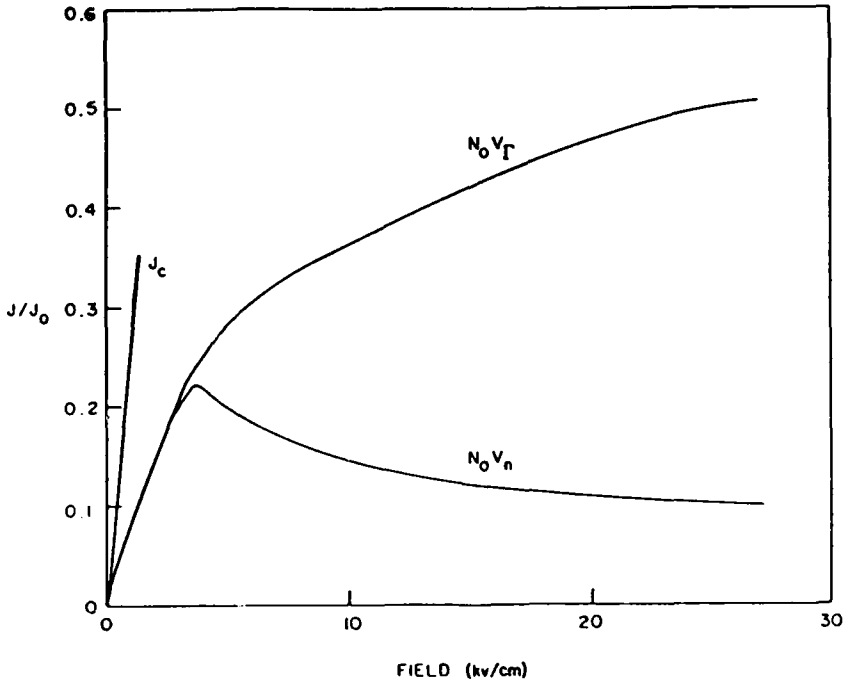


Fig. 9. Data of Fig. 5 plus the cathode current field relation for the accumulation layer boundary. $J_0 = 8 \times 10^4$ A/cm^2.

Gray *et al.* [11].) Furthermore, under conditions of finite bias in which current is transported through the device, the influence of the contact is expected to affect the entrance velocities [2]. In the discussion that follows, a very simple set of boundary conditions is imposed to represent the effects of the physical boundary. The importance of these boundary conditions is to create nonuniform fields. As will be seen, the boundary conditions chosen are not the result of an exhaustive study. Rather, they are associated with an initial effort. For example, in the following discussion, the initial sharing of carriers between the Γ and L portions of the conduction band is controlled

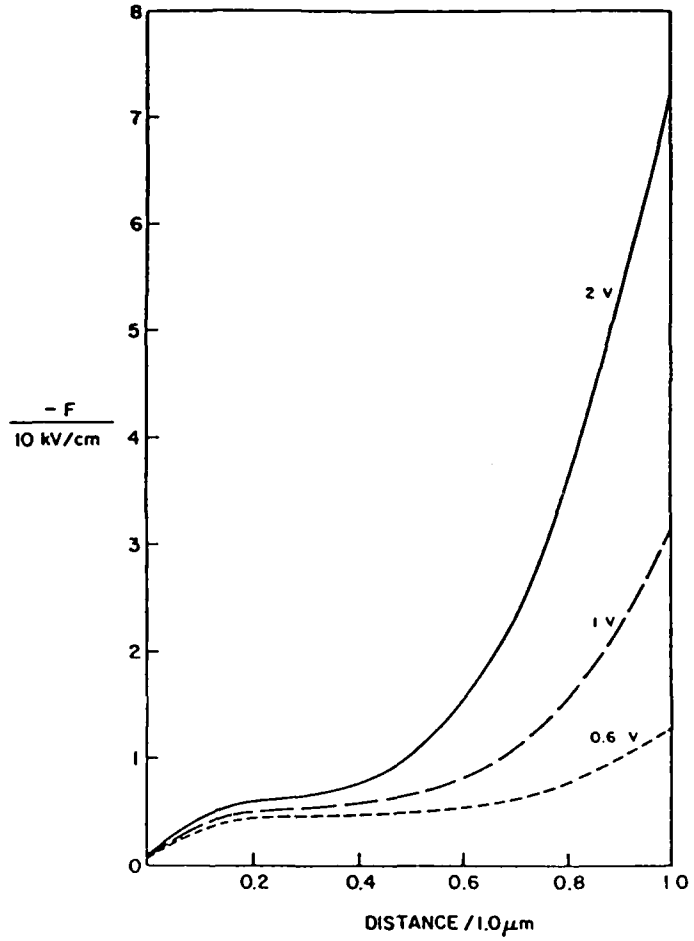


Fig. 10. Steady state distribution of field within a 1-μm-long GaAs element at three bias values; $T = 300$ K. (See appendix for boundary conditions.) Electron transfer occurs downstream from the cathode resulting in a downstream accumulation of carriers.

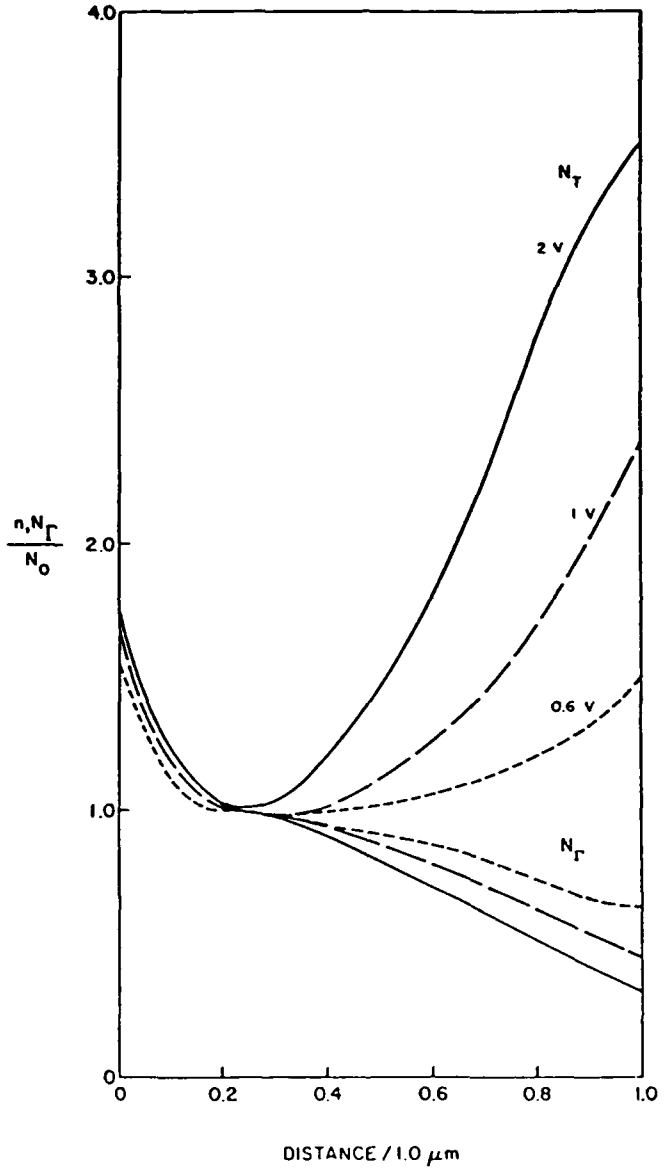
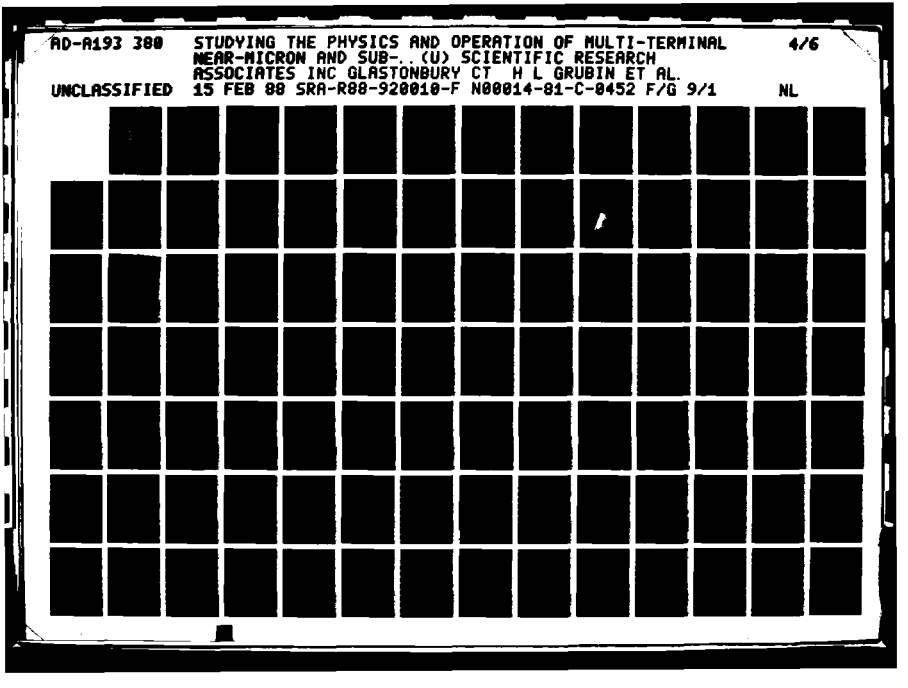
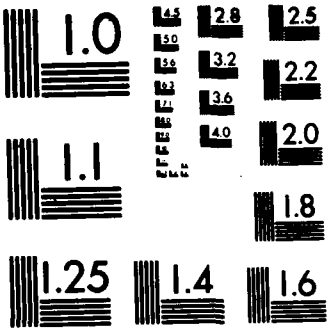


Fig. 11. Distribution of total and Γ-valley carrier density for the parameters of Fig. 10. Electron transfer begins within 0.2 μm downstream from the cathode. By comparing Figs. 2 and 11, it is noted that electron transfer at 6, 10, and 20 kV/cm significantly lags the uniform field value.

AD-A193 388 STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL NEAR-MICRON AND SUB-.. (U) SCIENTIFIC RESEARCH ASSOCIATES INC GLASTONBURY CT H L GRUBIN ET AL. 4/6

UNCLASSIFIED 15 FEB 88 SRA-R88-920010-F N00014-81-C-0452 F/G 9/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

by specifying a value for the electron temperature at the cathode boundary. In addition, a representation of the entrance velocity is through a cathode contact mobility. This is identified in the calculation beginning with Fig. 9.

Figures 9–12 are calculations performed for a gallium arsenide structure with the same material parameters as that of the uniform field calculations. Here, however, the boundary conditions are different. At the cathode

$$n_{xx} = n_{1xx} = 0, \qquad V_1 = -15{,}625\,F, \qquad V_{2x} = 0, \qquad T_1 = 300\,K, \qquad T_{2x} = 0, \tag{40}$$

and at the anode

$$n_{xx} = n_{1xx} = V_{1xx} = V_{2xx} = T_{1xx} = T_{2xx} = 0, \tag{41}$$

where the double x subscript denotes a second derivative. The consequences of this set of boundary conditions is that the Γ-valley electrons enter the structure with a velocity in excess of the steady state uniform field value.

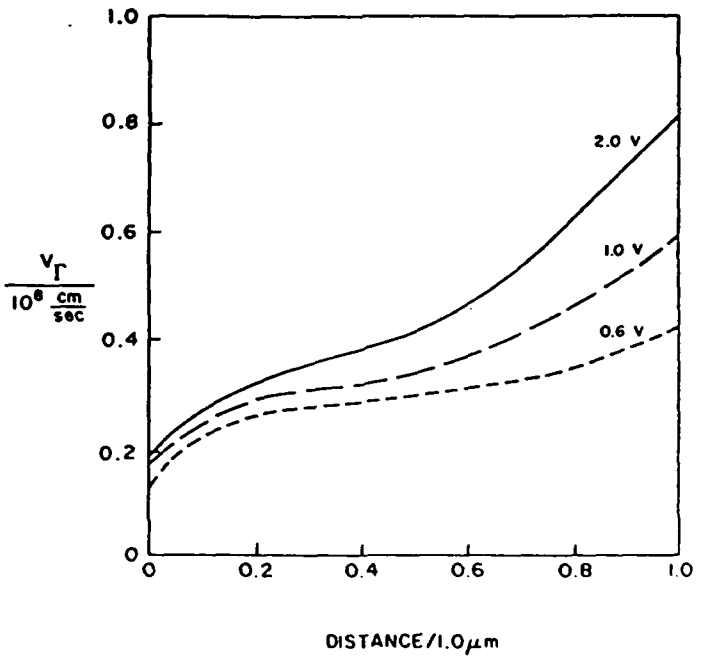


Fig. 12. Distribution of Γ-valley velocity for the parameters of Fig. 10. At a bias of 2 V and a field of 20 kV/cm the Γ valley velocity is slightly in excess of the uniform field calculation. At a bias of 1 V and a field of 10 kV/cm, the difference between the nonuniform and uniform field velocity is even greater. This excess is a consequence of a lower value of electron temperature at these given field values.

Specification of the Γ-valley temperature at 300K ensures that the relative cathode carrier contribution of the L valley is negligible. Furthermore, the fact that the mean velocity of the L-valley carriers is significantly below that of the Γ-valley carriers provides the demonstration that the cathode current field relation is dominated by the Γ valley carriers:

$$J_c = -e[nV_{1c} + (n - n_1)V_{2c}] \approx ne\mu_c F. \tag{42}$$

While Eq. (42) is significant in providing a description of the dominating carrier at the cathode, alone it will not determine whether the cathode is carrier depleted, neutral, or accumulated. The moment equations coupled to Poisson's equation must be solved. Qualitative information, however, can be obtained for the specific set of boundary conditions given by Eq. (40) through use of the mobility approximation. Because of the inherent limitations of the mobility approximation the consequences of its use must be regarded as relevant only if insight is provided in the interpretation of the exact solution.

The qualitative information is obtained through a calculation of the transit time of a carrier within the vicinity of the cathode. Because the transit time is necessarily a positive quantity, inequalities arise which express cathode depletion, neutrality, and accumulation. The transit time between the cathode and an interior point X is

$$t(X) = \int_0^x \frac{dX'}{V(X')}. \tag{43}$$

Assuming a constant mobility for the Γ-valley carriers and the significant approximation

$$nV \approx n_1 V_1. \tag{44}$$

then for carriers within the vicinity of the cathode the arguments leading to Eq. (42) imply that $V \approx -\mu_1 F$. This last statement, when coupled to Poisson's equation yields

$$t(X) = \tau_0 \log\left\{\frac{J - n_0 e\mu_1 F_c}{J - n_0 e\mu_1 F(X)}\right\}, \tag{45}$$

where

$$\tau_0 = \epsilon / n_0 e\mu_1 \tag{46}$$

is the dielectric relaxation time of the Γ-valley carriers. For a cathode boundary condition consistent with $J = n_c e\mu_c F_c$

$$t(X) = \tau_0 \log\left\{\left(1 - \frac{n_0\mu_1}{n_c\mu_c}\right) \Big/ \left(1 - \frac{n_0\mu_1 F(X)}{n_c\mu_c F_c}\right)\right\}. \tag{47}$$

Since the requirement that the transit time be positive must be met, two inequalities emerge:

$$n_c\mu_c > n_0\mu_1, \qquad F(X) < F_c, \tag{48a}$$

$$n_c\mu_c < n_0\mu_1, \qquad F(X) > F_c. \tag{48b}$$

For Eq. (48a), local charge accumulation is present at the cathode. Note that the condition $F(X) < F$ is stronger than necessary and requires that $N_c > N_0$. In Eq. (48b) cathode depletion occurs with $N_c < N_0\mu_1/\mu_c$. The results of the following simulation discussion are consistent with Eq. (48a) as demonstrated in Fig. 9. For Eq. (48b), reference is made to the discussion of Grubin and Kreskovsky [12]. Figure 9 is a plot of the computed dc current versus field relationship at the cathode boundary. It is approximately linear with only a marginal variation in field. The cathode field is effectively pinned. For reference purposes, the current–field relationship for the uniform field structure is also shown.

The characteristics of the uniform field curve and the cathode current–field relationship are different and for a constant current through the semi-

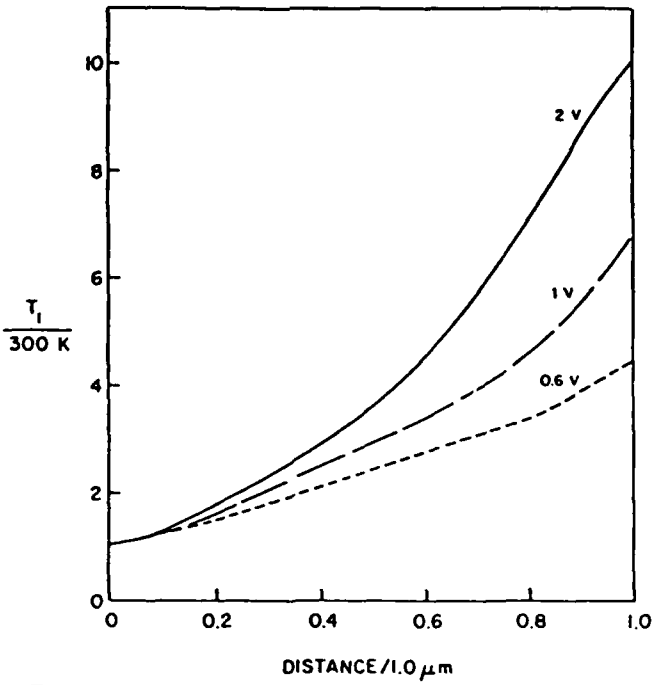


Fig. 13. Temperature distribution within the Γ valley for the parameters of Fig. 10. See comments associated with Fig. 12.

conductor at least two different field values result at intersections. The cathode boundary field is lower than that of the neutral field intersection, a result that is consistent with cathode accumulation.

The steady state time independent distributions of electric field, carrier density, Γ-valley velocity, and electron temperature are displayed in Figs. 10–13 for various bias levels. While the calculation displays the excess carrier velocity at elevated bias levels, there is also an enhanced electron transfer and the dc current shows saturation. The clear consequence of the transfer is that the current does not scale the velocity. This latter feature is reflected in the current–voltage relationship shown in Fig. 14.

With regard to the current–voltage characteristic, while the current does not scale the velocity and thus does not fully reflect overshoot contributions, its high bias level is above that associated with the equilibrium steady state velocity field relationship, while below that associated with the Γ-valley velocity. The excess above V_s is due predominantly to the cathode boundary condition that allows for a high level of injected charge. The depression below V_Γ is due to electron transfer. It is also noted that there is virtually no electron transfer near the cathode. Most of it occurs near the anode, and the effect of electron transfer leads to saturation in the current density. Another feature of the nonuniform field calculation lies in the clear *absence* of negative differential conductivity, a phenomenon present in uniform field calculations.

The significant qualitative differences between the steady state uniform field characteristics and those associated with nonuniform fields suggest some differences in the transient characteristics. This is indeed the case as discussed below.

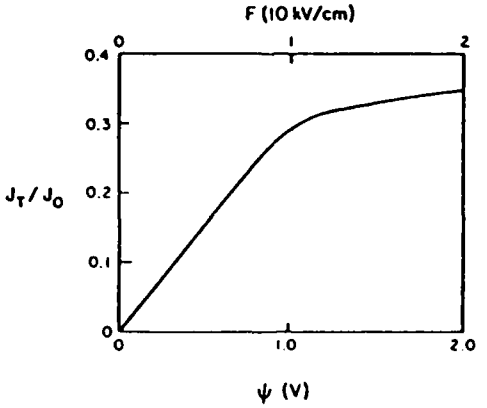


Fig. 14. Steady state current density versus applied voltage and average field for the 1.0-cm-long structure with the parameters of Fig. 10. $J_0 = 8 \times 10^4$ A/cm^2.

Figure 15 displays the current transient following application of a voltage pulse. The first point we emphasize is that the plot consists of current rather than velocity. The second point is that the current transient is ostensibly similar to that associated with velocity overshoot. There is, however, a fundamental difference between the two. For uniform and nonuniform fields during the first time step, the field throughout the structure is increased by an amount equal to the change in applied voltage divided by device length. This introduces a one-time-step displacement current whose magnitude is computationally dependent and therefore nonphysical. For uniform fields, all displacement current contributions cease after the initial time step. For nonuniform fields all time-dependent field evolution is accurately calculated following the initial time step. Here, with the cathode boundary introducing a cathode adjacent accumulation layer, the time dependence introduces a layer that propagates toward the anode boundary. This propagation is accompanied by field rearrangement and internal point-by-point displacement current contributions.

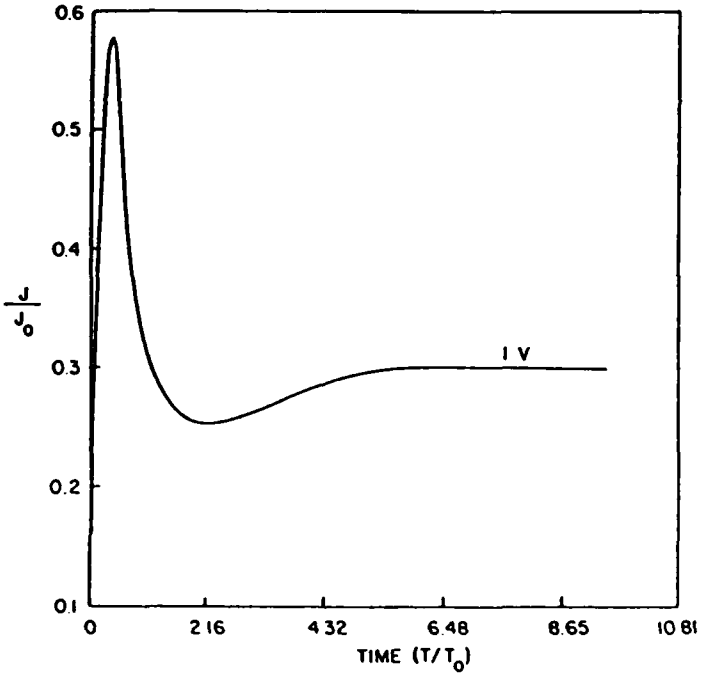


Fig. 15. Magnitude of current transient following application of a step change in bias to 1.0 V for a 1.0-μm-long device at 300 K with the parameters of Fig. 10. Current peak is similar to that of Fig. 1. Steady state velocity is above that of the uniform field case. It is noted that the time to steady state is approximately 50% longer than that associated with the steady state calculation of Fig. 1. $J_0 = 8 \times 10^4$ A/cm^2; $T_0 = 1$ psec.

Figure 16 shows the space- and time-dependent evolution of the electric field within the device. The effect of the boundary condition is to introduce a propagating accumulation layer originating at the cathode, while downstream from the anode the field is approximately uniform during the first 0.5 psec becoming highly nonuniform as steady state is approached. The early time transients dictate that displacement current contributions will be significant within the vicinity of the propagating accumulation layer, as shown in Fig. 17 and will be insignificant downstream from the layer. In the latter regions, the familiar velocity transients obtained from uniform field calculations arise. At later times propagation continues but is accompanied by electron transfer. The long-time transient differs from that of uniform fields. (We note from Figs. 18 and 19 the absence, for $t < 0.5$ psec, of any significant transfer downstream from the moving space charge layer. Hereto, the carrier velocity (Fig. 20) downstream from the moving space charge layer sustains high values common to overshoot).

There are three dominating features of the preceding calculations. The first two are the boundary conditions on the Γ-valley temperature and mean carrier velocity. The third is the length of the structure. As discussed earlier, the specification of the Γ-valley electron temperature provides dominant control in these calculations of the relative population of the Γ-valley carriers at the cathode. For the calculations of Figs. 10–17, specifying T_1 at 300K resulted in virtually the entire sea of cathode carriers as Γ-valley carriers. In a study performed earlier [12] in which the device length was 2.0 μm and the Γ-valley velocity was subject to a mobility boundary condition as in Eq. (40), the results were qualitatively similar for $T_1 = 300$K and a Γ-valley boundary mobility greater than that of the low-field steady state mobility of the Γ-valley carriers. In those calculations, the total set of boundary conditions was somewhat different than those employed in the discussion of Figs. 10–20 but there were several definite trends. For example, by retaining a suitably high cathode mobility and by elevating the electron temperature, space charge accumulation at the cathode was retained, but the relative proportion of Γ-valley carriers at the cathode decreased. Again, on the obverse side, retaining a cathode temperature of $T_1 = 300$K but reducing the boundary mobility of the Γ-valley carriers to a value below that of the low-field mobility of the Γ-valley carrier in steady state results in a partial depletion of carriers at the cathode and a concomitant increase in the cathode field to values in excess of that within neutral regions interior to the device. Each of these results is consistent with the qualitative arguments contained in Eqs. (43)–(48).

The immediate conclusion that can be drawn from the set of referenced results is that the presence of space charge accumulation or depletion at the cathode is dominated by the field dependence of the entering carrier velocity

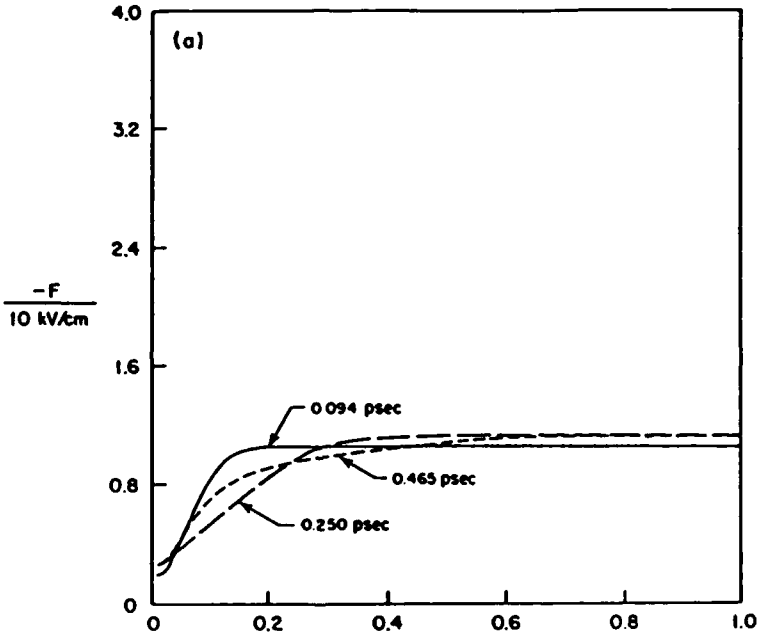


Fig. 16. Distribution of electric field at successive instants of time following application of a step change in voltage for the parameters of Fig. 15. During the first time step, the field increases from its steady state value at a bias of 0.01 V by an amount equal to 9.99 kV/cm [(1 - 0.01) V/ 1 μm]. Subsequent time dependence shows a space charge layer propagating toward the anode. (a) During the first 0.5 psec the field downstream from the propagating accumulation layer is spatially uniform. Within this region transients are governed by the uniform field velocity overshoot transients. (b) During the long-time transients, electron transfer occurs and relaxation differs from that of the uniform field transient.

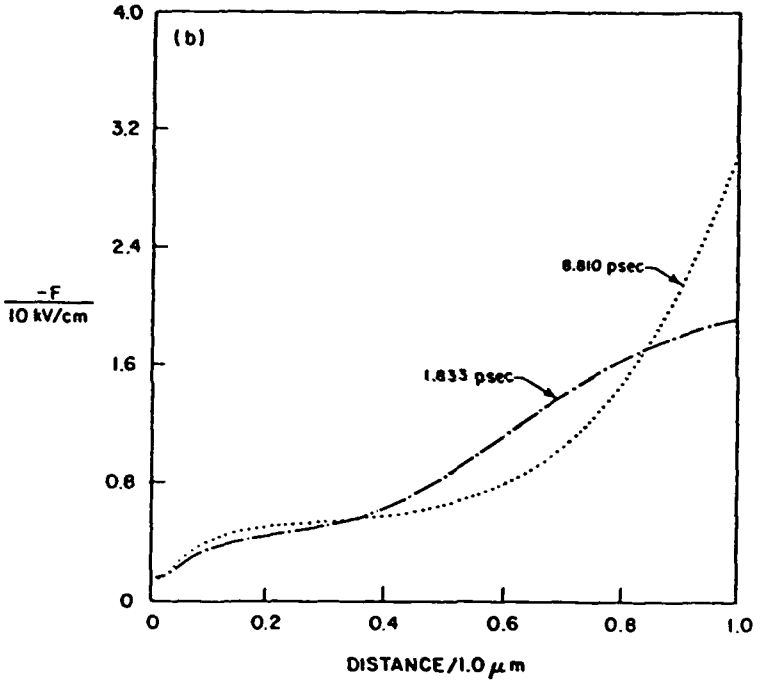


Fig. 16. *(Continued)*

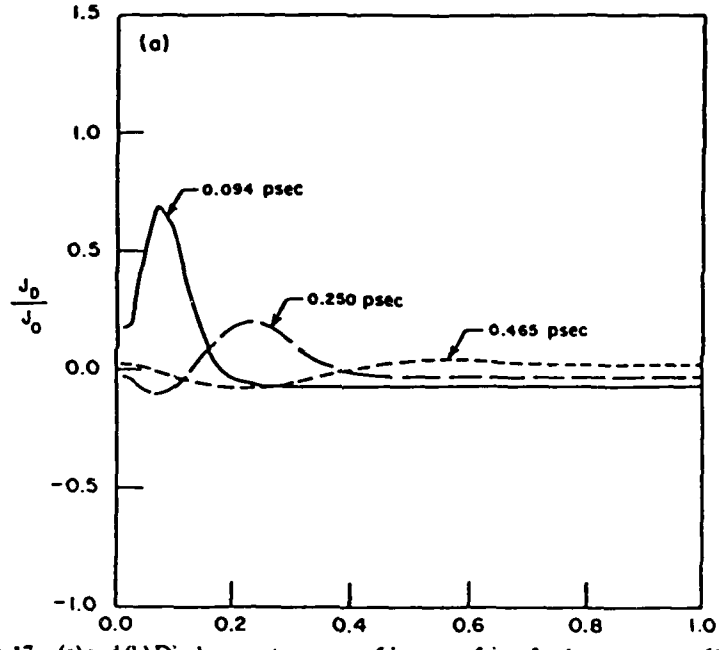


Fig. 17. (a) and (b) Displacement current at 5 instants of time for the parameters of Fig. 15. Initial displacement currents are strong and accompany the moving accumulation layer. $J_0 = 8 \times 10^4$ A/cm^2

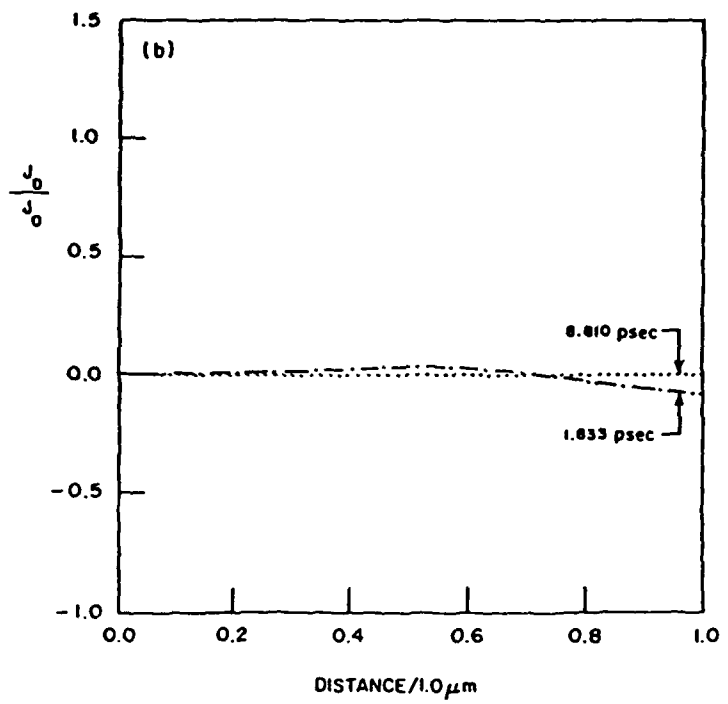


Fig. 17. *(Continued)*

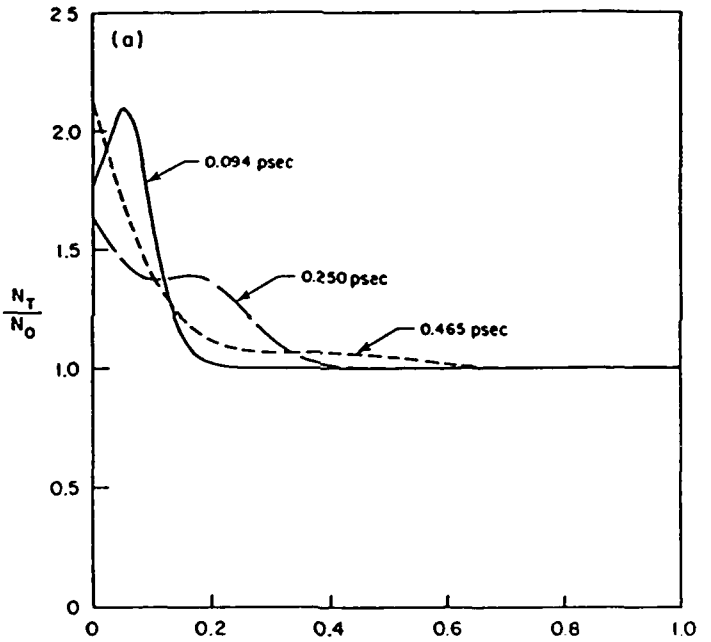


Fig. 18. (a) and (b) Transient distribution of total charge following application of a step change in potential for the parameters of Fig. 15. Note that downstream from the propagating accumulation layer the charge distribution is flat as reflected, additionally in the flat field profile of Fig. 16a. Space charge accumulation occurs during the longer time interval.

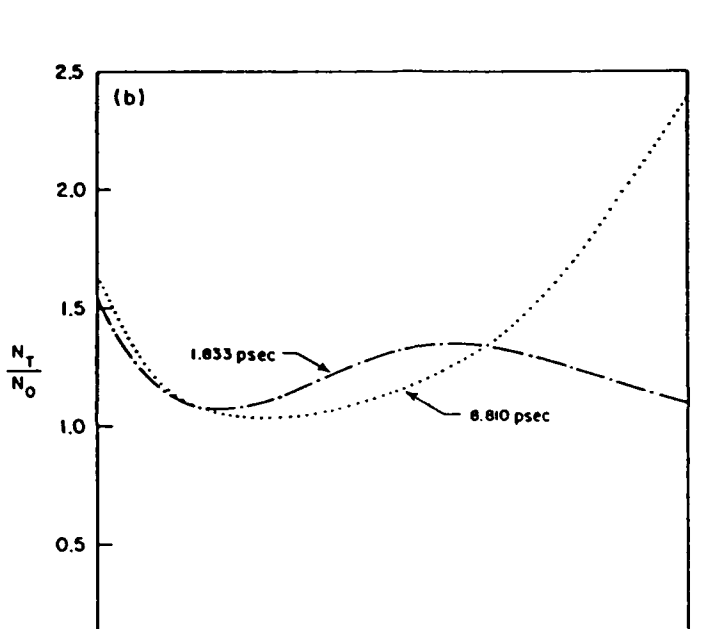


Fig. 18. *(Continued)*

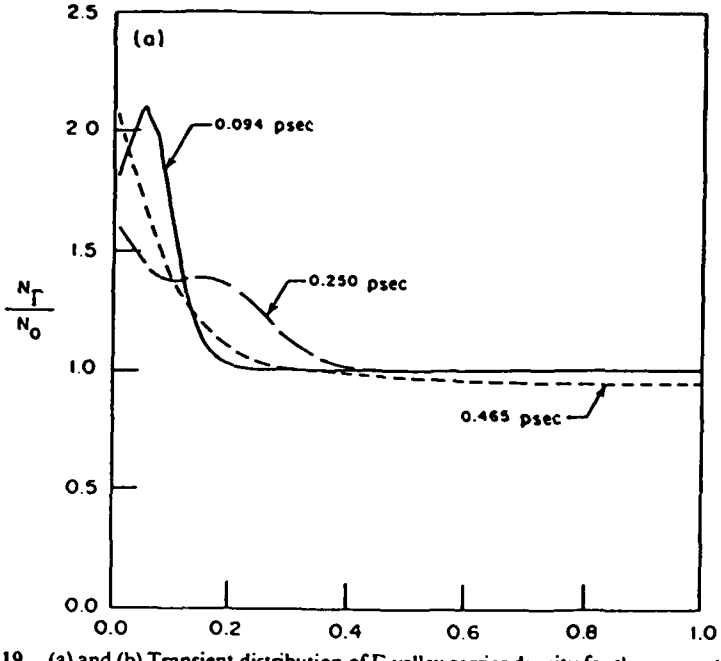


Fig. 19. (a) and (b) Transient distribution of Γ-valley carrier density for the parameters of Fig. 15. Note that within the first 0.5 psec, very little transfer occurs.

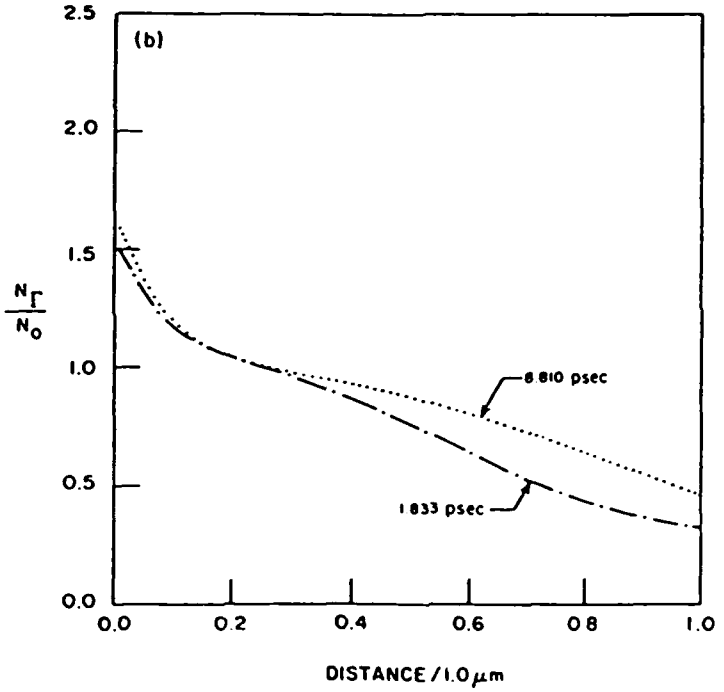


Fig. 19. *(Continued)*

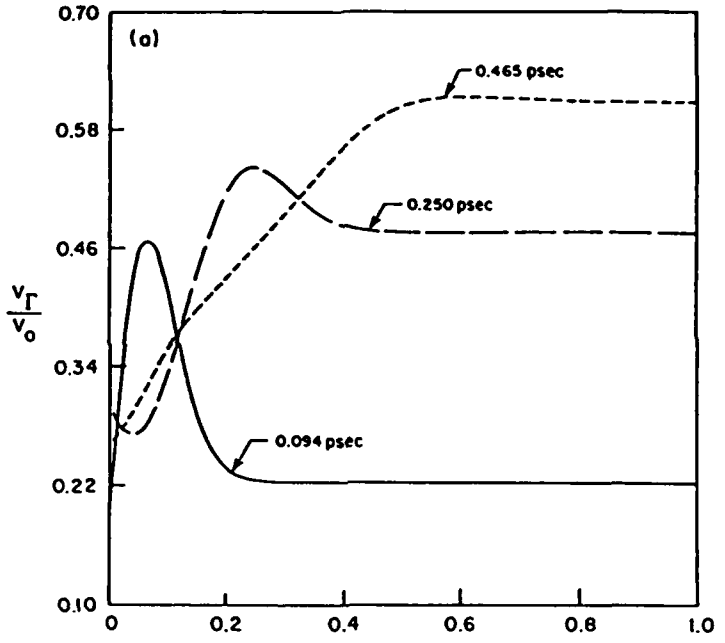


Fig. 20. (a) and (b) Transient distribution of the Γ-valley velocity following application of a voltage pulse for the parameters of Fig. 15. The velocity layer propagates and shows a tendency to lead the transient changes in the Γ-valley carrier density. Downstream, the velocity transient is relatively uniform for $t < 0.5$ psec and tends to follow the uniform field transient of Fig. 4. Differences from the uniform field calculations occur during the long-time transient. $V_0 = 10^8$ cm/sec.

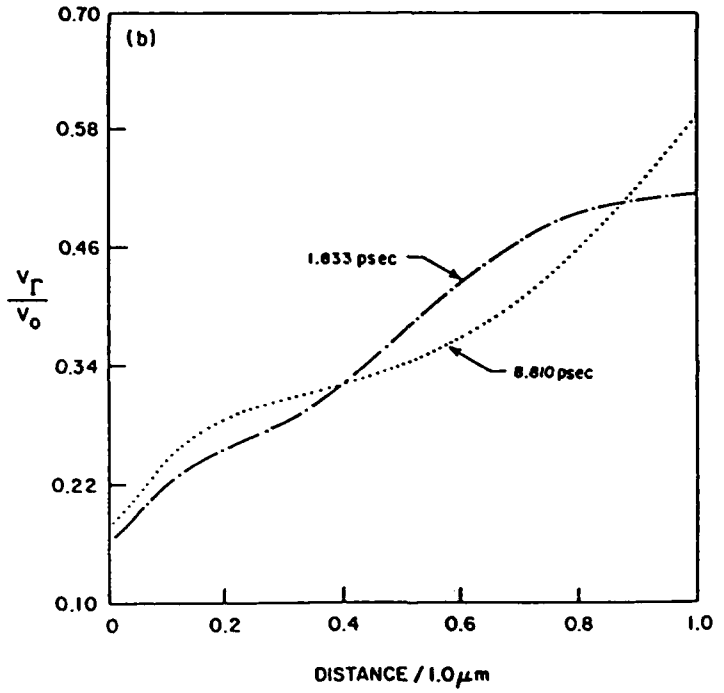


Fig. 20. *(Continued)*

vis-a-vis, that within the interior of the device. This conclusion is applied to a problem of high-visibility, transit-time Gunn domain instabilities in GaAs.

E. Steady State and Transient Behavior and Partially Blocking Cathode ($L = 5.00\ \mu$m) Gunn Oscillations

The structure under consideration is "long" with respect to submicrometer dimension; the device length is 5.00 μm. The boundary conditions here are different than those used for Figs. 10–20. In this case, those of Eq. (30) are repeated, with two critical variations:

$$V_1 = -4000F \quad \text{and} \quad T_1 = 1200\ \text{K} \tag{49}$$

As in the case of the accumulated cathode, the situation represented by Eq. (49) can be described qualitatively by Eqs. (43)–(47) with the modification

$$nV \approx n_1 V_1 \cong \beta n V_1, \tag{50}$$

where β represents an average of the fraction of Γ-valley to total carriers

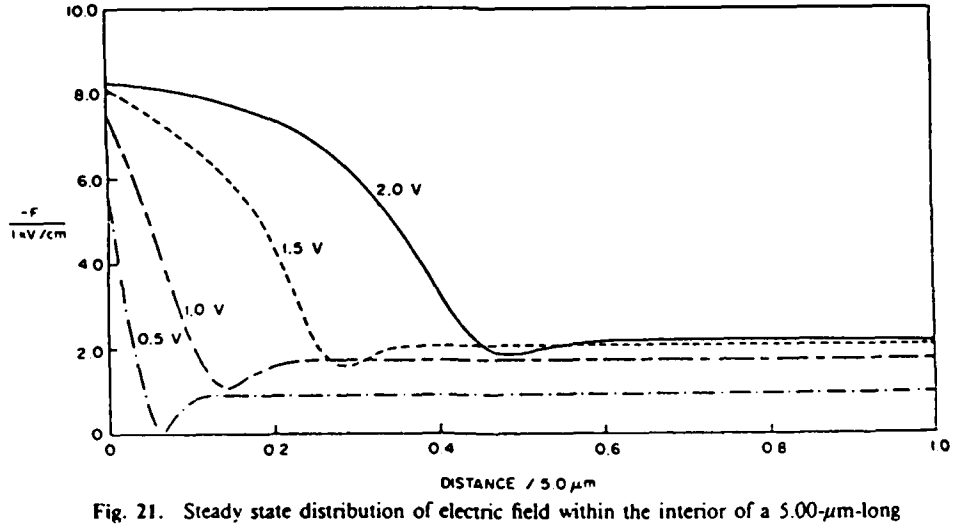


Fig. 21. Steady state distribution of electric field within the interior of a 5.00-μm-long uniformly doped GaAs structure at various bias levels. Parameters are given in the appendix. It is noted that the field at the cathode, in response to the boundary conditions, is qualitatively different than that associated with Figs. 10–20. Here, the field decreases from the cathode to the anode. In the vicinity of the anode the field is uniform. The net decrease in field is consistent with a cathode region partially depleted of carriers. Note that prior to reaching the downstream portion of the structure, the field displays a minimum followed by a change in slope. This change in slope represents the presence of a region of local charge accumulation.

within the cathode region. With this change

$$t(X) = \tau_0 \log\left\{\left(1 - \frac{n_0\mu_1}{\beta n_c \mu_c}\right) \Big/ \left(1 - \frac{n_0\mu_1 F(X)}{\beta n_c \mu_c F_c}\right)\right\}. \tag{51}$$

The requirement that the transit time be positive leads to the inequality

$$\beta n_c \mu_c < n_0 \mu_1, \qquad F(X) > F_c, \tag{52}$$

for a restricted range of field values. Equation (52) is represented in detail below and shows the presence of cathode depletion.

Figure 21 displays the field distribution, whose most obvious characteristic is that of a broad depletion region adjacent to the cathode. The characteristics of this depletion region are that with increasing bias, the depletion zone broadens, the cathode field increases, and the downstream field begins to approach a constant value. This latter feature manifests itself as hard saturation in current versus voltage.

Figure 22 is a display of the carrier density in the Γ valley as well as total carrier density. It is first noted that for all of the bias values chosen, the Γ-valley carrier density displays partial depletion in the vicinity of the cath-

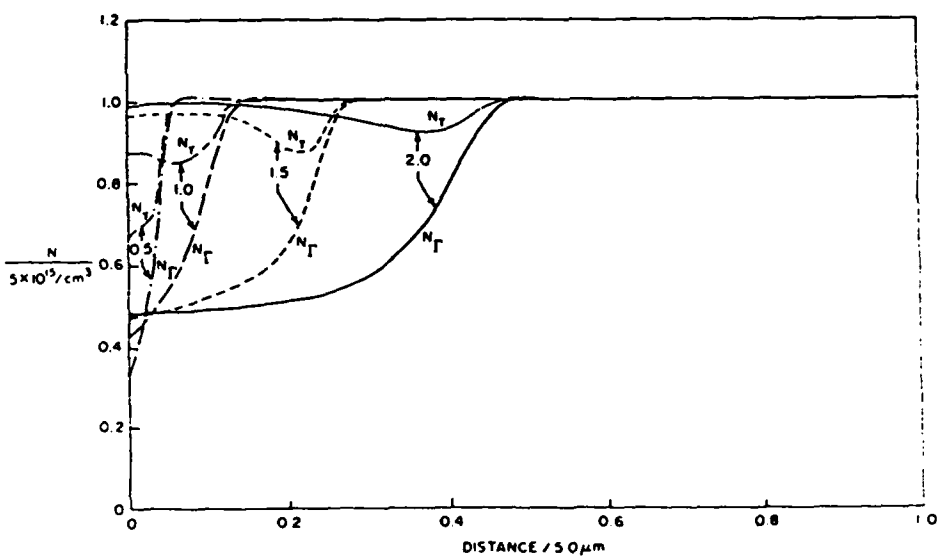


Fig. 22. Distribution of total (N_T) and Γ-valley (N_Γ) carriers within the device for the parameter of Fig. 21. Note that there is a net depletion of total charge within the vicinity of the cathode and that this partial depletion is reduced as the bias level is raised. For all bias levels the Γ-valley population is below that of the total carrier density.

ode boundary. This, it may be anticipated, will manifest itself as an excess carrier velocity at the cathode (Fig. 23). It is also noted that as the bias level increases, the total charge at the cathode shows a diminished depletion, while downstream there is a weak region of charge accumulation. With regard to the Γ-valley velocity, this follows the pattern dictated by current continuity and cathode adjacent charge depletion. The carrier velocity at the cathode sustains values in excess of that within the neutral interior regions of the semiconductor.

Figure 24 displays the dc current–voltage relationship for this structure. Several points are noteworthy. The first point is the absence of negative differential conductivity even though the neutral interior region is characterized by a region of negative differential mobility. The second point to note is that current saturation occurs at values below that associated with the 1-μm-long device.

The cathode current–field relationship is displayed against the neutral field characteristic in Fig. 25. In addition, the cathode boundary neutral field characteristic is also shown. The curves display an apparent tendency to intersect within the region of negative differential mobility, resulting in two approximately neutral regions sustaining different values of field and velocity. Under a well-defined set of conditions, this configuration is electrically unstable and leads to the nucleation and propagation of high-field domains. For the configuration under consideration, an increase in bias level from 2.0

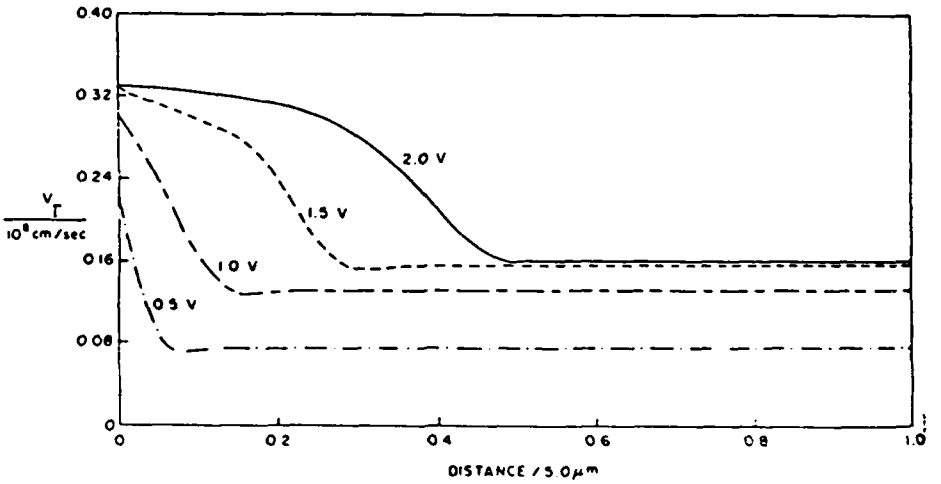


Fig. 23. Steady state velocity distribution of Γ-valley carriers for the parameters of Fig. 21. Note that unlike the velocity distribution for the Γ-valley electrons of an injecting contact where the carrier velocity is greatest at the anode, for this length structure the Γ-valley velocity is greatest at the cathode. Note further that the change in cathode velocity with increased bias is very small at high bias levels and reflects the presence of current saturation.

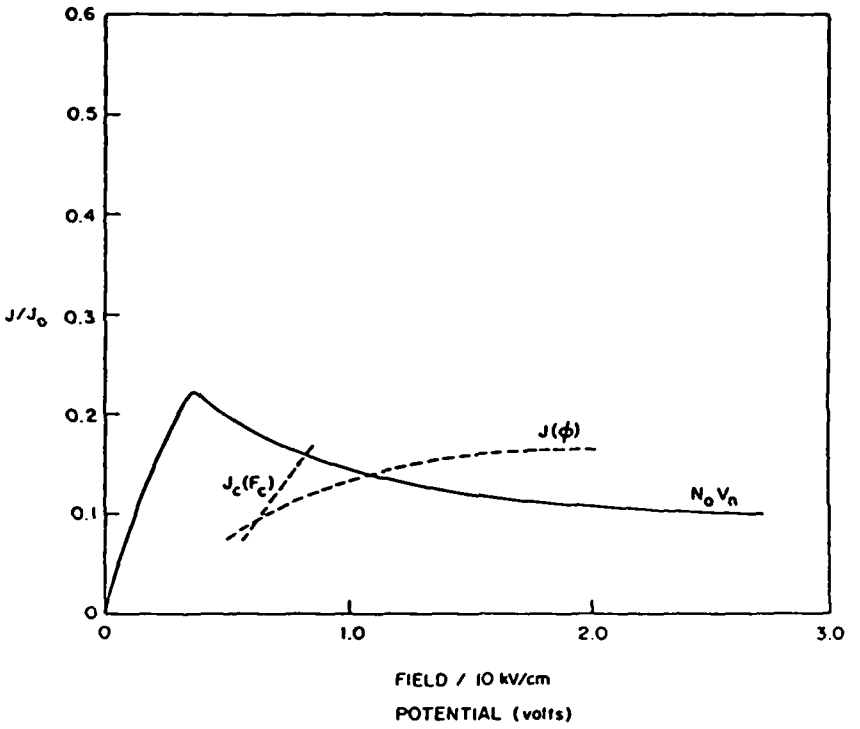


Fig. 24. Steady state current voltage characteristics $J(\phi)$ for the partially depleted cathode structure with a length of 5.00 μm. Parameters are those of Fig. 21. Of significance here is the fact that saturation in current occurs at an average field significantly below that of the 1.0-μm device and that the current in saturation is approximately $\frac{1}{2}$ that of the 1.0-μm-long device. Also shown is the cathode current field relation $J_c(F_c)$ and the neutral field characteristic N_0V_0; $J_0 = 8 \times 10^4$ A/cm^2.

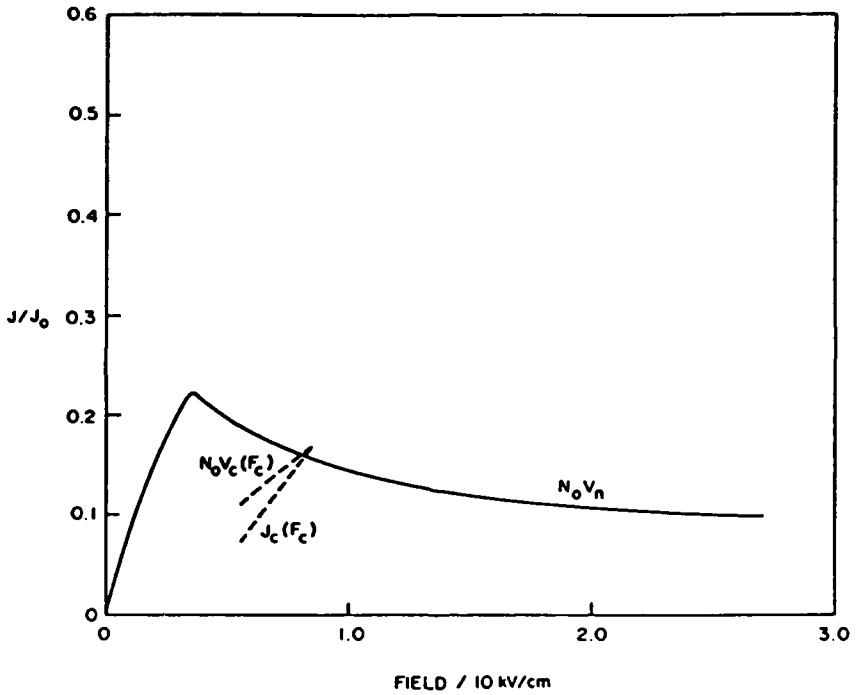


Fig. 25. Neutral current field relationship for gallium arsenide n_0V_n, the cathode current field relation n_cV_c, and the neutral cathode field relation n_0V_c. Parameters are those of Fig. 21; $J_0 = 8 \times 10^4$ A/cm^2

to 3.0 V results in transient local cathode adjacent accumulation and subsequent dipolar propagation as displayed in Fig. 26. The details of Fig. 26 show the transient transformation of the space charge layer (as reflected in the electric field distribution) from a depletion layer to a dipole propagating layer. The dipole layer is quenched at the anode boundary and repeated transit time oscillations occur. The time-dependent oscillations are displayed in Fig. 27 and occur after an initial transient that is qualitatively similar in structure to that associated with the accumulation cathode and the uniform field transients. Indeed, the peak current is both a reflection of overshoot and the influence of the cathode boundary condition which reduces its value to a level below that of the uniform field transient. It is important to note that while the development of a set of conditions for initiating a propagating domain is of clear technological significance, it plays a secondary role to the thrust of this chapter, which is that conditions at the cathode are likely to be the single most pervasive influence on near- and submicrometer-length semiconductor devices, much as they are for longer devices.

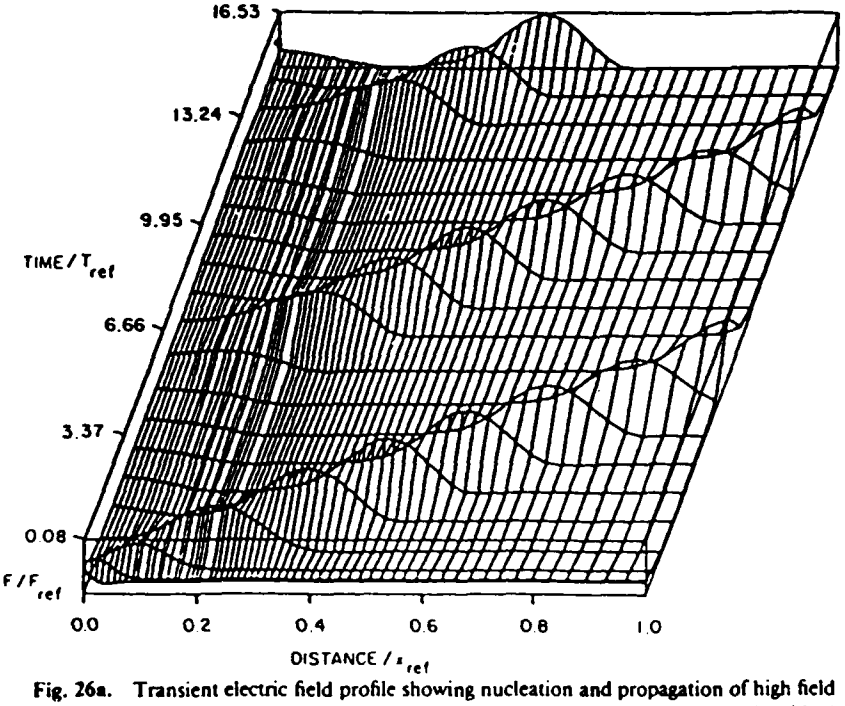


Fig. 26a. Transient electric field profile showing nucleation and propagation of high field domain. Note that propagation is accompanied by low downstream field values and residual cathode adjacent depletion. Parameters are those of Fig. 21; $F_{ref} = 2$ kV/cm; $T_{ref} = 5$ psec, $x_{ref} = 5$ μm, and $\psi_{ref} = 3$ V.

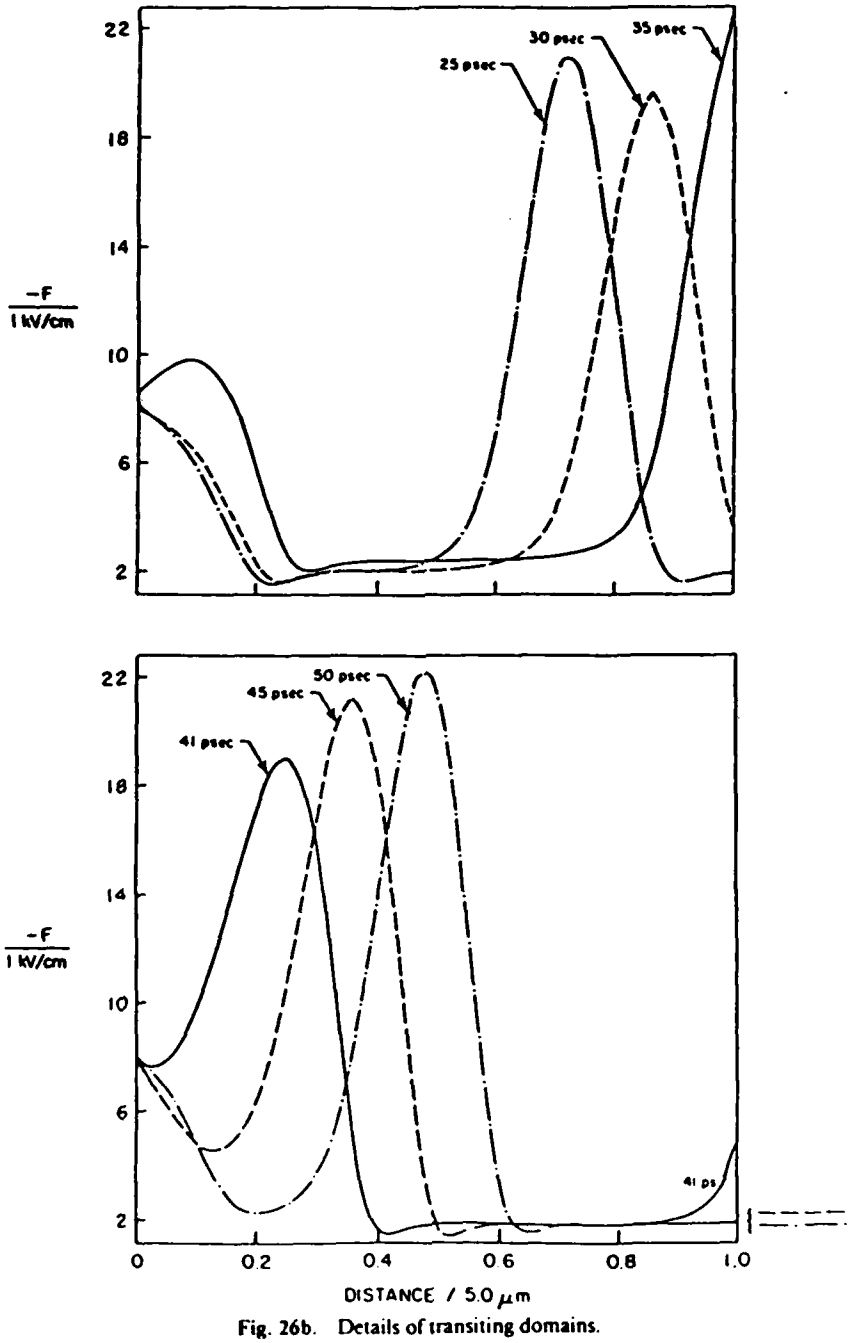


Fig. 26b. Details of transiting domains.

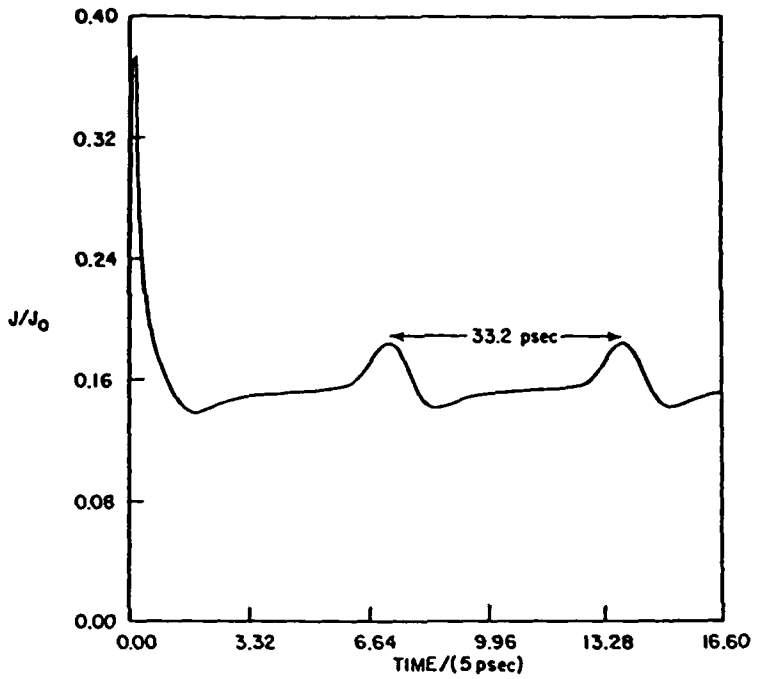


Fig. 27. Current transient following application of a step change in potential to 3.0 V. Initial transient is a reflection of nonequilibrium transient transport. The steady long-time transient reflects the nucleation, propagation, and quenching of a propagating high-field dipole layer. Parameters are those of Fig. 21; $J_0 = 8 \times 10^4$ A/cm^2, $\psi = 3$ V.

F. Nonuniform Fields and Length Scaling

While the calculations in Subsections III.D and III.E were for structures of different length, the emphasis was on the effects of the boundary. However, the effects of length scaling, insofar as they affect the velocity field relationship, (vis-a-vis Fig. 8) will influence the electrical transient and the steady state field profiles. This is illustrated for two situations. The first situation is for a uniformly doped structure with the same boundary conditions as given by Eq. (40) but with a length of 0.25 μm. The second structure considered is that of a $n^+ - n^- - n^+$ device, with a 1-μm cathode-to-anode spacing but with a variable-length n^- region.

The calculations for the 0.25-μm-long device are displayed in Figs. 28 - 36. The steady state electric field distribution is displayed in Fig. 28 for the indicated bias levels. Note that although the average fields for the 0.25-μm device and the 1.0-μm device are the same, the field distributions are quantitatively different. The difference lies in the fact that at the lower bias levels

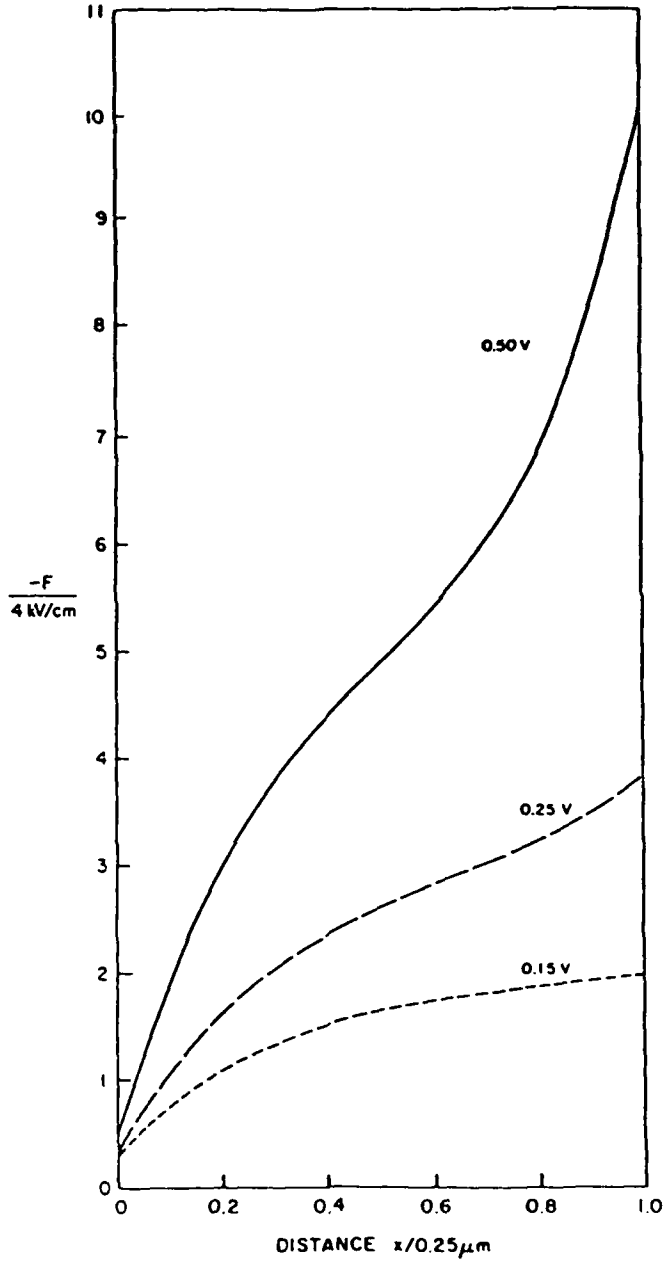


Fig. 28. Steady state distribution of electric field for 0.25-μm uniform structure with injecting cathode contacts. Average field across structure is the same as that of Fig. 10 for 1.0-μm-long device. Parameters for this calculation are listed in the appendix.

only a marginal amount of electron transfer occurs within the shorter structure. Note also that the electric field at the cathode is low, as for the 1-μm-long device.

Figure 29 displays the steady state population of the Γ-valley as well as the total carrier density. The first point to note here is that the density of carriers for the given bias level exceeds that for the 1-μm-long element. The second point is that considerably less electron transfer occurs downstream from the cathode. There is, however, a far more significant aspect to the quantitative differences between the results of the 0.25- and 1.0-μm devices. The carrier and velocity distributions for the two structures are different. These differences are, in part, a result of the fact that conditions at the upstream bound-

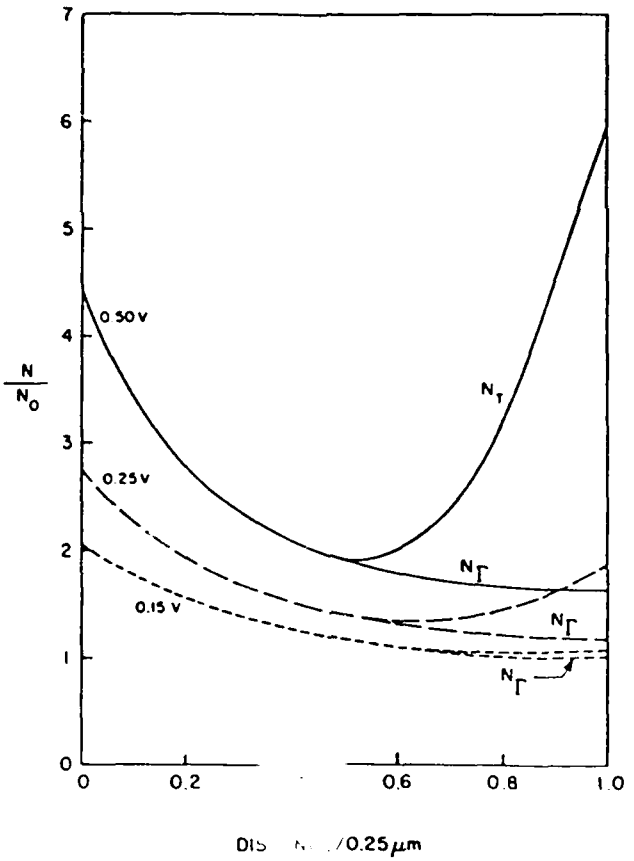


Fig. 29. Distribution of total (N_T) and Γ-valley (N_Γ) carrier density. Only marginal transfer occurs for the lowest bias level. Substantial transfer occurs at the higher bias levels. At all bias levels, injection level is extremely high, and N_Γ exceeds N_0. ($N_0 = 5 \times 10^{15}/cm^3$.)

ary are sensitively dependent upon the proximity of the collecting contact. Further evidence for this is provided by the velocity distribution displayed in Fig. 30 which shows higher entrance velocities, but lower exit velocities.

Figure 31 is a plot of current versus voltage for the 0.25-μm-long device. Again, two points are emphasized. The first point shows the absence of any negative differential conductance. The second point is that the presence of increased levels of charge injection yield an increase in the drive current over that of the 1-μm-long device.

The transient characteristics at 0.25 μm are displayed in Figs. 32–36. The results are quantitatively different from that associated with the 1-μm device. The first difference is displayed in the current transient (Fig. 32) which shows a higher peak current and a smaller current dropback. As revealed in the time dependent distributions of field (Fig. 33) the higher peak current (greater than 25%) is in large part due to displacement current contributions. The higher long-time steady state current level reflects the increased injection level (Figs. 34 and 35) over that of the 1.0-μm calculation. In this regard, it is again pointed out that the exit velocity for the 0.25-μm structure is below that for 1.0 μm (Fig. 36). The final point of interest involves the time to

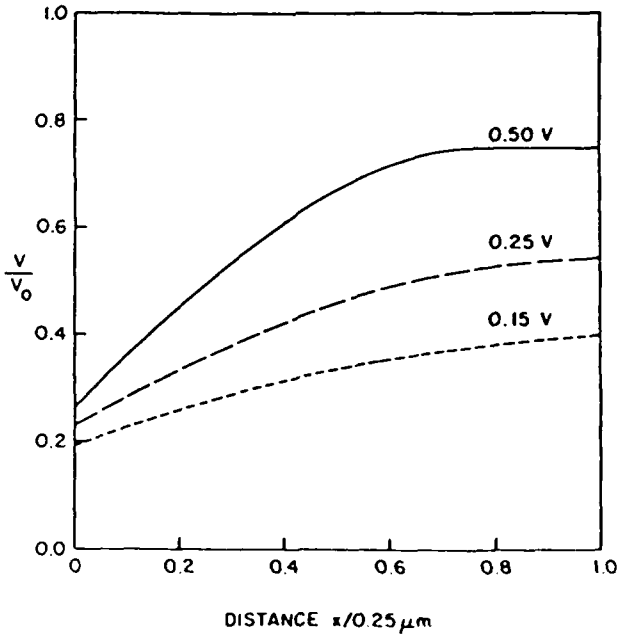


Fig. 30. Steady state distribution of the Γ-valley carrier velocity at three values of bias. Note that the velocity increases from cathode to anode, corresponding to a decrease in the Γ valley carrier density. Parameters are those of Fig. 28; $V_0 = 10^8$ cm/sec.

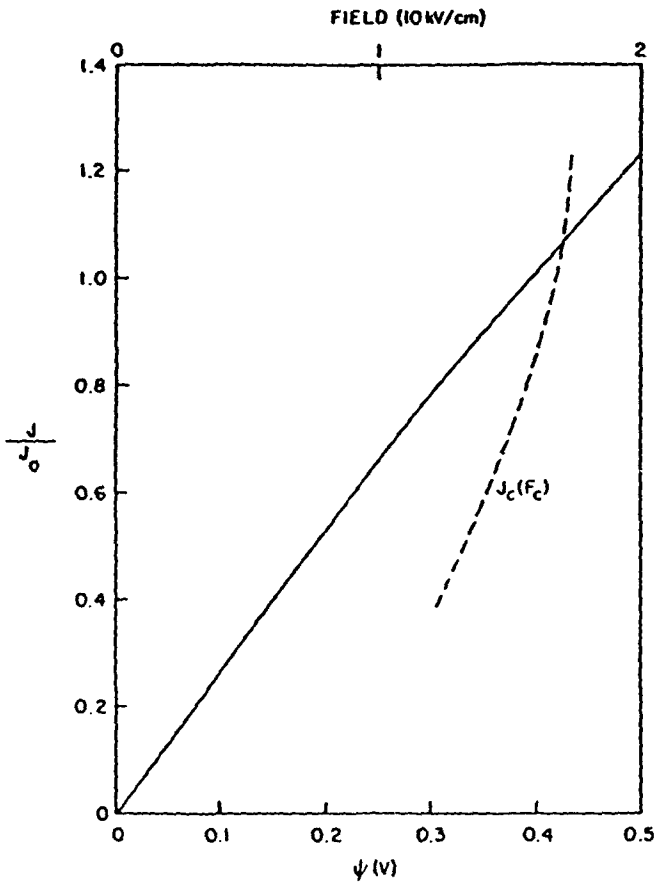


Fig. 31. Steady state current density versus applied voltage and average field for a 0.25-μm-long structure with injecting contacts. Excessively high current levels are primarily due to high levels of space charge injection. Also shown is the cathode current field relationship $J_c(F_c)$. Parameters are those of Fig. 28; $J_0 = 8 \times 10^4$ A/cm^2.

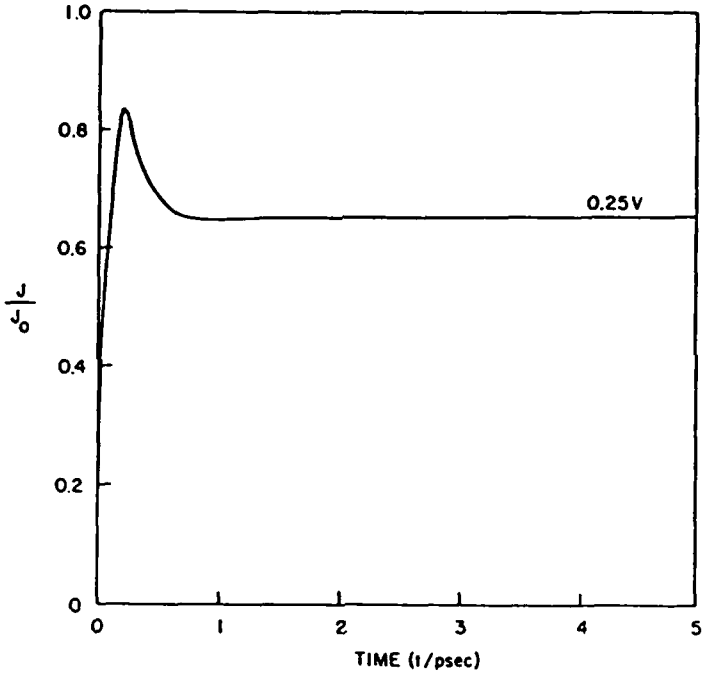


Fig. 32. Magnitude of current transient following application of a step change in potential for a 0.25-μm-long device with injecting contacts. Current peak exceeds that of both the uniform field structure and the 1.0-μm device with injecting contacts. Steady state current level is above that of the 1.0-μm device with injecting contacts. The time to relaxation is 4.9 psec, which is approximately 40% less than that of the 1.0-μm device with the same average field. Given the fact that the structure is 0.25 μm in length, this result provides evidence that the relaxation effects are influenced by nontransit time effects. Parameters are those of Fig. 28; $J_0 = 8 \times 10^4$ A/cm^2.

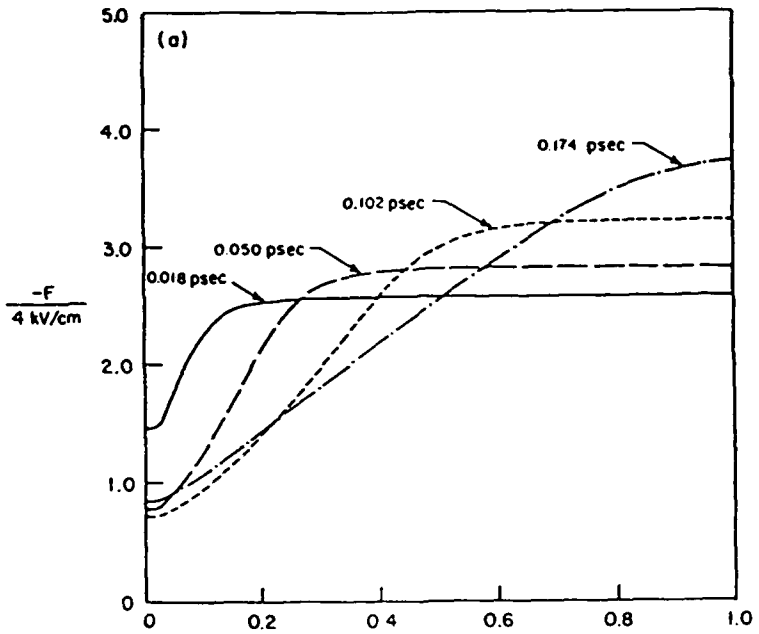


Fig. 33. (a) and (b) Time-dependent evolution of electric field distribution for the 0.25-μm device subject to a step change in bias of 0.25 V. During the first 0.1 psec, the field propagates downstream from the cathode, indicating a propagating accumulation layer. The field downstream from the cathode is relatively uniform. To satisfy the constraints of constant voltage across the device there are displacement current contributions at the bottom half of the structure that account for much of the difference in the peak currents associated with the 0.25-μm and 1.0-μm devices. Parameters are those of Fig. 32. *(Continued on next page.)*

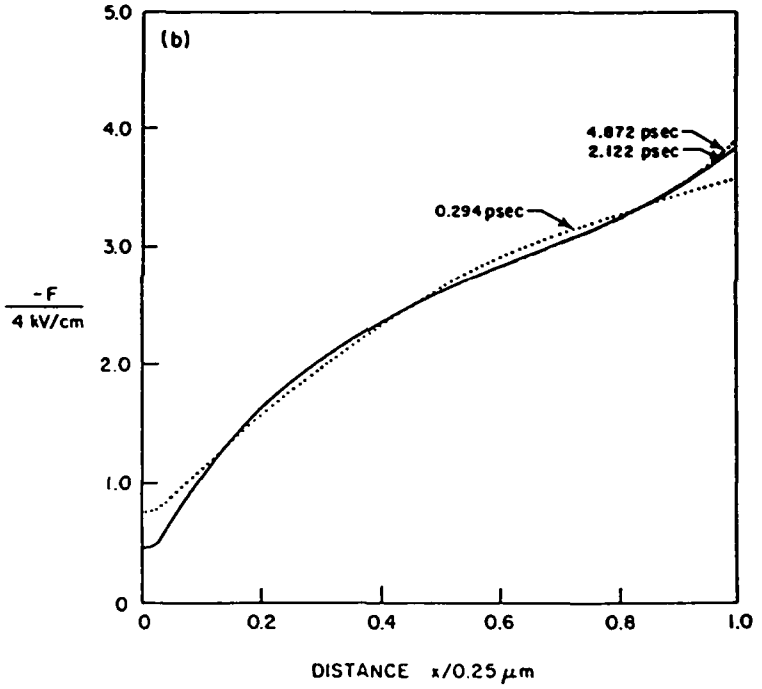


Fig. 33. *(Continued)*

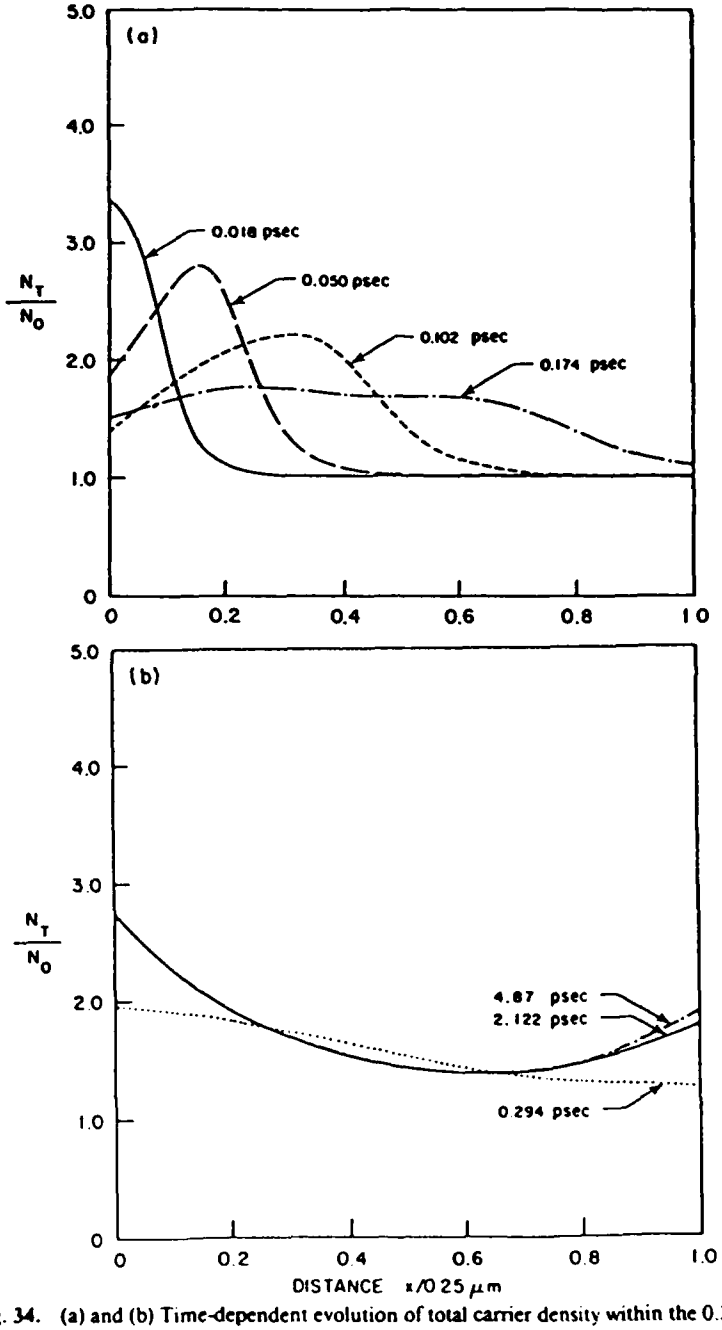


Fig. 34. (a) and (b) Time-dependent evolution of total carrier density within the 0.25-μm device. Initial propagation characteristics are similar to those of the 1.0-μm devices. Proximity effects are introduced after 0.1 psec and differences in the 0.25-μm and 1.0-μm calculations arise. Parameters are those of Fig. 32; $N_0 = 5 \times 10^{15}/cm^3$.

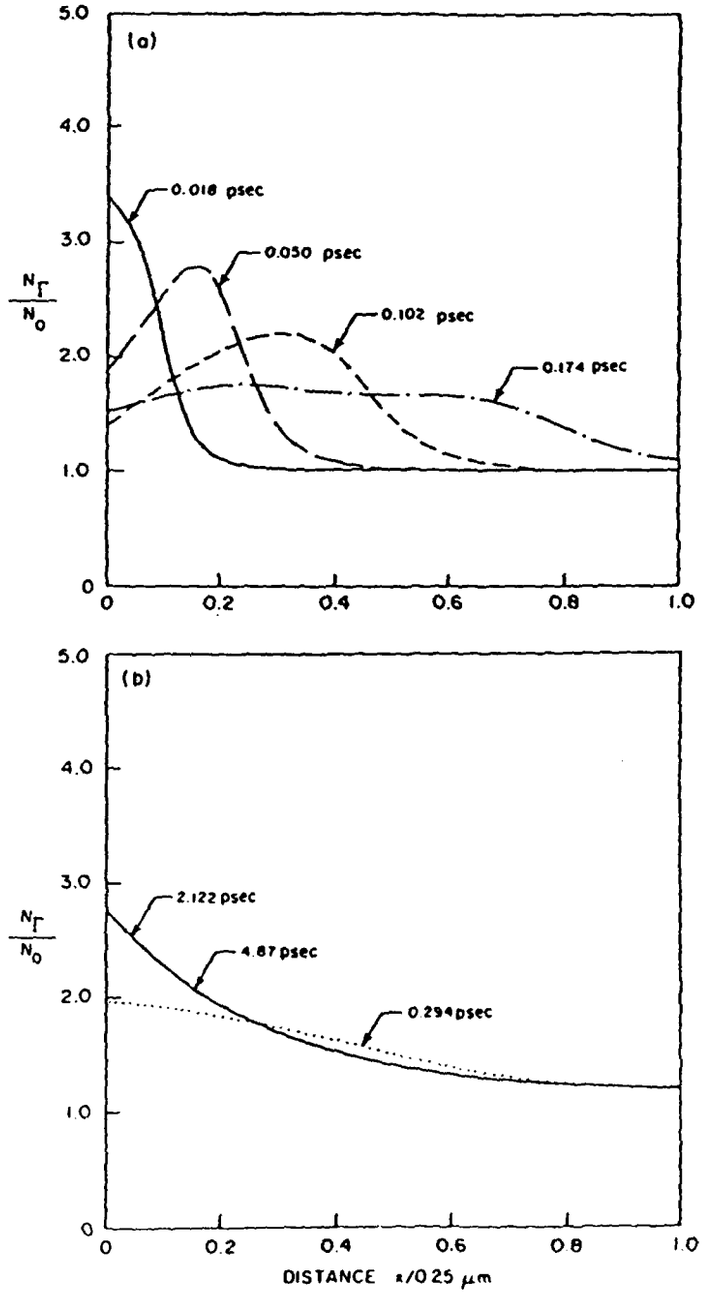


Fig. 35. (a) and (b) Time-dependent evolution of the Γ valley for the parameters of Fig. 32. Note the electron transfer after 2 psec.

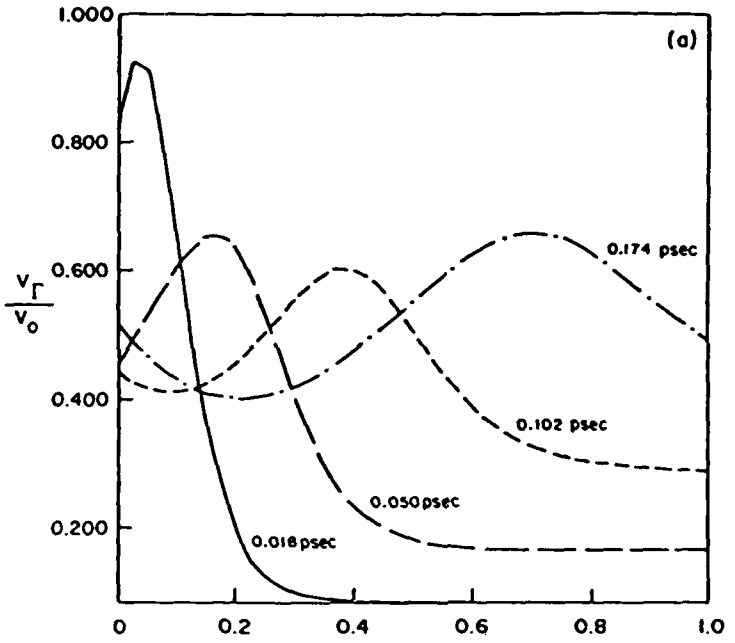


Fig. 36. (a) and (b) Distribution of Γ-valley carrier velocity following application of a step change in bias of 0.25 V. The initial velocity distribution is similar to that found in the 1.0-μm transient study. Downstream velocity values during the first 0.1 psec are higher than that of the 1.0-μm calculation and correspond in part to the presence of slightly higher downstream fields. Parameters are those of Fig. 32; $V_0 = 10^8$ cm/sec. *(Continued on next page.)*

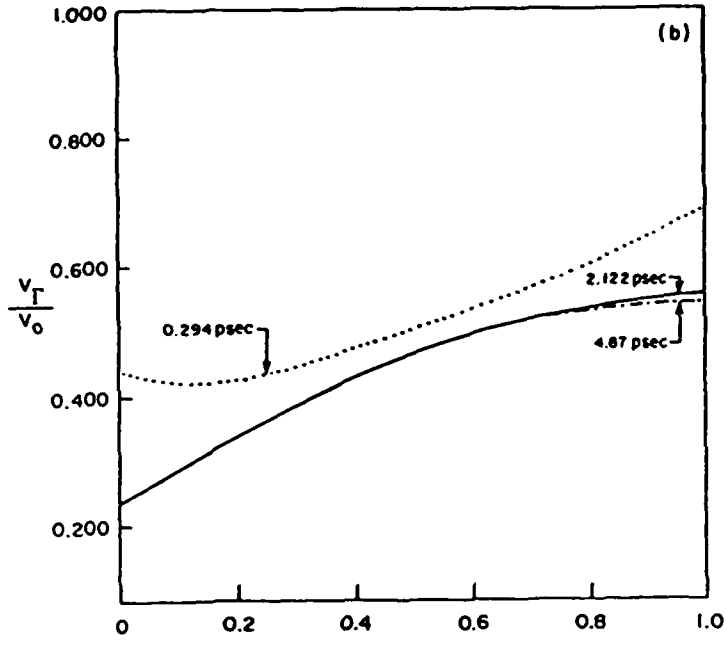


Fig. 36. *(Continued)*

steady state: This time is shorter for the 0.25-μm electron, but not a factor of four shorter. The time scales involved in the approach to steady state involves nontransit-time contributions.

G. Transients in $n^+-n^-n^+$ Structures and Length Scaling

The final two-terminal structure considered is the $n^+-n^-n^+$ device, and there are several key features to note. The first is that the dominant interfaces for this structure, the n^+-n^- and n^-n^+ interfaces, are not the physical boundaries of the device and are thus likely to have a different effect on the electrical behavior of the device. The second feature of importance lies in the fact that the electric field profile is highly nonuniform in the steady state and may dominate the transient and completely camouflage all submicrometer effects. Third, for a sufficiently small n^- regions, the influence of the n^+-n^- and n^-n^+ interfaces for carrier confinement may be less prominent. Thus, this last two-terminal structure offers the most serious example of the interplay of the interface and the length of the critical submicrometer region on

the electrical characteristics of the submicrometer structures. To avoid conflict with the influence of the true metal confining contacts, the physical boundary conditions at the cathode were taken as

$$n = n_0, \quad n_1 = n_{1eq}, \quad V_{1x} = 0, \quad V_{2x} = 0, \quad T_1 = 300\ \text{K}, \quad T_{2x} = 0. \tag{53}$$

At the anode all second derivatives were set to zero.

The $n^+-n^--n^+$ calculations performed were for the one-dimensional structure of Fig. 37, in which the n^- region was assigned a nominal doping level of $10^{15}/\text{cm}^3$ and the n^+ region was at $10^{17}/\text{cm}^3$. The length of the n^-

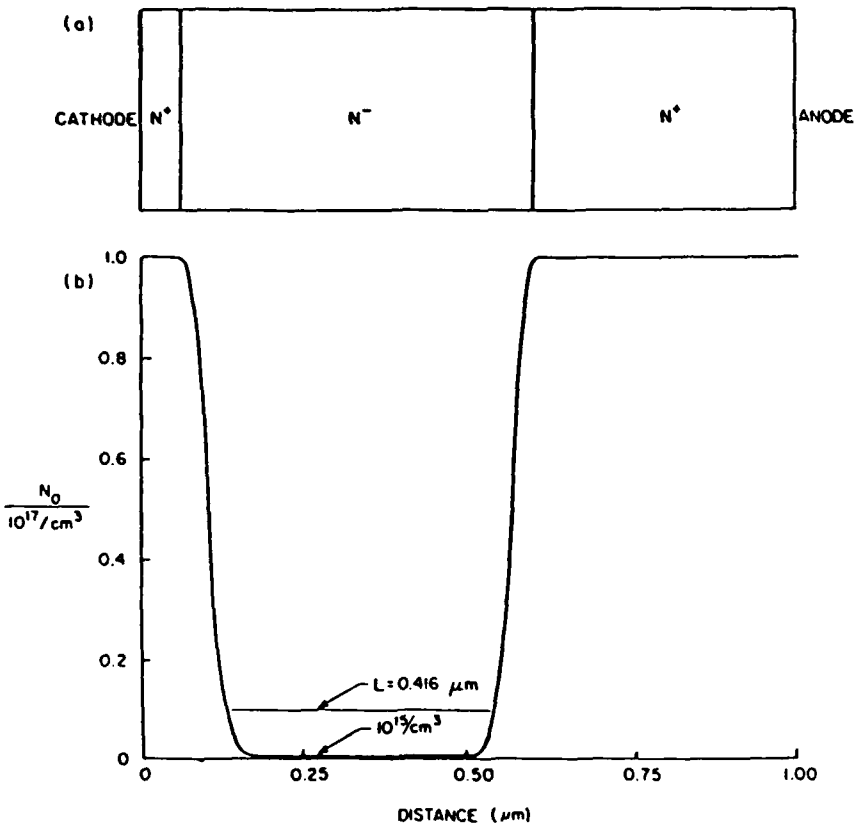


Fig. 37. (a) One-dimensional structure used for calculations. (b) Donor distribution of the $n^+-n^--n^+$ structure used in the study. In the calculations, the width of the n^- region (defined at a donor level of $10^{16}/\text{cm}^3$) varied from 0.416 to 0.116 μm. In all calculations the width of the upstream n^+ region was unchanged.

region is specified at the doping level of $10^{16}/cm^3$, and varied from 0.416 to 0.116 μm. The entire structure was fixed at a length of 1.0 μm. The design of the structure dictates that nonuniform fields and charge densities form within it. Thus, again the relevant experimental quantity is current density, rather than velocity. The first set of results is shown in Fig. 38. Figure 38 displays the total current flowing through the device following application of a voltage pulse of magnitude 1.0 V.

As in the uniform N_0 studies, the calculation is performed in two stages. The first involves obtaining a steady state solution at 0.01 V. For the second, using this as an initial condition the bias is raised, in one time step, to 1.0 V. Application of the bias in one time step replicates the procedure of most of the uniform field calculations.

As seen in Fig. 38, the current displays an initial peak at approximately 0.15 psec, followed by a drop in current and a subsequent rise toward a

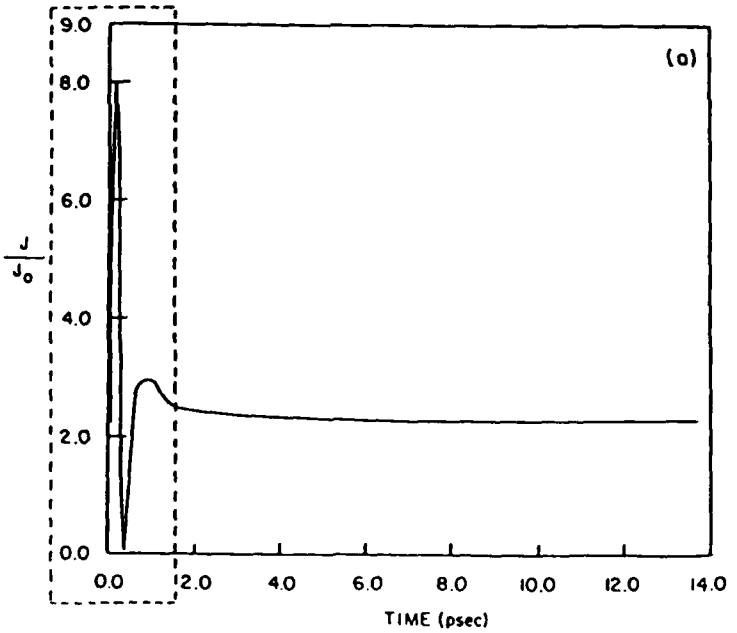


Fig. 38. (a) Time-dependent current following application of a step change in bias to 1.0 V for the $n^+-n^--n^+$ structure with an n^- region of 0.416 μm. The structure of the current profile displays significant quantitative differences from that of the uniform donor calculations. First, the peak in the current occurs within 0.10 psec, which is below that of the uniform donor calculations. Second, there is a strong current minimum, followed by relaxation. Steady state requires approximately 15 psec. Parameters for the calculation are listed in the appendix. (b) Magnification of dotted area. (From Grubin and Kreskovsky [13].)

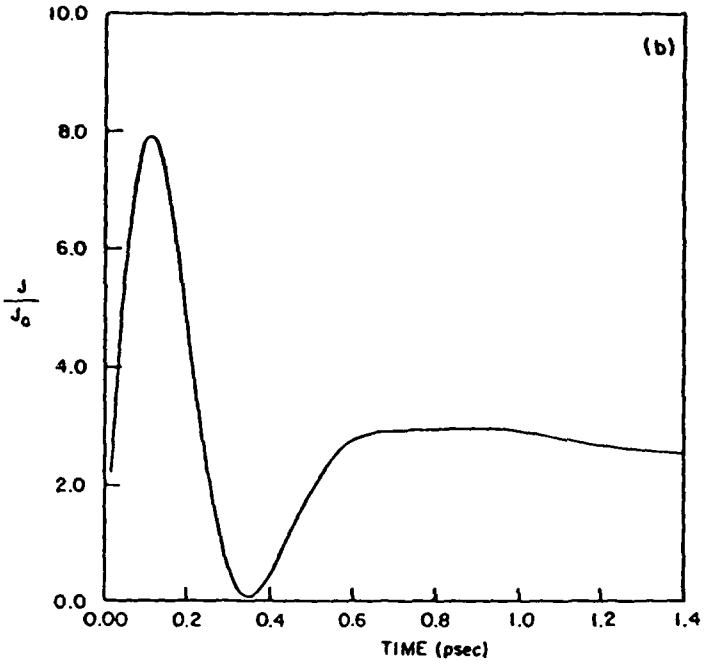


Fig. 38. *(Continued)*

steady state value. For uniform field calculations in which the voltage is increased in one time step as discussed earlier, there is an initial displacement current whose magnitude is determined entirely by the computational time step. Thereafter, all displacement currents are zero and all transients are particle current transients. (Note that with a load line, displacement currents would exist.) The situation with the nonuniform field calculation and displacement current contributions is different. Figure 39 displays the particle current through the device at select instants of time. A comparison of the magnitude of the particle and total current indicates that within certain key regions of the device (particularly near the n^+-n^- and n^--n^+ interface regions) that the displacement current dominates the current level. The general conclusion of this calculation is that since the initial transient is strongly influenced by displacement current contributions it would be inappropriate to assume that the initial current transient is a measure of velocity overshoot.

The details of the transient, specifically as it relates to displacement current contributions, are reflected in the time dependence of the electric field and potential profiles (Figs. 40 and 41) and the spatially dependent charge

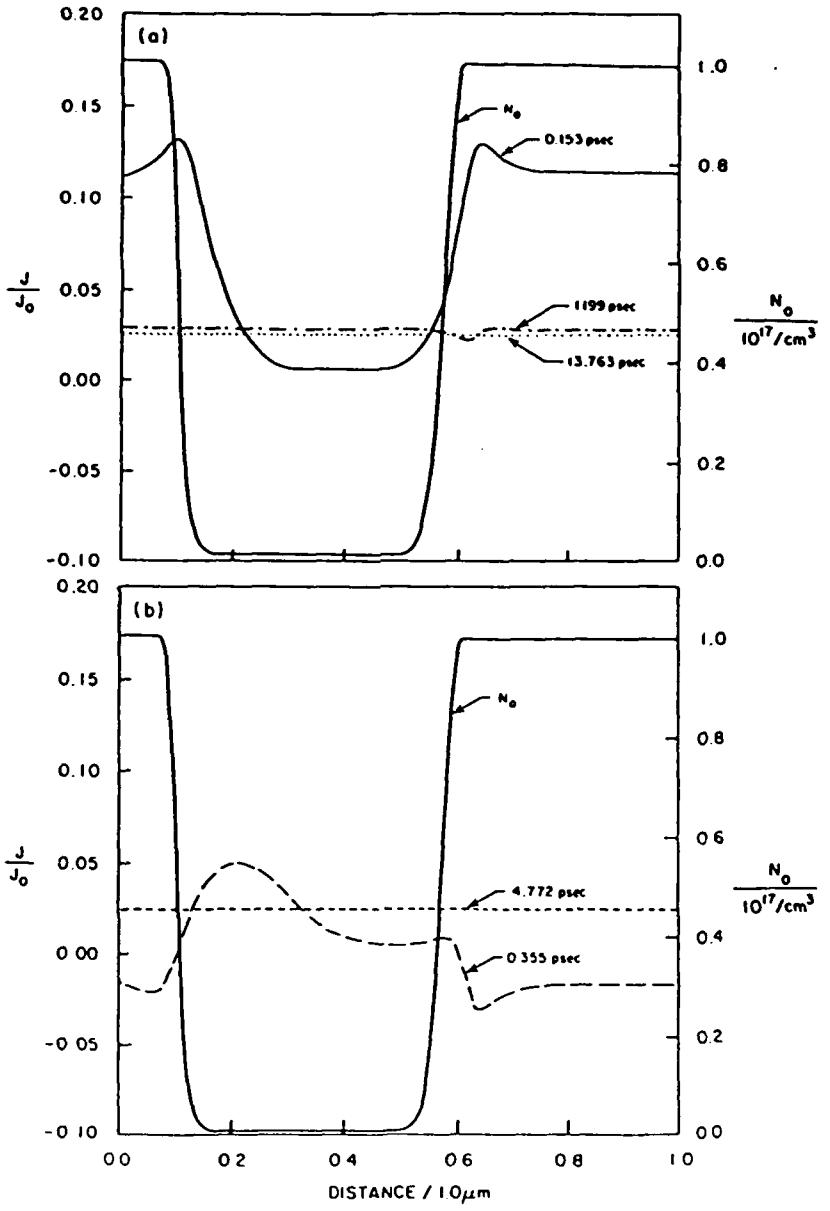


Fig. 39. (a) and (b) Spatial distribution of particle current at different instants of time for the parameters of Fig. 38. Also shown is the donor distribution N_0. The largest spatial variation in particle current occurs near the interfacial boundaries.

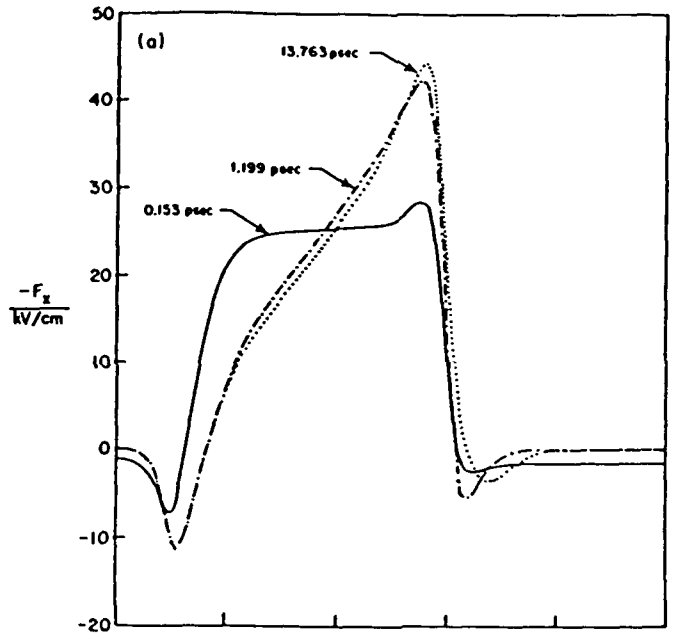


Fig. 40. (a) and (b) Distribution of electric field following application of a bias pulse. Note the strong temporal variation in field at the upstream interface and within the n^- region. The propagation characteristics associated with the electric field distribution under uniform donor conditions are camouflaged here by the spatial rearrangements within the interface region [13]. In addition, note the presence of the strong retarding field, one that is characteristic of $n^+-n^--n^+$ structures (see also Cook and Frey [14]). Parameters are as in Fig. 38. *(Continued on next page)*

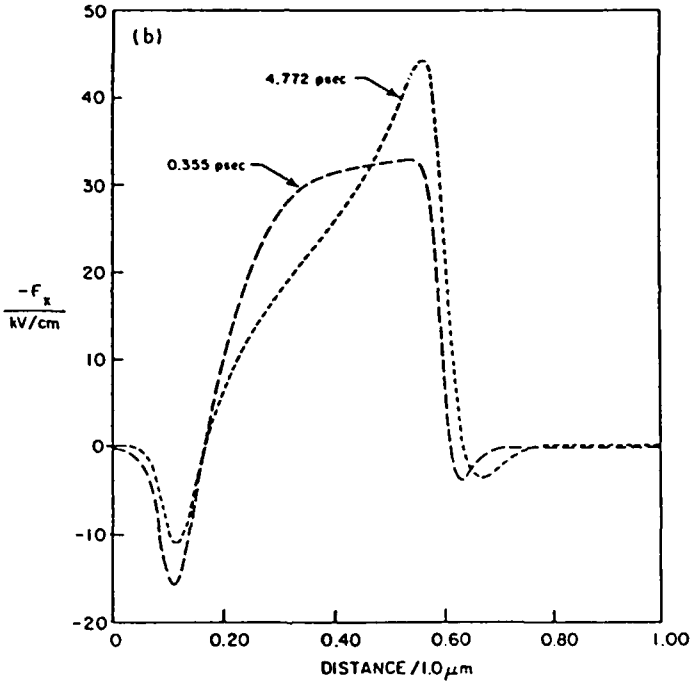


Fig. 40. *(Continued)*

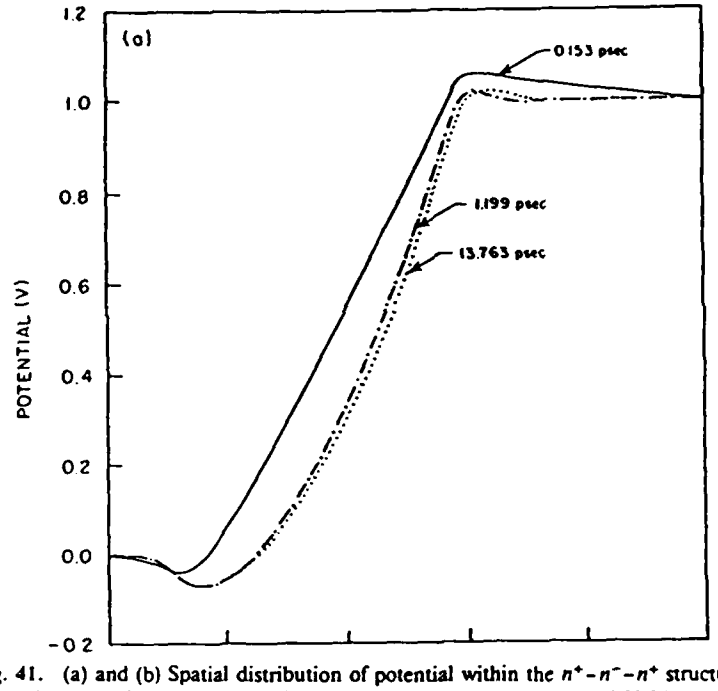


Fig. 41. (a) and (b) Spatial distribution of potential within the $n^+-n^-n^+$ structures at different instants of time. Note that in the steady state approximately 1.0 V falls across the 0.41-μm-long region. It may be anticipated that this will lead to large-scale injection into the n^- region. Parameters are as in Fig. 38; $L = 0.416\ \mu$m. *(Continued on next page.)*

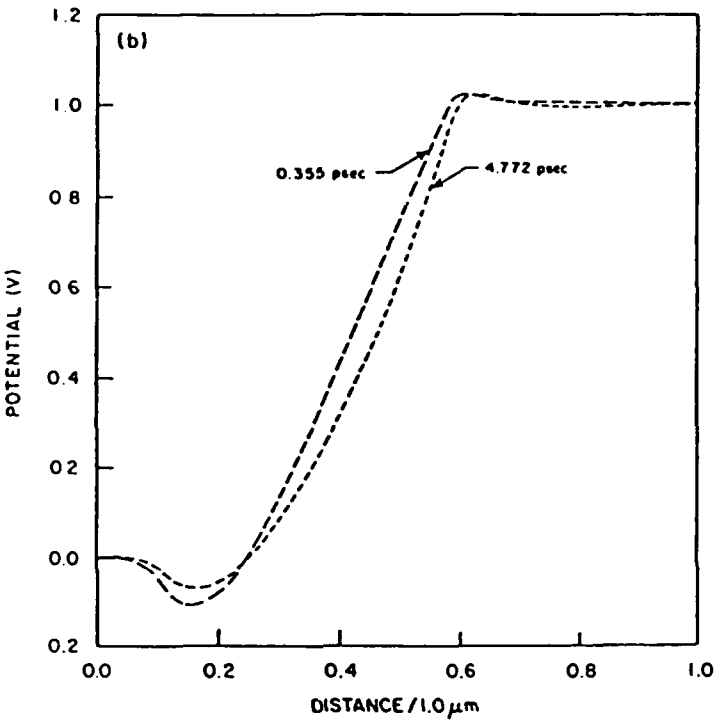


Fig. 41. *(Continued)*

density profiles (Figs. 42 and 43). It is noted, however, that as in the uniform field calculations, immediately following the voltage step, the electric field increases everywhere by the ratio of the applied bias to the length of the structure (in this case 10 kV/cm). This initial increase introduces a displacement current whose magnitude does not correctly represent the physical transient but rather the impulsive change in the applied potential over a single, small but finite time step. Physically accurate calculations follow the initial time step and are discussed below.

Prior to the application of the step potential a retarding field is formed at the upstream $n^+ - n^-$ interface limiting further injection of space charge into the n^- region. This retarding field, which at its maximum is positive in sign, is significantly reduced following application of the step potential; carrier injection into the n^- region is thereby resumed. Two events accompany this enhanced injection. First, to accommodate the increased charge within the n^- region Gauss's law dictates that the electric field within the region must become increasingly negative. Second, the space charge injection is self-limiting in that as the process of injection proceeds, the retarding field begins to

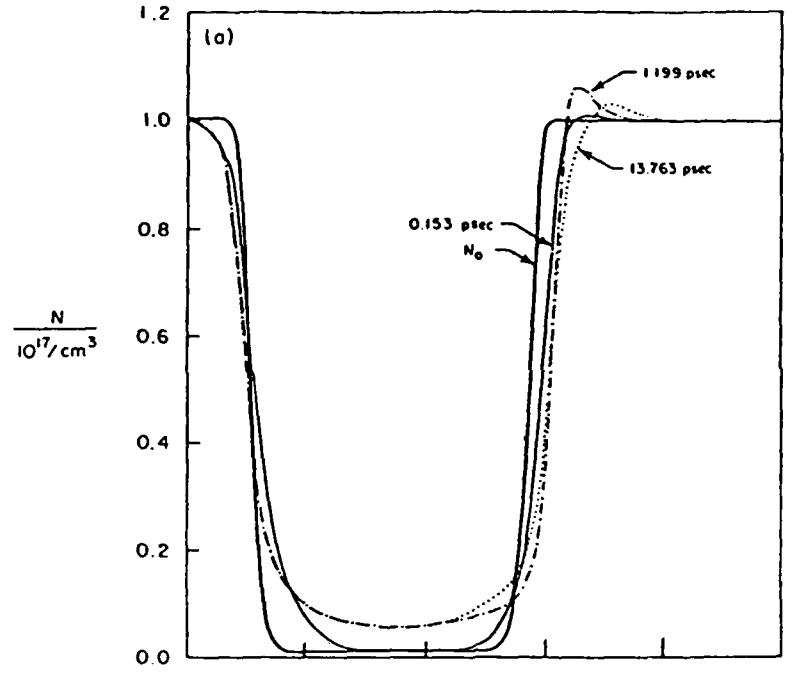


Fig. 42. (a) and (b) Time-dependent evolution of the total carrier concentration within the $n^+-n^--n^+$ structure [13]. During the early transient, the space charge layer displays a characteristic propagation downstream from the cathode, as seen in the uniform donor calculations. Insofar as the form of the electric field profile is controlled by differences between n and n_0, these propagation characteristics lose the distinction that emerges from the uniform donor calculations. The injection level is almost an order of magnitude higher than the n^- donor level. Parameters are as in Fig. 38. *(Continued on next page.)*

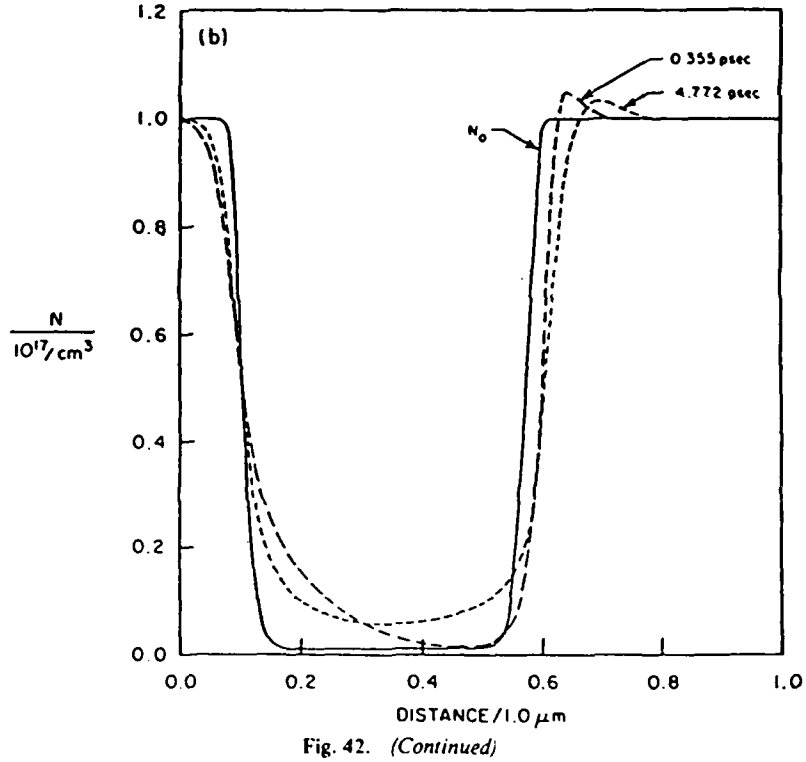


Fig. 42. *(Continued)*

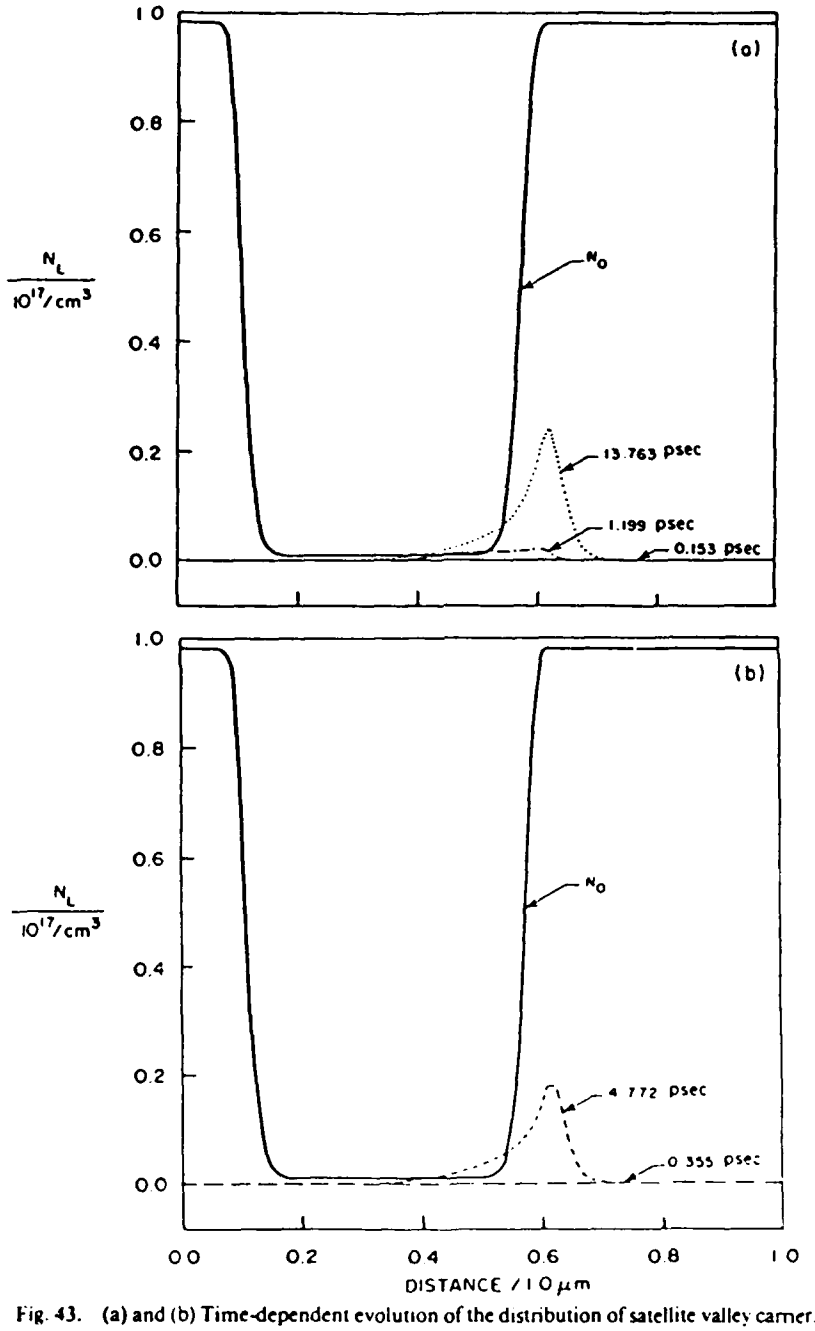


Fig. 43. (a) and (b) Time-dependent evolution of the distribution of satellite valley carrier. An inconsequential number of carriers is scattered into the satellite valley during the early transient ($t < 0.45$ psec). Electron transfer is apparent in the steady state. Parameters are as in Fig. 38.

reform and positive field values result at the upstream n^+-n^- interface. The process of an increased and positive retarding field accompanied by an increasingly negative field within the n^- region and the constraint of the fixed potential leads to the spatially dependent displacement currents inferred from Fig. 39. Note that in steady state a significant amount of injected charge resides at the downstream n^--n^+ interface and results in the presence of a downstream retarding field.

Notwithstanding the displacement current contributions, it is necessary to determine the extent to which the carrier velocity can exceed the equilibrium values. For the structure considered herein with $L_{n^-} = 0.416$ μm, most of the current is transported by the Γ-valley carriers. For this case, the mean carrier velocity thereby exceeds the steady state value. This is seen in Figs. 44 and 45, the latter displaying a plot of the Γ-valley temperature. The steady state distribution is qualitatively similar to that reported in a number of different studies [14–16]. In particular, the presence of a local cooling at the n^+-n^- interface is noted [14].

There are several noteworthy features associated with Fig. 44. First, there is the progressive movement of the velocity layer toward the downstream n^--n^+ interface. This migration is associated with the spatial and temporal derivatives on the left-hand side of Eq. (2). Second, there is a progressive decrease in the velocity in the n^+ regions as dictated by the decreasing field within these regions. Indeed, the possibility exists for the carriers to sustain a transient separation at the n^+-n^- interface, with carriers on the upstream portion of the interface moving toward the cathode and carriers at the downstream portion of the interface moving toward the downstream boundary. This separation is accompanied by compensating displacement current contributions.

The results clearly indicate the presence of velocity overshoot under nonuniform field conditions. Under uniform field conditions, the transient following the peak velocity is dominated by electron transfer. The question then becomes: do similar phenomena occur when nonuniform fields are present? Figure 46 is a plot of the time evolution of the total carrier density and the Γ-valley carrier density at two points within the $n^+-n^--n^+$ structure. It is seen that as in the uniform donor calculations at the upstream portion of the structure very little transfer occurs. Most of the transfer is at the downstream portion of the structure. One necessary conclusion as before is that the uniform field calculations bear little resemblance to the transients occurring in the $n^+-n^--n^+$ structure.

We next consider the dependence of the results on the length of the n^- region and note the expectation that the shorter the active region the higher the dc drive current [10]. For the $n^+-n^--n^+$ structure, as in the uniform N_0 structure with injecting contacts, a significant contribution to the current

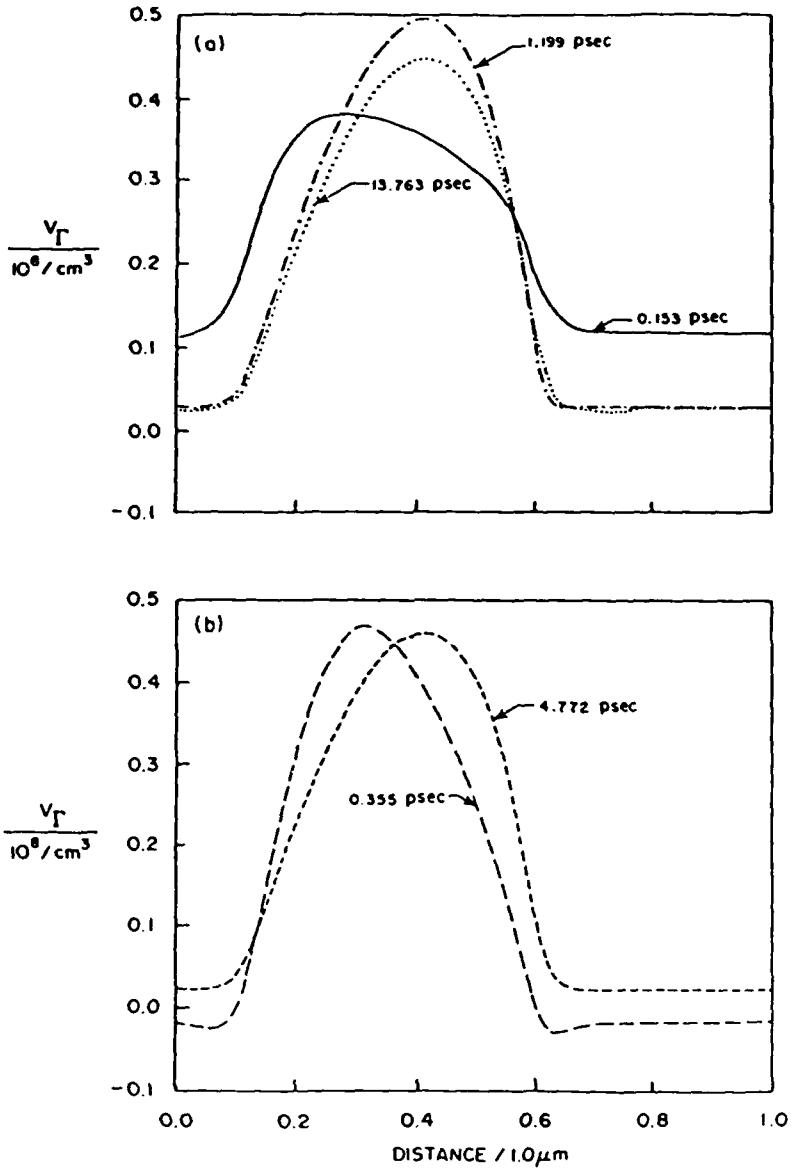


Fig. 44. (a) and (b) Time-dependent evolution of the spatial distribution of the Γ-valley carrier velocity. Note the initial propagation of the velocity layer downstream from the $n^+ - n^-$ interface. While the peak field for this calculation is approximately 44 kV/cm, the peak velocity is only near 5×10^7 cm/sec, a value considerably below that of the 0.25-μm-long uniform donor calculations. Also note that during the early time transient the carriers outside the n^- region drift in a direction opposite to carriers within the n^- region. Here, displacement current contributions dominate. Parameters are as in Fig. 38.

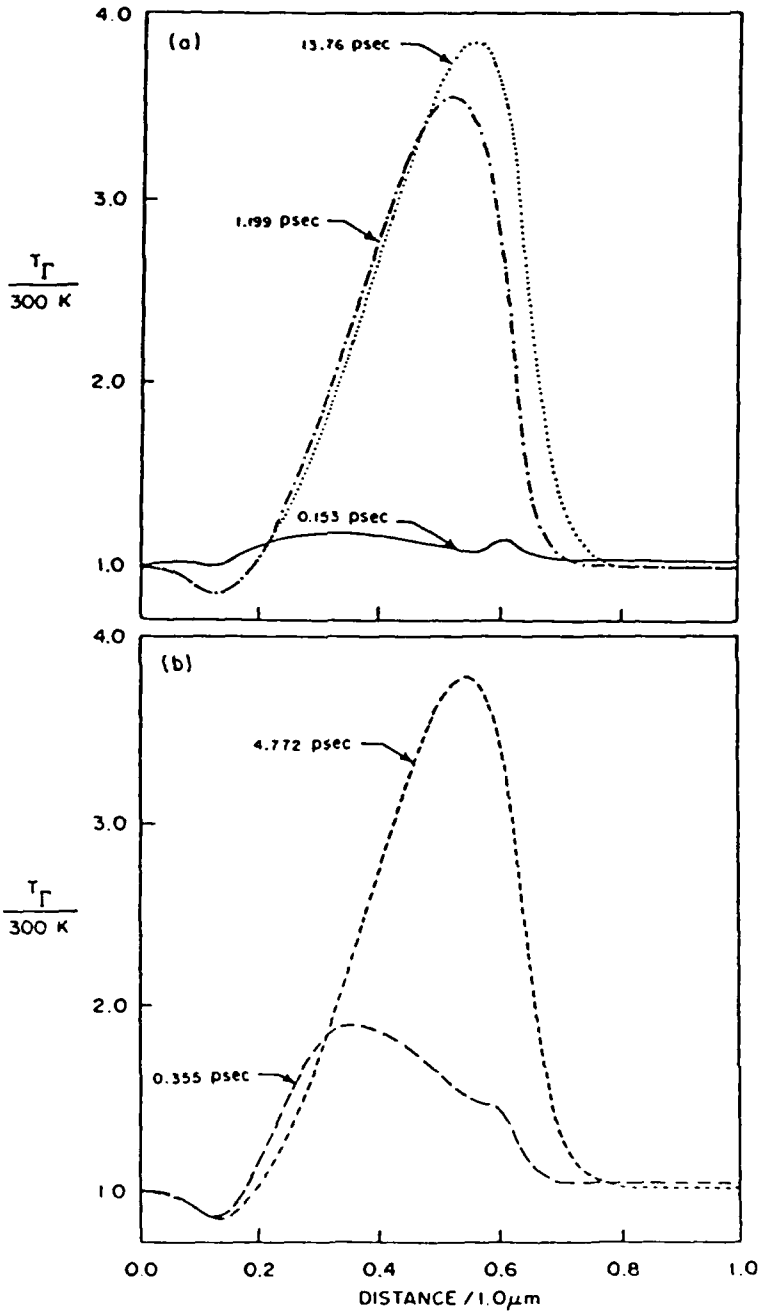


Fig. 45. (a) and (b) Time-dependent evolution of the spatial distribution of Γ-valley temperature within the $n^+-n^--n^+$ structure. Note that in steady state an apparent cooling of the carriers occurs within the region near the n^+-n^- interface where a large excess of carriers is present (see also Cook and Frey [14]). Parameters are as in Fig. 38.

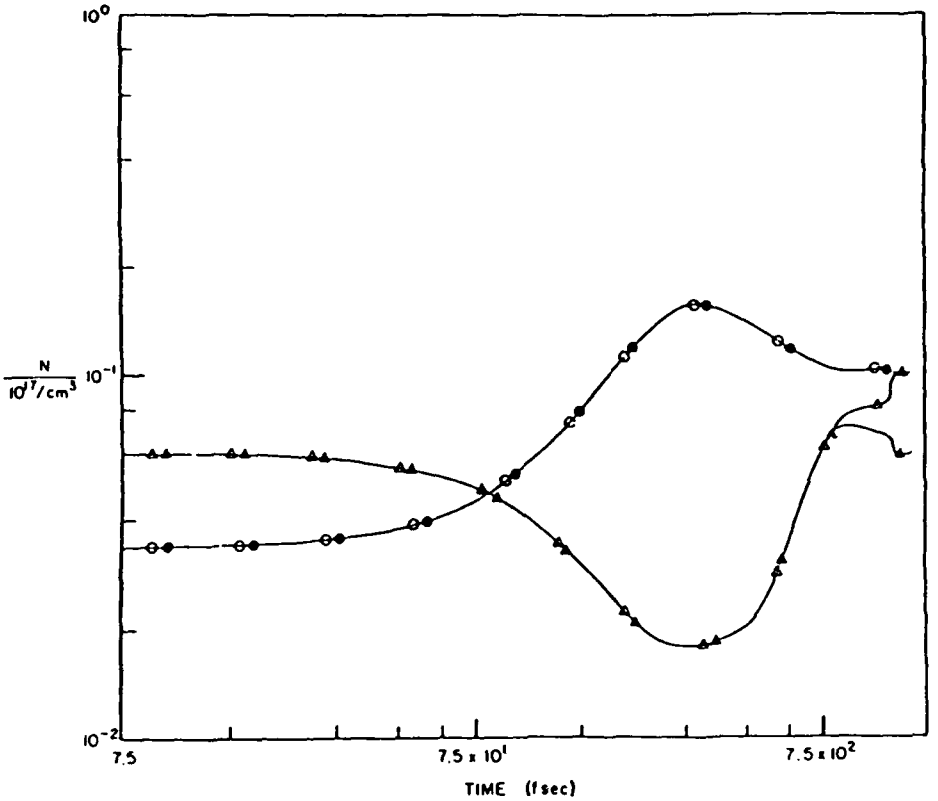


Fig. 46. Transient distribution of total (○, △) and Γ-valley (●, ▲) carrier density at two points within the $n^+-n^-n^+$ structure. At 0.2 μm (○, ●), there is no electron transfer of any significance. At 0.5 μm (△, ▲), electron transfer occurs at the end of the transient. Parameters are as in Fig. 38.

arises from the excess charge injected into the n^- region. This point was also made by East and Blakey [16], who also examined the dependence of current and voltage on n^- region length. A second point of importance here concerns determining which portion of the structure dominates its transport. It may be intuitively expected that for the structure considered it is the n^- region that dominates. This appears to be the case for the preceding discussion. But we may expect that for a sufficiently small n^- region, no single region dominates. In the calculations reported here, the absence of a single dominating region becomes apparent at higher voltage levels and for the case when $L_{N^-} = 0.116$ μm. These results are illustrated in Figs. 47–53, with particular attention paid to voltage sharing and electron transfer in the n^+ region as the n^- region is reduced in size. Figure 47 is a sketch of the background doping

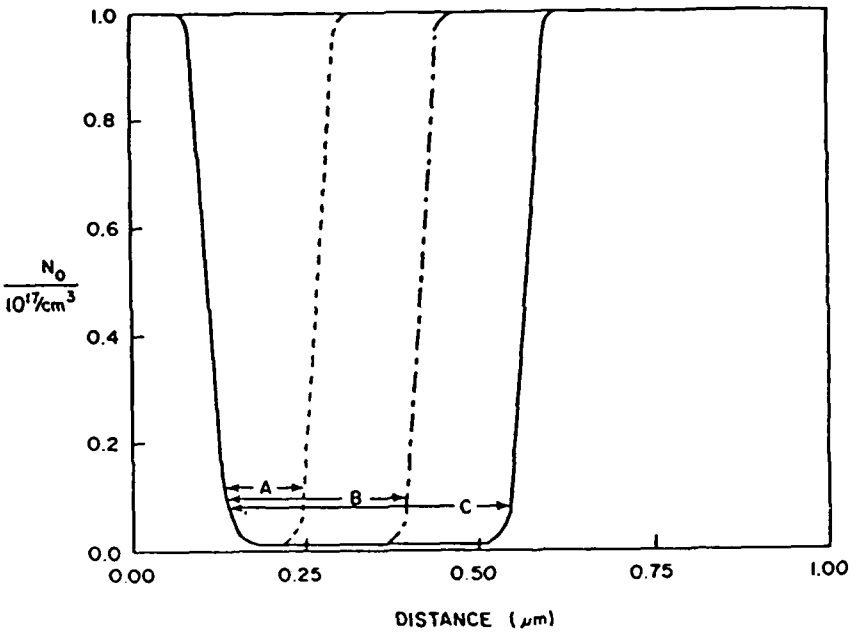


Fig. 47. Donor concentration for $n^+-n^--n^+$ structure with three different n^- region lengths. *A*, $L_{n^-} = 0.116\ \mu$m; *B*, $L_{n^-} = 0.266\ \mu$m; and *C*, $L_{n^-} = 0.416\ \mu$m. The calculations for *C* have already been presented and are included for completeness.

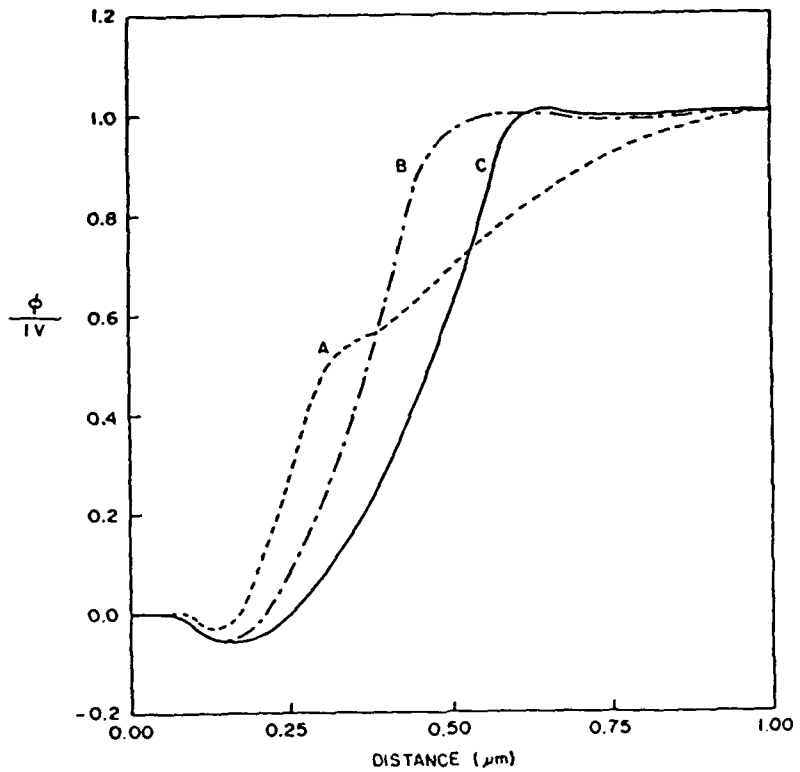


Fig. 48. Steady state distribution of potential for structures *A*, *B*, and *C* subject to a bias of 1.0 V. For structures *B* and *C* the potential drop is confined mainly to the n^- region. For structure *A* a significant fraction of potential falls across the n^+ region.

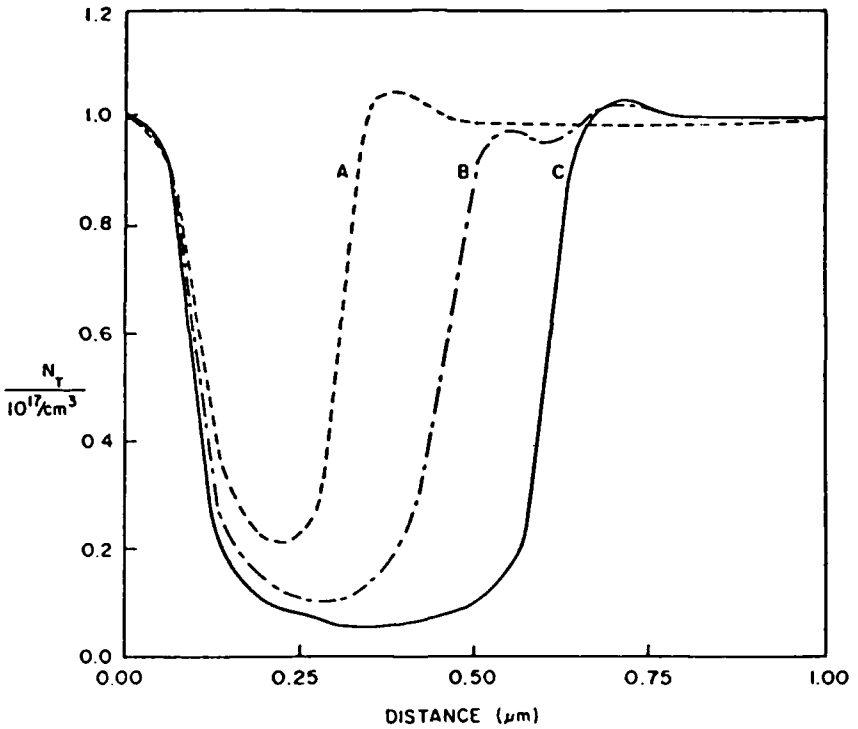


Fig. 49. Steady state distribution of total carrier concentration for structures A, B, and C. Note, for all three structures, the free carrier concentration closely traces the donor variation.

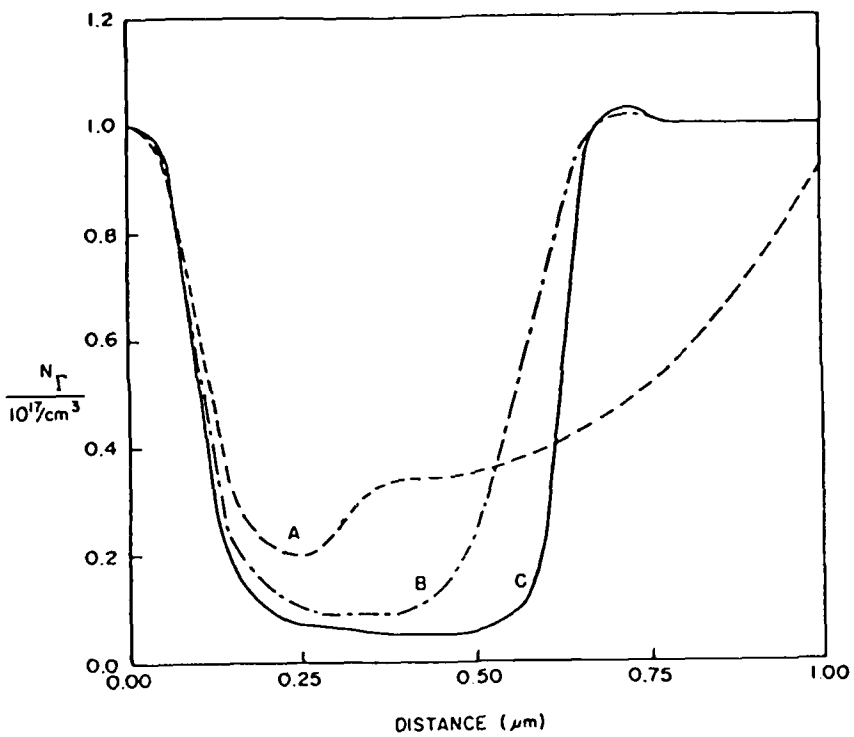


Fig. 50. Distribution of Γ-valley carriers. The least amount of electron transfer occurs for the widest structure *C*. For structure *A*, transfer continues to the anode contact and is a consequence of a large potential falling across the downstream n^+ region.

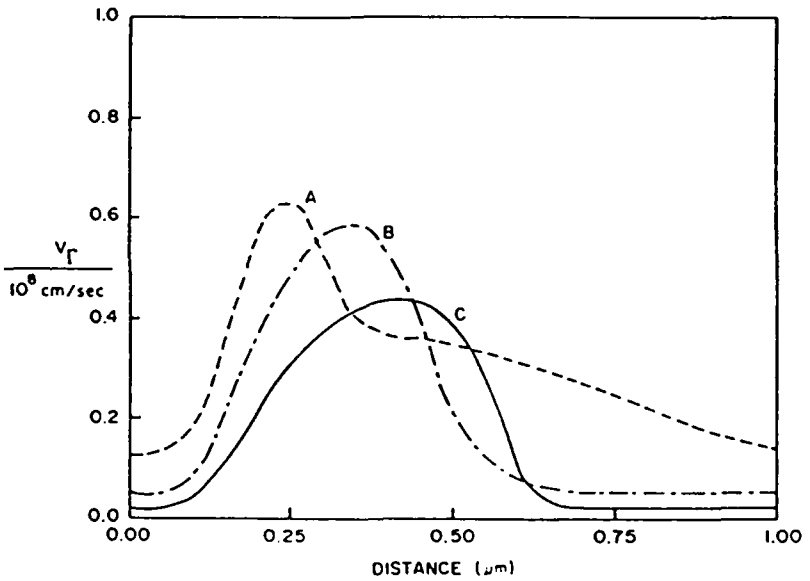


Fig. 51. Distribution of Γ-valley carrier velocity for structures *A*, *B*, and *C*. Peak velocity gradually increases as the n^- region decreases in length. Additionally, the upstream and downstream carrier velocities increase as the n^- region decreases in length.

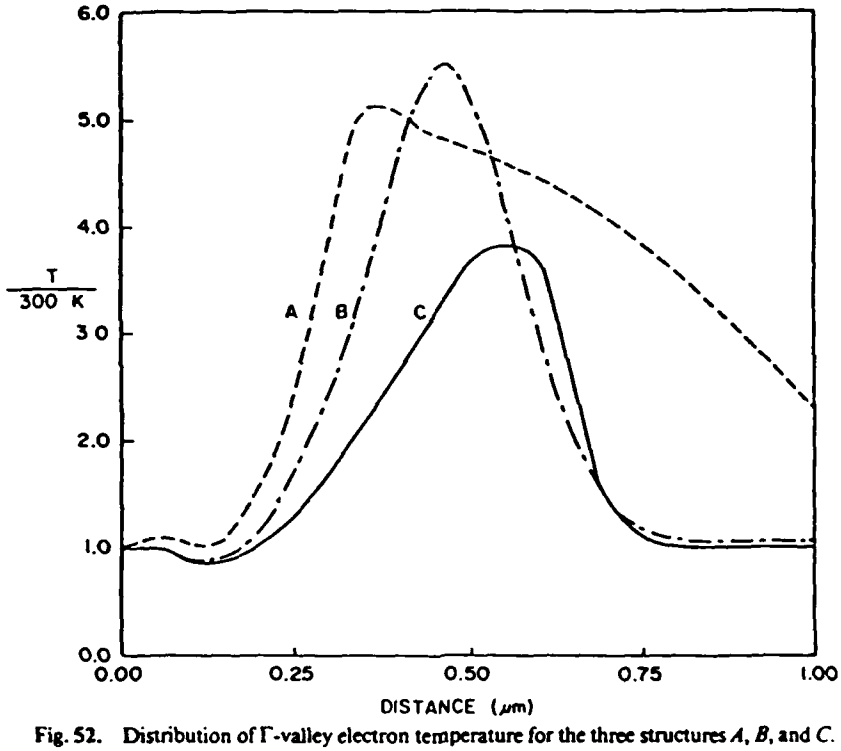


Fig. 52. Distribution of Γ-valley electron temperature for the three structures *A*, *B*, and *C*. Electron temperature distribution is qualitatively different for structure *A*. A longer downstream n^+ region is needed before the temperature approaches 300 K.

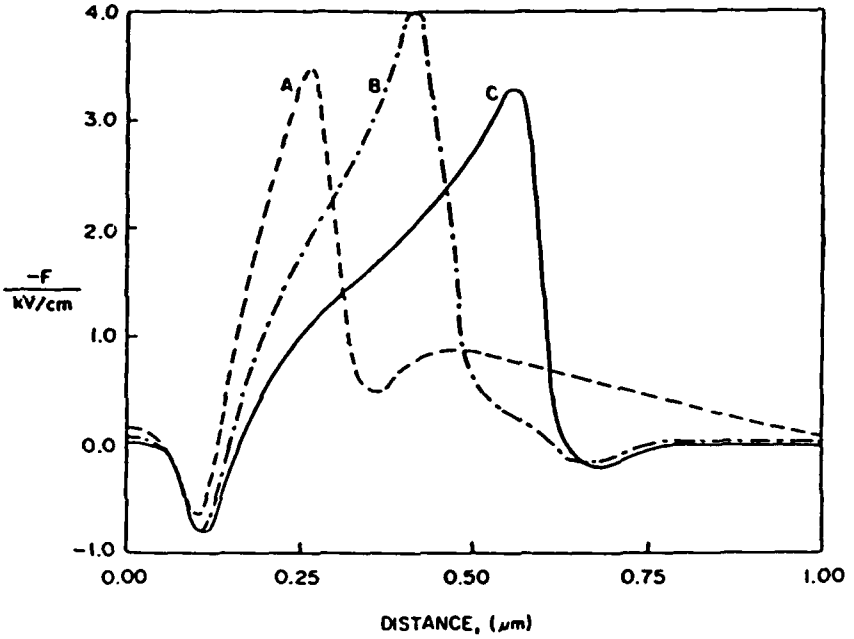


Fig. 53. Steady state distribution of electric field for the three structures *A*, *B*, and *C*. Note that for structure *A*, a large residual field remains across the downstream N^+ layer.

level associated with the variable n^- region. Within these regions and at a bias of 1 V, the potential is calculated self-consistently and is displayed in Fig. 48. It is noted that for n^- regions of length 0.266 and 0.416 μm, most of the potential drop is across the n^- region. For the smallest region a substantial potential drop falls across the n^+ region. The origins of this enhanced potential drop may be found in examining the self-consistently computed charge distribution (Fig. 49) which shows the presence of an excess charge accumulation at the downstream $n^- - n^+$ interface, resulting in a change in sign of the curvature of the potential. The distribution of Γ-valley carriers is displayed in Fig. 50, where the presence of a substantial electron transfer in the n^+ region is noted. The carrier velocity (Fig. 51) and electron temperature (Fig. 52) within the Γ valley display the expected increases for the shorter n^- region. The electric field distribution, shown in Fig. 53, displays higher field values within the n^+ region.

The significance of the preceding result is that while variations in the total charge density tend to screen variations in the doping profile of the structure, the potential drop across the n^- region may be small enough to allow a substantial drop across the downstream n^+ regions thereby permitting electron transfer to occur *away* from the n^- region. This, of course, is not

unexpected. It is implicit in the design of Gunn oscillators with doping variations assigned the task of domain nucleation sites. The current–voltage characteristics, therefore, are expected to reflect a complex set of electrical phenomena.

Figure 54 displays a series of current-versus-voltage curves for $n^+-n^--n^+$ structures with the indicated n^- region length. Each curve displays J/J_{ref} versus ϕ/ϕ_{ref}. The J_{ref} is the computed value of current at $\phi_{ref} = 0.25$ V. The valued J_{ref} is indicated in the figure caption. Because of the intuitive relationship between the space charge injection properties of the submicrometer $n^+-n^--n^+$ structure and those associated with Child's law, a power law $J\alpha\phi^\gamma$ was extracted. Note that J_{ref} increases as the n^- region decreases in length. At low bias levels the current–voltage relationship appears to follow a power

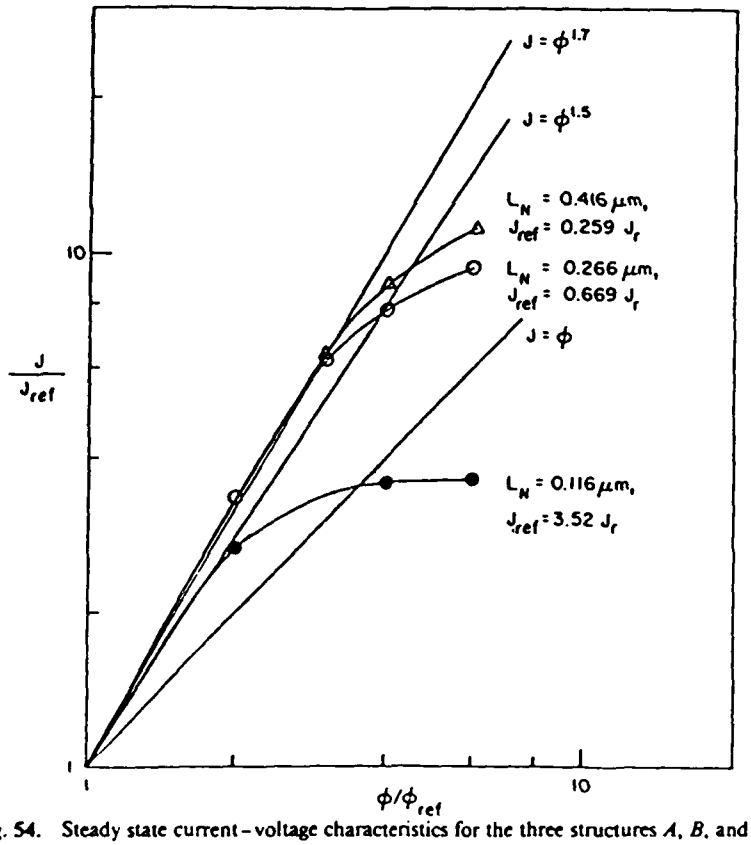


Fig. 54. Steady state current–voltage characteristics for the three structures *A*, *B*, and *C*. The current level for structure *A* is higher than that of *B* which in turn is higher than *C*. Note that the low-field resistance of structure *A* is the lowest of the three. Also included for reference are the Child's law $J = \phi^{1.7}$, $J = \phi^{1.5}$, and $J = \phi$ curves. $J_r = 1.6 \times 10^4$ A/cm^2; $\phi_{ref} = 0.25$ V.

relationship that is slightly less than $J/J_{ref} = (\phi/\phi_{ref})^{\gamma}$ with $\gamma = 1.7$ (as compared to a Child's law relationship where $\gamma = 1.5$). At higher values of bias there is enhanced sublinearity in the current–voltage relationship, due in part to electron transfer to the satellite valleys.

As indicated above, a considerable amount of electron transfer occurs in the downstream portion of the n^+ region when the n^- region is decreased in length. Indeed, the detailed calculations indicate that the relative amount of electron transfer *increases* as the n^- region *decreases* in length. At first glance, this result appears to contradict all that has been discussed about transport in submicrometer devices. But it is not unusual when it is realized that as the n^- region decreases in length a greater fraction of the voltage drop falls across the n^+ regions of the device. It is this latter feature that is responsible for the enhanced transfer. To place this in different terms, the active region length of the device increases as the n^- region becomes insignificantly small.

IV. Conclusions

The major technological interest in transient transport arises from the predictions of unusually high mean carrier velocities. The initial discussions of these high velocity values was for uniform space charge distributions, but the results were thought to be relevant for those situations where the mean carrier energy was insufficient to lead to substantial electron transfer in gallium arsenide. Thus the trend developed toward submicrometer-scale devices. The complication that arises in submicrometer devices is that the boundary conditions will be the determinant as to whether high velocities will be attained. Additionally, the constraints of current continuity dictate whether high velocities will be accompanied by high carrier densities. For example, in the case of injecting contacts the velocity of the entering carriers was significantly below that within the interior of the semiconductor. The situation was reversed for the case of partially blocking contact conditions.

Several critical results emerged from the discussion: (1) Transient overshoot in submicrometer structures reflects the presence of velocity overshoot and displacement current effects. It is not possible, in a simple way, to separate the two, with the result that transient measurements of overshoot require extreme care in interpretation. (2) Relaxation times to steady state are dominated by the dominating boundary; e.g., either the metal contact or the critical interface. Relaxation times do not scale linearly with device length. The relaxation time scales monotonically with length. (3) Transient overshoot effects are dependent upon rise times and the time for relevant field rearrangement within the structure.

ACKNOWLEDGMENTS

The authors benefited greatly from many discussions with their colleagues, particularly, G. J. Iafrate, D. K. Ferry, K. Hess, R. F. Greene, M. Silver, and M. P. Shaw. This work was supported by ONR and DARPA to whom the authors are grateful.

REFERENCES

1. K. Hess and G. J. Iafrate *in* "VLSI Electronics" (N. G. Einspruch, ed.). Vol. 9, Chapter 6, Academic Press, New York,
2. M. P. Shaw, P. R. Solomon, and H. L. Grubin, *IBM J. Res. Dev.* **13**, 587 (1969).
3. H. L. Grubin, *IEEE Trans. Electron. Devices* **ED-25**, 511 (1978).
4. A. Zur, T. C. McGill, and D. L. Smith, *Surf. Sci.* **132**, 456 (1983).
5. A. W. DeGroot, G. C. McGonigal, D. J. Thomson, and H. C. Card, *J. Appl. Phys.* **55**, 312 (1984).
6. M. P. Shaw, P. R. Solomon, and H. L. Grubin, "The Gunn Hilsum Effect." Academic Press, New York, 1979.
7. A. Sommerfield, "Thermodynamics and Statistical Mechanics." Academic Press, New York, 1956.
8. J. G. Ruch, *IEEE Trans. Electron. Devices* **ED-19**, 652 (1972).
9. H. Kroemer, *IEEE Trans. Electron. Devices*, **ED-15**, 819 (1968).
10. H. L. Grubin, D. K. Ferry, G. J. Iafrate, and J. Berker, *in* "VISI Electronics" (N. G. Einspruch, ed.), Vol. 3, p. 198. Academic Press, New York, 1982.
11. K. W. Gray, J. Pattison, H. D. Rees, B. A. Prew, R. C. Clarke, and L. D. Irving, *Proc. Bienn. Cornell Univ. Conf., 5th* 215 (1975).
12. H. L. Grubin, and J. P. Kreskovsky, *Surf. Sci.* **132**, 594 (1983).
13. H. L. Grubin, and J. P. Kreskovsky. *J. Vac. Sci. Technol.* (in press).
14. R. K. Cook, and J. Frey, *IEEE Trans. Electron. Devices* **ED-28**, 951 (1981).
15. K. Tomizawa, Y. Aweno, N. Hashizume, M. Kawashima, *IEEE Proc.* **129**, 131 (1982).
16. J. R. East, and P. A. Blakey, *Phys. Submicron Struct. Proc.* 1982 *Workshop* (H. L. Grubin, ed.) p. 287. Plenum, New York, 1984.

APPENDIX. DIMENSIONLESS EQUATIONS USED IN THE NUMERICAL SIMULATIONS

The continuity equations in dimensionless form are as follows:

(i) Equation (3):

$$\frac{\partial n_1^*}{\partial t^*} = -\frac{\partial n_1^* V_{1j}^*}{\partial x_j^*} - n_1^* f_1 + (n^* - n_1^*) f_2.$$

(ii) Equation (5):

$$\frac{\partial n^*}{\partial t^*} = -\frac{\partial}{\partial x_j^*}(n_1^* V_1^{*j} + (n^* - n_1^*) V_2^{*j}).$$

The dimensional terms are identified in the following tabulation:

$$n_1^* = n_1/n_{\text{ref}}, \qquad V_1^* = V_1/V_{\text{ref}},$$
$$n^* = n/n_{\text{ref}}, \qquad V_2^* = V_2/V_{\text{ref}},$$
$$f_1 = \Gamma_1/\Gamma_{\text{ref}}, \qquad f_2 = \Gamma_2/\Gamma_{\text{ref}},$$
$$x^* = x/x_{\text{ref}}, \qquad t^* = t/t_{\text{ref}},$$

with

$$t_{\text{ref}} = x_{\text{ref}}/V_{\text{ref}}, \qquad \Gamma_{\text{ref}} = 1/t_{\text{ref}}.$$

The momentum balance equations in dimensionless form are as follows:

(i) Equation (23):

$$\frac{\partial n_1^* V_1^{*i}}{\partial t^*} = -\frac{\partial n_1^* V_1^{*j} V_1^{*i}}{\partial x_j^*} + Pf\frac{n_1^*}{m_1^*}\frac{\partial \phi^*}{\partial x_i^*} - \frac{1}{\gamma M^2}\frac{\partial}{\partial x_i^*} n_1^* R_1^* T_1^* + \frac{\mu_1^*}{Re \cdot m_1^*}\frac{\partial^2 V_1^{*i}}{\partial x_j^{*2}} - n_1^* V_1^{*i} f_3.$$

(ii) Equation (24):

$$\frac{\partial (n^* - n_1^*) V_2^{*i}}{\partial t^*} = -\frac{\partial}{\partial x_j^*}(n^* - n_1^*) V_2^{*i} V_2^{*j} + Pf\frac{(n^* - n_1^*)}{m_2^*}\frac{\partial \phi^*}{\partial x_i^*} - \frac{1}{\gamma M^2}\frac{\partial}{\partial x_i^*}(n^* - n_1^*) R_2^* T_2^* + \frac{\mu_2^*}{Re \cdot m_2^*}\frac{\partial^2 V_2^{*i}}{\partial x_j^{*2}} - (n^* - n_1^*) V_2^{*i} f_4.$$

The dimensionless terms and parameters are identified in the following tabulation:

$$\phi^* = \phi/\phi_{\text{ref}},$$
$$m_1^* = m_1/m_{\text{ref}}, \qquad m_2^* = m_2/m_{\text{ref}},$$
$$R_1 = k_B/m_1, \qquad R_2 = k_B/m_2,$$
$$R_{\text{ref}} = k_B/m_{\text{ref}},$$
$$R_1^* = R_1/R_{\text{ref}} = \frac{1}{m_1^*}, \qquad R_2^* = R_2/R_{\text{ref}} = \frac{1}{m_2^*},$$
$$f_3 = \Gamma_3/\Gamma_{\text{ref}}, \qquad f_4 = \Gamma_4/\Gamma_{\text{ref}},$$
$$\mu_1^* = \hat{\mu}_1/\hat{\mu}_{\text{ref}}, \qquad \mu_2^* = \hat{\mu}_2/\hat{\mu}_{\text{ref}}.$$

with

$$Pf = e\phi_{ref}/m_{ref}V_{ref}^2, \qquad M = V_{ref}/V_a,$$
$$V_a = (\tfrac{5}{3}R_{ref}T_{ref})^{1/2} \qquad \gamma = \tfrac{5}{3},$$
$$Re = x_{ref}V_{ref}n_{ref}m_{ref}/\hat{\mu}_{ref}.$$

The energy balance equations in dimensionless form are as follows:

(i) Equation (30):

$$\frac{\partial n_1^* T_1^*}{\partial t^*} = -\frac{\partial n_1^* T_1^* V_1^{*j}}{\partial x_j^*} - (\gamma - 1)n_1^* T_1^* \frac{\partial V_1^{*j}}{\partial x_j^*} + \frac{1}{Re \cdot Pr}\frac{1}{m_1^* c_{v_1}^*} \times \frac{\partial}{\partial x_j^*}\left(\kappa_1^* \frac{\partial T_1^*}{\partial x_j^*}\right) - n_1 T_1 f_5 + (n - n_1)T_2 f_6 + \gamma(\gamma - 1)M^2 \frac{V_1^{*j}V_1^{*j}}{2c_{v_1}^*}[2n_1^* f_3 - n_1^* f_1 + (n^* - n_1^*)f_2].$$

(ii) Equation (31):

$$\frac{\partial (n^* - n_1^*)T_2^*}{\partial t^*} = -\frac{\partial (n^* - n_1^*)T_2^* V_2^{*j}}{\partial x_j^*} - (\gamma - 1)(n^* - n_1^*)T_2^* \frac{\partial V_2^{*j}}{\partial x_j^*} + \frac{1}{Re \cdot Pr}\frac{1}{m_2^* c_{v_2}^*}\frac{\partial}{\partial x_j^*}\left(\kappa_2^* \frac{\partial T_2^*}{\partial x_j^*}\right) + \gamma(\gamma - 1)M^2 \frac{V_2^{*j}V_2^{*j}}{\partial x_j^*} \times [2(n^* - n_1^*)f_4 + n_1^* f_1 - (n^* - n_1^*)f_2] - (n^* - n_1^*)T_2^* f_7 + n_1^* T_1^* f_8.$$

The dimensionless terms and parameters are identified in the following tabulation:

$$f_5 = \Gamma_5/\Gamma_{ref}, \qquad f_6 = \Gamma_6/\Gamma_{ref},$$
$$f_7 = \Gamma_7/\Gamma_{ref}, \qquad f_8 = \Gamma_8/\Gamma_{ref},$$
$$\kappa_1^* = \kappa_1/\kappa_{ref}, \qquad \kappa_2^* = \kappa_2/\kappa_{ref},$$
$$C_{v_1} = \tfrac{3}{2}R_1, \qquad C_{v_2} = \tfrac{3}{2}R_2, \qquad C_{v_{ref}} = \tfrac{3}{2}R_{ref},$$
$$C_{v_1}^* = C_{v_1}/C_{v_{ref}}, \qquad C_{v_2}^* = C_{v_2}/C_{v_{ref}},$$
$$P_r = C_{v_{ref}}\hat{\mu}_{ref}/\kappa_{ref}.$$

Poisson's equation in dimensionless form:

(i) Equation (1):

$$\frac{\partial^2 \phi^*}{\partial x_j^{*2}} = Sn[n_1^* + (n^* - n_1^*) - n_0^*].$$

The dimensionless terms and parameters are

$$Sn = \frac{x_{\text{ref}}^2 e n_{\text{ref}}}{\phi_{\text{ref}} \epsilon}, \qquad n_0^* = \frac{n_0}{n_{\text{ref}}}.$$

The boundary conditions in dimensionless form are:

(i) Equation (32):

$$\frac{\partial n^*}{\partial x^*} = \frac{\partial n_1^*}{\partial x^*} = \frac{\partial V_1^*}{\partial x^*} = \frac{\partial V_2^*}{\partial x^*} = \frac{\partial T_1^*}{\partial x^*} = \frac{\partial T_2^*}{\partial x^*} = 0$$

at $x = 0, \quad \phi^* = 0,$

$x = L, \quad \phi^* = \phi_a.$

(ii) Equation (40):

$$\frac{\partial^2 n^*}{\partial x^{*2}} = \frac{\partial^2 n_1^*}{\partial x^2} = 0, \qquad V_1^* = \mu_c^* \frac{\partial \phi^*}{\partial x^*}, \qquad \mu_c^* = \frac{\mu_c \phi_{\text{ref}}}{x_{\text{ref}} V_{\text{ref}}},$$

$$\frac{\partial V_2^*}{\partial x^*} = 0, \qquad T_1 = T_c^*, \qquad \frac{\partial T_2^*}{\partial x^*} = 0, \qquad \phi^* = 0.$$

(iii) Equation (41):

$$\frac{\partial^2 n^*}{\partial x^{*2}} = \frac{\partial^2 n_1^*}{\partial x^{*2}} = \frac{\partial^2 V_1^*}{\partial x^{*2}} = \frac{\partial^2 V_2^*}{\partial x^{*2}} = \frac{\partial^2 T_1^*}{\partial x^{*2}} = \frac{\partial^2 T_2^*}{\partial x^{*2}} = 0, \qquad \phi^* = 0.$$

(iv) Equation (53):

$$n^* = n_0^*, \qquad n_1^* = \frac{f_2}{f_1 + f_2} n_0^*, \qquad \frac{\partial^2 V_1^*}{\partial x^{*2}} = \frac{\partial V_2^*}{\partial x^*} = 0,$$

$$T_1^* = T_c^*, \qquad \frac{\partial T_2^*}{\partial x^*} = 0, \qquad \phi^* = 0,$$

where f_1 and f_2 are evaluated at $T = T_{\text{ref}}$. Common parameters and dimensionless reference quantities are given in Table A-I.

TABLE A-I
Parameters

Dimensionless Reference Quantities					
	Figure number				
Parameter	1–4	9–19	21–27	28–36	37–53
Device length (μm)	1.0	1.0	5.0	0.25	1.0
x_{ref} (μm)	1.0	1.0	5.0	0.25	0.75
n_{ref} (cm^{-3})	$5. \times 10^{15}$	$5. \times 10^{15}$	$5. \times 10^{15}$	$5. \times 10^{15}$	$1. \times 10^{17}$
t_{ref} (psec)	1.0	1.0	5.0	0.25	0.75
Γ_{ref} (10^{12}/sec)	1.0	1.0	0.2	4.0	1.33
κ_{ref} (J/K cm sec)	2.0×10^{-6}	2.0×10^{-6}	2.0×10^{-6}	2.0×10^{-6}	4.14×10^{-5}
$\hat{\mu}_{ref}$ (gm/cm · sec)	5.74×10^{-11}	5.74×10^{-11}	5.74×10^{-11}	5.74×10^{-11}	1.15×10^{-9}
Re	53.17	53.17	265.86	13.29	39.88
Sn	8.16	8.16	204.00	0.51	91.84
ϕ^*	0.6	0.6	0.5	0.15	1.0
	1.0	1.0	1.0	0.25	
	2.0	2.0	1.5	0.50	
			2.0		
			3.0		
T_c^*	—	1.0	4.0	1.0	1.0
μ_c^*	—	1.56	0.08	6.25	—

Common parameters		
$V_{ref} = 10^6$ cm/sec	$m_1^* = 1$	$\kappa_1^* = 1.0$
$\phi_{ref} = 1.0$ V	$m_2^* = 3.31$	$\kappa_2^* = 1.0$
$m_{ref} = 6.10 \times 10^{-29}$ g	$R_1^* = 1.0$	$Pr = 9.40$
$k_B = 1.38 \times 10^{-23}$ J/K	$R_2^* = 1.0$	$Pf = 2.62$
$R_{ref} = 2.26 \times 10^5$ J/K · gm	$\mu_1^* = 1.0$	$M = 2.97$
$T = 300$ K	$\mu_2^* = 1.0$	
$V_s = 3.36 \times 10^7$ cm/sec		

Reprint 16
Physica 134B (1985) 67-71
North-Holland, Amsterdam

BAND STRUCTURE DEPENDENT TRANSIENT TRANSPORT IN NEAR AND SUBMICRON LENGTH SEMICONDUCTOR DEVICES*

H.L. GRUBIN and J.P. KRESKOVSKY
Scientific Research Associates, Inc., Glastonbury, Connecticut 06033, USA

Scaling principles are applied to the moments of the Boltzmann transport equation to establish guidelines for submicron device/material selection.

INTRODUCTION

The rapid advances in numerical methods to calculate band structure parameters on demand (Dow[1]), and the need to provide choices for device applicable semiconductors has spurred interest in the development of guidelines for material selection. The guidelines must address high speed, hot carrier, requirements, and determine the constraints these materials impose. In the past, device structures, such as the field effect transistor, submitted to figures of merit which taught that high mobility and/or saturated drift velocity were beneficial. These figures of merit were derived from the steady state field dependent velocity relationships. The situation, as is now known, is that the equilibrium transport parameters are not relevant for submicron devices. To treat these devices a set of scaling guidelines, discussed by Thornber[2], is introduced. These guidelines result from an analysis of the Boltzmann transport equation (BTE), in which length and time scales, and scattering rates are altered. The application of these guidelines to submicron devices was introduced by Grubin[3,4], et al., to uniform field transport, and is continued below for non-uniform fields. The basic conclusion of the study is that all relevant device figures of merit are dependent upon their critical feature size, and the results obtained for micron length structures are not necessarily appliable to submicron structures.

*This [illegible] and ONR.

RELEVANT TRANSPORT EQUATIONS AND SCALING

The BTE for electrons in a parabolic band with effective mass m is:

$$\frac{\partial f}{\partial t} + \frac{\hbar\vec{k}}{m}\cdot\nabla_r f - \frac{e\vec{F}}{\hbar}\cdot\nabla_k f = -\frac{1}{4\pi^3}\left\{\int d^3k' W(\vec{k},\vec{k}')f(\vec{k})(1-f(\vec{k}')) - \int d^3k W(k',k)f(k')(1-f(k))\right\} \quad (1)$$

where W(k,k') is the probability of a carrier undergoing a transition from the state k to the state k'. As used in submicron transport W(k,k') is usually separated into contributions from each band. Scaling as discussed by Thornber is typified by the following examples: [1] Under uniform field time independent conditions, scaling W by a constant λ, i.e., $W(k,k')=\lambda W(k,k')$, results in an alteration of the mean carrier velocity $V(F)=V_\lambda(\lambda F)$ where the latter is the scaled mean velocity and

$$V \equiv \int \frac{\hbar k}{m} f d^3k \Big/ \int f d^3k \quad (2)$$

If $\lambda=1$ represents GaAs, then, as seen in Ref. 4, the field dependent velocity relations is, apart from the region of negative differential mobility, remarkably similar to that of InP. Additionally the $\lambda=1/2$ curve bears a strong resemblence to the field dependent velocity of InGaAs. [2] If a constant scaling is included in the time dependent portion of the BTE, then the transient response of

system is altered, with $V_\lambda[F(t),t]=V[F/\lambda(\lambda t),\lambda t]$. The implication is that relaxation effects are more rapid for λ greater than or equal to unity, as illustrated below. [3] The situation with spatial scaling is similar. If the BTE is solved for a spatially varying sinusoidal field, and recomputed for a constant scattering scaling, it is found that similar velocity variations will occur for the scaled device but over the distance L/λ, providing the amplitude of the field is increased by the factor λ.

Apart from the fact that practical constant scattering scaling is not possible, as discussed below, the physical situation is more complex then indicated because self-consistency has been ignored. Thus, Poissons equation:

$$\nabla^2\phi = \frac{e}{\epsilon}(n-n_0) \qquad (3)$$

must be solved. If x is altered by the constant λ, the right hand side of equation (3) is increased by the factor λ^2, and the donor and carrier density must be altered.

The relevance of altering all scattering rates by the same constant is moot, and is introduced as a starting point for discussion. It differs from the classical scaling as discussed, e.g., by Bar-Lev[6], insofar as it focuses attention on alterations in the mobility rather than the extrinsic characteristics of the structure, although the latter will clearly enter the picture.

In this study specific semiconductors are identified (apart from band structure) by a set of scattering rates, e.g., LO phonon, intervalley phonon, etc., and these are varied. For GaAs, assuming two levels of transferred electron (Γ and L) scattering rates [illegible] the parameters identified in Grubin[5], et al. In calculating transport the first three moments of the BTE are utilized. The scaling arguments as applied to the BTE are applicable to the moment equations and the conceptual results are interchangable.

MOMENTS OF THE BOLTZMANN TRANSPORT EQUATION (MBTE)

The three moment equations are: continuity, momentum and energy balance; and are written for two species of carriers. Thus, $n=n_1+n_2$, where n_1 and n_2 designate the population of the Γ and L valley, respectively. Poisson's equation is coupled to the moment equations, the first set of which involves continuity:

$$\frac{\partial n_1}{\partial t} = -\frac{\partial}{\partial x_j}\frac{n_1\hbar k_1^j}{m_1} - n_1\Gamma_1 + (n-n_1)\Gamma_2 \qquad (4)$$

Here Γ designates the scattering rate. It is noted for parabolic bands $\hbar k^j_1/m_1 = v^j_1$. An equation similar to (4) describes transient population changes in the L valley. When the two are combined, a global continuity equation results.

$$\frac{\partial n}{\partial t} = -\frac{\partial}{\partial x_j}\left[n_1\frac{\hbar k_1^j}{m_1} + (n-n_1)\frac{\hbar k_2^j}{m_2}\right] \qquad (5)$$

The quantity in the brackets of equation (4) is the velocity flux density, Q^j, of the system and is related to a mean spatially dependent drift velocity as $V_d = \frac{Q}{n}$. The total current density, including displacement currents, is conserved. For momentum relation:

$$\frac{\partial}{\partial t} n_1\hbar k_1^j + \frac{\partial}{\partial x_i}\frac{\hbar}{m} n_1\hbar k_1^j = \qquad (6)$$

$$en_1\frac{\partial \xi}{\partial x_j} - \frac{\partial}{\partial x_i}\zeta_1^{ij} - n_1\hbar\frac{\Gamma}{}$$

where ζ^{ij} represents the [illegible] pressure tensor. For a displaced [illegible] ζ^{ij}_1. For [illegible]

$$\frac{\partial}{\partial t} W_i = -\frac{\partial}{\partial x_j}\frac{\hbar k_2^j}{m_i} W_i + n_i e \frac{\hbar k_i^j}{m_i}\frac{\partial \phi}{\partial x_j} - \frac{\partial}{\partial x_j}\frac{\hbar k_i^j}{m_i}\psi_i^{ij} - \frac{\partial}{\partial x_j} Q_i^j - n_i U_i \Gamma_5 + (n-n_i)U_2\Gamma_6 \quad (7)$$

where

$$U_i = \frac{3}{2} k_B T_i \qquad W_i = n_i\left[\frac{1}{2}\frac{\hbar^2 k^2}{2m_i} + U_i\right] \quad (8)$$

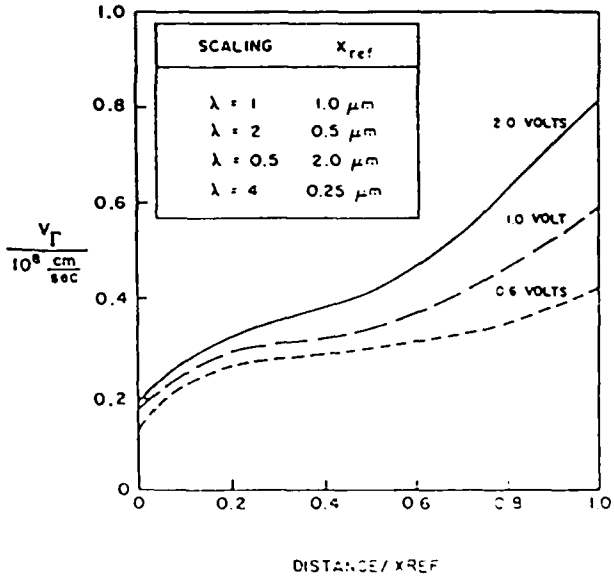
The quantity Q_i^j is zero for spherically distribution functions; more generally it represents a flow of heat and is treated phenomenologically through analogy to Fourier's law. Similar equations occur for the L valley.

CALCULATIONS WITH CONSTANT SCALING PARAMETERS

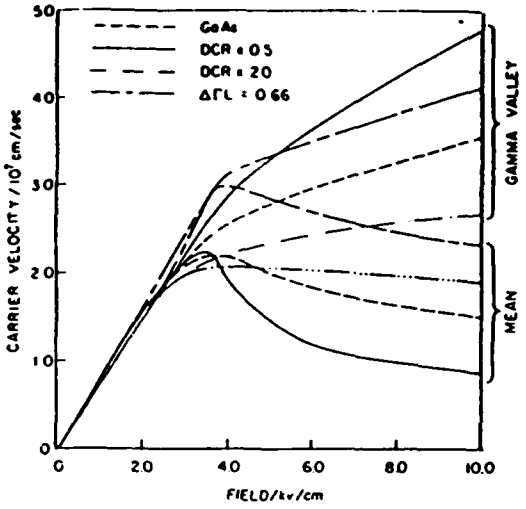
To illustrate, the governing equations are altered through the prescriptions of second section. For this case the unscaled GaAs device sustains spatially non-uniform transport arising from injecting cathode boundary conditions.

For the first set of calculations all scattering rates are altered by the constant λ. With all velocities measured against 10^8cm/sec, all reference times and lengths are scaled as: $t \to t/\lambda, x \to x/\lambda$. For a constant potential, scaling 'x' requires an alteration in the reference density. Thus, $n \to \lambda^2 n$. (It is noted that momentum balance contains a viscous contribution, which in the scaled system is increased by the λ. For energy balance equation, the inclusion of a thermal conductivity, requires that it be increased by the same scaling constant (Grubin[5], et al)).

Two figures illustrate figure 1 displays the Γ valley velocity computed for the unscaled GaAs element. The structure is 1.0 microns long, and subject to the indicated values of bias. The inset indicates that for λ=2, an identical result is be obtained for a length of 0.5 microns with an average field twice that of the unscaled calculation. This result indicates the nonuniformities to be expected in InP, once results for GaAs are known. Similarly, the results for λ=1/2 are representative of InGaAs. Figure 2 shows the current transient for the GaAs structure subject to an instantaneous change in bias to 1.0 volts. There is a current overshoot that is related in part to velocity overshoot. The significance of figure 2 is the indication that InP is likeky to be faster than GaAs, whereas that GaInAs is likely to be slower. Note results with Gunn diodes appear to indicate an InP upper frequency limit that is higher than that of comparably prepared GaAs diodes; although the evidence is not unequivocal that the origin of this difference



1. Γ-valley velocity for GaAs and scaled materials subject to injecting contacts, Ref. 5

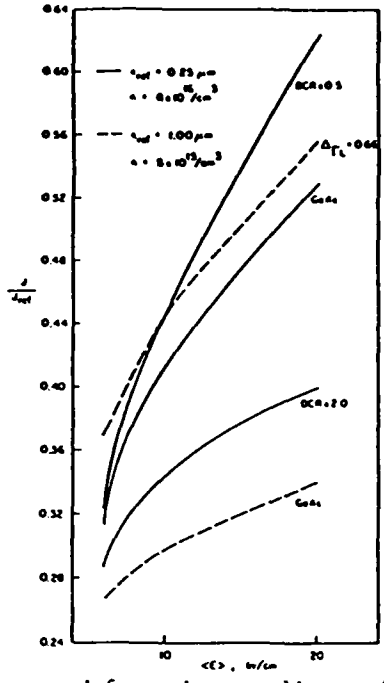


3. V(F) versus deformation coupling coefficient and intervalley separation, Ref. 4.

CONCLUSIONS

The above study indicates that a set of guidelines can be established for scaling materials and devices, and used for making broad comparisons between such materials as, e.g., InP and InGaAs, Si, and GaAs. The principle conclusion from this scaling study is that for submicron devices any alteration in the band structure that increases either the number of carriers in the Γ valley, increases the carrier mobility, will yield an improved figure of merit for submicron feature size devices. Saturated drift velocity, as obtained from the mean velocity of the carriers is not a significant figure of merit for submicon devices. In the absence of thermal considerations, this places a material like InGaAs in a favorable position for small signal device operation. On the other hand for large signal transient operation, GaAs is superior to InGaAs.

The above results are obtained without consideration of density dependent effects. This study has also been performed, and the results summarized. The effects of ionized impurity scattering are well known and



4. I-V versus deformation coupling coefficient and intervalley separation, Ref. 4.

of field. But of more importance, is that an increased density serves to increase the nonuniformity of the space charge distribution resulting in reductions in current level that are significantly higher than those arising from ionized impurity scattering. This result tends to camaflouge interpretation of mesurements concerning the effects of impurity scattering.

REFERENCES

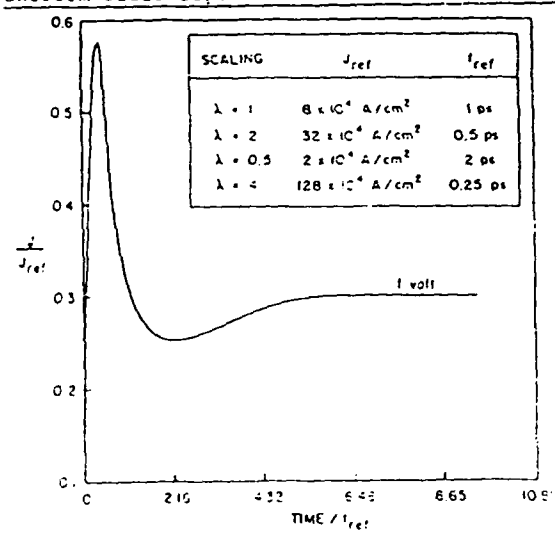
1. J.Dow, ARO Workshop, June 1985
2. K.Thornber, J.Appl.Phys. 51, 2127 (1980)
3. H.L.Grubin, J.P.Kreskovsky, G.J.Iafrate, D.K.Ferry and R.F.Greene, in Phys. Submicron Struct. 63,Plenum Press (1984)
4. H.L.Grubin et al, to be published
5. H.L.Grubin and J.P.Kreskovsky, Microstructure Sci. and Tech. 14 ,Acad.Press (1985)
6. A.Bar-Lev, Semiconductors and Electronic Devices (2 ed) Prentice-Hall Int. NJ (84).
7. J.R.Hauser, T.H.Glisson and M.A.Littlejohn, Solid State Elect. 23, [illegible]

is completely accounted for by the results of scaling.

The situation above is artificial but there are relevant extrapolations. For example, the effects of intervalley scattering increases as the length of the structure increases. This result suggests that intervalley scattering effects at scaled field values becomes more prominant in smaller structural lengths in the λ=2 material than in either the λ=1/2, or λ=1 material. It is noted that in scaling these structures, an increasing scaling parameter is matched by a decreasing background doping. Should scaling be performed without a comparable decrease in doping level field nonuniformities would be more pronounced.

CALCULATIONS WITH VARIABLE PARAMETERS

The relevance of scaling is emphasized in those cases where the scattering rates are nonuniformly changed. Here two types of calculations illustrate, one in which all but one of the scattering elements of GaAs, the intervalley deformation potential coefficient, is constant. In the second, only the intervalley energy separation is altered. The uniform field dependent mean carrier velocity



3. Effect of parameter variation on the current transient, Ref. 4.

are shown in figure 3. The increased deformation potential coupling coefficient results in an increased saturated drift velocity, but a corresponding decrease in the Γ valley velocity at high values of electric field. The results have the following significance: Present choice of semiconductor materials for such devices as the field effect transistor are based upon several critical material parameters, one of which is the saturated drift velocity. Indeed, part of the Si/GaAs controversy is based on the value of the electron saturated drift velocity, which is higher for Si. The situation for submicron devices is that Γ valley transport dominates device behavior, as revealed by supplementary one and two dimensional studies (see Hauser,[7] et al). The effect, of this parameter variation is shown in figure 4 which displays the current voltage relations for a 1/4 micron structure with a doping level of $8 \times 10^{16}/cm^3$, variable intervalley separation; and for a 1.0 micron structure variable coupling coefficient, with a doping of $5 \times 10^{15}/cm^3$. Both calculations are for nonuniform fields arising from injecting contacts. Note: For the variable coupling coefficient calculations the scaled material with the highest mobility is that yielding the highest current levels. The turn around to a saturated drift velocity dominated material occurs at longer device lengths. The situation with the variable intervalley separation is shows higher current levels for the larger separation, again corresponding to a higher Γ valley mobility. The results are clear: the current level is highest for the material with the higher intervalley separation. In this case more carriers are being retained in the central valley, which is again dominating the sport.

Reprint 17
Physica 134B (1985) 499-501
North-Holland, Amsterdam

A STUDY OF NEAR-MICRON InP TRANSFERRED ELECTRON DEVICES

J. CZEKAJ and M.P. SHAW, Dept. of Electrical & Computer Engineering, Wayne State University, Detroit, MI 48202*; J. EAST, Dept. of Electrical & Computer Engineering, University of Michigan, Ann Arbor, MI 48109; P.A. BLAKEY, Dept. of Electrical Engineering, Arizona State University, Tempe, AZ; H.L. GRUBIN, Scientific Research Associates, Glastonbury, CT 06033

We have experimentally investigated a large number of InP Transferred Electron Devices ranging from 1 to 2 microns in thickness, and have numerically simulated their behavior using both a drift and diffusion model and a Boltzmann transport equation approach. We were able to fit the static I(V) curves and threshold conditions rather well by a proper choice of ideality factor, n, in the control characteristic, which we assumed to be that of a Schottky diode.

1. INTRODUCTION

We have obtained good agreement between our experimental results and numerical simulations of transport in transferred electron InP devices as small as 1.0 micron . The amplification and oscillatory characteristics of the devices were examined in high-Q resonant cavities in the 75-100 Ghz region. Theoretically, we used a standard drift and diffusion model to predict both the static and dynamic characteristics of the device, and calculated the velocity-field curve using the Boltzmann transport equation (BTE). Our work has centered on the vital role that contacts play in influencing hot electron phenomena. The cathode contacts were characterized as Schottky diodes, and the ideality factor, n, was the single parameter employed to fit the prethreshold current-voltage (I(V)) characteristics. From the temperature dependence of the dc to ac conversion efficiency, we have also been able to estimate the manner in which electrons enter into the active region for specific contacts.

2. EXPERIMENT

The devices, provided by the Varian Solid State Microwave Division, were constructed with the active region doped n-type. The carrier concentration varied from 7.6 to 9.5 * $10^{15}/cm^3$ with active region lengths varying from 1.60 to 1.96 microns. The cathode contacts consisted of Au/Ni/Ge-InP junctions while the anode contact was an n-n+ junction with the n+ region doped to about $10^{17}/cm^3$. The devices were packaged on a threaded stud suitable for operation in the 75-100 GHz frequency range.

The oscillatory characteristics were examined while the device was inserted in a high-Q narrow bandwidth resonant cavity using a radial hat bias choke. While mounted in this environment the following data were collected: pulsed (to prevent burnout) I(V) characteristics, oscillation threshold point, frequency of oscillation, and output power. (Because of the relatively high rate of burnout, once oscillations were observed (on a power meter) the bias was not increased by more than a few tenths of a volt.)

All of the devices displayed some degree of current saturation prior to the onset of oscilation. The frequency of oscillation was determined primarily by the radius of the radial

*Supported by NSF Grant ECS 82-11841

0378-4363/85/$03.30 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

hat used. When microwave power was observed, the E-H tuner, serving as an impedance transformer, was adjusted to give a maximum reading. The region of power observed was, in general, a sharp peak with almost no power readings greater than a milliwatt being observed at frequencies of more than ±0.5 GHz away from the center frequency of oscillation (F_0). Efficiencies as high as 5% at a duty cycle of 10% were observed (40 mW RF power at 85.6 GHz). In general, back short tuning was not able to shift F_0 along the frequency spectrum in any systematic manner.

It is interesting to note that some devices with low current levels would not oscillate in the pulsed mode but did oscillate in the CW regime. This behavior can be attributed to heating, which lowers the velocity-field (V(E)) curve and increases the cathode current levels. This provides for an intersection of the V(E) curve with the cathode control characteristic; domain nucleation results[1,2].

Investigation of the amplification characteristics were carried out using a reflection amplifier configuration. The tests were run in the CW mode in order to obtain more accurate readings from the power detector. The observed behavior fell into several catagories (again depending on which choke was used): 1) oscillation only, with no prethreshold amplification; 2) amplification (increasing with increasing bias) followed by oscillation; 3) amplification increasing then decreasing with increasing bias followed by (in some cases) oscillations. While the highest efficiency device did give the highest gain (12dB at 90 GHZ), the higher efficiency devices generally did not produce gains higher than the lower efficiency devices. Furthermore, the highest gains were obtained just prior to reaching the threshold current and voltage conditions. The narrow-band circuits displayed a shift in the gain curve along the frequency axis (90 to 96 GHz) as the back short was moved with almost no decrease in gain at the peak gain curve.

3. THEORY

Theoretical calculations were carried out assuming the cathode junction was a modified Schottky barier having the form (3)

$$I = I_r\,(\exp(-qV/nkT)-\exp(((1/n)-1)qV/kT))$$

where I_r is the reverse saturation current, q the electronic charge, and k the Boltzmann constant. The analysis utilized the drift and diffusion equation, which was solved by means of finite diferences. I(V) curves were obtained and fit to the experimentally generated curves by varying the ideality factor n. They predicted the threshold point to better than 10%, and also the region of bias corresponding to gain and loss[2,3]. The high efficiency devices were found to have the intersection of the cathode control characteristic and V(E) curve in the region near velocity saturation, as shown in Figure 1. The influence of the cathode contact caused most of the device field to fall within the NDM region, with the field decreasing gradually from cathode to anode. Figure 2 compares an experimentally measured and numerically determined I(V) curve for a particular sample.

Solutions of the first 3 moments of the BTE[4] for uniform field devices predicts V(E) trajectories, for the large signal case, that vary markedly from the static V(E) relationship. Regions of velocity overshoot, NDM, and a region of velocity saturation as the RF voltage decreases are observed. The situation for small signals is different. The V(E) trajectory follows the static curve more closely. Calculations utilizing non-uniform field

configurations are presently being carried out. The optimum small signal gains seem to occur in devices with relatively uniform field configurations (near-threshold); the results in the small signal regime may not differ greatly from what is being reported here.

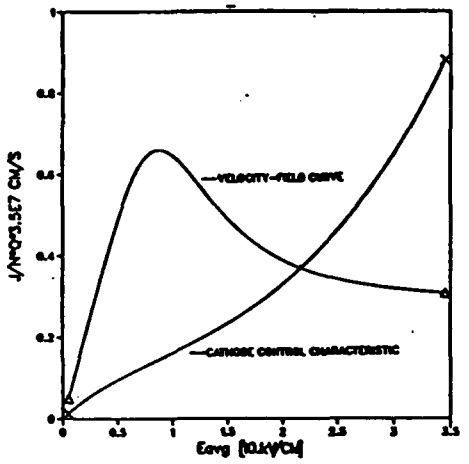


FIGURE 1
Cathode control characteristic and V(E) curve

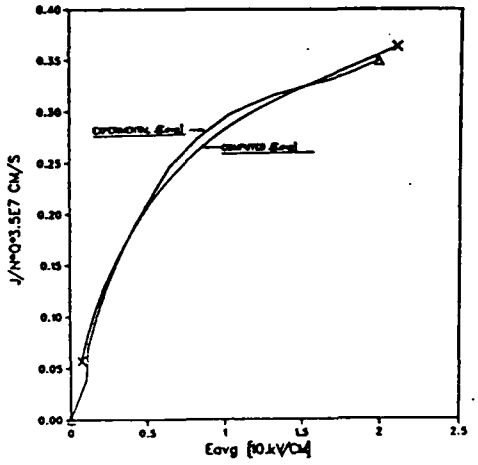


FIGURE 2
J(E) characteristic of sample (experimental and computed)

4. CONCLUSIONS

In summary, near-micron InP TED's were studied experimentally and theoretically. A model using the drift and diffusion equation accurately models the static I(V) characteristic as well as predicts the threshold values. The highest values of computed small-signal negative resistance were obtained for devices with nearly uniform field configurations with the highest fields at the cathode in the shallow NDM regime. Since the electric field becomes highly non-uniform and increases sharply from cathode to anode for biases greater than threshold, high eficiency devices will have to be stabilized by external circuitry to achieve maximum gain.

Although it has been shown that in the large signal regime the V(E) trajectory does not follow the static curve, a drift and diffusion analysis can still be utilized to evaluate the potential performance of these devices as oscillators and amplifiers. We have seen that high efficiency devices have most of the device in the NDM region at threshold, providing for a larger overall "growth-factor" for these devices over other field configurations. Finally, it has again been demonstrated that it is the cathode contact that controls not only the oscillatory but also the amplifying behavior of the device.

ACKNOWLEDGEMENTS

We would like to thank B. Fank and J. Crowley of the Varian Solid State Microwave Division for supplying the devices and for providing experimental facilities and expertise in the amplification measurements.

REFERENCES

1. M.P. Shaw, H.L. Grubin and P.R. Solomon, The Gunn-Hilsum Effect, (Academic Press, N.Y., 1979)
2. H. Kroemer, IEEE Trans. Electron. Devices ED-11, 819 (1968).
3. V.L. Rideout, Sol. St. Elec. 18, 54 (1975).
4. R. Bosch and H.W. Thim, IEEE Trans. Electron. Devices ED-21, 16 (1974).

Reprint 18

TRANSPORT AND MATERIAL CONSIDERATIONS FOR SUBMICRON DEVICES

H.L. GRUBIN

Scientific Research Associates Inc.

Glastonbury, Connecticut USA

INTRODUCTION

The purpose of this lecture series is to identify the important areas of device physics and the questions related to them. In addressing these problems we will emphasize an understanding of transport within these devices and the role of the circuit, the boundaries and the material variations. We will also scrutinize some of the formal underpinnings of device physics. Thus, the lectures will be primarily theoretical. With regard to experimental device physics studies, direct evidence of high speed, submicron scale effects is sparse, and is likely to remain so for the near future. Although, here, recent activities by workers at Bell[1], Los Alamos[2], and the University of Paris--Orsay[3], on submicron, high speed device phenomenology, may provide some important keys to device related effects.

In the absence of a broad experimental program in the area of submicron devices, we are likely to see an increased reliance on numerical simulation as the principle vehicle for determining the physics of device operation. This, of course, places a considerable burden on those who perform device simulations. But the general availability of numerical programs to those who do not develop them will widen the use of numerical studies, particularly in a direct coupling to experiment. In any case, the accumulated studies of the past half century has tended to provide a natural classification of device physics into three groupings: (1) classical device physics, (2) submicron device physics, and (3) ultrasubmicron device physics, where quantum effects may dominate.

The domain of classical semiconductor device physics is one where the temporal scales of interest are generally large enough that accelerative effects are often ignored. These approximations find their most common application in the drift and diffusion equations, where the relation between carrier velocity and applied field is in terms of a local mobility.

The first departure from classical device physics introduces the idea of non-locality where the response of a particle at a point "x" and time "t" is dependent on disturbances at earlier times $t'<t$ and at points $x'<x$. This is the critical feature of the Boltzmann transport equation (BTE), and is one that has found expression in Chambers[4] path integral solutions as well as some initial work of Pippard[5]. While non-local effects have appeared only recently in semiconductor device studies, they have been the object of serious study for the past twenty-five years. Such phenomena as the anomalous skin effect[6], magneto-acoustic attenuation[7] are based on these effects. They are also critical to hydrodynamic studies.

Other departures arise when spatial scales are of the order of the deBroglie wavelength associated with a quasi-particle, and arise when scattering events can no longer be regarded as occurring instantaneously. Approximate spatial and temporal scales associated with these three device divisions are identified in the first two tables and figure 1. Further, each of the above device groupings are identified by a specific set of generic device equations. The key point that must be kept in mind, is that the development of these equations is keyed to the development of high frequency, high speed electronic devices. This will be the common thread through our entire discussion, and also forms the basis for this NATO ASI.

TABLE 1. Approximate spatial scales (GaAs)

Term	Value
Active region length	10^{-5}cm, 10^{-4}cm
Impurity separation ($N_o^{-1/3}$) 10^{-6}cm @ $10^{18}/cm^3$	10^{-5}cm @ $10^{15}/cm^3$
Mean free path	10^{-5}cm
Thermal deBroglie wavelength	2.6×10^{-6}cm

TABLE 2. Approximate temporal scales (GaAs)

Term	Value
Transit time (for $v=2\times10^{7}$/cm/s) 5×10^{-12}s @ $L=10^{-4}$cm	5×10^{-13}s @ $L=10^{-5}$cm
Momentum relaxation	3×10^{-13}s @ 300K
Collision duration	2×10^{-14}s

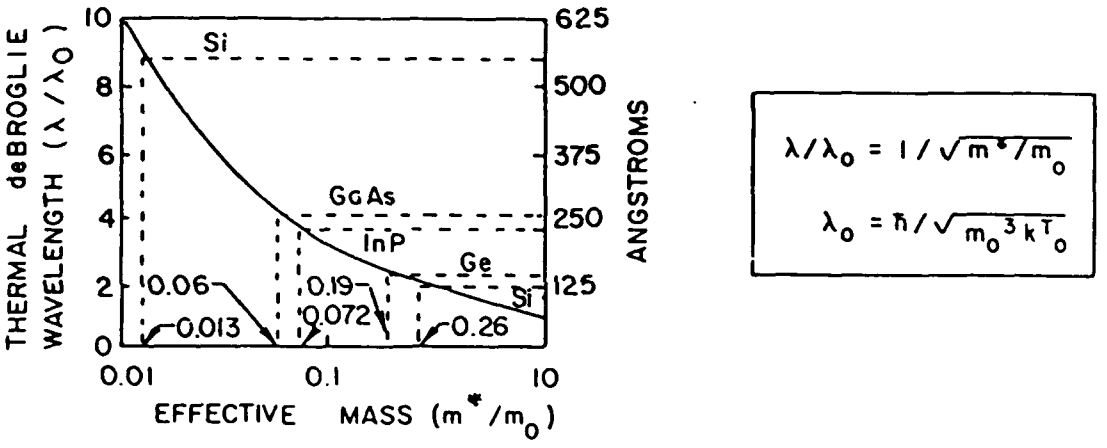


FIGURE 1 . Thermal deBroglie wavelength vs effective mass.

We begin the discussion of submicron device physics with a simple description of the Boltzmann transport equation and its approximate solutions, as well as those areas of such common interest as overshoot phenomena.

THE BOLTZMANN TRANSPORT EQUATION

The conceptual picture most frequently used to describe conduction in semiconductors is that of a gas of carriers interacting with external forces. The gas is characterized by the fact that the state of any of its constituents, e.g., electrons, are independent of all others except at the instant of collision. The state of each constituent is described by specifying its

position and momentum at each instant of time. Rather than treating the transport of individual particles, transport within the semiconductor is discussed statistically. Volume elements within phase space are identified as

$$d^3r = dx_1 dx_2 dx_3 \tag{1}$$

$$d^3p = dp_1 dp_2 dp_3 \tag{2}$$

The differentials above are chosen large enough to contain a large number of carriers and small enough to permit the neglect of density variations within the differential elements.

The number of particles at a point (r, p) is taken as

$$dN = f(r,p,t)d^3rd^3p \tag{3}$$

which may be taken as the definition of the distribution function. Several other definitions are in order. The total number of particles is given by

$$N = \int f(r,p,t)d^3rd^3p \tag{4}$$

The mean value of a function G(r,p) is

$$\ll G \gg = \int G(r,p)f(r,p,t)d^3rd^3p/N \tag{5}$$

In much of the discussion below, we will be interested in local mean values.

If $\phi(p)$ denotes any function of momentum, the local mean value of $\phi(p)$ is given by

$$\langle \phi \rangle = \int \phi(p)f(r,p,t)d^3p/n \tag{6}$$

where

$$n = \int f(r,p,t)d^3p \tag{7}$$

We note that < > is a function of position and time through the distribution function f(r,p,t).

To determine the variation of the distribution function with time, we assume the presence of external and self-consistent

forces, as well as inter-particle forces. We consider time intervals which are large, compared to the duration of a collision, modifications of which have been considered by Ferry and Barker[8]. We also assume that the time interval Δt is small compared to the mean time interval between collision.

If no collisions occur during the time interval Δt, the following transformation applies

$$\mathbf{r} \rightarrow \mathbf{r}' = \mathbf{r} + (\mathbf{p}/m)\Delta t \tag{8}$$

$$\mathbf{p} \rightarrow \mathbf{p}' = \mathbf{p} + \mathbf{F}\Delta t \tag{9}$$

and

$$f(\mathbf{r},\mathbf{p},t)d^3\mathbf{r}d^3\mathbf{p} \rightarrow f(\mathbf{r}',\mathbf{p}',t + \Delta t)d^3\mathbf{r}'d^3\mathbf{p}' \tag{10}$$

Application of Louville's theorem concerning elements of volume in phase space

$$d^3\mathbf{r}d^3\mathbf{p} = d^3\mathbf{r}'d^3\mathbf{p}' \tag{11}$$

leads to the expansion

$$f(\mathbf{r},\mathbf{p},t) \cong f(\mathbf{r},\mathbf{p},t) + \left(\frac{\partial f}{\partial \mathbf{r}} \cdot \frac{\mathbf{p}}{m} + \frac{\partial f}{\partial \mathbf{p}} \cdot \mathbf{F} + \frac{\partial f}{\partial t}\right)\Delta t + \cdots (\ )(\Delta t)^{\ell} \tag{12}$$

In examining transport in semiconductors we generally neglect terms of order $(\Delta t)^{\ell}$ with $\ell>1$; the assumption being that the distribution function does not change appreciably during the interval Δt. According to Sommerfeld[9], this assumption is compatible with considerable changes in $f(\mathbf{r},\mathbf{p},t)$ within one mean free path, because $\Delta t<\tau$, where τ is the mean time between collisions. This latter result is extremely important when examining overshoot phenomena in devices. Thus within this approximation, we obtain the collisionless Boltzmann transport equation

$$\frac{\partial f}{\partial \mathbf{r}} \cdot \frac{\mathbf{p}}{m} + \frac{\partial f}{\partial \mathbf{p}} \cdot \mathbf{F} + \frac{\partial f}{\partial t} = 0 \tag{13}$$

Collisions cause carriers to leave different elements of phase

space with the following balance holding:

$$f(r,p,t)d^3pd^3r - f(r',p',t' + \Delta t)d^3p'd^3r' - \left[\frac{\partial f}{\partial t}\right]_{coll} d^3pd^3r, \qquad (14)$$

from which we obtain the BTE with collisions,

$$\frac{\partial f}{\partial r} \cdot \frac{p}{m} + \frac{\partial f}{\partial p} \cdot F + \frac{\partial f}{\partial t} = \left[\frac{\partial f}{\partial t}\right]_{coll} \qquad (15)$$

Among the various problems in semiconductor device physics, one involves the proper formulation of the collision integral. We will not concern ourselves with this during the following lectures. Rather, we will borrow the necessary terms for establishing a suitable submicron device physics.

Before proceeding to an actual description of submicron device transport, which will necessarily involve solving a complicated set of equations derived from the BTE, it is worthwhile presenting a formal solution to the BTE within the framework of the relaxation time approximation. This formal solution will highlight the significant differences between the more standard approach to examining device physics, and what is currently required.

Within the framework of the relaxation time approximation, the collision integral takes the form

$$\left[\frac{\partial f}{\partial t}\right]_{coll} = -\left[\frac{f-f_o}{\tau}\right] \qquad (16)$$

in which we assume that τ may be a function of energy. The formal solution to the above equation basically involves transforming the BTE, as a partial differential equation into an ordinary differential equation

$$\frac{df}{ds} + \frac{f}{\tau} = \frac{f_o}{\tau} \qquad (17)$$

where "s" denotes a "characteristic" path variable. There is, of course, considerable history associated with this, see e.g., Chambers[4], and Thornber[10]. Basically we consider the variables

t(s), r(s) and p(s) such that

$$\frac{df}{ds} = \frac{\partial f}{\partial t} \cdot \frac{dt}{ds} + \frac{\partial f}{\partial r} \cdot \frac{dr}{ds} + \frac{\partial f}{\partial p} \cdot \frac{dp}{ds} \tag{18}$$

where we require

$$\frac{dt}{ds} = 1 \quad , \quad \frac{dr}{ds} = \frac{p}{m} \quad , \quad \frac{dp}{ds} = F \tag{19}$$

For the simple case[10] where F is taken to be independent of t, r and p

$$\begin{aligned} t(s) &= s - s_o + t_o \\ r(s) &= \frac{p}{m}(s - s_o) + r_o \\ p(s) &= F(s - s_o) + p_o \end{aligned} \tag{20}$$

Equation (18) has the formal solution

$$f\Big[r(s),p(s),t(s)\Big] = f\Big[r(s_o),p(s_o),t(s_o),\Big] \exp - \int_{s_o}^{s} \frac{ds'}{\tau[p(s')]}$$

$$+ \int_{s_o}^{s} \frac{ds' f_o[r(s'),p(s'),t(s')]}{\tau[p(s')]} \exp - \int_{s'}^{s} \frac{ds''}{\tau[p(s'')]} \tag{21}$$

When the following identification is made

$$s_o = t_o \ , \qquad t(s) = s \tag{22}$$

The formal solution takes the form

$$f(r,p,t) = f\left[r - \frac{p}{m}(t-t_o), p - F(t-t_o), t_o\right] \exp\left[-\int_{t_o}^{t} dt''/\tau[p - F(t-t'')]\right] \tag{23}$$

$$+ \int dt' \, \frac{f_o[r - \frac{p}{m}(t-t'), p - F(t-t'), t']}{\tau[p - F(t-t')]} \exp\left[-\int_{t'}^{t} \frac{dt''}{\tau[p - F(t-t')]}\right]$$

The physical significance of the above solution is contained in the exponential contributions. Here the probability that an electron initially at

$$r - \frac{p}{m}(t-t_o), \quad p-F(t-t_o), \quad t_o \tag{24}$$

has arrived at

$$r, \quad p, \quad t \tag{25}$$

without scattering is

$$\exp - \int_{t'}^{t} \frac{dt''}{\tau[p-F(t-t'')]} \tag{26}$$

Thus, all transport is governed by non-local events. This is the point we wish to emphasize. We now apply this to a very specific and familiar problem.

We consider the situation where a system of carriers is subjected to an applied external field and scattering centers, and that the field represents a small departure from equilibrium:

$$f = f_o + f_1 \tag{27}$$

where f_o represents an equilibrium time independent distribution function. Under the typical small perturbation approximation, the BTE becomes

$$\frac{\partial f_1}{\partial t} + \frac{p}{m} \cdot \frac{\partial f_1}{\partial r} + \frac{f_1}{\tau} = - \left(F \cdot \frac{\partial f_o}{\partial p} + \frac{p}{m} \cdot \frac{\partial f_o}{\partial r} \right) = g \tag{28}$$

If we now solve this equation using the method of characteristics and concentrate only on the particular solutions

$$f_{1,p} = \int_{t_o}^{t} dt' g[r - \frac{p}{m}(t-t'), t'] \exp - \left(\int_{t'}^{t} \frac{t-t''}{\tau(p)} \right) \tag{29}$$

we see an explicit dependence of the perturbed distribution function on past history. Now while this will be important for

submicron devices, there exists an entire class of devices whose spatial scales are longer than a mean free path and whose temporal scales are longer than the mean time between collisions. On these longer spaces and time dependent scales the spatial and temporal variation of the integral in equation (29) is not important, and can be taken out of the integral. Under these circumstances the perturbed distribution function, for times longer than the mean time between collisions, becomes

$$f_{1,p} \approx -\tau\left(F\cdot\frac{\partial f_0}{\partial p} + \frac{p}{m}\cdot\frac{\partial f_0}{\partial r}\right) \tag{30}$$

Equation (30) is the typical starting point for most semiconductor device transport descriptions. It requires for its validity that only insignificant variations occur over a mean free path, or over a mean time between collisions. It has nevertheless been used for situations where there are strong field gradients over very short distances, such as that within the vicinity of a strongly asymmetrical PN junction. It may be expected that many of the more common device calculations such as those associated with bipolar silicon devices, MOS devices will require significant scrutiny to determine if the more commonly used assumptions are valid in design on a submicron scale.

The above description of transport as governed by the non-local feature of the BTE will now be applied to examining submicron high field transport. The requirements for this discussion have at their basis, the need for a set of governing differential equations. The equations that we use to describe submicron and near micron transport are based upon the moments of the BTE. Typically, these moments are often truncated after the first three moments although in principle, an infinite number of moment equations will provide an exact solution to the BTE. In the discussion that follows we will apply the use of the displaced Maxwellian distribution function to obtain a set of device equations.

THE DISPLACED MAXWELLIAN DISTRIBUTION AND THE MOMENT EQUATIONS

As indicated in the above discussion, within the relaxation time approximation, a momentum relaxation time may be defined, where $\tau(p)$ depends on energy through momentum. As discussed by Keyes[11], we can add a term to equation (16) which describes the effects of electron-electron scattering by observing that if the external fields and scattering mechanisms (exclusive of e-e scattering) were suddenly removed, the e-e collisions would provide a mechanism for transferring energy and momentum between different

parts of the distribution function. This would bring f to some other equilibrium function, f_o^*. f_o^* must be different from f_o, since with f_o the electrons have zero total momentum. Since e-e collisions cannot alter the total electron momentum, f_o^* must have the same total momentum as the nonequilibrium distribution function f, at the time the field and scattering centers were removed.

The determination of f_o^*, has been carried out by the methods of statistical mechanics, and it has been shown that f_o^* is a Boltzmann distribution moving with velocity v=p*/m, i.e.,

$$f_o^*(r,p,t) = ae^{-|p-p^*|^2} \tag{31}$$

The prominence of the displaced Maxwellian lies in its successful application to examining transport in III-V materials, particularly gallium arsenide (GaAs). Its use in transport is primarily through the moments of the BTE.

More generally, the moments of the BTE are obtained by multiplying equation (15) by successive powers of p, i.e.,

$$\phi^{(\ell)}(p) = p^{\ell}, \ell = 0,1,2, \cdots \tag{32}$$

The result is

$$\frac{\partial}{\partial t}(n\langle\phi^{(\ell)}\rangle) + \frac{\partial}{\partial r}\cdot\left[\frac{n}{m}\langle p\phi^{(\ell)}\rangle\right] = Fn\cdot\left\langle\frac{\partial\phi^{(\ell)}}{\partial p}\right\rangle + \frac{\partial}{\partial t}(n\langle\phi^{(\ell)}\rangle)_{coll} \tag{33}$$

where

$$\frac{\partial}{\partial t}(n\langle\phi^{(\ell)}\rangle)_{coll} = \int\phi^{(\ell)}\left(\frac{\partial f}{\partial t}\right)_{coll}d^3p \tag{34}$$

In the absence of collisions, an exact solution to the BTE can be obtained by solving the infinite set of coupled equations. This is not satisfactory, particularly when the effects of collisions must by accounted for. The situation then calls for an estimate of the distribution function. The often used assumption is that of a displaced Maxwellian which cuts off the moment equations at the number "3". The degree to which the distribution function is well represented by the displaced Maxwellian has been examined by a number of workers with the general conclusion that as far as bulk transport is concerned, it is likely to be a reasonable assumption at high carrier densities[12]. It is likely to be seriously

unrealistic to assume that the distribution function is spherical in momentum space. Such a situation would tend to ignore important thermal gradient effects as well as viscous effects, which are known to appear in fluid flow and in plasma physics. The situation is such that the displaced Maxwellian needs generalization. One such approach, which we consider below was discussed by Sommerfeld[9]. Here the distribution function of interest is expressed as

$$f(r,p,t) = \left[1 + A_k \frac{\partial}{\partial p_k} + A_{k\ell} \frac{\partial^2}{\partial p_k \partial p_\ell} + A_{k\ell m} \frac{\partial^3}{\partial p_k \partial p_\ell \partial p_m} + \cdots\right] f_o^*(r,p,t) \tag{35}$$

with the coefficients to be determined below. Additionally, the following convention is assumed:

$$A_k \frac{\partial}{\partial p_k} = A_x \frac{\partial}{\partial p_x} + A_y \frac{\partial}{\partial p_y} + A_z \frac{\partial}{\partial p_z} \tag{36}$$

The constraints on this expansion are that the mean carrier density

$$n = \int f(r,p,t) d^3p = \int f_o^*(r,p,t) d^3p \tag{37}$$

and that the mean momenta

$$n\langle p\rangle = \int f(r,p,t) p d^3p = \int f_o^*(r,p,t) p d^3p \tag{38}$$

which yields

$$\langle p\rangle = p^* \tag{39}$$

It is further required that the mean, isotropic, thermal kinetic pressure satisfy the following requirement

$$\mathcal{P} = \frac{2}{3m} \int (p-p^*)\cdot(p-p^*) f d^3p = \frac{2}{3m} \int (p-p^*)\cdot(p-p^*) f_o^* d^3p = \frac{n}{2\gamma m} \tag{40}$$

The identification of a suitable electron temperature model is

through an assumed gas law, which is taken as

$$\mathcal{P} = nk_BT \tag{41}$$

Thus,

$$\gamma = \frac{1}{2mk_BT} \tag{42}$$

And along with a suitable normalization that includes spin the displaced Maxwellian is of the following form

$$f_o^* = \frac{n}{(2\pi mk_BT)^{3/2}} \exp\frac{-|p-p^*|^2}{2mkT} \tag{43}$$

The expansion coefficients, which identify the important nonspherical contributions of the distribution function are identified as follows: (1) the linear expansion coefficient is zero; (2) the expansion coefficient A_{kl} is identified as a stress tensor; and (3) the expansion coefficient A_{klm} is associated with a heat flux. The incorporation of these terms leads to the following set of balance/moment equations, from which the numerical simulation of the submicron devices is obtained.

Carrier Balance:

$$\frac{\partial n_r}{\partial t} + \frac{\partial}{\partial x_i} v_r^i n_r = - n_r\Lambda_r + n_s\Lambda_s \tag{44}$$

Momentum Balance:

$$\frac{\partial p_r^j}{\partial t} + \frac{\partial}{\partial x_i} v_r^i p_r^j = -en_rF^j - \frac{\partial}{\partial x_i} n_r k_B T_r + \mu_r \underline{\frac{\partial^2}{\partial x_i \partial x_i} v_r^i} - p_r^j \Pi_r \tag{45}$$

Energy Balance:

$$\frac{\partial W_r}{\partial t} + \frac{\partial}{\partial x_i} v_r^i W_r = -en_r v_r^i F^i - \frac{\partial}{\partial x_i} v_r^i k_B T_r + \underline{\frac{\partial}{\partial x_i} \kappa_r \frac{\partial}{\partial x_i} T_r}$$

$$- \frac{3}{2} n_r k_B T_r \Sigma_r + \frac{3}{2} n_s k_B T_s \Sigma_s \tag{46}$$

Equations (44) through (46) are the governing equations used below to discuss submicron transport. In these equations

$$p_r = n_r p_r = n_r m_r v_r \tag{47}$$

and

$$W_r = n_r\left[\frac{1}{2}m_r v_r^i v_r^i + \frac{3}{2}k_B T_r\right] \tag{48}$$

the subscripts are used to emphasize the fact that we are considering multi-valley species transport. The underlined terms in these equations identify the stress term contributions (equation (45)) and the thermal heat transport contribution (equation (46)). Stress contributions in equation (46) have been ignored. The coefficient in the momentum balance equation has dimensions of viscosity, and this term behaves as a viscous contribution.

We examine the above equations in detail below. In order to do this effectively, we will demonstrate how, for the case of the dipole transit, the results obtained from these equations differ from those obtained using the standard drift and diffusion

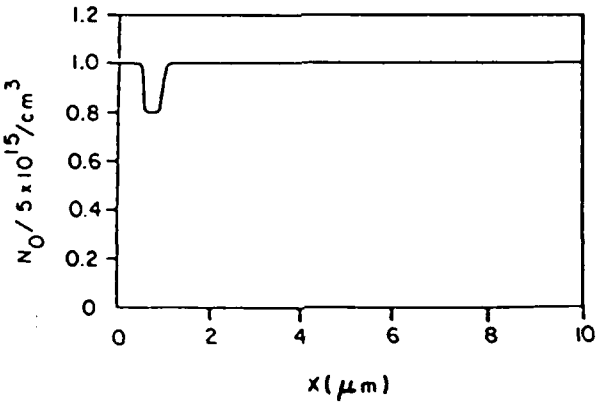


FIGURE 2. Doping profile used for dipole simulations.

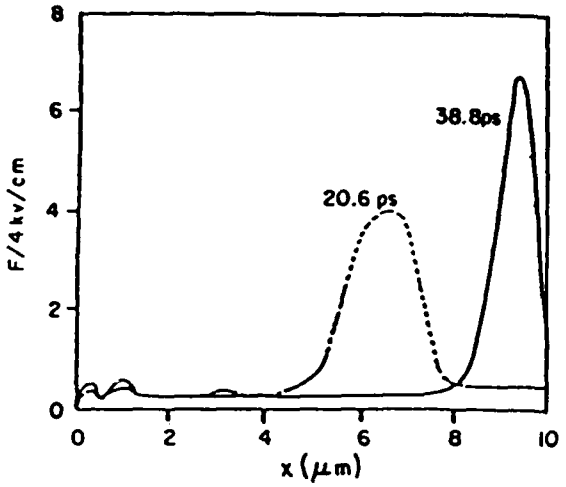


FIGURE 3a. Electric field vs distance vs time, for a propagating layer.

equations. To illustrate this, we consider a 10μm-long GaAs element with a doping level of $5x10^{15}/cm^3$ and a doping notch near the cathode boundary, as shown in figure 2. The device is clamped at 4 volts (there is no load line). At this voltage level domain oscillations occur. The oscillations are displayed in figures 3 and 4. (We note: In this simulation Poisson's equation is solved, along with the moment equations.)

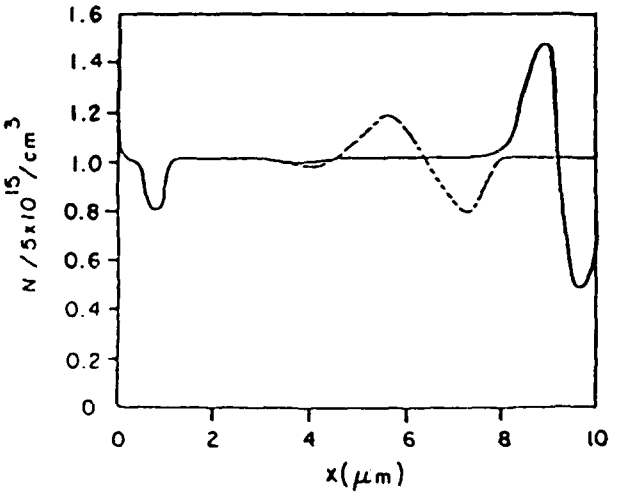


FIGURE 3b. Mobile carrier density at the time of figure 3a.

Figures 3a through 3d display, respectively, the time dependent behavior of electric field, total carrier density, gamma valley carrier density and gamma valley velocity, for a propagating domain (additional development is shown in figure 4). The domain is first shown approximately halfway down the element at a time of 20.6ps; and later at the anode at 38.8ps.

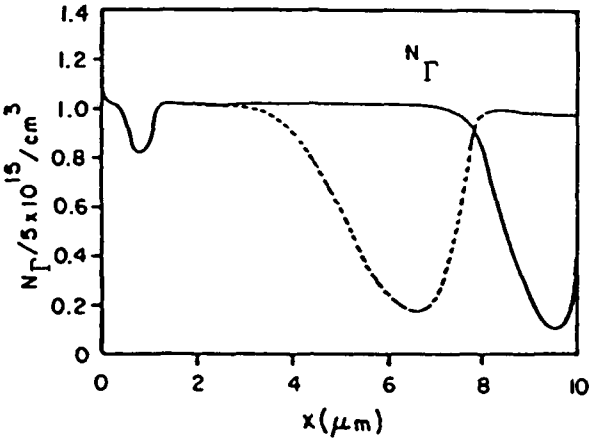


FIGURE 3c. Γ-valley carrier density at the time of figure 3c.

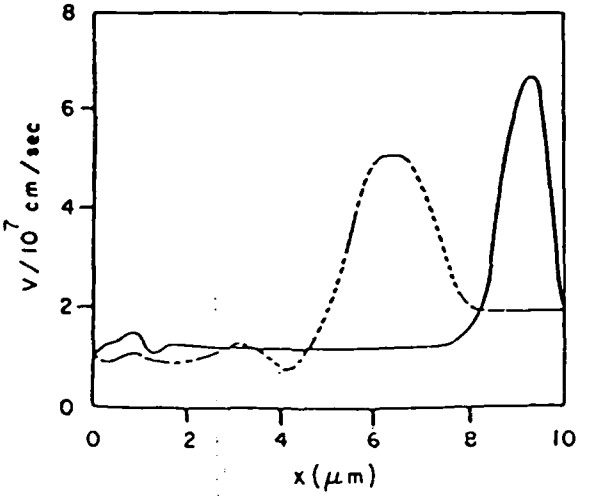


FIGURE 3d. Γ-valley carrier density at the time of figure 3a.

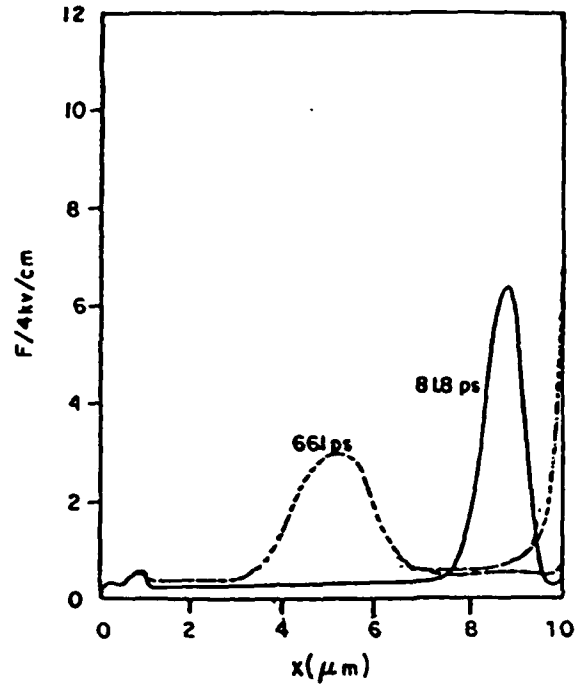


FIGURE 4a. As in figure 3a but at later times.

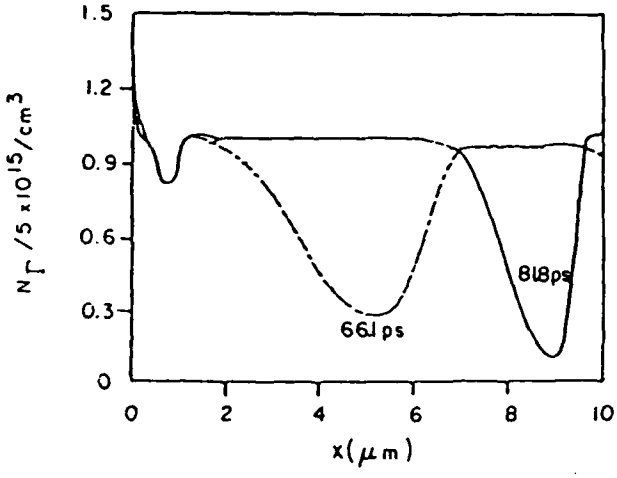


FIGURE 4b. As in figure 3c but at later times.

The travelling domain has all the appearance of a classical transit time oscillation. This, however, is deceptive because immediately upstream from the domain where the field has decreased to below the threshold for negative differential resistance (NDR), and where semiconductor drift and diffusion would indicate that relaxation to quiescent values of carrier density and near ambient electron temperatures is complete, the moment equations indicate otherwise. The moment equations show a substantial satellite valley population even after the high field region has passed and only low values of field remain. This significant relaxation effect was first discussed by Rees[13]. Its numerical origin arises from the first order spatial and temporal derivatives in the left hand side of the moment equations (45) and (46). In the absence of these terms, we, of course, recover the drift and diffusion equations. The results of this calculation are significant and cannot be described in any reasonable way by the drift and diffusion concepts. Further differences arise here, because at these moderate values of electric field, the satellite valleys carry only a negligible amount of current. It is as though in an approximate sense, once the carriers have been transferred to a satellite valley, they are effectively out of the conduction process. (This is clearly not the case at high bias levels). Thus is a real sense, stable domain theory must be reexamined. Of interest here is that at the peak field where the total carrier density is equal to the background, i.e., $N=N_0$, the Γ valley population is less than N/4, yet the velocity of these carriers, figure 3d, is greater than twice the mean velocity, obtained from the relation

$$\langle v \rangle = J/n_0 e \qquad (49)$$

Nonequilibrium transport as depicted above is of more than academic interest. Classical modeling of a 10μm-long device is predicated on the notion that proximity effects are negligible. Based upon the size of the high field domain as depicted in figure 3a, this would appear to be a valid assumption. On the other hand, transport in the wake of these domains, at least for doping levels, near $5\times10^{15}/cm^3$ extends over a distance of nearly twice the width of the domain, figures 3c and 4d. This factor of two should be cause of considerable concern when transit time devices below 5μm in active region length are designed.

PARAMETRIC DEPENDENCE OF THE MOMENT EQUATIONS

At this point, it is necessary to examine several parametric aspects of the moments of the BTE. To facilitate this parametric

study we introduce the following dimensionless primed variables:

$$n' = n/n_{ref} \qquad v' = v/v_{ref}$$
$$t' = t/t_{ref} \qquad \Psi' = \Psi/\Psi_{ref}$$
$$x' = x/x_{ref} \qquad t' = t/t_{ref}$$

with similar descriptions for other variables, and where

$$x_{ref} = v_{ref}\, t_{ref}$$

Additionally, Ψ is obtained from a solution to Poisson's equation

$$\frac{\partial^2 \Psi}{\partial x_i \partial x_i} = + \frac{e}{\epsilon}(n-n_o) \tag{50}$$

for a permittivity ϵ, and we are observing the summation convention. The moment equations and Poisson's equation in the primed variables become

Carrier Balance:

$$n_r' = n_s' \frac{\Lambda_s'}{\Lambda_r'} - \frac{1}{\Lambda_r'} \left\{ \frac{\partial n_r'}{\partial t'} + \frac{\partial}{\partial x_i'} v_r^{i'} n_r' \right\} \tag{51}$$

Momentum Balance:

$$n_r' m_r' v_r^{j'} =$$
$$a n_r' \frac{\partial \Psi'}{\partial x_j'} - \frac{1}{\Pi_r'} \left\{ b \frac{\partial n_r T_r}{\partial x_j'} - c \frac{\partial^2 \mu v_r^i}{\partial x_i' \partial x_i'} + \frac{\partial}{\partial t'} n_r' m_r' v_r^{i'} + \frac{\partial}{\partial x_i'} n_r' m_r' v_r^{i'} v_r^{j'} \right\} \tag{52}$$

Energy Balance:

$$n_r' T_r' = n_s' T_s' \frac{\Sigma_s'}{\Sigma_r'} + d n_r' v_r^i \frac{\partial \Psi'}{\partial x_i'} \cdot \left[\frac{2}{3b}\right]$$
$$- \frac{1}{\Sigma_r'} \left\{ b \frac{\partial}{\partial x_i'} n_r' v_r^{i'} T_r' + e \frac{\partial}{\partial x_i'} \kappa' \frac{\partial}{\partial x_i'} T_r + \frac{\partial}{\partial t'} W_r' + \frac{\partial}{\partial x_i'} v_r^{i'} W_r' \right\} \cdot \left[\frac{2}{3b}\right] \tag{53}$$

Poisson's equation:

$$\frac{\partial^2 \Psi'}{\partial x_i' \partial x_i'} = -\frac{\partial F_i'}{\partial x_i'} = f(n' - N_0') \tag{54}$$

In the above

$$W_r' = \frac{1}{2} n_r' m_r' v_r^{i'} v_r^{i'} + \frac{3}{2} b n_r' T_r' \tag{55}$$

The primed scattering rates in the above equations are, $\Lambda'_r = \Lambda_r \cdot t_{ref}$ with similar expressions for the other scattering rates. For dimensional reasons, as well as to establish a dialogue concerning the dissipative terms and the thermal conductivity contribution, the viscosity has been expressed via the Maxwell relation

$$\mu = \frac{1}{3} n_{ref} \, m_{ref} \, \langle v \rangle \ell \tag{56}$$

where $\langle v \rangle$ is a characteristic mean velocity and ℓ an effective mean free path. The thermal conductivity has been expressed in terms of a Widemann-Franz relation

$$\kappa = \frac{\mathcal{L} \sigma T k_B^2}{e^2} \tag{57}$$

where

$$\sigma = \frac{n_{ref} \, e^2 \langle \tau \rangle}{m_{ref}}$$

and $\langle \tau \rangle = \ell / \langle v \rangle$. In these terms the bold coefficients in equations (51) through (54) are

$$a = \frac{e t_{ref}}{m_{ref} \Pi_r'} \frac{F_{ref}}{v_{ref}} \tag{58}$$

where

$$F_{ref} = \frac{\Psi_{ref}}{x_{ref}} , \tag{59}$$

$$b = \frac{k_B T_{ref}}{m_r v_{ref}^2} \tag{60}$$

$$c = \frac{1}{3} \frac{\langle v \rangle \ell}{v_{ref} x_{ref}} \tag{61}$$

$$d = a \frac{\Pi_r}{\Sigma_r} \tag{62}$$

$$e = b^2 \frac{\ell \langle r \rangle}{t_{ref}} \tag{63}$$

$$f = \frac{e}{\epsilon} \frac{x_{ref} n_{ref}}{F_{ref}} \tag{64}$$

The connection between these bold letter coefficients and dimensionless hydrodynamic quantities is

$$b = \frac{1}{M^2 \gamma} \qquad (\gamma = \frac{3}{5}) \tag{65}$$

where M is the Mach number, and

$$c = \frac{1}{Rey} \tag{66}$$

where Rey is the Reynolds number. We consider the broad consequences of the normalization.

The first point that should be noticed is that the effects of the nonspherical distribution, as represented by thermal conductivity and viscosity are likely to be more pronounced at the shorter device lengths. The situation with regard to the thermal conductivity is illustrated in figures 5 and 6.

In figure 5, we display the distribution of field and mobile carrier density for a 2μm-long element with the indicated structure and a high thermal conductivity consistent with Eq. (57)

[14]. We note the presence of highly nonuniform fields, figure 5b, marginal electron transfer, figure 5c, and local Γ-valley carrier velocities in excess of $5x10^7$ cm/sec, figure 5d.

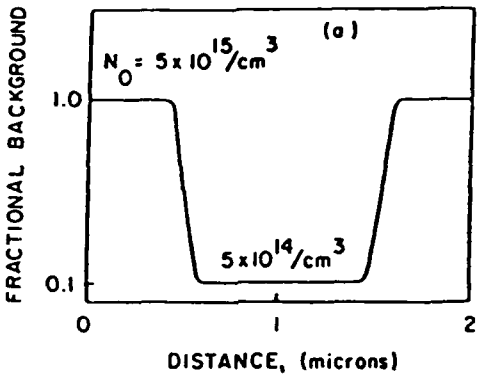


FIGURE 5a. Doping profile for high thermal conductivity calculations.

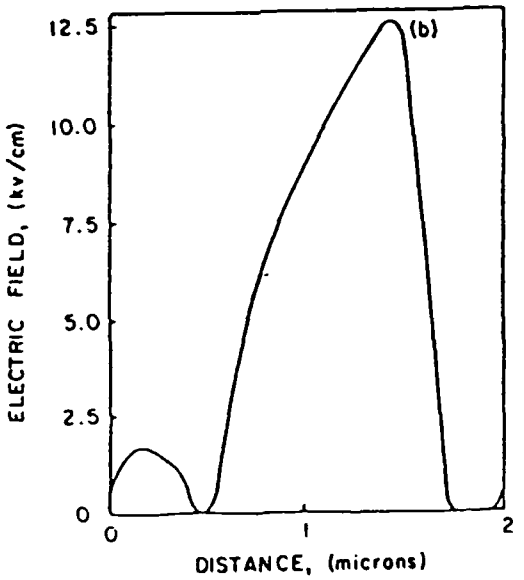


FIGURE 5b. Field profile for high thermal conductivity calculations.

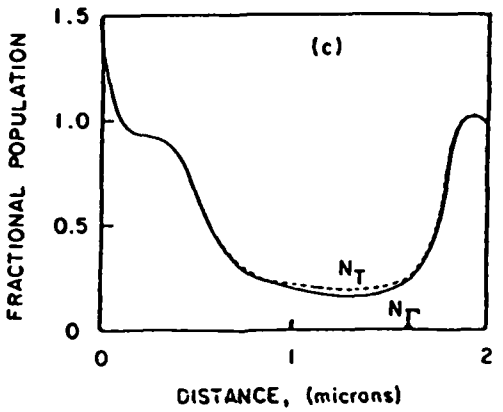


FIGURE 5c. Carrier distribution for high thermal conductivity calculations.

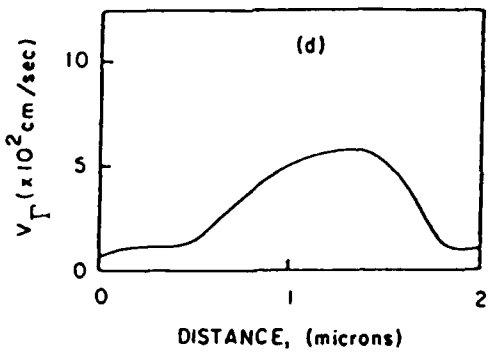


FIGURE 5d. Velocity distribution for high thermal conductivity calculations.

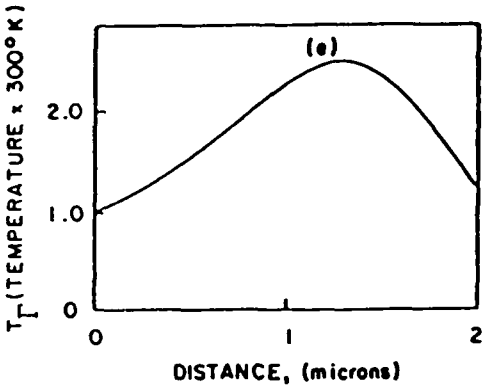


FIGURE 5e. Temperature distribution for high thermal conductivity calculations.

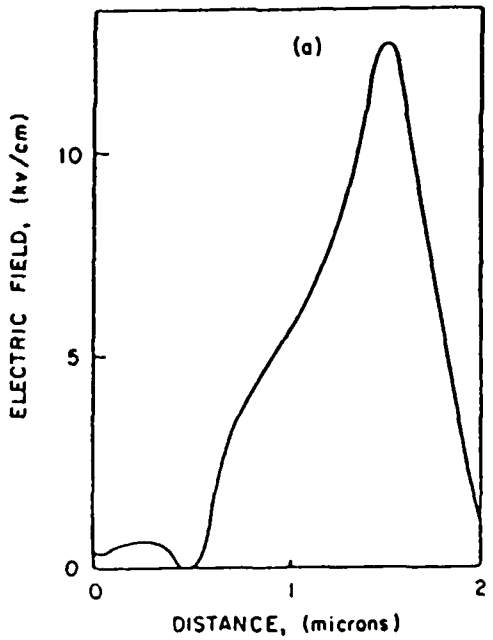


FIGURE 6a. Field profiles for a lower thermal conductivity of.

Figure 6 is for a lower value of the thermal conductivity [14]. A decrease in the thermal conductivity by an order of magnitude results in an expected increase in the overall carrier temperature and an increase in the number of carriers that have

transferred to subsidiary valleys, figure 6b. The results suggest that further study is required with respect to the importance of the thermal conductivity with regard to device design.

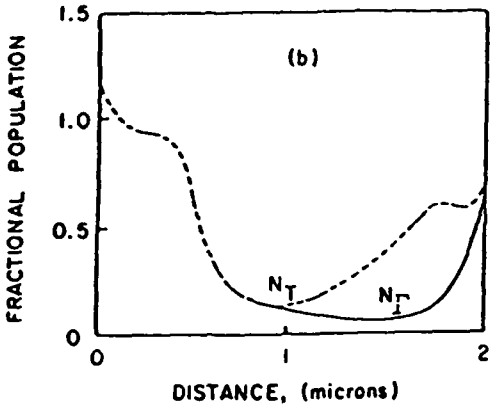


FIGURE 6b. Free carrier density for a low thermal conductivity.

The situation with regard to the viscous contribution is considered next. First, viscous contributions are basically dissipative. This is displayed in figures 7a and 7b, where for the structure of figure 2 and a bias level of two volts, the steady

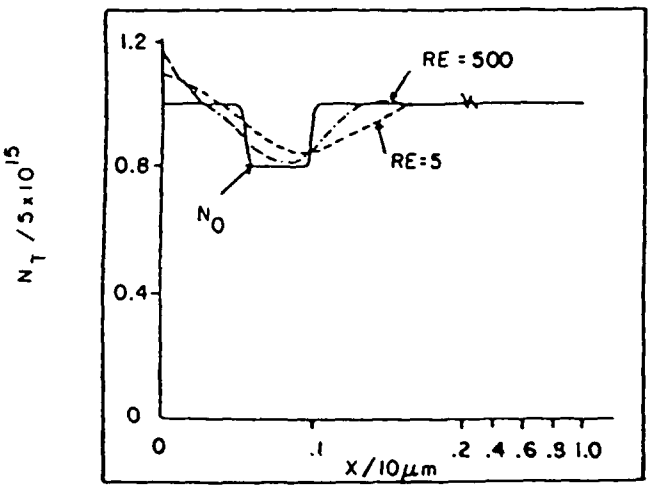
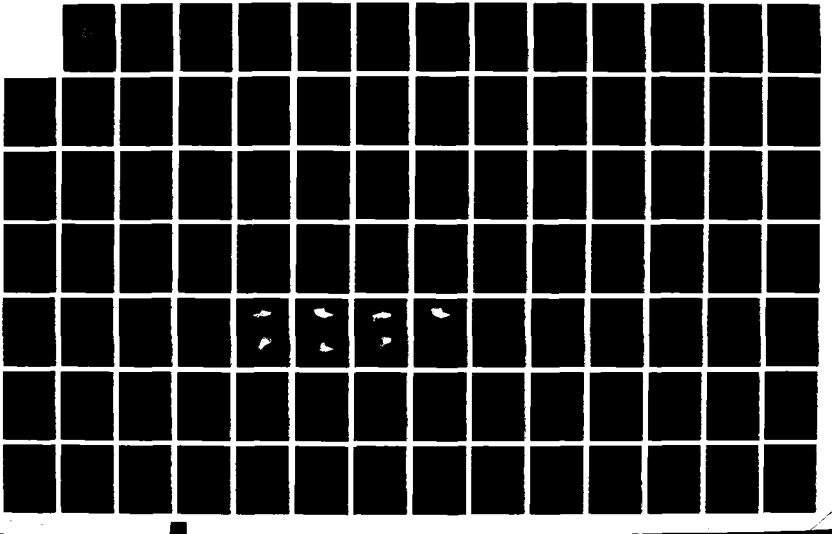
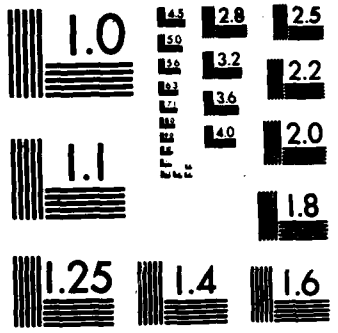


FIGURE 7a. Dependence of charge density on Reynolds number. (Ψ = 2 volts)

AD-A193 380 STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL NEAR-MICRON AND SUB-..(U) SCIENTIFIC RESEARCH ASSOCIATES INC GLASTONBURY CT H L GRUBIN ET AL. 5/6

UNCLASSIFIED 15 FEB 88 SRA-R88-920010-F N00014-81-C-0452 F/G 9/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

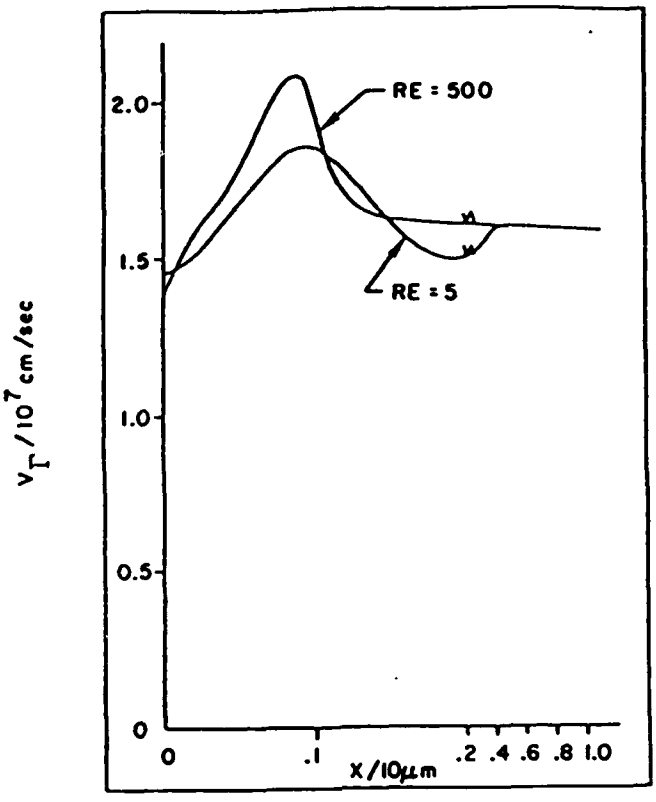


FIGURE 7b. Dependence of Γ-valley velocity on Reynolds number. (Ψ = 2 volts).

state carrier density and Γ valley velocity are sketched for two different values of the Reynolds number - the viscous contributions are different. It is seen from these figures that the smaller the Reynolds number, the smoother the carrier density and velocity profiles. This of course, is expected. The smooth profile, however, tends to result in a larger variation in electric field, as seen from figure 7c as well as from Poisson's equation. Thus, simple predictions about the effect of viscosity on the detailed performance of a device cannot yet be made. More study is clearly in order. We will come back to these equations and their scaling implications later. Meanwhile, we consider the detailed transient behavior of submicron devices.

SPATIALLY INDEPENDENT TRANSIENT TRANSPORT

We have emphasized in these lectures that one of the more important features of transient micron and submicron transport in

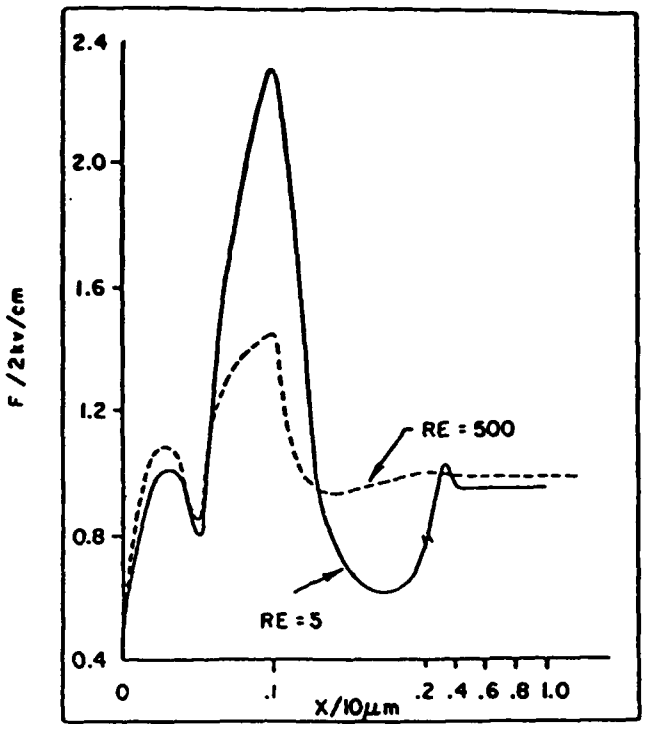


FIGURE 7c. As in Figure 7a, but for electric field distribution.

semiconductors is the non-local contribution. Thus, events at time t depend on an earlier time history, and transient phenomena at a point x depend on disturbances at a point x' and on the propagation characteristics from x' to x. In addition, as dimensions shrink and proximity effects become pronounced, reflections and, hence, time dependent contributions from the downstream, as well as the upstream boundary comes into importance. While these effects are implicit in the discussion of dipole transport through solution to the moments of the Boltzmann transport equation, a fuller discussion is given below. We begin with a discussion of nonequilibrium high field transient effects.

Transient nonequilibrium transport, or velocity overshoot as we have come to know it, is prominently identified with the studies of Ruch[15], although historically one of the earliest contributions is due to Butcher et al[16]. The phenomena, as we begin to understand it better, arises from at least two effects: (1) the incorporation of acceleration in the governing equations, and (2) energy dependent relaxation rates. Additionally, in the III-V

there is also intervalley transfer. Let us first identify the overshoot and then go through a description of its origin.

Figure 8 is a plot of velocity versus time for a GaAs uniform field device subject to a sudden change in electric field. In this curve the long time asymptotic value is that appropriate to steady state values. The interest in this curve lies in it's short time transient which has a peak velocity in excess of it's steady state value. In the case of Si, the peak velocities are only marginally higher than their steady state values, and nothing truly dramatic occurs. The situation as represented by figure 8 is for GaAs. Indeed, it is this difference that is, in good part, responsible for the high activity in the III-V compounds.

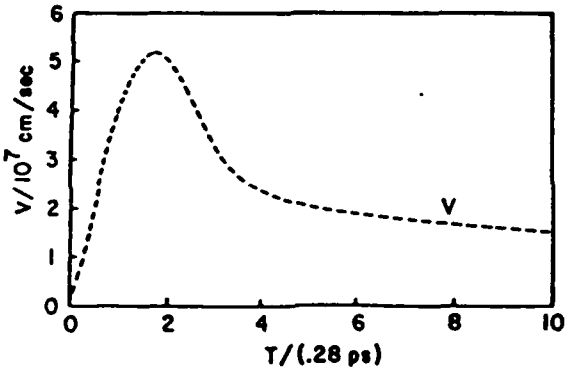


FIGURE 8. Velocity transient for GaAs, from the moment equations. The applied field is 9.7kv/cm.

While much of the early work on velocity overshoot was obtained within the framework of Monte Carlo calculation, the discussion below will be through application of the Moment/Balance equations. These have been discussed in fairly general terms earlier. For spatially uniform fields these equations reduce to a set of simpler coupled ordinary differential equations, which will be used to identify significant nonequilibrium phenomena. For uniform fields, the continuity equation, as well as that for momentum and energy

balance reduce to:

$$\frac{\partial n_r}{\partial t} = - n_r \Lambda_r + n_s \Lambda_s \tag{67}$$

$$\frac{\partial p_r}{\partial t} = - en_r F - p_r \Pi_r \tag{68}$$

$$\frac{\partial W_r}{\partial t} = - en_r v_r F - \frac{3}{2} n_r k_B T_r \Sigma_r + \frac{3}{2} n_s k_B T_s \tag{69}$$

The first equation states that the time rate of decrease of the population of the "r"th conduction (or valence) band is proportional to the rate at which the particles leave, less the rate at which they are replenished. The second equation is, of course, Newton's law with a driving force, the electric field, and dissipative term (momentum relaxation). The third equation describes energy balance. We note that with regard to the acceleration equation, nothing other than the driving term introduces gain. With regard to the energy balance equations, the carriers gain energy from the field when

$$sgnv_i = -sgnF \tag{70}$$

and return energy when the signs are the same. There is also the possibility of either an energy loss through intravalley electron-phonon interaction and energy gain or loss through intervalley transfer.

For uniform fields the entire description of the transient transport through the semiconductor is governed by a balance between the driving field and the scattering rates. These rates identify all of the relevant physics associated with electron-impurity scattering, electron-hole and hole-impurity scattering, electron-phonon and hole-phonon scattering, etc. The scattering rates fold-in the matrix elements and are obtained through several layers of approximation. One approximation is the use of the BTE for situations where inelastic scattering occurs. Another approximation is associated with the fact that we do not generally obtain matrix elements from first principle arguments, and that there does not exist a set of corresponding experiments that allow us to confidently use the results of these matrix elements. Generally, the strength of a specific scattering matrix element is identified by a coupling coefficient, e.g., the

deformation potential, and a scattering rate is computed. Similar procedures are applied to different scattering mechanisms. The resulting scattering rates are then summed. The assigned scattering rates are deemed satisfactory when the calculated steady state field-dependent velocity curve agrees with experiment. The difficulty with this approach is that (1) it is not unique, a variety of different coefficients are likely to give similar results, (2) steady field dependent velocity curves are not obtained experimentally under steady state conditions, (3) consistency does not exist between alternative methods of calculation. Here Monte Carlo and displaced Maxwellian calculations yield somewhat different results (although they are qualitatively similar). The above statements are not meant to suggest that all transport studies should be delayed until all of

TABLE 3. GaAs Parameters Used in Calculations

Parameter	Γ	*L*	X	Common
Number of equivalent valleys	1	4	3	
Effective mass (m_e)	0.063	0.222	0.58	
Γ-*L* separation (eV)		0.33	0.522	
Polar optical scattering				
Static dielectric constant				12.90
High-frequency dielectric constant				10.92
LO phonon (eV)				0.0354
Γ-*L*, X scattering				
L-X scattering				
Coupling constant (eV/cm)		0.800×10^9	0.200×10^9	
Phonon energy (eV)		0.100×10^9		
Γ-*L*, X		0.0278	0.0299	
L-X		0.0293		
L-*L*, X-X scattering				
Coupling constant (eV/cm)		2.0×10^9	2.0×10^9	
Phonon energy (eV)		0.0299	0.0290	
Acoustic scattering				
Deformation potential (eV)	7.0	9.2	9.2	
Nonpolar scattering (*L*)				
Coupling constant (eV/cm)		0.300×10^9	0.300×10^9	
Phonon energy (eV)		0.0343	0.0343	

the above problems are overcome. They are mentioned to indicate that caution is in order and that an over reliance on available parameters for describing scattering rates is not acceptable.

As indicated earlier, all scattering rates will be obtained within the framework of the displaced Maxwellian. The integrals as used in the discussion below are summarized in Table 3.We note that the scattering matrix elements for electron-phonon interaction were applied, within the early fifties to calculate the scattering rates via the displaced Maxwellian, by Froelich and coworkers[12]. They have since been repeated by a variety of other workers, in some cases with important generalizations. One of the earliest applications of the moment equation and the above scattering integrals was to the semiconductor GaAs. These calculations were performed under the assumption of transport within two portions of the GaAs conduction band, one at Γ, and the second at X. The parameters for this calculation are displayed in Table 4[16].

TABLE 4. GaAs Parameters Used in Calculations

Parameter	Γ	X	Common
Number of equivalent valleys	1	3	
Effective mass (m_e)	0.067	0.40	
Γ-X separation (eV)			0.36
Polar optical scattering			
Static dielectric constant			12.53
High-frequency dielectric constant			10.82
LO phonon (eV)			0.0354
Γ-X scattering			
Coupling constant (eV/cm)			0.621
Phonon energy (eV)			0.0300
X-X scattering			
Coupling constant (eV/cm)		1.064	
Phonon energy (eV)		0.0300	
Acoustic scattering			
Deformation potential (eV)	7.0	7.0	

In the late seventies experimental evidence pointed to the L valley as the lowest satellite valley, but with X sufficiently close to L to provide enough coupling at high fields. Thus GaAs was regarded as a three-level transfer semiconductor. Transient and steady state calculations with three levels of transfer have been considered by many workers. Results within the framework of the displaced Maxwellian will be illustrated below. Additionally, several two-level spatially dependent electron transfer calculations, with Γ-L ordering, using the parameters of Table 5.

For three-level transfer with the parameters of Table 3, the scattering integrals are shown in figure 9, where scattering rates are plotted as a function of electron temperature. For example,

TABLE 5. GaAs Parameters Used in Calculations

Parameter	Γ	*L*	Common
Number of equivalent valleys	1	4	
Effective mass (m_e)	0.067	0.222	
Γ-*L* separation (eV)			0.33
Polar optical scattering			
Static dielectric constant			12.90
High-frequency dielectric constant			10.92
LO phonon (eV)			0.0354
Γ-*L* scattering			
Coupling constant (eV/cm)			0.800×10^9
Phonon energy (eV)			0.0278
L-*L* scattering			
Coupling constant (eV/cm)		2.0×10^9	
Phonon energy (eV)		0.0354	
Acoustic scattering			
Deformation potential (eV)	7.0	9.2	
Nonpolar scattering (*L*)			
Coupling constant (eV/cm)		0.300×10^9	
Phonon energy (eV)		0.0343	

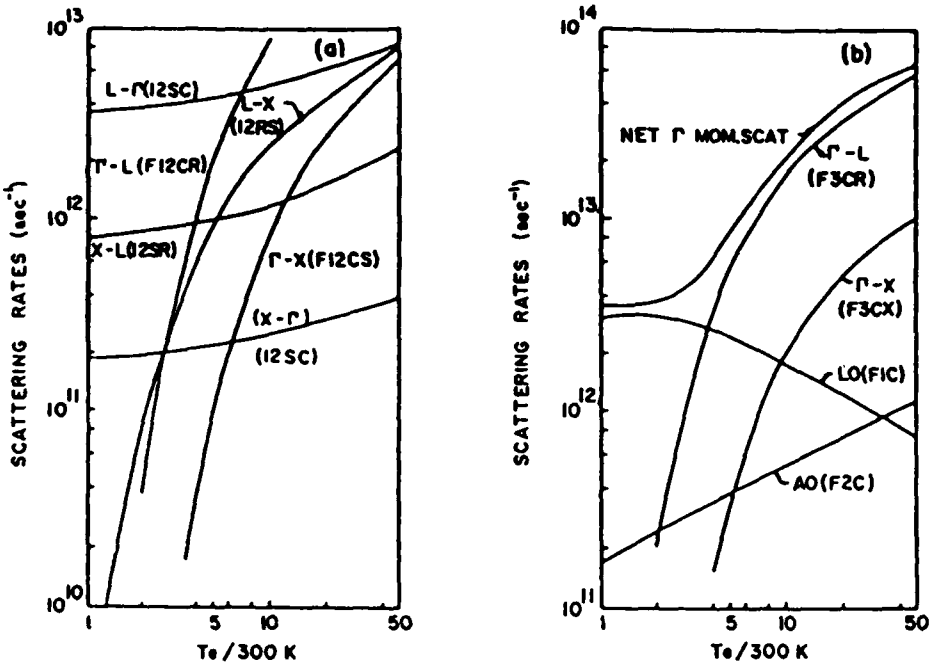


FIGURE 9. Three-level (a) carrier scattering rates for use in equation (44); and (b) Γ valley momentum scattering rates for use in equation (45).

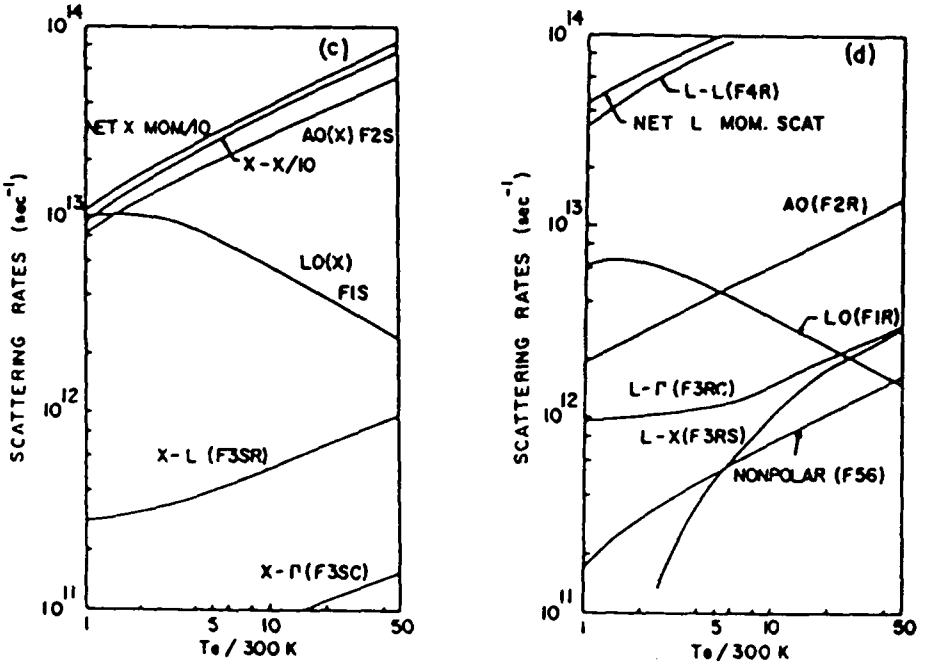


FIGURE 9. Three-level (c) L valley and (d) X valley momentum scattering rates, for use in equation 45.

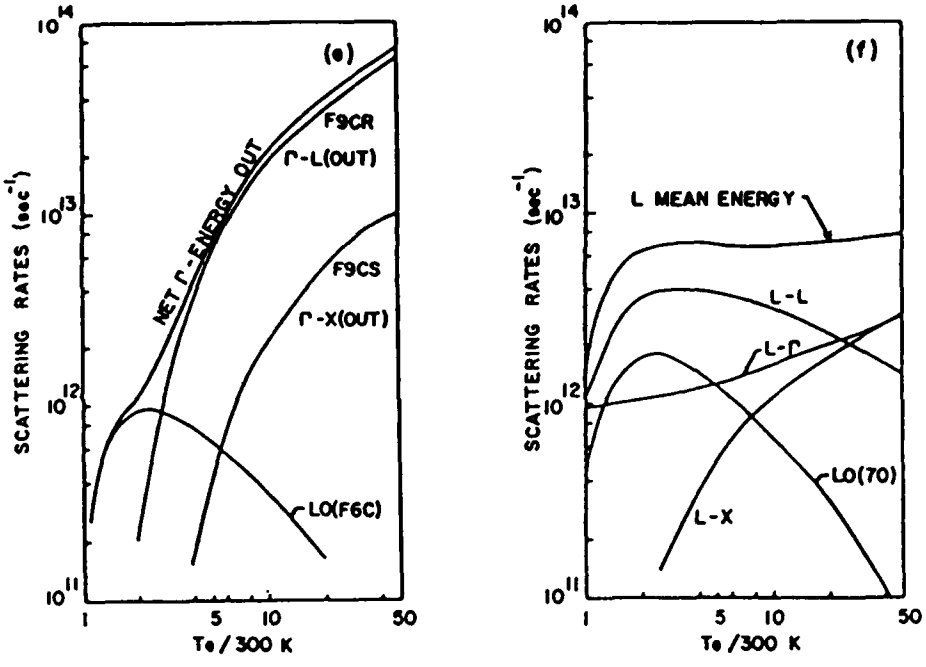


FIGURE 9. Three-level (e) Γ valley energy scattering, and, f) L valley energy scattering, for use in equation (46).

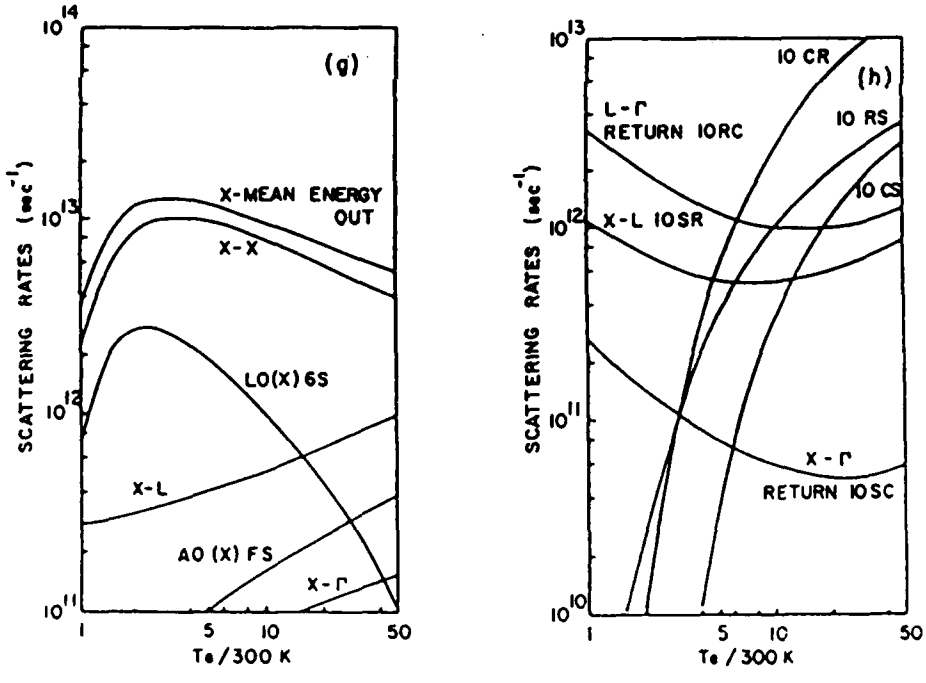


FIGURE 9. Three-level (g) X valley energy and (h) return energy scattering for use in equation (46).

figure 9a displays the carrier density scattering rates. Note that the ratio, at T_e=300K, of the LΓ carrier scattering to ΓL scattering is governed by the equilibrium distribution of the Γ and L valley carriers, as given by equation 71. The scattering contributions to the Γ valley momenta are shown in figure 9b. At low energies scattering is dominated by the LO phonon, whereas at high energies by the intervalley phonon. The scattering contributions to the L and X valleys momentum are shown in figures 9c and 9d, respectively. The dominant contribution is equivalent intervalley scattering. The Γ, L and X valley energy scattering rates are shown in figures 9e, 9f, and 9g, respectively. Note the increased scattering rate at high energies for the Γ valley carriers, and the decreased scattering rate for the X-valley carriers. The energy loss for the Γ valley carriers is dominated by transfer to valleys with higher density of states. The L valley energy loss is intermediate between Γ and X. The return energy scattering is shown in figure 9i.

Note that there are 42 scattering integrals, exclusive of ionized impurity scattering, to contend with. After summing, there are 18 scattering rates of interest. The steady state calculation arising from these integrals is shown in figures 10 and 11.

Figure 10 displays the field dependent mean velocity for Γ-L-X ordering. Figure 11a is the distribution of carriers as a function of applied field. The electric field versus Γ-valley electron temperature is displayed in figure 11b. In figure 11c and 11d, we sketch (e/mΠ)F for select Γ and L valley integrals, as well as the mean velocity for Γ and L-valley electrons. Note that the Γ valley velocity, whose contributions are dominated by LO phonons at low energies, and intervalley phonons at high energies, is at least an order of magnitude greater than the L-valley velocity.

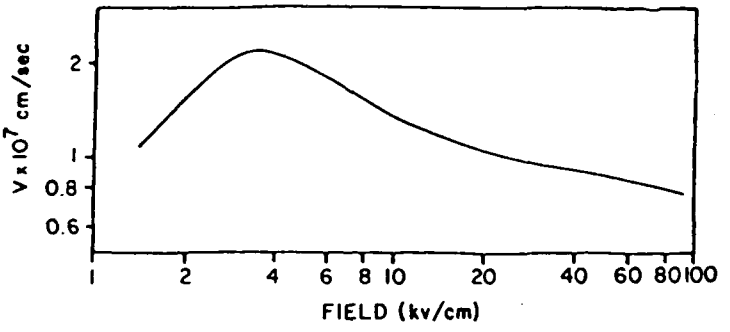


FIGURE 10. Steady state velocity vs field for Γ-L-X ordering in GaAs.

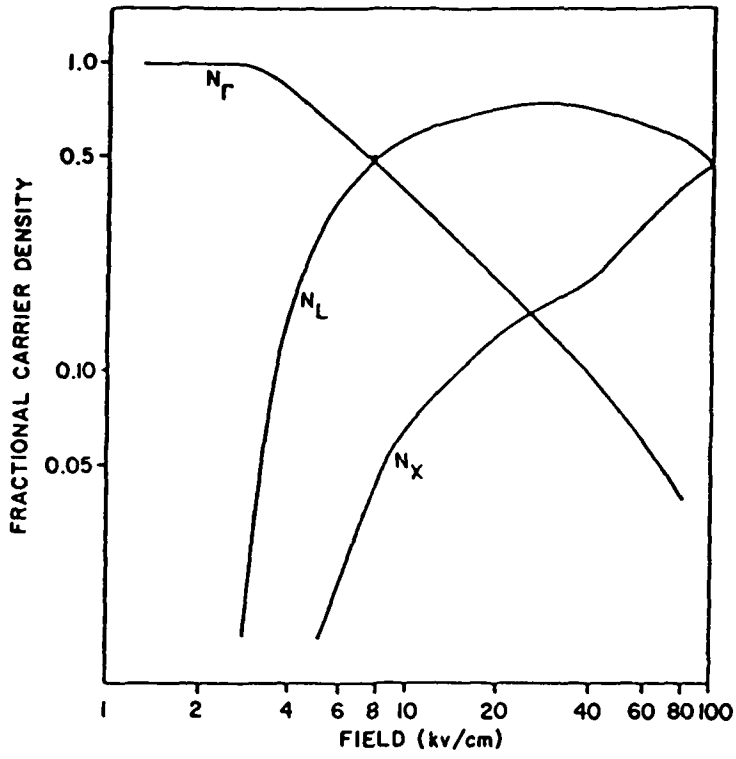


FIGURE 11a. Carrier density of Γ-L-X valleys for the calculation of figure 10.

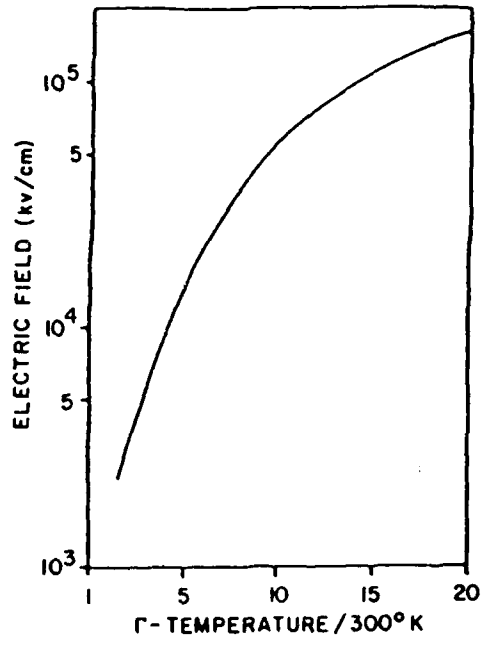


FIGURE 11b. Field vs Γ-valley electron temperature.

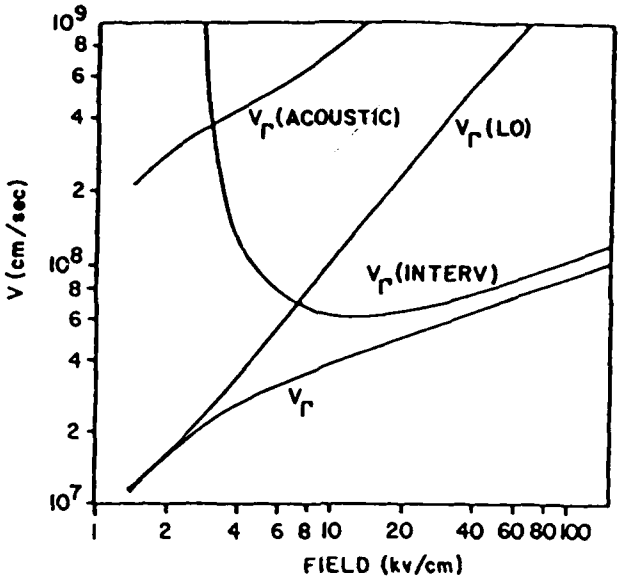


FIGURE 11c. Contributions to Γ-valley velocity.

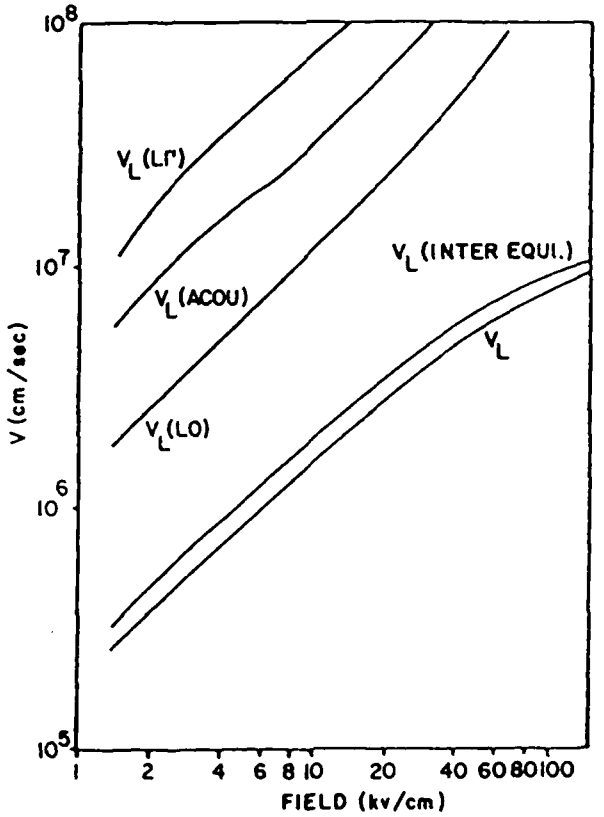


FIGURE 11d. Contribution to L-valley velocity.

There are several points to be made about these integrals. The first is that the energy scattering rates, particularly those associated with the central valley are several orders of magnitude less than those of the corresponding momentum scattering rates. Thus, while initially in an applied field the momentum increase without any significant relaxation, subsequent momentum relaxation is energy dependent and dominated by the longer energy relaxation times. Prior to significant energy relaxation, the momentum and carrier density scattering rates are being altered by the increasing electron temperature as well as the increasing carrier velocity. Thus in this case of momentum balance, there is a decrease in carrier momentum, and in the case of carrier balance, there is a decrease in central valley population. The consequences of this is a net decrease in carrier velocity. These results are born out numerically and are displayed in figure 12 for carriers

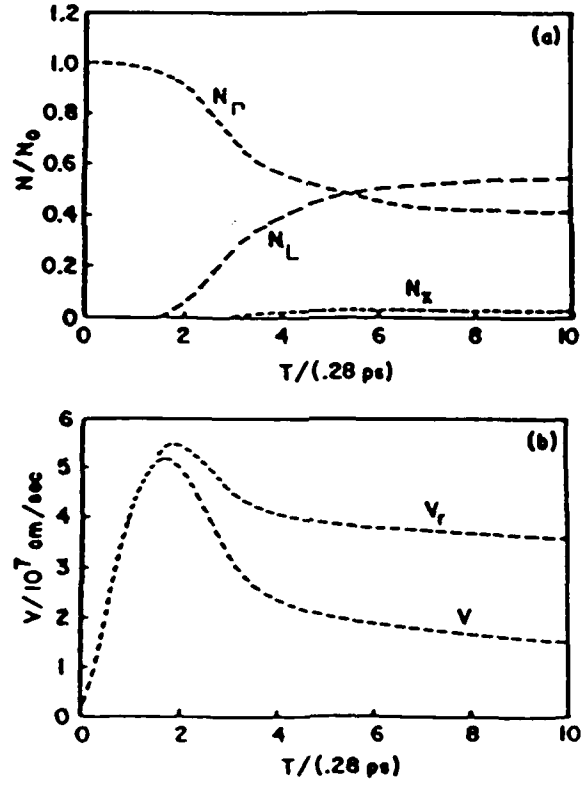


FIGURE 12. Transient uniform field response, <F> = 9.7Kv/cm μ.

with a zero initial velocity, an initial electron temperature in the central and satellite valley equal to the ambient temperature,

and an initial equilibrium distribution of carriers:

$$\frac{n_\Gamma}{n_L} = \left(\frac{m_\Gamma}{m_L}\right)^{\frac{3}{2}} \exp\Delta_{\Gamma L}/kT \tag{71}$$

$$\frac{n_\Gamma}{n_X} = \left(\frac{m_\Gamma}{m_X}\right)^{\frac{3}{2}} \exp\Delta_{\Gamma X}/kT \tag{72}$$

Note: There is overshoot in the Γ-valley carrier velocity, as well as the mean velocity.

As the discussion above indicates, the initial transients are dependent in a detailed way on the way the energy relaxes. Thus the details of transient transport are likely to depend significantly on at least the following items: (1) rise time of the external source, (2) initial conditions, (3) boundary conditions, and (4) device structure. The first two items are taken up next.

The study of the rise time is basically a matter of matching time constants. For example, if the time it takes electrons to reach steady state equilibrium is of the order of one picosecond, then rise times of the order of one picosecond or less will introduce strong nonequilibrium effects. These effects will decrease as the rise time increases. This is illustrated in figure 13 where we plot peak velocity versus rise time for a device subject to a specific value of bias field.

The significance of initial conditions for uniform field calculations is that it is thought to identify the properties of the entrance contact. In the strictest sense the properties of the entrance contact are determined by a set of boundary conditions, with the initial conditions on the boundary playing a less prominent role. Calculations with both will be illustrated below with more emphasis on the contact as a boundary conditions.

The first set of calculations is for uniform field and GaAs with three level transfer. The bias field for the case is 17.6 kv/cm and the initial carrier velocity is zero, figure 14, dashed line. Strong velocity overshoot is apparent from this calculation, and there is overshoot in the temperature. (Please note: that for this calculation capacitive effects have not been included. They are incorporated implicitly in the spatially dependent transient calculations). The situation when the entrance carrier velocity and temperature are varied is indicated by the solid curves of figure 14.

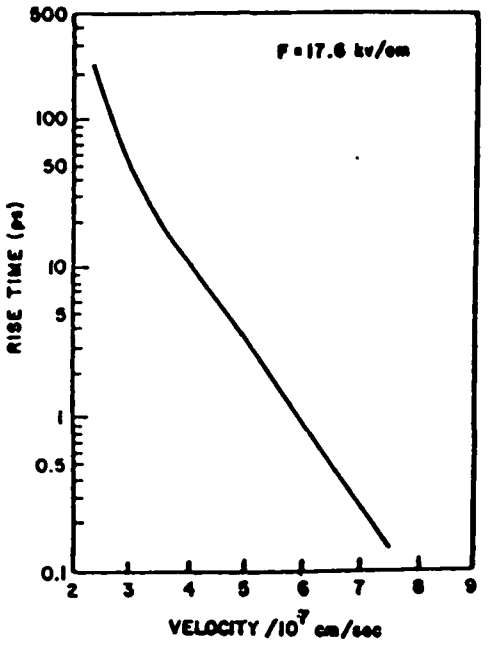


FIGURE 13. Velocity vs rise time.

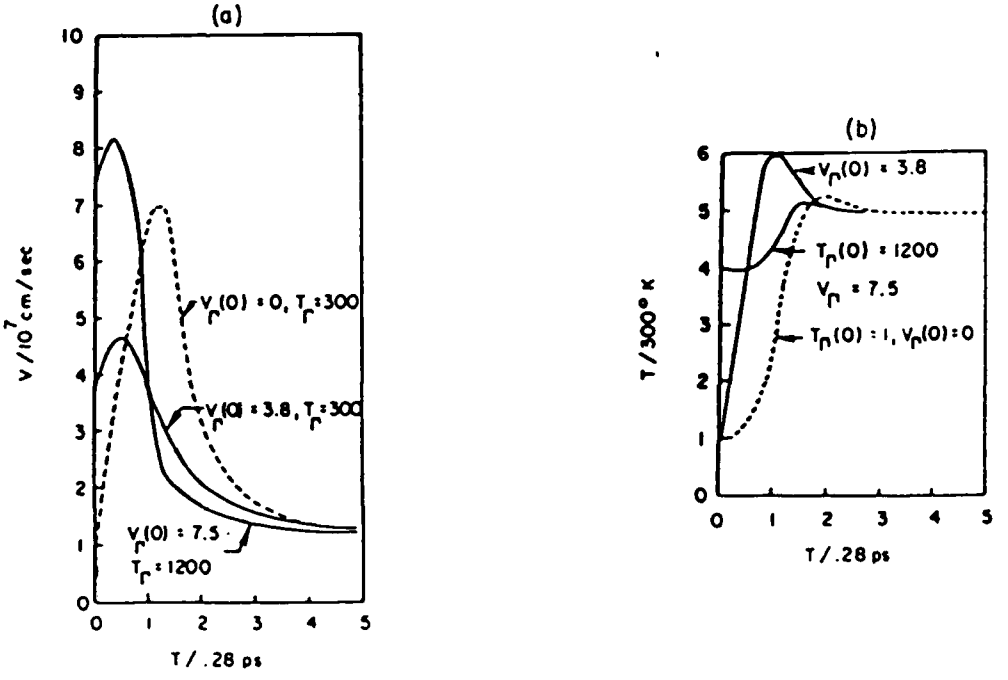


FIGURE 14. (a) Transient response as a function of initial velocity and temperature. (F=17.6kv/cm). Load line included; (b) As in 14a, but for Γ-valley temperature.

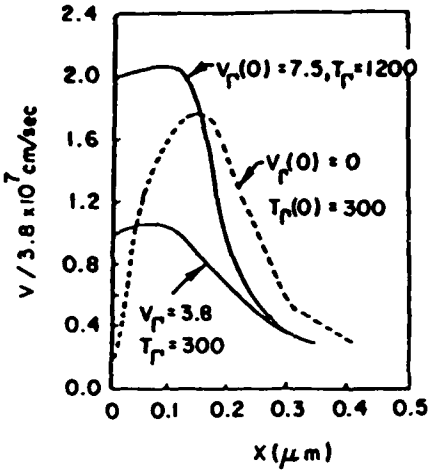


FIGURE 15. As in 14a, but for velocity vs distance.

Several things are apparent from these results. First, for an initial entrance temperature equal to the ambient temperature, an increase in the entrance velocity results in an increase in the peak velocity and the rate at which electron temperature increases. As a consequence of the latter, the momentum and density relaxation rates increase. Thus, the time to steady state is reduced. We can project that a tradeoff exists for an optimum set of initial conditions in which an electron will traverse a given distance in the shortest time interval.

A very similar set of results was presented by Iafrate et al.,[17] using Monte Carlo techniques. There are some very important differences in the assumptions associated with the Monte Carlo technique and the displaced Maxwellian, particularly with regard to the assumption concerning the distribution of the entering carriers. Nevertheless, the qualitative features of the two results are similar. The Monte Carlo studies are displayed in figures 16 and 17. Note, the F=20 curve of figure 16 is qualitatively similar to that of the $v_\Gamma(0)=7.5$ curve of figure 15, which is the same as figure 14a but with distance as the abscissa. Figure 17 indicates that the rate of carrier velocity decrease increases with increasing energy. Again, in qualitative agreement with figure 15a.

INFLUENCE OF SPATIAL NONUNIFORMITIES ON TRANSIENT TRANSPORT

The above discussion was concerned with uniform fields, a situation that is not expected to occur. Indeed if it did occur, we would necessarily conclude that submicron two-terminal GaAs devices would possess a region of dc negative differential conductivity. Further, the implication of the spatially uniform velocity overshoot calculations is that they are representative of

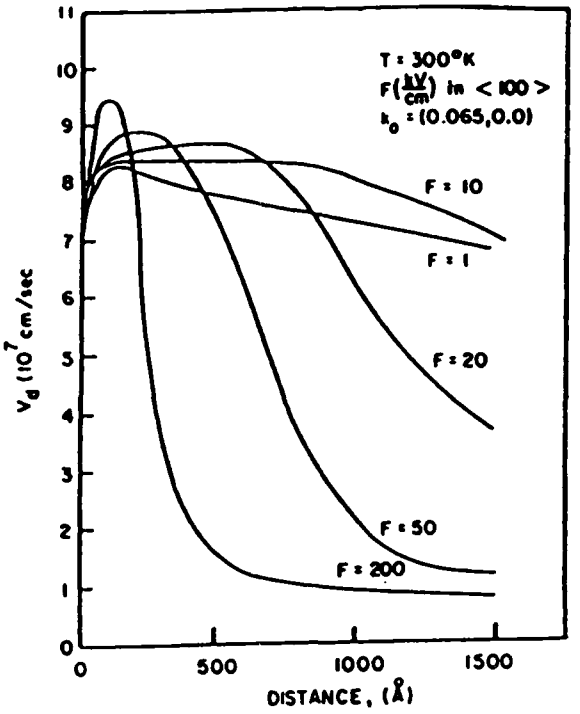
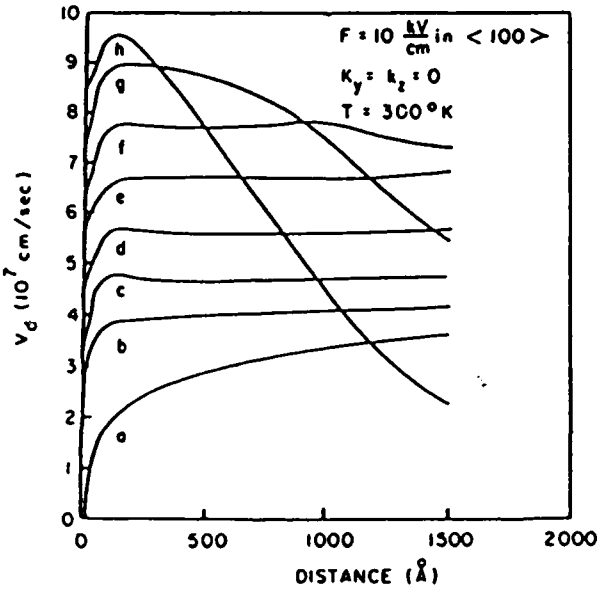


FIGURE 16. Velocity vs distance vs applied field.



	$k_x \frac{2v}{a}$ / 0	E_o(eV) / 0
a		
b	0.02	0.04
c	0.03	0.07
d	0.04	0.11
e	0.05	0.16
f	0.06	0.21
g	0.07	0.27
h	0.08	0.34

FIGURE 17. Velocity vs distance vs energy.

measured current transients profiles. As discussed below, this is not likely to be the case.

To illustrate the role of spatial nonuniformities on transient transport, a limited number of numerical calculations involving space and time dependent solutions to the first three moments of the BTE. These calculations have been performed as a function of device length, boundary conditions, doping level and material variations. None of these calculations display a region of dc NDC. Indeed, all calculations show considerable departure from the familiar homogeneous uniform field studies. All calculations are performed using a model for GaAs that contains two levels of transfer, and assumes a conductivity effective mass equal to the density of state effective mass. In addition, the boundary conditions are chosen to represent the effects of injecting contacts. It is possible to argue against the use of a specific set of boundary conditions to represent the properties of the physical contact. These arguments would, however, tend to remove us from the essential features of the problem, which are to

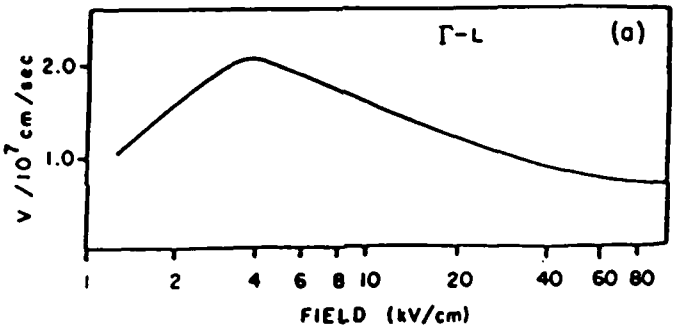


FIGURE 18a. Field dependent velocity for Γ-L two-level transfer.

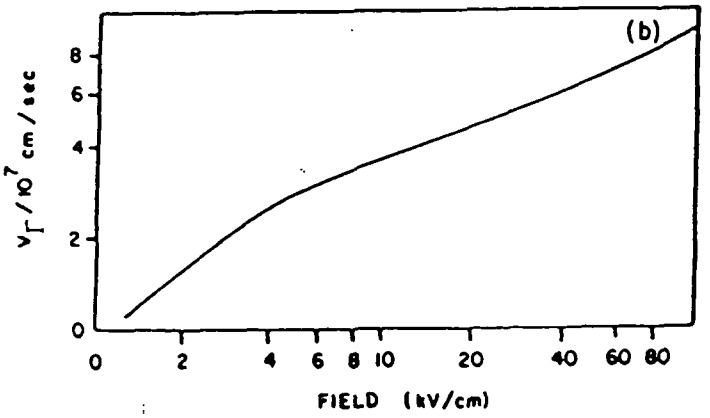


FIGURE 18b. As in 18a but for V_Γ.

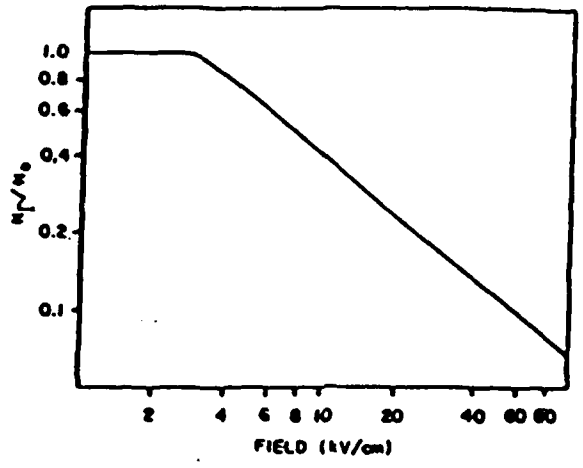


FIGURE 18c. As in 18a but for NΓ.

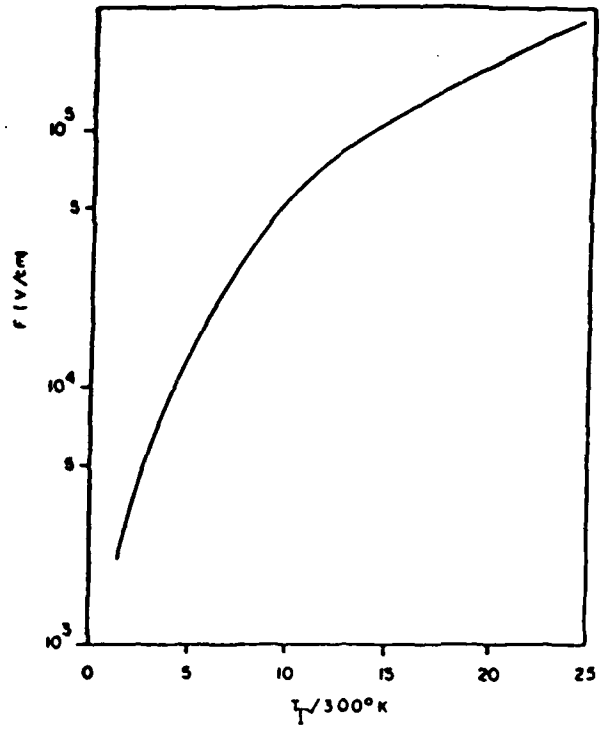


FIGURE 18d. Field vs T_Γ.

determine the role of the boundary and of nonuniform space charge effects on device performance. The parameters used for the calculations are displayed in Table 5. For reference, the steady

state field dependent transport results are displayed in figure 18.

The first set of calculations shown is for a uniform field. These calculations are performed for reference only. For the uniform field calculations, the device is in series with a load resistor and an inductor. The device is also in parallel with a capacitor. For the first set of calculations, the inductor is set to zero, and the load resistor is set to $0.01R_o$, where R_o is the low field resistance of the element. The capacitor was set to $C_o = \tau_R/R_o$, where τ_R is the dielectric relaxation time for the semiconductor element. For the space charge dominated device, the device was subjected to a specified time dependent voltage source and all capacitive effects were included self-consistently through application of the continuity equation. (An earlier study incorporated the load resistor). The two space charge dependent calculations were for two terminal devices, one with an active region length of 0.25μm, the second with a device length of 1.0μm. Both had doping levels of $5x10^{15}/cm^3$ and were subject to injecting boundary. Further in these calculations, ionized impurity scattering was ignored.

The uniform field calculations were performed: (1) at a bias level of 6kv/cm, and (2) at 10kv/cm. The familiar overshoot results discussed earlier are displayed here in figure 19.

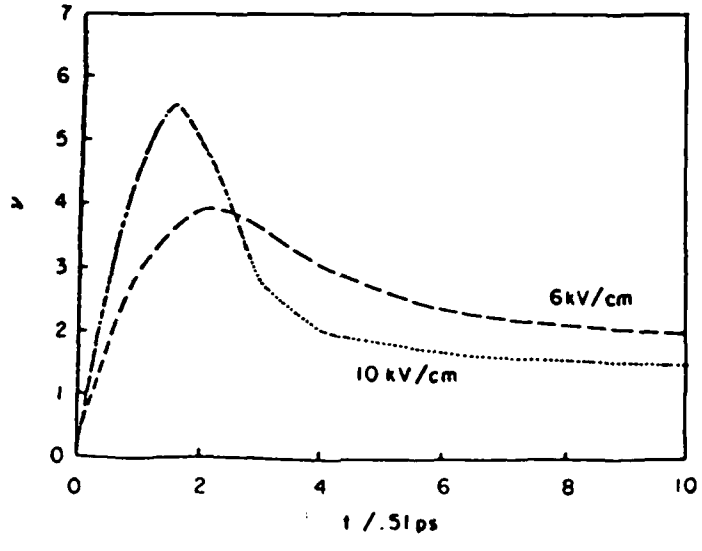


FIGURE 19. Transient transport uniform field for two-level Γ-L transport.

The situation with regard to the space charge dependent calculations with injecting boundary conditions is shown in the next set of figures. Figure 20 displays the current time transients for the 0.25 and 1.0μm devices at two different bias levels. The use of current rather than velocity is significant

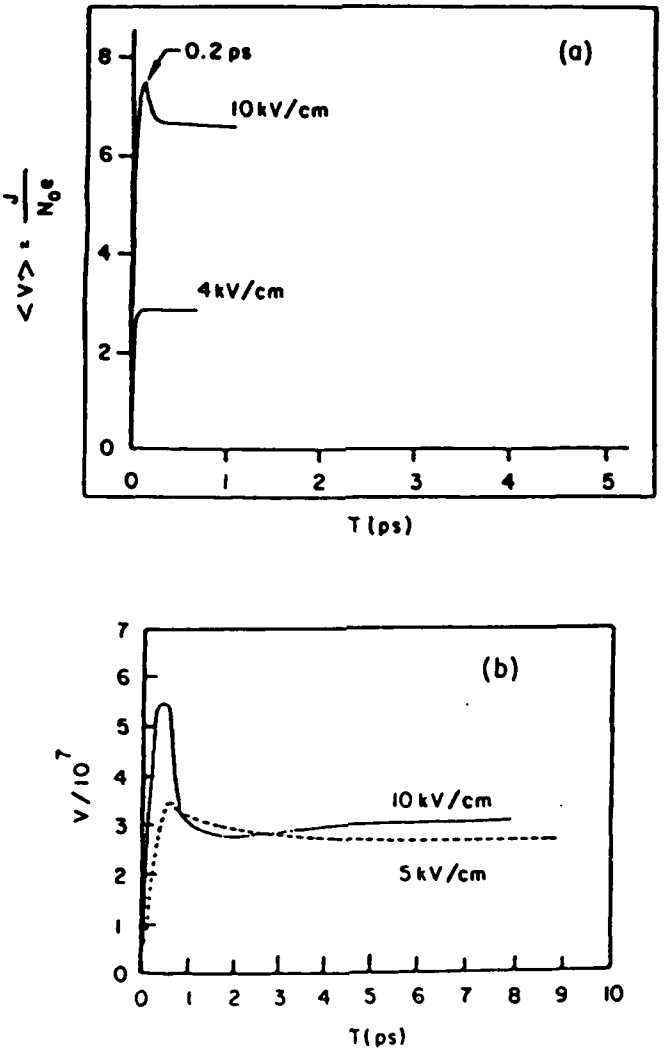


FIGURE 20. (a) <v> vs time for .25μm device with nonuniform fields; (b) <v> vs time for 1.0μm device with nonuniform fields. Each transient was taken from a non-zero initial state.

here – it is current that is conserved - not velocity. In one dimension the total current is independent of position. Note the significant result that at the higher bias levels the steady current level increases – a result opposite to that for the uniform field calculations.

The steady state distribution of electric field in these devices is nonuniform and an understanding of them is crucial for understanding device operation. The field profiles for the 0.25μm device are displayed in figures 21a and 22a, and show a

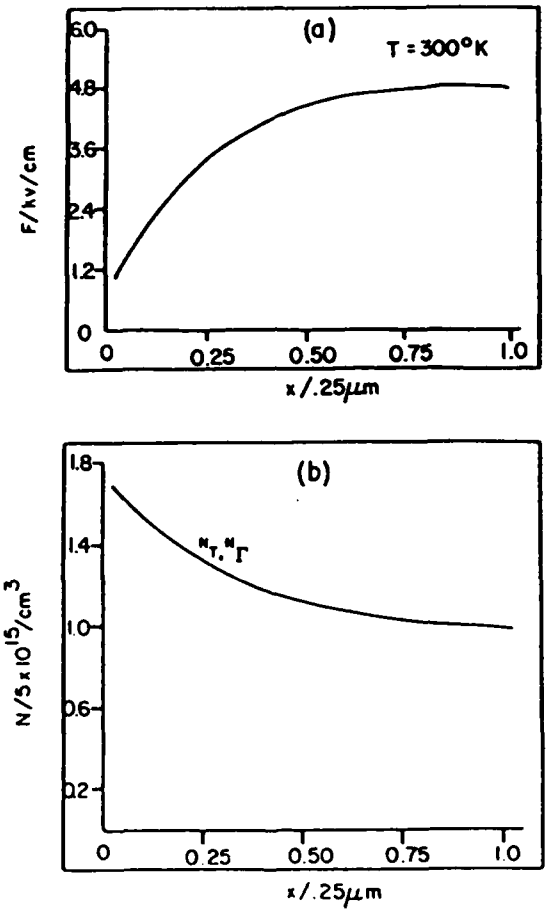


FIGURE 21. (a) Electric field vs distance for 0.25μm device with injecting cathode and an applied bias of 0.1 volts; (b) N-total and N_Γ carrier density.

characteristic increase in field as the downstream boundary is approached. In both cases the initial increase from the cathode is

a consequence of the injecting boundary condition. In the 0.25μm

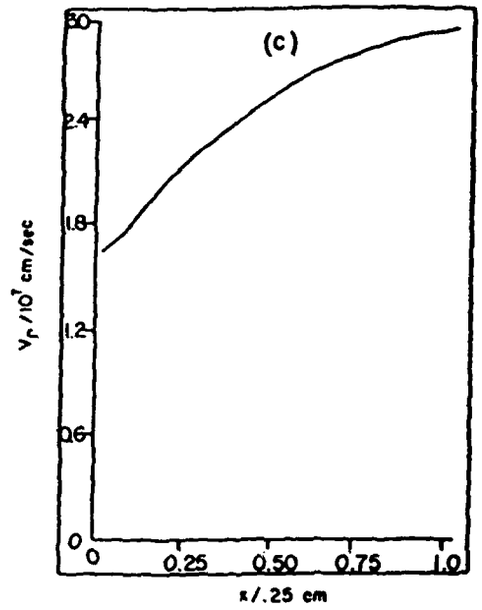


FIGURE 21c. Γ-valley velocity distribution.

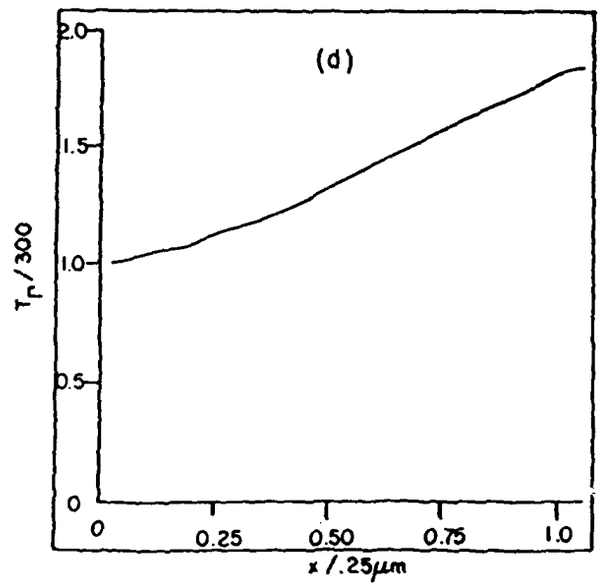


FIGURE 21d. Γ-valley temperature distribution.

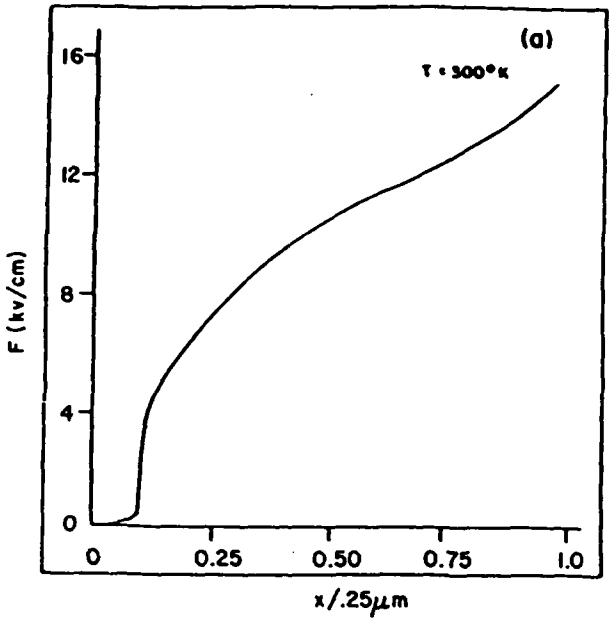


FIGURE 22(a). Electric field vs distance for 0.25μm device with injecting cathode and a bias of 0.25 volts.

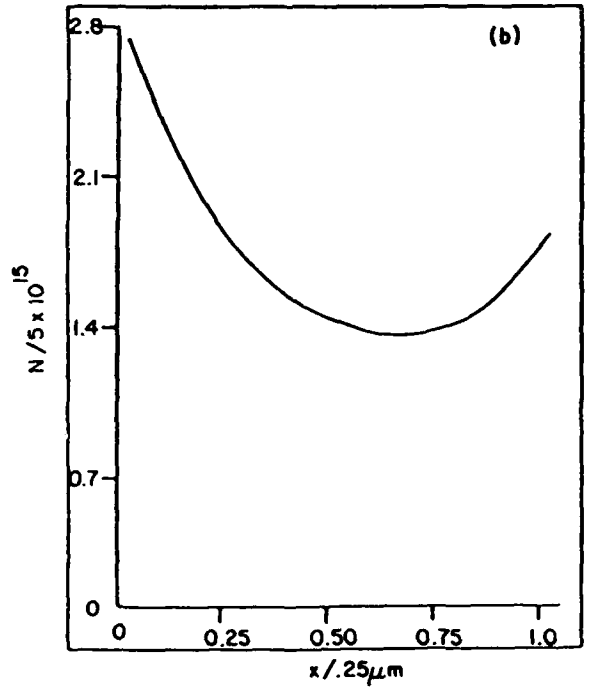


FIGURE 22(b). Total carrier density vs distance.

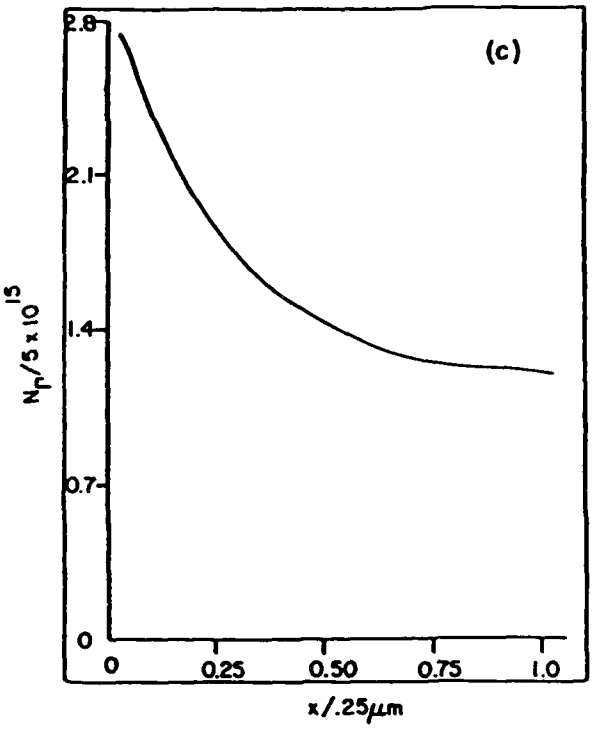


FIGURE 22(c). Γ-valley density vs distance.

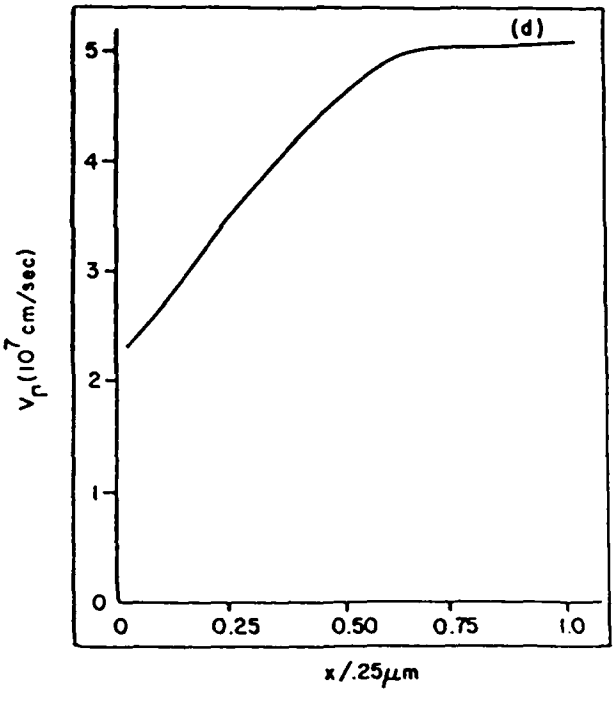


FIGURE 22 (d). Γ-valley velocity vs distance.

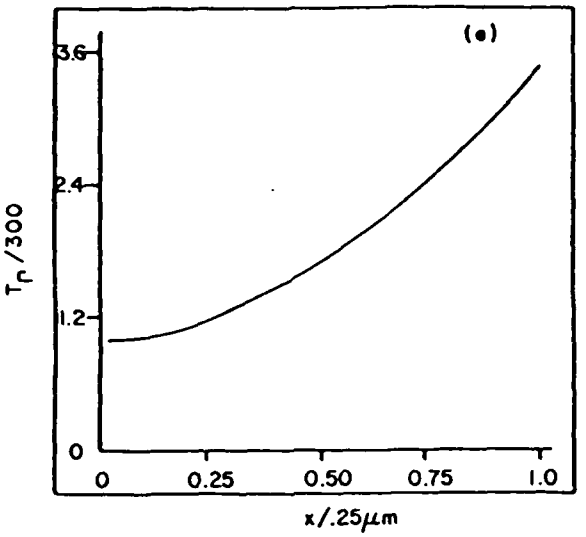


FIGURE 22(e). Γ-valley temperature.

structure and a bias of 0.1v the anode field is flat, while that for the 0.25v case the field continues to increase. There is little electron transfer at the lower bias level, as indicated in figure 21b, but a substantial number for the higher bias level, as indicated in figures 22b and 22c. It is anticipated that increasing the applied bias will cause an increase in the numbers of carriers scattered to the satellite valleys where the carrier velocity is reduced. Experimentally, however, in the absence of device heating an increase in current density with an increase bias field is expected, at least under dc conditions. The most likely reason for this to occur is through net carrier accumulation within the region of electron transfer. This, indeed, is what is occurring and is seen in figure 22b. We note that the decreasing carrier concentration for the short device and low bias requires, through current continuity an increase in carrier velocity, as seen in figure 21c. Further, the Γ-valley electron temperature increases only to 600°K, not enough to result in carrier transfer.

The increasing electric field at high bias levels in the vicinity of the anode is a feature of materials with a region of negative differential mobility and does not appear to be dependent upon the specific boundary conditions applied. It was predicted by Shockley[18], and its consequences have been detected experimentally[19]. We note that, as in the low bias calculation, the Γ-valley carrier velocity increases substantially toward the anode (figure 22d) and the electron temperature exceeds 900°K (figure 22e), forcing electron transfer.

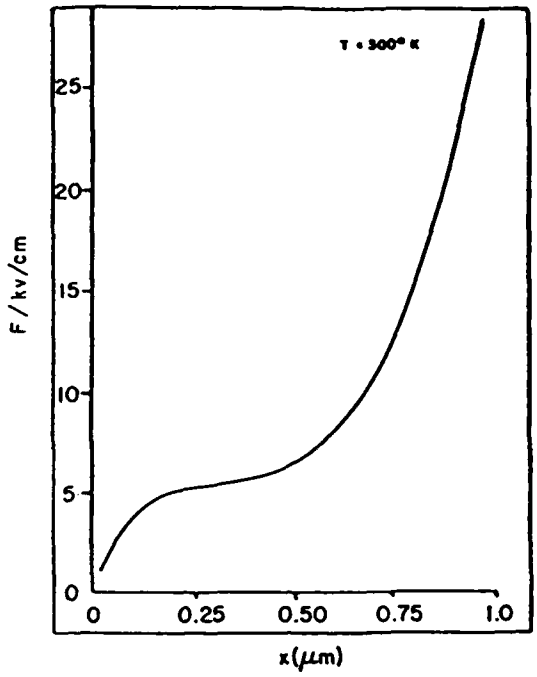


FIGURE 23. Electric field vs distance of 1μm device with an injecting cathode and a bias of 1.0 v.

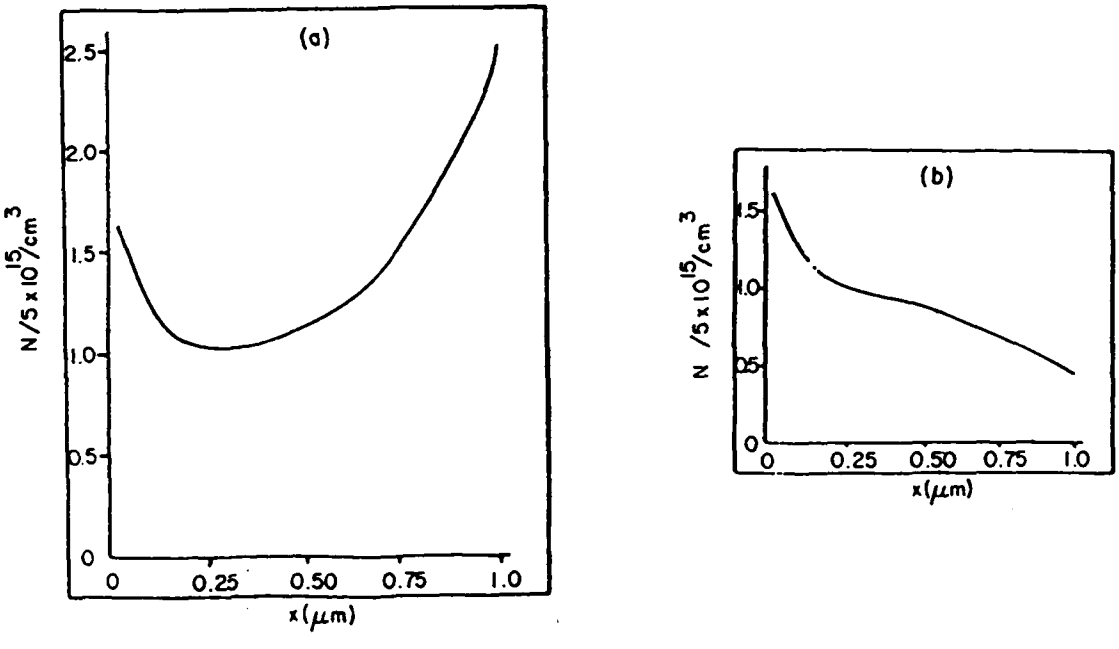


FIGURE 24(a). Total carrier density corresponding to the field distribution; **(b)** Γ-valley carrier density.

The features of the 0.25μm calculation are exaggerated for the longer 1.0μm calculation of figure 23. Note the high anode field near 30kv/cm, and the substantial electron transfer (figures 24a and b). The Γ-valley carrier velocity for this case is larger than that of the 0.25μm case as required by current continuity.

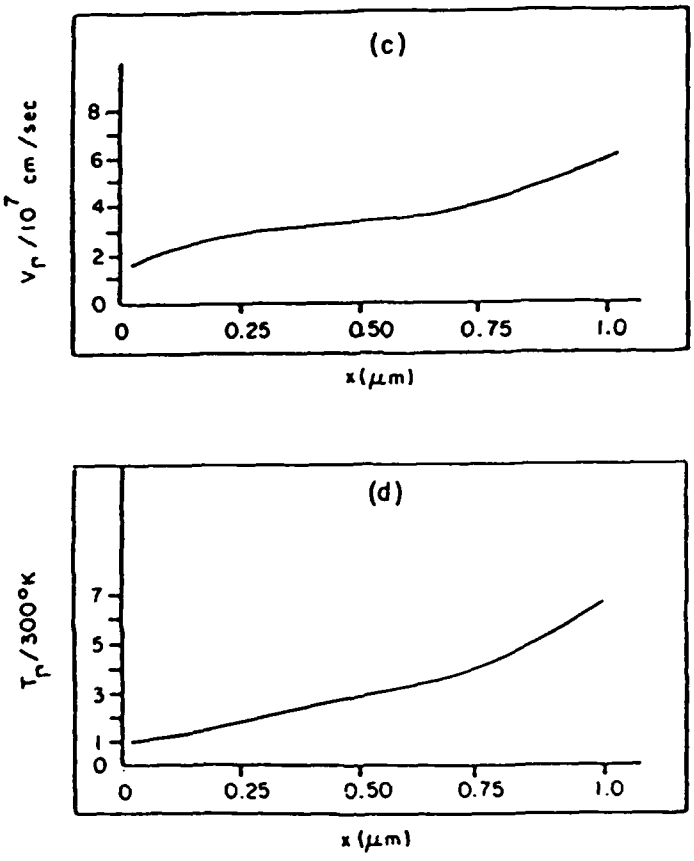


FIGURE 24(c,d). As in figure 24a, b, but for Γ-valley carrier velocity and electron temperature.

It is important to reiterate the significance of these results. It was originally anticipated that the uniform field calculations and the consequent predictions were relevant to high frequency and, therefore, submicron devices. With regard to the uniform field calculations, it must be noted that these are applicable to devices of any length, as long as the field is uniform. With regard to submicron devices with nonuniform fields, the time dependent uniform field results are not at all adequate. They do not predict the peak or asymptotic currents correctly.

The numerical results are somewhat model dependent. Certainly the details are dependent upon boundary conditions, device structure, length, doping, and etc. But there are several interesting features that should be highlighted as they appear in a number of different studies. The first point to be noted is that the carriers are entering the device at high velocities of 1.5×10^7 cm/sec. Within the device and near the downstream boundary, as indicated above, a large fraction of the carriers transfer to the satellite valley. These low mobility carriers contribute only an insignificant amount to the net current through the device. The high current drive is a consequence of the high level of injection throughout the active region of the device. (In longer "active region" devices, the advantages of charge injection are reduced and intervalley transfer leads to lower current levels).

The importance of space charge injection is not limited to the structure just discussed. An extremely interesting device that appears to exhibit many of the properties of the 0.25μm device is a one dimensional version of the space charge injection field effect transistor (Constant et al[20]). The geometry of this structure is shown in figure 25, and the space charge dependent

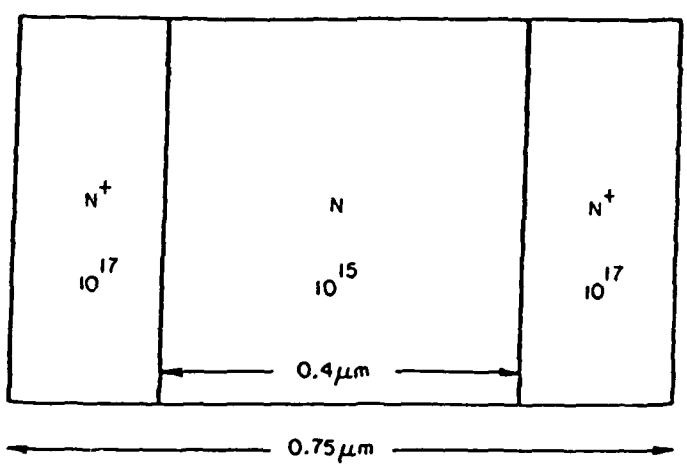


FIGURE 25. Structure for simulation of submicron space charge limited transport.

profiles for a one-volt bias across the device are displayed in figures 26a through 26c.

For the injection device depicted in figures 25c and 26, the space charge dependent profiles display considerable similarity to the results of the space charge limited diode. There is a potential minimum some distance away from the physical cathode, resulting in a retarding electric field. The consequence of this

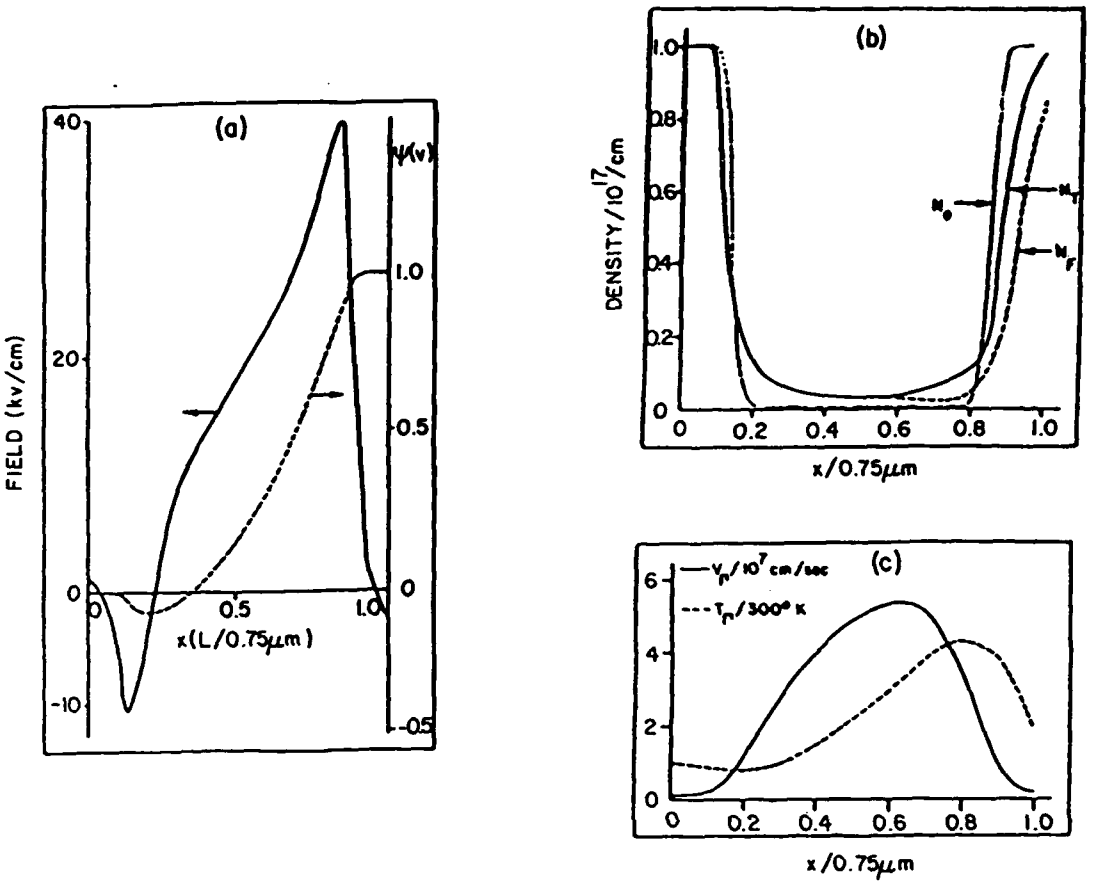


FIGURE 26. (a) Field and potential vs distance for the structure of figure 25; (b) carrier density; (c) carrier velocity and electron temperature.

retarding field is that the electrons are cooled as they approach the minimum[21]. (Note: we are not accurately representing the scattering at temperatures below the ambient. The results are therefore only qualitatively significant). The carriers are entering the active region at relatively high values of velocity and are exiting at even higher values. The field is nonuniform with high values near the downstream contact. There is considerable transfer which is compensated by an excess of injected carriers.

ASSESSING THE UPPER FREQUENCY LIMIT OF SMALL SIGNAL OPERATION

The next item of interest in connection with any of these devices, and certainly with regard to assessing material properties is the upper frequency limit of operation. The determination of

the upper frequency limit depends on the structure of interest and the mode of operation. If we confine ourselves to the transferred electron effect, then there are at least five items of interest: (1) large versus small signal operation, (2) device length, (3) boundary conditions, (4) doping variations, and (5) circuit. There has been only a limited amount of work done with regard to the last four items and here only with long devices. We will consider these later. At this point, we will confine our activity to large and small signal operation.

Consider first small signal high frequency operation, within the context of the following type of "experiment" (see figure 27). We assume a device is in a quiescent state at a voltage level Ψ_o and a current level I_o. The device is then subjected to a sudden perturbation in the form of a square wave pulse. The questions to be asked are: what is the output response, and what is the upper frequency limit for small signal negative conductance?

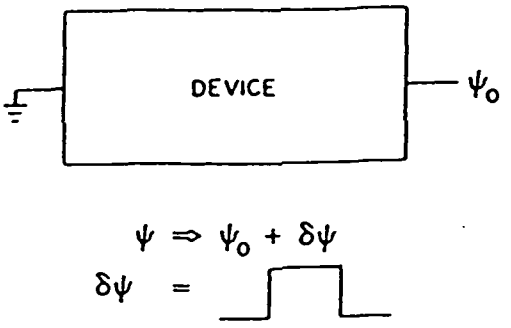


FIGURE 27. Small signal configuration.

Under uniform field conditions, NDR occurs when carriers transfer to the low mobility satellite valleys. Since the number of electrons transferring to subsidiary values requires a finite time to be completed, the extent of negative conductance is frequency dependent. This frequency dependence is complicated by the presence of nonuniform fields which completely eliminates dc NDR. Indeed for nonuniform fields, the most significant contribution to small signal NDR is the non-local transit time effect. (Note: this does not preclude the possibility of a transient "uniform-field"-like response to an external source). We first examine the way a material like GaAs responds to such an external stimulus. Consider figure 28, which displays the voltage pulse and a qualitative sketch of a current response (based on the

overshoot calculations). The current response includes an apparent overshoot and a relaxation to a current level below that of the unperturbed system. We are clearly considering uniform fields here. There could, additionally, be some nonequilibrium undershoot as represented by the dashed lines. When the pulse is turned off the system returns to its steady state value. Hereto, there may be some residual nonequilibrium overshoot.

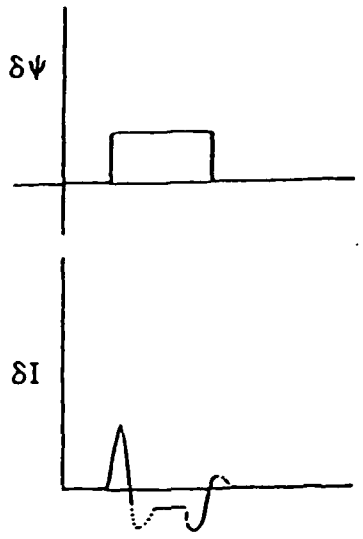


FIGURE 28. Small signal current response.

With reference to figure 28, the product $\delta\Psi\delta I$ represents the time rate of change of energy across the device. When it is positive, the system is lossy, when it is negative, the device is returning power. For uniform fields where an increase in the bias field results in a decrease in current, there is necessarily a region of small signal frequency dependent NDR[22]. The frequency dependence is due to several sources. First, there is over and undershoot prior to steady state (dashed lines in figure 28). Second, the extent of the frequency dependence is sensitive to the time duration of the disturbance. For example, if the excitation is turned off before the system had a chance to reach the NDR region, no power would be delivered to the load. Thus, the limitation to small signal negative resistance is principally at high frequencies.

The small signal NDR is dependent upon the scattering rates. These in turn are dependent upon the bias field. Thus, the magnitude and extent of the small signal NDR is bias dependent. Specifically, at sufficiently high bias levels a substantial amount of transfer occurs and the small signal NDR will decrease. There

should be an optimum bias level for small signal NDR, and this is displayed in figure 29a. The data summarized in 29a is obtained from dispersion relationships. The field dependence of the real part of the dispersion, yields the frequency dependent mobility, as is shown in figure 29b. Here, as the quiescent field increases,

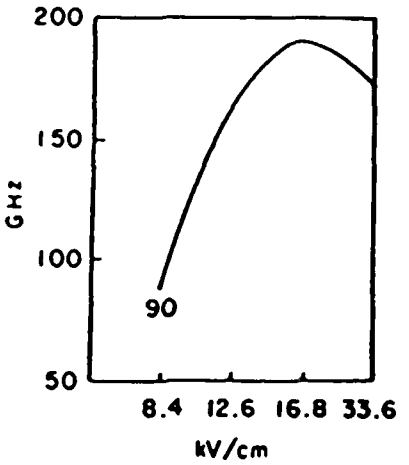


FIGURE 29a. Maximum frequency for small-signal oscillations (from uniform field MBTE).

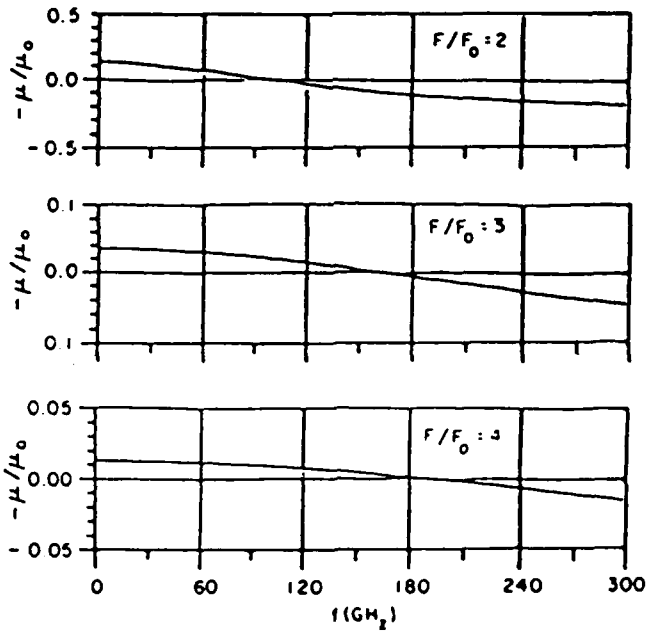


FIGURE 29b. Small signal differential mobility for two-level transfer (from uniform field MBTE). Here F_o = kv/cm.

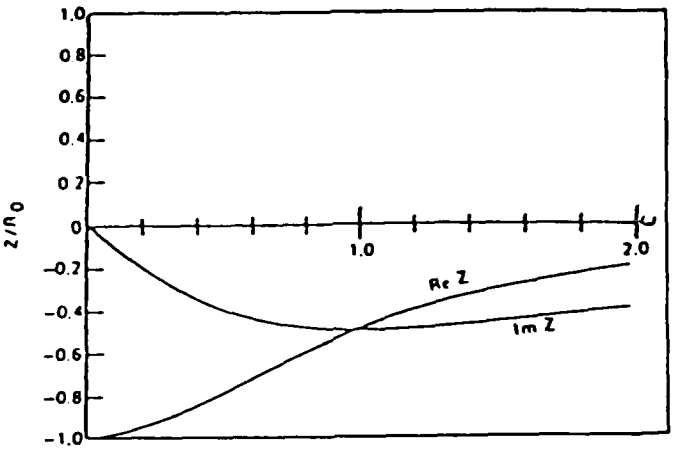


FIGURE 30. Small signal impedence as a function of normalized frequency from the drift and diffusion equations[19].

the magnitude of the small signal mobility decreases, while the maximum frequency increases.

It is important to note that the dispersion relations, obtained from the drift and diffusion equation, will not cross over to positive values at high frequencies. The absence of this cross over is shown in figure 30c[19].

The above small signal discussion has been for uniform fields. The question, of course, is what happens for nonuniform fields. For nonuniform fields and structures several microns in length, transit time contributions introduce significant phase delays. For submicron devices, the extent to which these transit time delays enter is not yet clear. What is apparent, however, is that if Shockley's positive conductance theorem is valid for submicron devices, then a residual pre-relaxation under- or overshoot is necessary. This was not apparent in the 0.25μm calculation of figure 20a, but it was present in the 1.0μm result (figure 20b).

THE UPPER FREQUENCY LIMIT OF SELF-EXCITED OSCILLATIONS

The above discussion was concerned with small signal NDR. The other extremes are the large signal oscillations, often identified as relaxation oscillations. For transferred electron devices, these oscillations received prominent attention when LSA devices were being studied. These self-excited oscillations require the presence of an inductor in the circuit. When dealing with

transferred electron devices, they are a fundamental phenomena to be dealt with, and the question, of course, is how to describe them.

The simplest description of the self-excited oscillator for uniform fields is through the coupled circuit equations (see figure 31)

$$\Psi_B = IR + L\frac{dI}{dt} + \Psi \tag{73}$$

$$I = I_c + C\frac{d\Psi}{dt} \tag{74}$$

where Ψ is the voltage across the device and I_c is the conduction current and is equal to $I_c = N_o ev$. These equations can, in turn, be represented by a second order ordinary differential equation:

$$\Psi_B = \left(1 + \frac{I_o}{\Psi}R\right)\Psi + \frac{Z_o}{R_o}\left(R_o\frac{dI_c}{d\Psi} + \frac{R_oC}{L/R}\right)\frac{d\Psi}{dt'} + \frac{d^2\Psi}{dt'^2}$$

$$Z_o = \left(\frac{L}{C}\right)^{\frac{1}{2}} \qquad t' = \frac{t}{(LC)^{\frac{1}{2}}} \tag{75}$$

For the situation where the relation between I_c and Ψ includes that of a region of NDR, we have essentially a generic form of a Van der Pol oscillator. Typically the NDR region displays no hysterisis and is not frequency dependent. The normalization identifies the quantity Z_o/R_o as the strength of the nonlinear element. Its effect on long devices is that it will determine when space charge contributions will dominate an oscillation.

The relevance of the dynamic properties of high frequency devices is that the origin of the self-excited oscillation lies in the presence of a region of NDR. Thus the upper frequency limit of oscillation is essentially defined by the intervalley transfer rates. Further, because the scattering rates are sampled differently from that of the small signal oscillation, the upper

frequency limit of the two will be different. Figure 31 summarizes the dependence of the upper frequency limit of the self-excited oscillations. Note the significantly lower maximum frequencies.

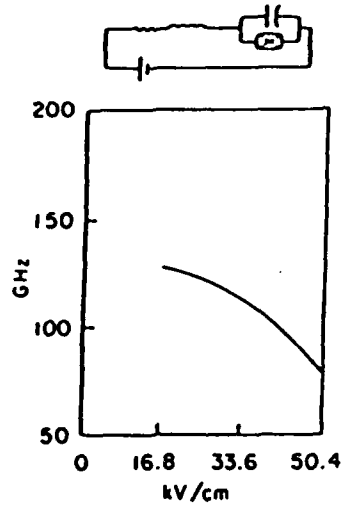


FIGURE 31. Maximum frequency for self-excited oscillations.

Figures 32a-32c illustrate a dynamic large signal self-excited oscillation at 78GHz as computed from the moments of the Boltzmann

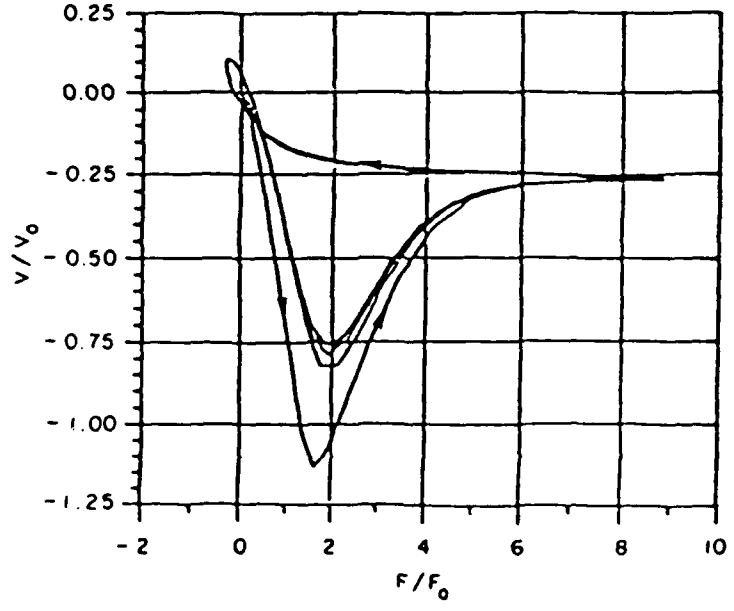


FIGURE 32a. Velocity vs field lissajous for 78 GHz osc.

transport equation. Figure 31a displays the velocity, not the magnitude of the velocity, as computed self-consistently from the MBTE and the circuit equation. Note that the velocity field figure arises from eliminating time between the velocity-time profile and the field-time profile. In figure 32a, there is an initial very high velocity arising from the initial overshoot transient. Under steady state, the time dependent voltage, or field, as displayed in figure 32b sustains a softer rise time and overshoot effects are reduced. Note the hysterisis in velocity arises because the response of the carrier density, velocity and temperature are not instantaneous. In the case of the carrier density, there is a delay following a field increase, before the electrons transfer to the satellite valley, as shown in figure 32c, and a delay following

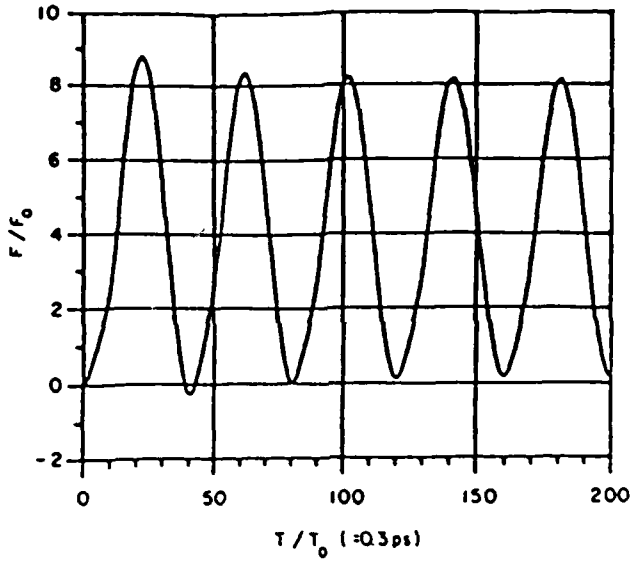


FIGURE 32b. Field vs time for the 78 GHz oscillation.

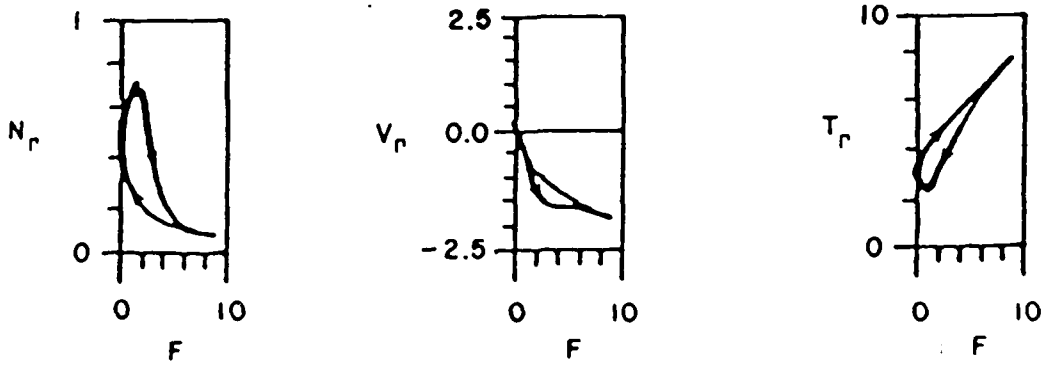


FIGURE 32c. Dynamic lissajous figures for Γ-valley carrier density, velocity and temperature.

a field decrease, before the electrons return to the Γ-valley. At higher frequencies, obtained by reducing the inductance, the finite intervalley relaxation times prevent sustained oscillations.

The question, of course, is how are space charge dependent effects going to enter the picture. For submicron devices, this has not yet been determined. Let us, however, go through some discussion to see if we cannot pick up a thread of prediction. To do this we rely on some analysis done earlier. Basically, we are concerned with three time constants: (1) an intervalley relaxation time, (2) a circuit frequency, (3) the propagation time of a space charge wave. In the case of low frequency operation, space charge control requires that the circuit frequency be sufficiently high to suppress residual space charge nonuniformities. (This will also require a sufficiently high Z_o/R_o circuit. See Eq. 75) The above arguments are also required for short submicron length devices. But if we look at the transit time associated with the smaller devices, particularly the 0.25μm device, the transit time is of the order of the LO phonon scattering time. It is unrealistic to suggest self-excited oscillations on this time scale for GaAs devices.

In addition to both the self-excited oscillation and the small signal oscillation, there is the driven oscillator, where a sinusoidal driving potential is imposed on the device as shown in figure 33. The dynamic situation associated with the above figure

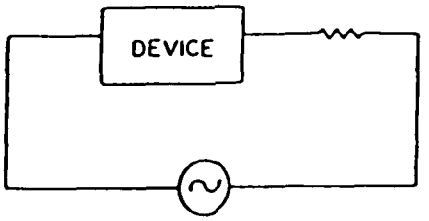


FIGURE 33. Driven oscillator configuration.

is as follows. When the power is delivered to the device-plus-load, power is dissipated to the load in an amount equal to

$$R\langle I^2\rangle \tag{76}$$

where the brackets denote a time average over one period. The power delivered to the device is

$$\langle I\Psi\rangle \tag{77}$$

When the above quantity is positive, the device is lossy. Interest in the transferred electron device arises from the fact that Eq. (77) can be negative and the device can deliver power to the load. The situation is clearly limited by the frequency of operation, and one particular set of results is displayed in figure 34. Which shows a Lissajous between velocity and applied field.

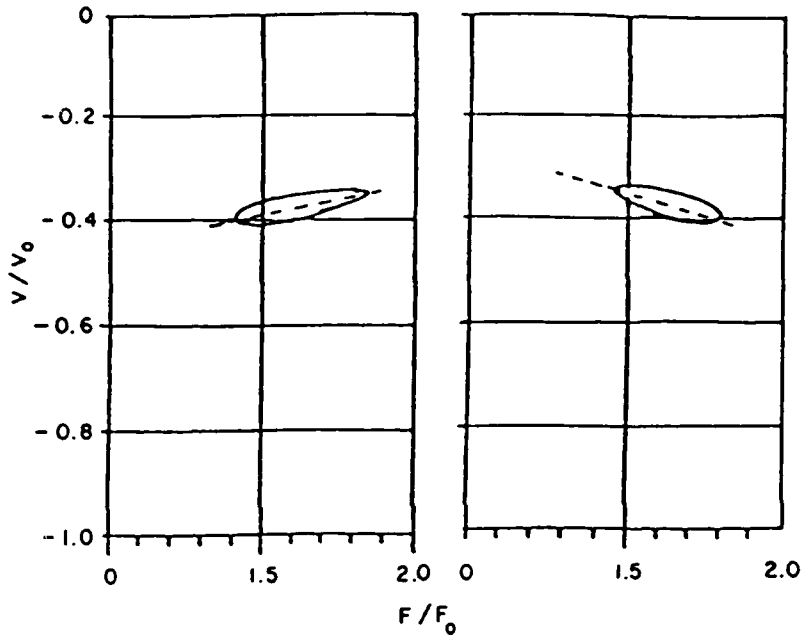


FIGURE 34. Velocity vs driving field; positive slope denotes gain, negative slope denotes loss.

When the net slope of the major axis is positive power is delivered by device. When the net slope is negative, power is dissipated. Both results are consistent with the phenomena of negative differential mobility. For nonuniform fields the question again arises as to how high in frequency we can go and still obtain amplification of the imposed signal. The situation is likely to be similar to that obtained from the small signal and large signal self-excited oscillation studies, and to depend critically upon device design. While the upper frequency limit for a particular device has not yet been obtained, steps have already been taken to couple solutions to the BTE with a driving sinusoidal source to design transferred electron amplifiers. These transferred electron amplifiers were designed using indium phosphide (InP)[23].

THREE-TERMINAL DEVICES: TWO-DIMENSIONAL DRIFT AND DIFFUSION STUDIES

The situation with three terminal elements is somewhat different. In this case, unless we specifically are interested in active oscillations as will occur in transferred electron logic elements, or with switching devices the relevant figures of merit are associated with the frequency at which a device amplifies the power of an incoming signal. For example, the frequency at which the current gain of an FET drops to unity is referred to as the cutoff frequency. In an approximate sense, the cutoff frequency is obtained from

$$f_T \sim \frac{g_m}{2\pi C_{gs}} \tag{78}$$

and is usually reduced by the presence of unwanted parasitic elements. The quantity f_T, when it is sufficiently high is usually the driving force behind the development of new device structures. The general expressions for f_T which are used in the design of a device depend in large part on the mobility model used. This is clearly unsatisfactory, and the situation requires assessment, which is difficult at this time. Frey, et al., using a highly approximate approach to examining submicron transport in two dimensions estimated f_T for a 2-micron long, 0.25-micron gate length device at approximately 60GHz.

More commonly, In its simplest description, f_T is expressed in terms of specific material quantities. For example, in the Shockley model, it is proportional to the carrier mobility and inversely proportional to the gate length.

$$f_T \approx \frac{\mu}{Lg^2} \tag{79a}$$

Under conditions of saturation in velocity, this expression has been estimated as

$$f_T = \frac{V_s}{2\pi Lg} \tag{79b}$$

and for $Vs = 10^7$cm/sec, $Lg = 0.25\mu$m, $f_T \approx 66$ GHz. Note: velocity overshoot contributions should improve f_T, as well as the transconductance from which these expressions are obtained.

The experimental situation on the gate-length dependence of FETs is also somewhat uncertain on this matter. For example, a recent study by Chao et al.[25], showed (1) an increased drain current with shorter gates, (2) significantly larger pinch-off voltage for shorter gate devices, (3) a slight increase in the transconductance of the shorter gate devices, and (4) the output conductance of the short gate device was not necessarily larger than that of the long gate device. The general conclusion of the study was that by keeping other device material parameters constant, no improvement would necessarily occur by going to short gate length devices.

Whether the conclusions of the Chao et al., study are going to be confirmed by a number of other studies remains uncertain. But diagnostic techniques are required for determining the effect of the space charge distribution on device performance and, in particular, when the speed enhancement can be utilized. One such set of measurements, which appears to be sensitive to the distribution of space charge within the device are the small signal circuit parameters. This was first demonstrated by Engelmann et al.[26], for longer gate length InP and GaAs devices. We review this below for long devices and suggest this as a method of examining the space charge structure of short gate length devices. The method involves obtaining the self-consistent solutions for the small signal microwave parameters of a device and identifying their frequency dependence with specific space charge profiles within the device. The calculation does not stem from the moments of the BTE, which, of course, should be used. Rather, it is derived from the standard drift and diffusion equations. In this case using parameters appropriate to GaAs.

CALCULATION OF THE SMALL SIGNAL "Y" PARAMETERS

This calculation is a numerical one. The equations solved are the two-dimensional continuity equation

$$- e \frac{\partial n(x,t)}{\partial t} + \text{div} \cdot J = 0 \tag{80}$$

the drift and diffusion equation

$$J = - e \left\{ n(x,t)\ v(F) - D(F) \nabla n(x,t) \right\} \tag{81}$$

Poisson's equation

$$\nabla^2\Psi(x,t) = + \frac{e}{\epsilon}\left[n(x,t) - n_0(x)\right] \tag{82}$$

and the relevant circuit equations. In the above X denotes the two dimensional position vector. The FET configuration is shown in figure 35, for a structure with a doping of $10^{15}/cm^3$, the low doping was chosen to reduce computational time.

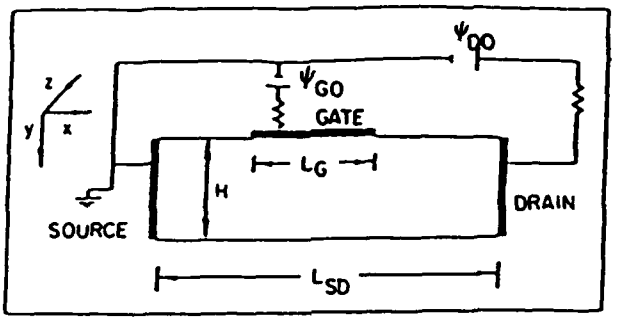


FIGURE 35. Two-dimensional device circuit configuration for the small signal calculations. Relevant dimensions are $L_G = 1.2\mu m$, $L_{SD} = 10\mu m$ and $H = 1.95\mu m$.

The small signal calculations are obtained as small time dependent perturbations of time independent steady state solutions. Since we are treating GaAs, the possibility of large signal domain transit time oscillations must be considered. Since this would only serve to complicate matters and render the concept of small signal parameters specious, attention was concentrated on thin "ND" product FET's that do not sustain instabilities [Grubin et al., 1980][27].

The steady state or dc characteristics of the device were self-consistently computed and are displayed in figure 36. Typical carrier density profiles are represented in figures 37 and 38.

Figure 37 displays results for a bias sufficiently high to generate a high field domain under the gate contact.

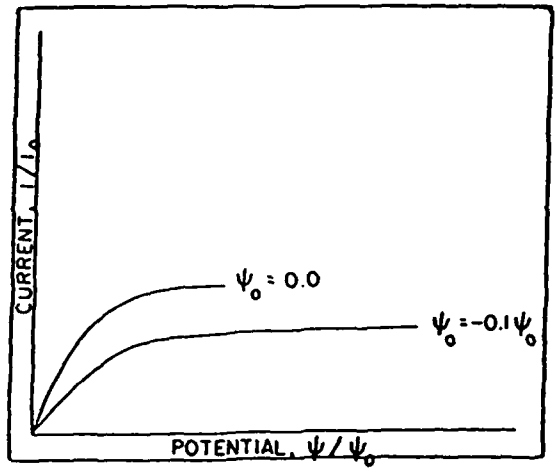


FIGURE 36. Steady state drain current vs drain potential for two values of gate bias. Here $I_o = N_o evpA$ and $\Psi_o = FpL$. Where v_p is the peak gallium arsenide carrier velocity and F_p is the threshold field for negative differential mobility.

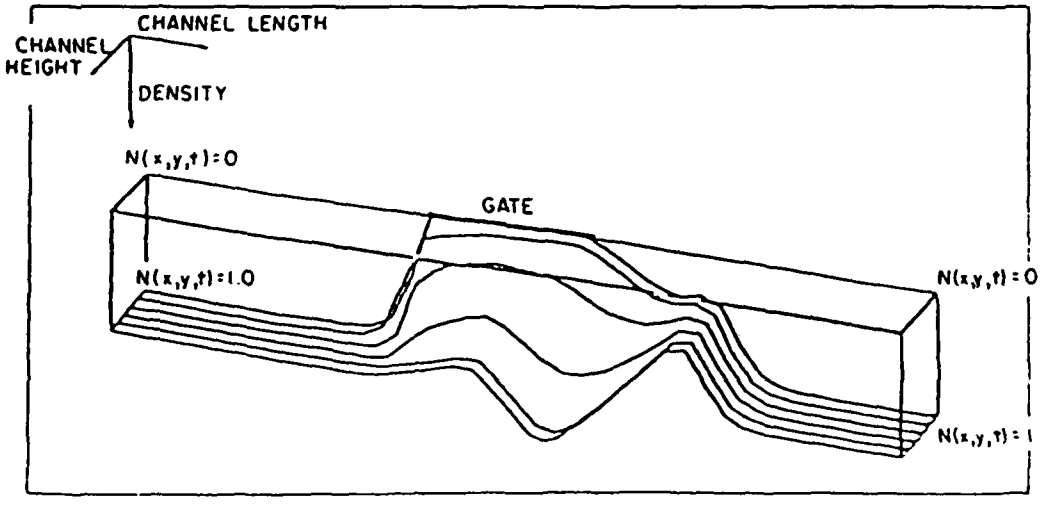


FIGURE 37. Projection of electron density within the FET. Channel height increases toward the bottom of the channel. Channel length is along a direction of increasing x. Density, which is in normalized units, increases in the downward direction. For this calculation $\Psi_G = 0.0$. and $\Psi_D = 1.0\Psi_o$.

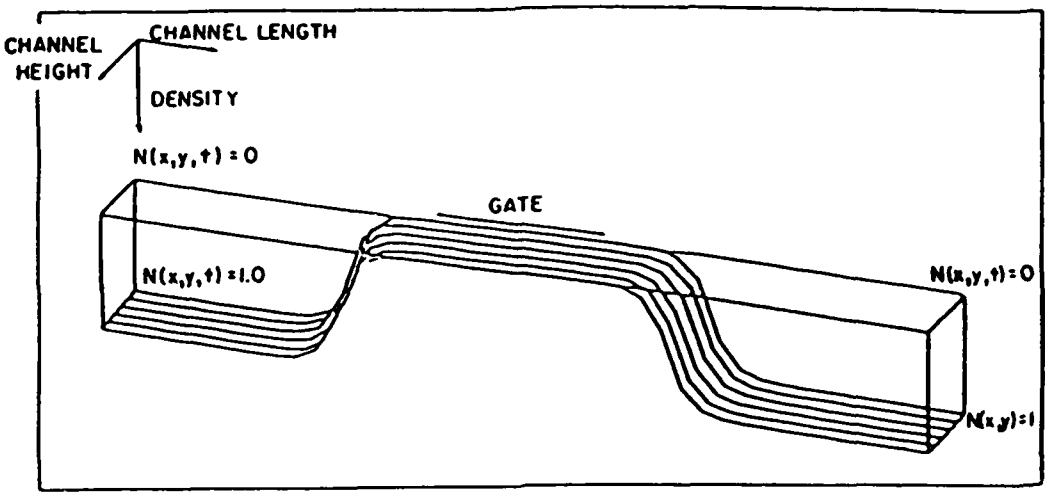


FIGURE 38. Carrier density for $\Psi G = -0.6\ \Psi_o$ and $\Psi_D = 0.4\ \Psi_o$. The current density for this case is $I_D \approx 0.0$.

Figures 37 and 38, and their respective current, and voltage values represent the starting point of the calculation. Subsequently, a small signal square wave voltage (see e.g., figure 39) was superimposed on the drain contract potential, with the gate contact potential fixed at its steady state value. The resulting

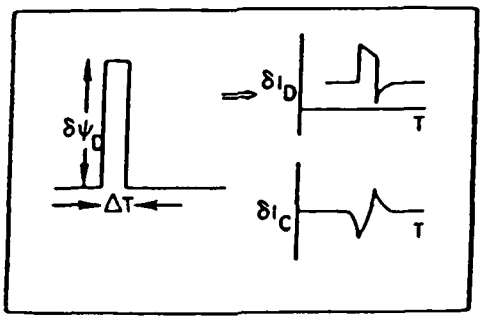


FIGURE 39. Representation of the small signal disturbance. A square wave pulse is applied to the drain contact, altering the drain contact boundary condition: $\Psi_D = \Psi_D + \delta\Psi_D$. There is a subsequent change in the drain and gate currents, δI_D and δI_C.

change in source, gate and drain currents are then computed subject to the constraint

$$\delta I_s(T) = \delta I_d(T) + \delta I_g(T) \tag{83}$$

The gate and drain current, and the drain potential are then Fourier analyzed. The ratios for a fixed gate potential are identified as the admittance parameters:

$$Y_{12}(\Omega) = \frac{\delta I_G(\Omega)}{\delta \Psi_D(\Omega)} \tag{84}$$

$$Y_{22}(\Omega) = \frac{\delta I_D(\Omega)}{\delta \Psi_D(\Omega)} \tag{85}$$

A similar exercise is performed for a perturbation on the gate contact, with the identification of two additional admittance parameters

$$Y_{11}(\Omega) = \frac{\delta I_G(\Omega)}{\delta \Psi_G(\Omega)} \tag{86}$$

$$Y_{21}(\Omega) = \frac{\delta I_D(\Omega)}{\delta \Psi_G(\Omega)} \tag{87}$$

The "Y" or admittance parameters are dependent on the space charge and potential profiles and are thus bias dependent. In addition, it is generally assumed that, about a given bias point, the small signal currents add in a linear way. Then:

$$\delta I_G = (Y_{11} + Y_{12})\ \delta \Psi_G - Y_{12}(\delta \Psi_G - \delta \Psi_D) \tag{88}$$

$$\delta I_D = (Y_{21} - Y_{12})\ \delta \Psi_G + Y_{12}(\delta \Psi_G - \delta \Psi_D) + (Y_{22} + Y_{12})\ \delta \Psi_D \tag{89}$$

with the equivalent circuit as shown in figure 40. The terms

$Y_{11} + Y_{12}$ are generally identified as source-gate admittance parameters, while $-Y_{12}$ is referred to as the gate-drain admittance.

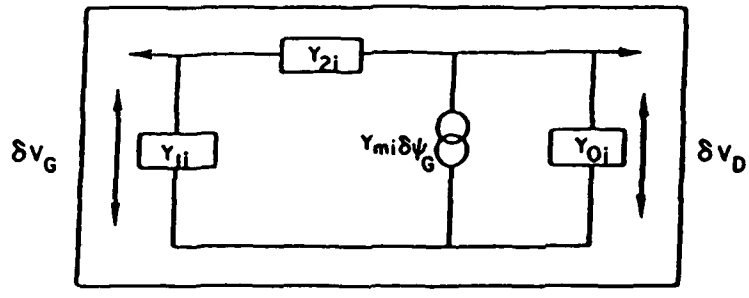


FIGURE 40. General form of the FET intrinsic circuit model with reference to equations 88 and 89: $Y_{0i} = Y_{22} + Y_{12} \cdot Y_{11} - +Y_{12} \cdot Y_{21} - \nabla^{12}$ and $Y_{mi} - \nabla_{21} - \nabla_{12}$. See also Cobbold (1970).[28] Figure 5.3.

The advantage of this type of representation is that a heuristic connection is often made to circuit elements as shown and intuitive design models may be expanded upon as shown in figure 41, which represents the first order equivalent circuit for small signal calculations. This circuit, while commonly used, does not adequately express the presence of high field dipole layers; as we show below.

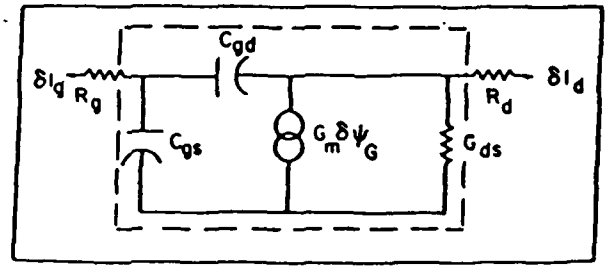


FIGURE 41. First order equivalent circuit terms within dashed border represent zeroth order circuit.

Selective "Y" parameters versus frequency are displayed in figures 42 through 44 for a range of drain bias values and a moderate value of gate bias[29].

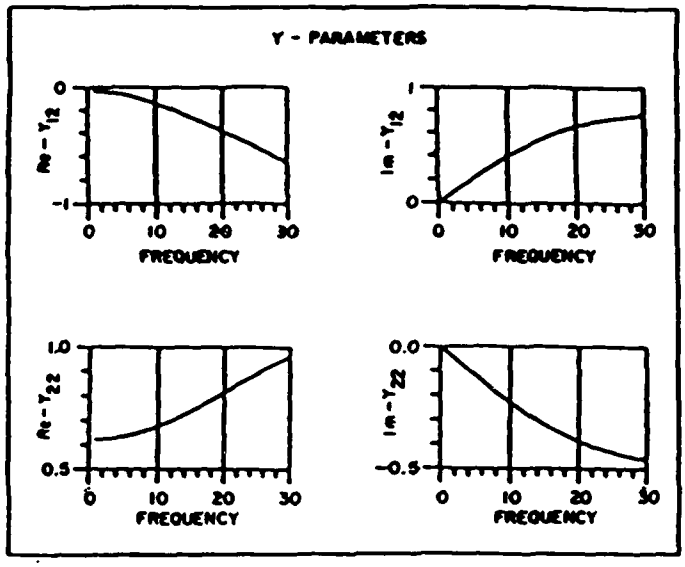


FIGURE 42a. Frequency dependent admittance parameters Y_{12} and Y_{22} obtained by Fourier analysis. The dc bias level for this calculation are $\Psi_{Do} = 0.1\Psi_o$. $\Psi_{Go} = -0.1\Psi_o$.

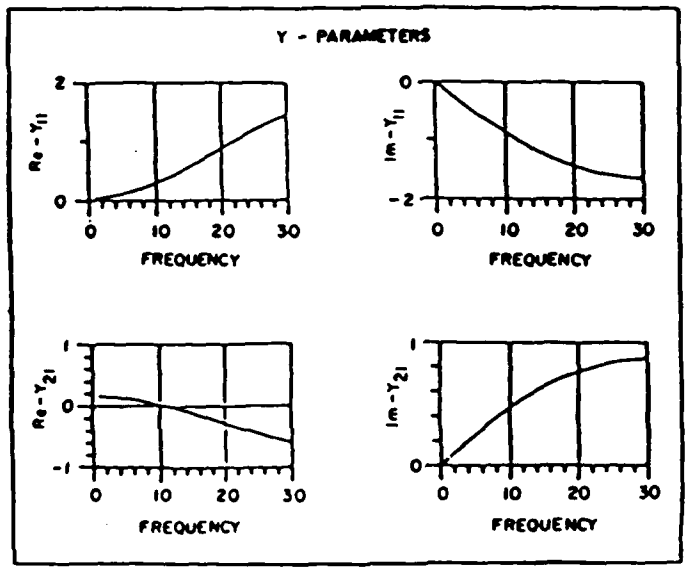


FIGURE 42b. Frequency dependent admittance parameters Y_{11} and Y_{21}.

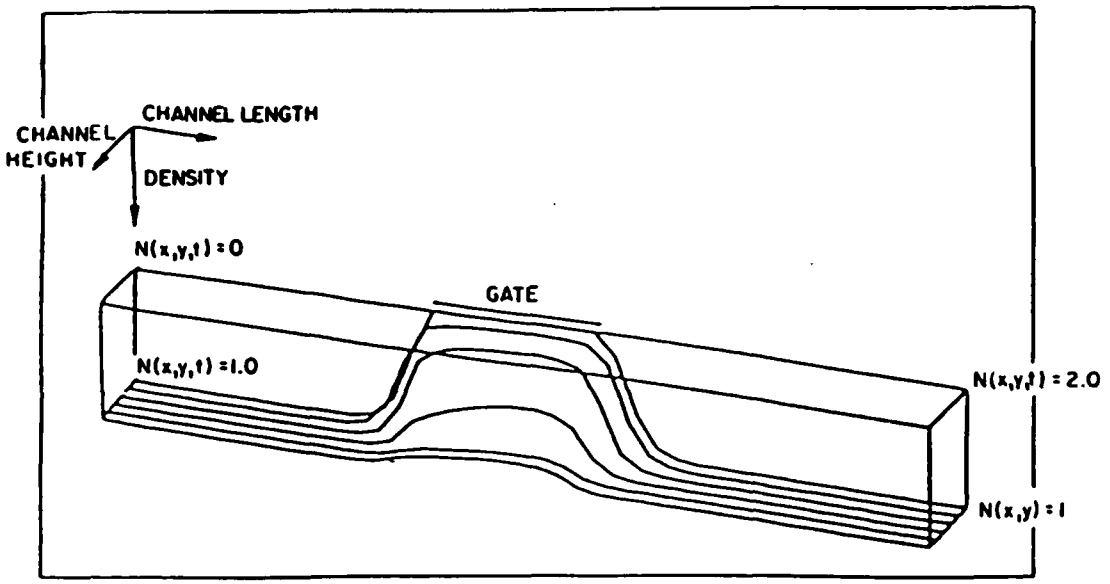


FIGURE 42c. Projection of steady state electron density for this bias level at which the admittance parameters were obtained.

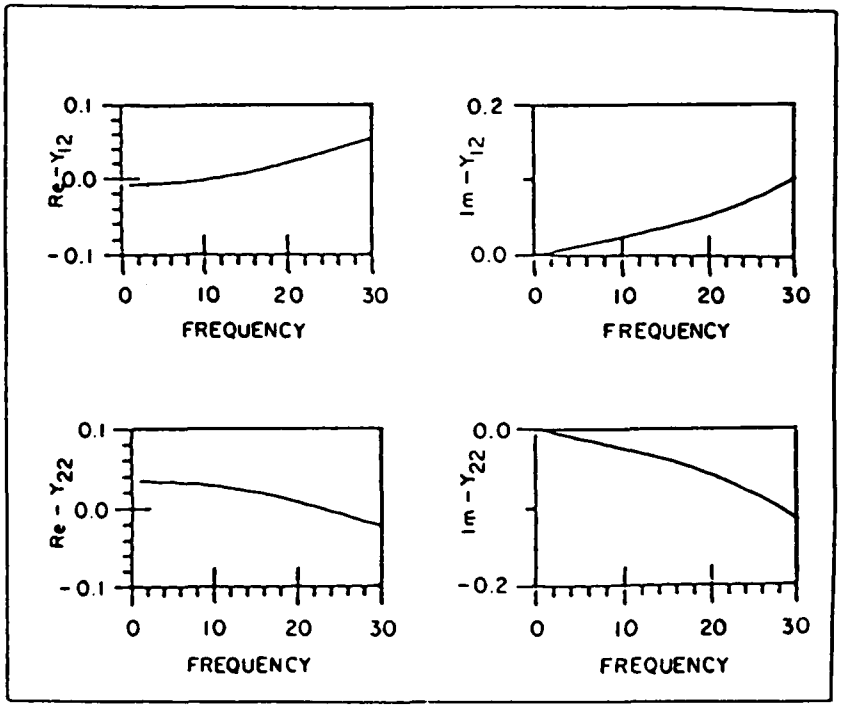


FIGURE 43a. As in 43a but for $\Psi_{Do} = 1.0\Psi_o$, $\Psi_{Go} = -0.1\Psi_o$.

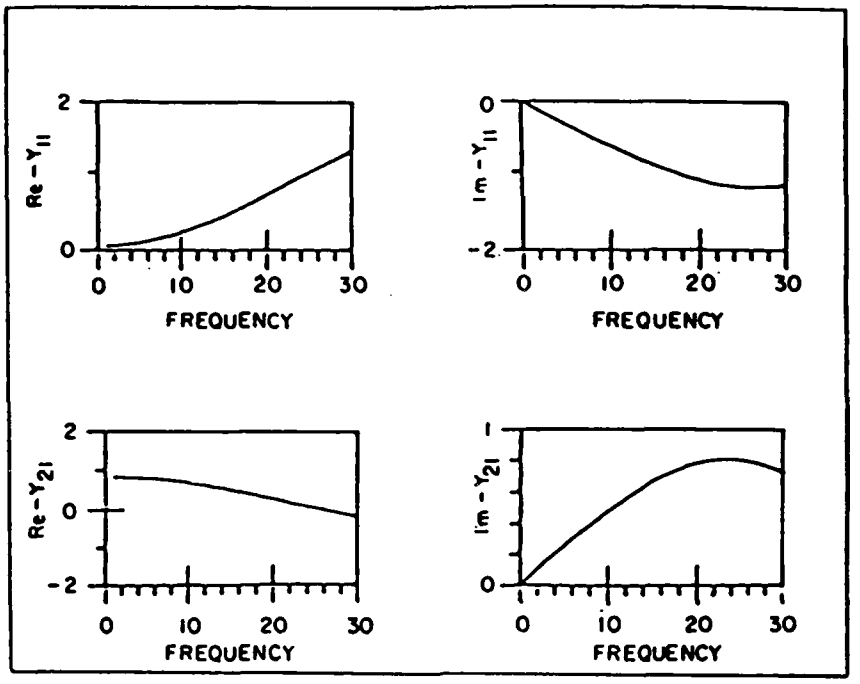


FIGURE 43b. As in 42a but for $\Psi_{Do} = 1.0\Psi_o$, $\Psi_{Go} = -0.1\Psi_o$.

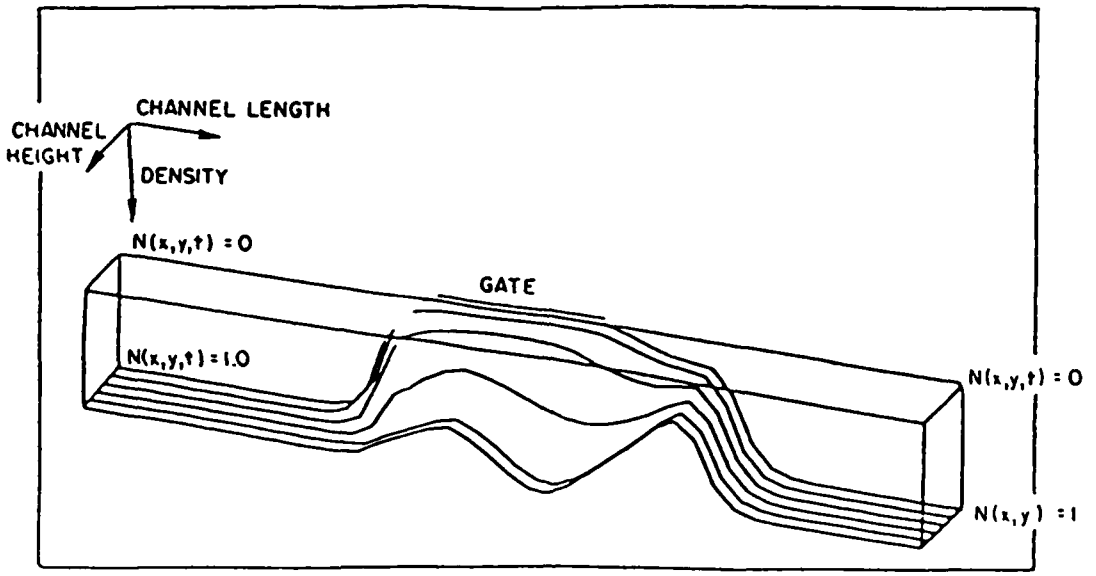


FIGURE 43c. Carrier density projections.

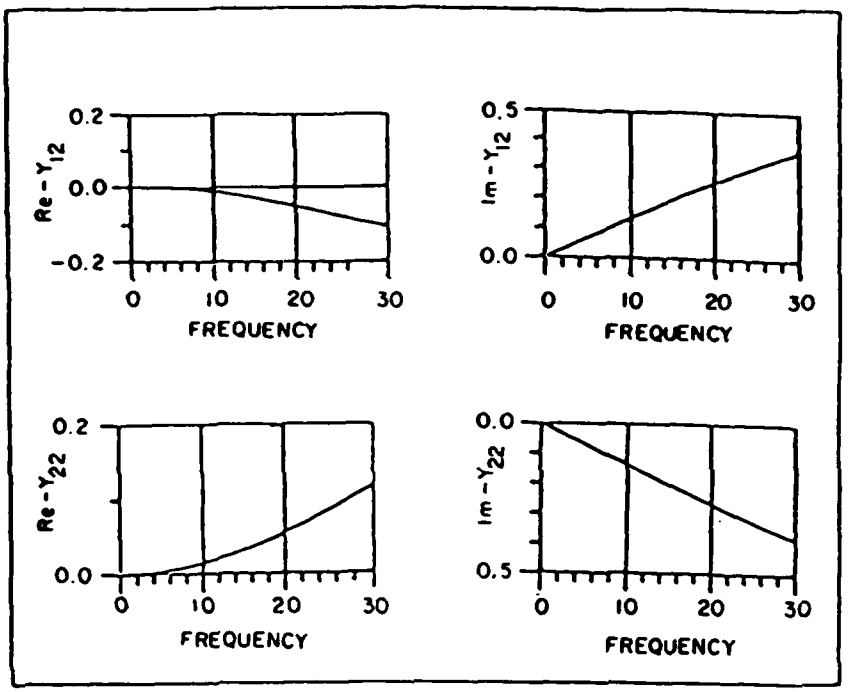


FIGURE 44a. As in figure 42, but for $\Psi_{Ho} = -1.0\Psi_o$.

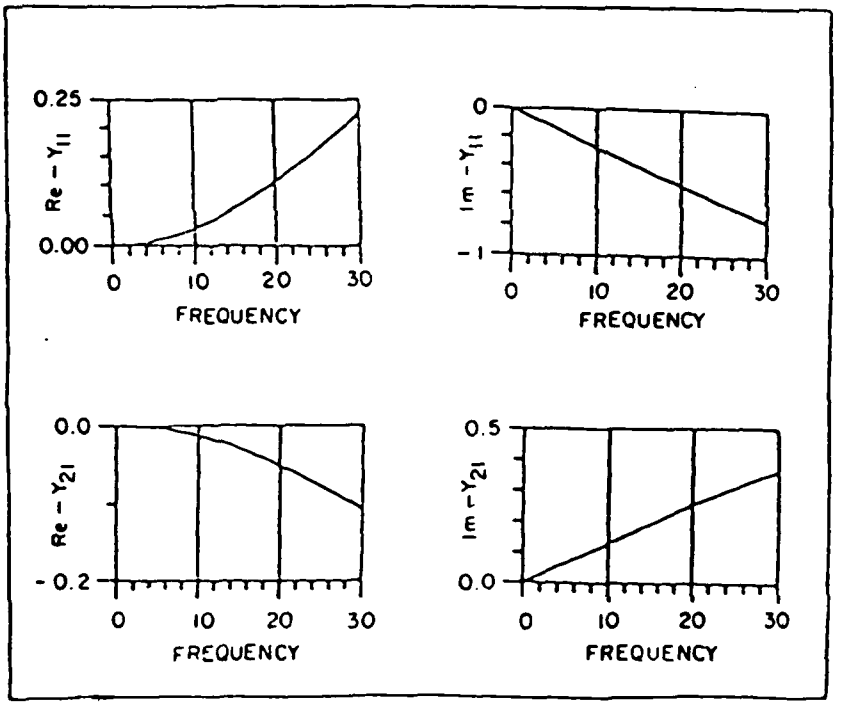


FIGURE 44b. As in figure 42, but for $\Psi_{Ho} = -1.0\Psi_o$.

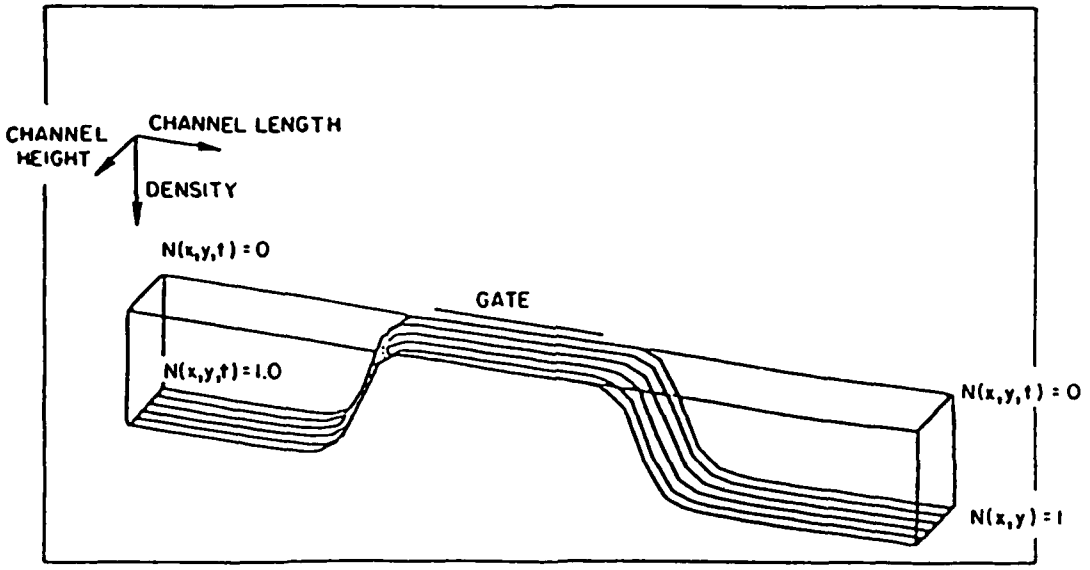


FIGURE 44c. As in figure 42, but for $\Psi_{H0} = -1.0\ \Psi_0$.

In these figures, the admittance parameters are expressed units of G_0, where

$$G_0 = \frac{N_0 e \mu_0 HW}{L} \tag{90}$$

the frequency is in multiples of f_0

$$f_0 = 0.542\text{GHz} \tag{91}$$

and the bias levels are in multiples of Ψ_0=3.2 kv/cmx10^{-4}cm=.32v. Some broad general features are clear and these are expressed in terms of the zeroth order circuit parameters of figure 41.

$$Y_{110} = -\ j\Omega C_{gs} - j\Omega C_{gd} \tag{92}$$

$$Y_{120} = j\Omega C_{gd} \tag{93}$$

$$Y_{210} = -G_m + j\Omega C_{gd} \tag{94}$$

$$Y_{220} = -G_{ds} + j\Omega C_{gd} \tag{95}$$

We begin a discussion of the "Y" parameter with an examination of drain components Y_{12} and Y_{22}. For a fixed gate bias of -0.1 normalized units and at low values of drain bias, $|\text{Im } Y_{22}| < |\text{Im } Y_{12}|$. This situation remains until a high field domain forms within the channel and $|\text{Im } Y_{22}| \geq |\text{Im } Y_{12}|$. The situation in which the drain bias is fixed and the gate bias is increased, results in movement of the depletion layer toward the bottom of the channel, with $|\text{Im } Y_{22}| \geq |\text{Im } Y_{12}|$. Insofar as the zeroth order element requires $|\text{Im } Y_{22}| = |\text{Im } Y_{12}|$, at the very least, the

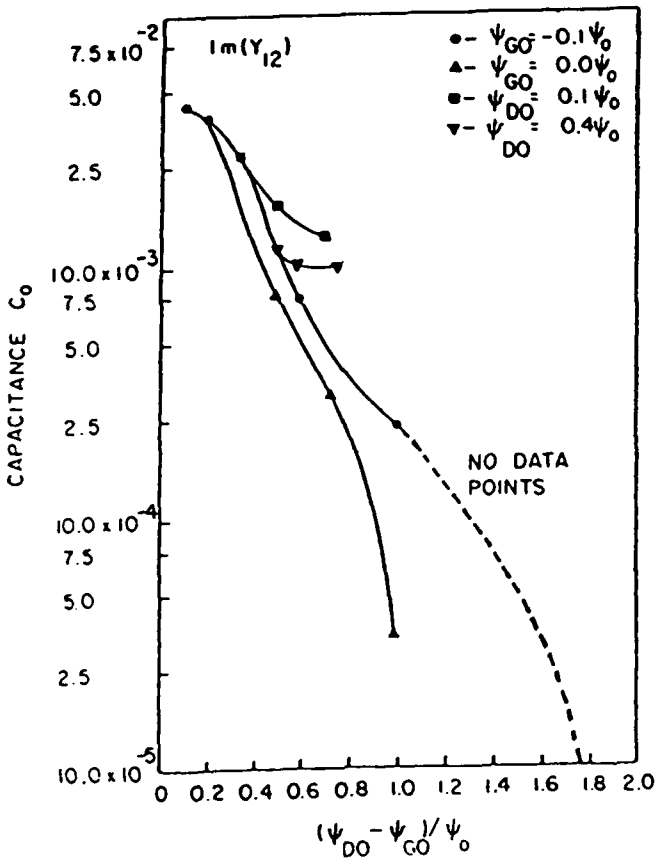


FIGURE 45. Gate to drain capacitance, as obtained from Y_{12}, vs $(\Psi_{DO} - \Psi_{GO})/\Psi_o$. Circular and triangular results are for a fixed gate bias. Square and inverted triangular data are for a fixed drain bias. For this calculation $C_o = 3.25 \times 10^2 \epsilon HW/L$.

drain element Y_o appears to require the inclusion of a bias dependent capacitor.

Additional study of the imaginary part of Y_{12} is shown next. As seen in figure 45, the capacitance of the system undergoes a precipitous drop. A feature also seen experimentally (Engelmann et al., 1977)[26]. The broad features of this result were discussed by Engelmann, et al, (1977)[26], and is represented in figure 46. Essentially, the presence of the high field domain is cutting off the gate drain coupling.

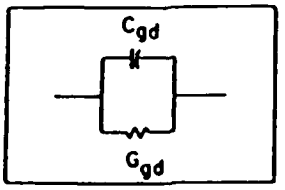


FIGURE 46. Schematic representation of the gate-drain admittance in the presence of a high field domain. C_{gd} represents a small signal conductance which exhibits a frequency dependent region of negative differential conductivity when a high field domain is present.

We recall that C_{gd} represents the gate-to-channel capacitance on the drain side of the channel as indicated in figure 47, and is a measure of the change in channel depletion charge as a result of changes in drain bias. Now, while a capacitance change is expected

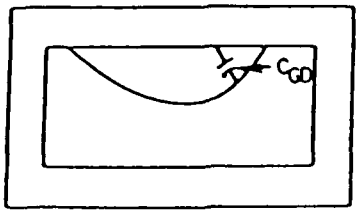


FIGURE 47. Schematic of the gate-to-drain capacitance as a measure of the change in channel depletion charge resulting from changes in drain bias.

as the edge of the depletion layer moves toward the bottom of the channel, the enhanced drop in capacitance suggests that most of the modulation is across the dipole layer. This is illustrated in figure 45.

There are basically two sets of data in figure 45. The bold line is primarily for a moderate value of gate bias. The most significant drop in capacitance occurs when a domain forms. The situation when the increase in gate to drain voltage is a result of an increased gate bias does not result in the dramatic decrease in capacitance. For the latter case, dipole domain do not form. We note that the qualitative features of this calculation are in agreement with the experimental results of Engelmann, et al., (1977)[26].

For the parameter Y_{22}, we simply note that at low bias levels $G_{ds}>0$, whereas at high values of bias, the results strongly suggest a frequency dependent region of small signal negative conductance, a feature consistent with the presence of high field domains.

We next consider Y_{11} and note again that to account for the high frequency high bias behavior of Y_{11}, it is necessary that C_{ds} exhibit a frequency dependence which allows for a range of small signal negative conductance. With regard to Im Y_{11}, which we write as

$$\text{Im } Y_{11} = -\Omega\, C_{11}(\Omega) \tag{96}$$

we note from figures 42 and 44 that at low bias levels C_{11} is somewhat greater than twice C_{gd}. At high bias levels when domains form $C_{11}(\Omega)$ is at least an order of magnitude greater than C_{gd}. Under low or moderate gate bias levels $C_{11}(\Omega)$ does not exhibit a precipitous drop in value. Rather, at first the capacitance decrease corresponding to a movement of the depletion layer toward the bottom of the channel. This was also the initial behavior of the gate to drain capacitance. Further increases in drain bias result in domain formation and space charge injection into the depleted zone. The effective capacitance shows a corresponding increase corresponding to an increase in the stored charge. This is displayed in figure 48. For the situation where the net voltage increase is due to an increase in gate bias, where no domains form there is the expected drop in capacitance, as also reflected in the gate drain capacitance without domains. This is also shown in figure 48. The experimental situation shows broad agreement with the numerical results.

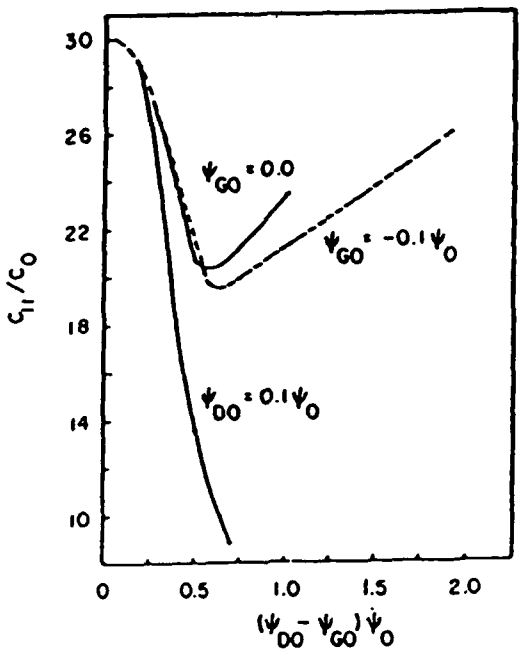


FIGURE 48. Capacitance C_{11}, obtained from Y_{11}, vs $(\Psi_{DO}-\Psi_{GO})/\Psi_o$. For this calculation $C_o = cHW/L$. Note $C_{11} = C_{gs} + C_{gd}$.

The two remaining items of interest here are theoretical/experimental comparison of the transconductance, and current-gain cutoff frequency. Figure 49 displays the transconductance, Re Y_{21}, vs $V_{DS} - V_{GS}$, for two values of gate bias. We see

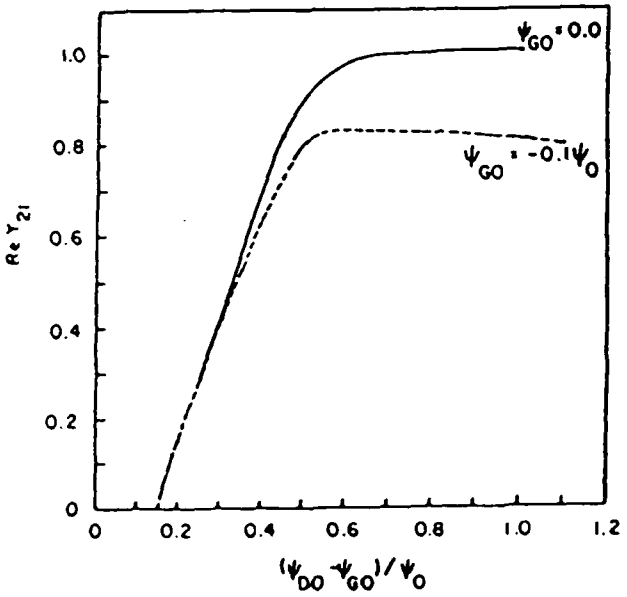


FIGURE 49. Re Y_{21} (in multiples of C_o) vs $(\Psi_{DO} - \Psi_{GO})\Psi_o$. There is saturation at high bias levels.

the presence of near saturation in both sets of data. This is also seen experimentally. Our data does not extend to high enough drain bias levels to determine whether a corresponding decrease in transconductance occurs. The current-gain cutoff frequency is obtained from the expression

$$f_t = \frac{\Omega \, \mathrm{Re}\, Y_{21}}{2\pi \left| \mathrm{Im}\, (Y_{11} + Y_{12}) \right|} \tag{97}$$

and is shown in figure 50. It is seen that saturation in f_T occurs under the presence of domain formation. The decrease in f_T at the higher drain bias levels appears to be associated with an increase in the source-gate capacitance. Experimental observations again show qualitative agreement with theory.

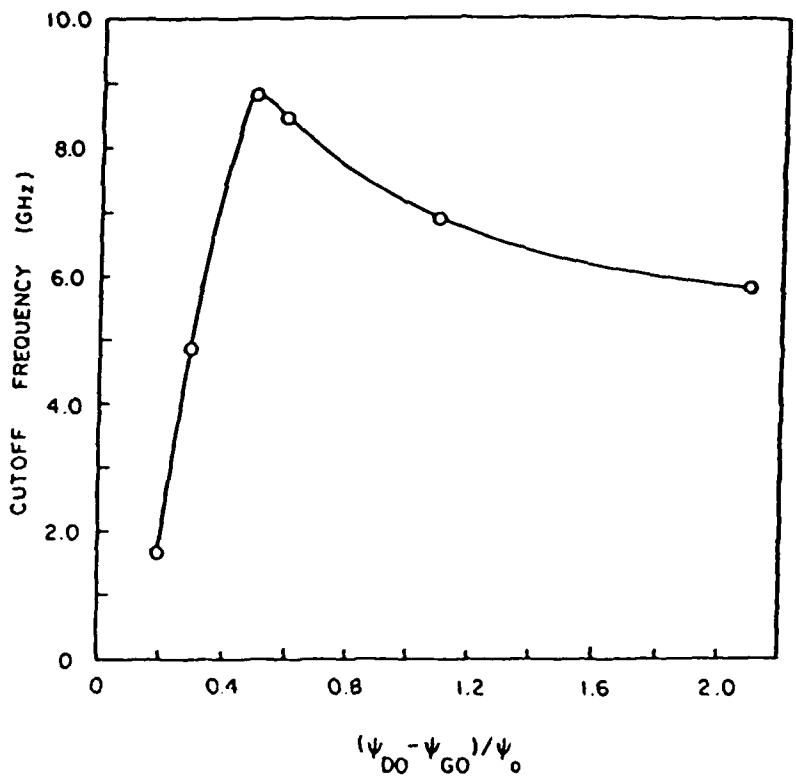


FIGURE 50. Cutoff frequency vs $(\Psi_{DO} - \Psi_{GO})/\Psi_o$ for $\Psi_{GO} = -0.1\Psi_o$.

THREE-TERMINAL DEVICES: TWO-DIMENSIONAL MOMENTS OF THE BOLTZMANN TRANSPORT EQUATION

The two-dimensional simulations are new and as a result there has not been much experience to rely upon. The boundary conditions are more complicated insofar as parallel components of velocity are also required to characterize the boundary and the contact. The result is that the familiar "simple" one-dimensional contact descriptions need rethinking. The next few figures will illustrate some of the present observed features. The simulations have been performed for a two-micron long device, with a gate length of 0.5μm with structure of figure 51. Because most of the

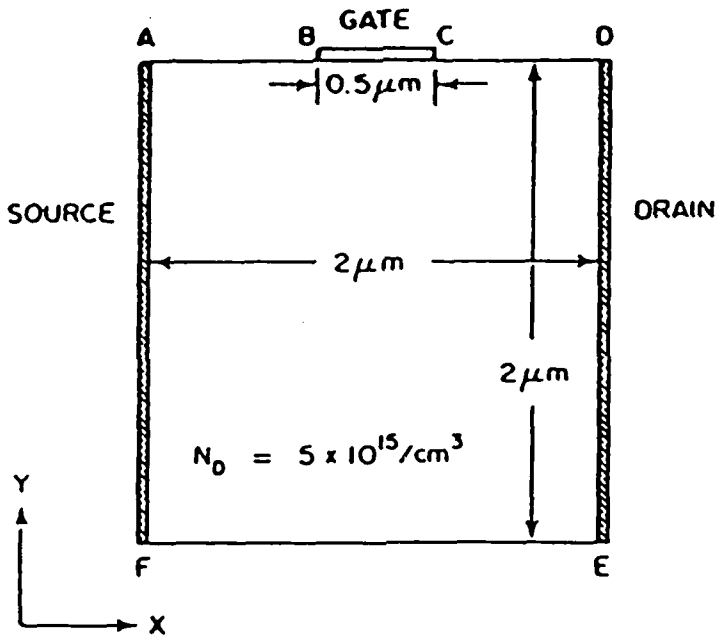


FIGURE 51. Schematic of "horizontally" placed submicron gate MESFET used in submicron dimensioned BTE "moment" simulations

one-dimensional calculations were performed for doping levels of $5x10^{15}/cm^3$, the doping levels of the two-dimensional calculation were taken at this value. The boundary conditions are listed in Table 6.

Figure 52a displays the distribution of total charge within the two-micron long device with a gate potential kept at ground, and a drain potential at 1 volt. The source contact was an injecting contact. The normal component boundary conditions were similar to those obtained from the one-dimensional case. For the Schottky

TABLE 6. Boundary Conditions

AB, CD, EF	$V_1^y - V_2^y = 0$, otherwise derivatives are zero
BC	$V_1^y - V_2^y = 0$, $\frac{\partial V_1^x}{\partial Y} - \frac{\partial V_2^x}{\partial Y} = 0$, $\frac{\partial T_2}{\partial Y} = 0$, T_1 = constant
	$\frac{\partial}{\partial Y}N_1N_2 - \frac{\partial}{\partial T}N_2T_2 = 0$
DE	$\frac{\partial^2(\text{all})}{\partial X^2} = 0$, Ψ = constant
FA	$V_1^x = \text{constant}\cdot\frac{\partial \Psi}{\partial X}$, $\Psi = 0$, $\frac{\partial V_2^x}{\partial X} = 0$, $T_1 = 1$, $\frac{\partial T_2}{\partial X} = 0$
	$V_1^y - V_2^y = 0$, $\frac{\partial^2 V_1}{\partial X^2} = 0$, $\frac{\partial^2 N_2}{\partial X^2} = 0$, T_1 = constant

gate, the normal component of velocity was set to zero, and the Γ-valley component of electron temperature was set to 900°K. Thus some transfer was effected at the gate. The parallel component of velocity was set to zero. What is apparent from these calculations is the space charge accumulation at the drain boundary, a feature already observed in the one-dimensional simulations. This downstream charge accumulation is a consequence of electron transfer and is likely to lead to reduced current drives. An increase in drain bias serves to enhance the level of space charge accumulation without any real improvement in current level. The Γ-valley charge density is displayed in figure 52b, and for this bias level, there is a certain degree of electron transfer. Within the channel near the bottom of the device the fractional component of $N_\Gamma \approx 0.9N_o$. Again, most of the current is carried by the Γ-valley electrons. The distribution of satellite valley carriers is shown in figure 52c. It tends to emphasize that a larger fraction of L-valley carriers are partaking in transport immediately downstream from the gate contact than within the channel, but their net contribution to the total current is negligible. (There are also fewer total carriers in this

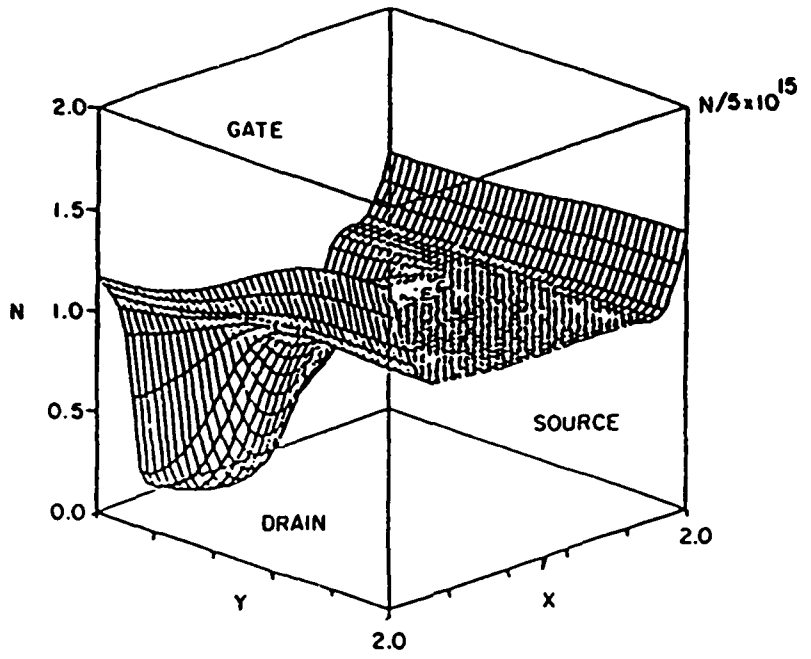


FIGURE 52a. Distribution of total charge within a three-terminal device (distance scales are nonlinear, the depletion region is approximately 1000A deep), for a potential of 1 volt on the drain.

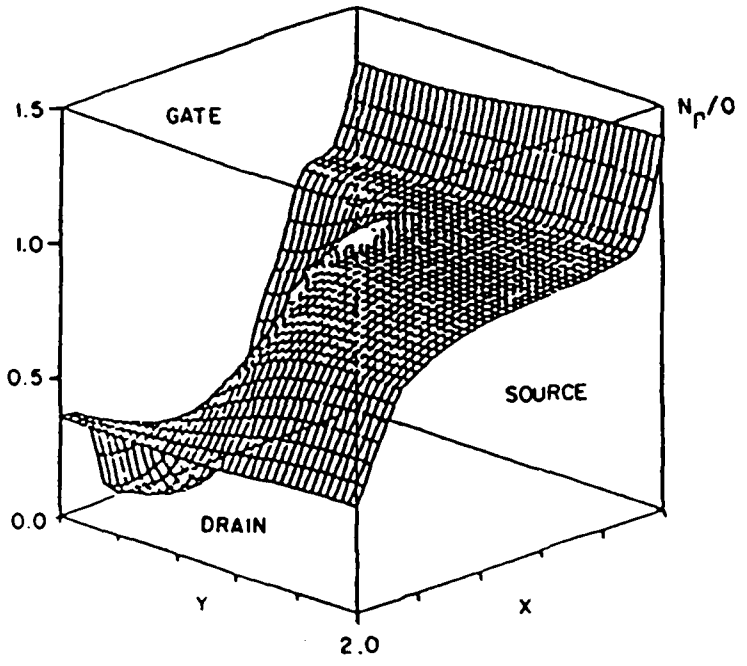


FIGURE 52b. As in (a), but for Γ-valley.

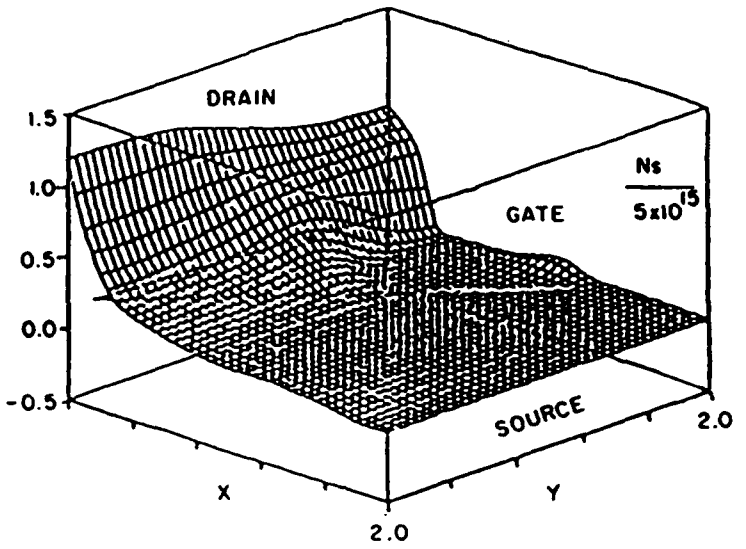


FIGURE 52c. As in (a), but for satellite valley.

region.) In figure 53 we have included the result of a calculation in which the boundary condition on the gate electron temperature was increased to 1500°K to show how this alters the charge distribution. At these bias levels no significant alteration in current level was observed. The total charge distribution for 1.25v on the case is also displayed in figure 54a. The observation is made of increased carrier accumulation at the drain boundary and enhanced electron transfer (figures 54b and 54c).

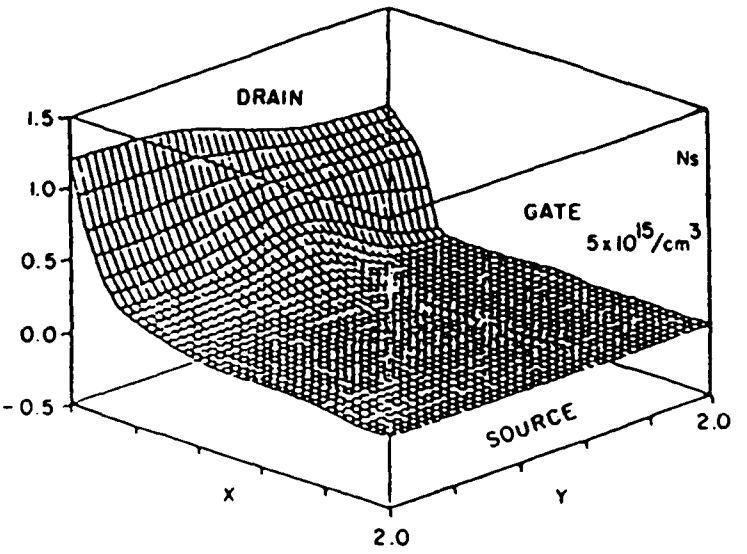


FIGURE 53. As in 52c. But for Tg = 1500°K.

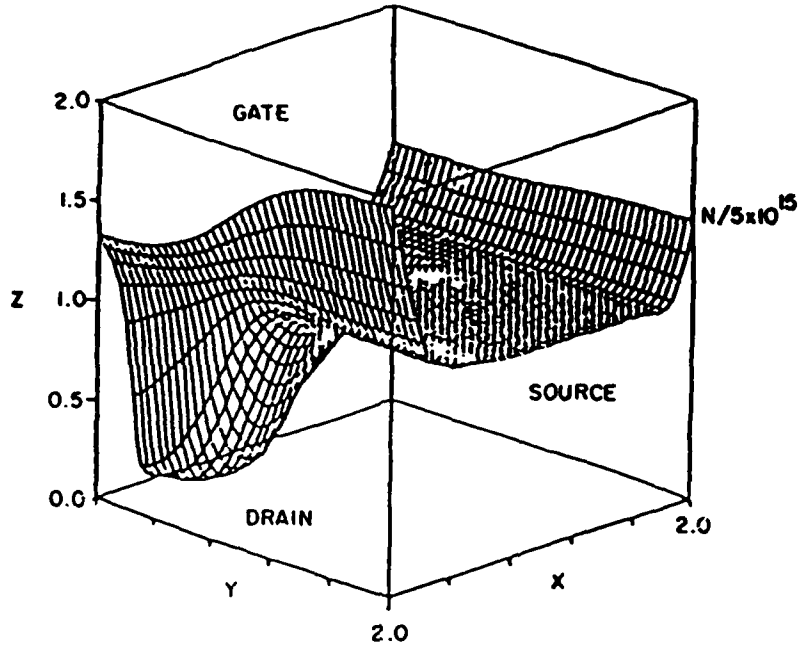


FIGURE 54a. As in figure 52a, but for a drain potential of 1.25V.

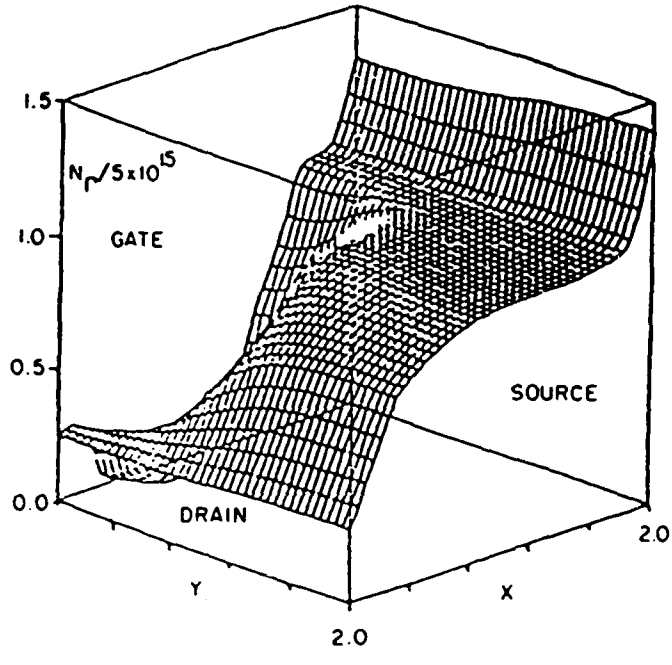


FIGURE 54b. Γ-valley carrier density.

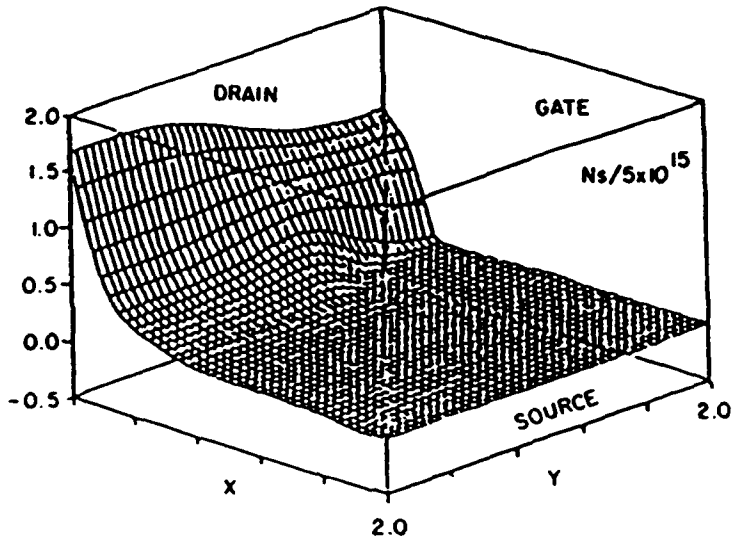


FIGURE 54c. Satellite valley population.

SCALING AND MATERIAL CONSIDERATIONS

There are two questions I want to address, and they have to do with the kinds of materials we should be looking for. The guide to this answer will come from the specific device goals. The question of interest concerns the significance of super velocity materials and how do we balance high speeds with sometimes reduced mobilities. To examine these issues we return to the moment equations as written in equations (51), (52) and (53). The first thing we will consider is what we will refer to as intrinsic scaling. This type of scaling is not dependent on device length explicitly, although it enters the discussion through the relation

$$x = \int v(t)dt \tag{98}$$

We turn to equations (51) through (53), assume a constant field across the device and set all other spatial derivatives to zero. Then

$$n_r' = n_s' \frac{\Lambda_s'}{\Lambda_r'} - \frac{1}{\Lambda_r'} \frac{\partial n_r'}{\partial t'} \tag{99}$$

$$n_r' m_r' v_r^{i'} = a n_r' \frac{\partial \Psi'}{\partial x_i'} - \frac{1}{\Pi_r'} \frac{\partial n_r' m_r' v_r^{i'}}{\partial t'} \tag{100}$$

energy:

$$n'_r T'_r = n'_s T'_s \frac{\Sigma'_s}{\Sigma'_r} + \left(dn'_r v'_r \frac{\partial \Psi'}{\partial x'_i} - \frac{1}{\Sigma'_r} \frac{\partial w'_r}{\partial t} \right) \cdot \frac{2}{3b} \tag{101}$$

We consider the situation as in Ref. 30 where all of the scattering rates are altered by the multiplicative constant λ, e.g.,

$$\Lambda_\Gamma \Rightarrow \lambda \Lambda_\Gamma \tag{102}$$

while the primed scattering rates are kept constant. Thus, e.g.

$$t_{ref} \Rightarrow t_{ref}/\lambda \tag{103}$$

We note that such materials as GaAs and InP are identified by their scattering rates, and constant scaling is directly relevant to a comparison of the two. We next consider the case where the coefficients a and d are invariant. For a constant v_{ref}, this is achieved by assuming the product

$$t_{ref} \, F_{ref} \tag{104}$$

to be constant. This requires that

$$F_{ref} \Rightarrow \lambda F_{ref} \tag{105}$$

The first consequence of this scaling is associated with the steady state velocity field curves. If

$$v_o(F)$$

represents the steady state field dependent velocity, taken as reference; and

$$v_\lambda(F)$$

represents the scaled velocity, then

$$v_\lambda(f) = v_o(F/\lambda) \tag{106}$$

In simplest terms this equations states that the magnitudes of the steady state and saturated drift velocities are unchanged by this

scaling. The low field mobilities differ by the factor λ. In a qualitative sense, many of the III-V NDR materials are subject to this type of scaling. It would not stand up in a quantitative sense because of variations in the phonon frequencies etc. In figure 55 we sketch the select scattering rates for InP and GaAs at room temperature. In figure 56 we display the scaled GaAs and InP

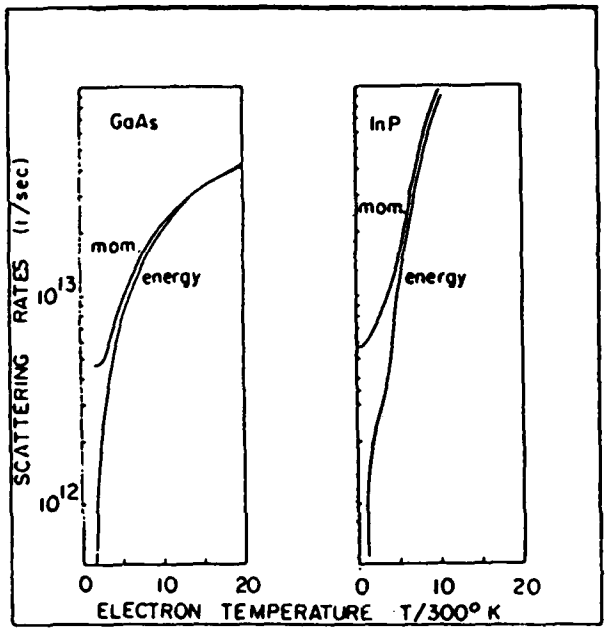


FIGURE 55. Comparison of Γ-valley scattering rates for GaAs and InP.

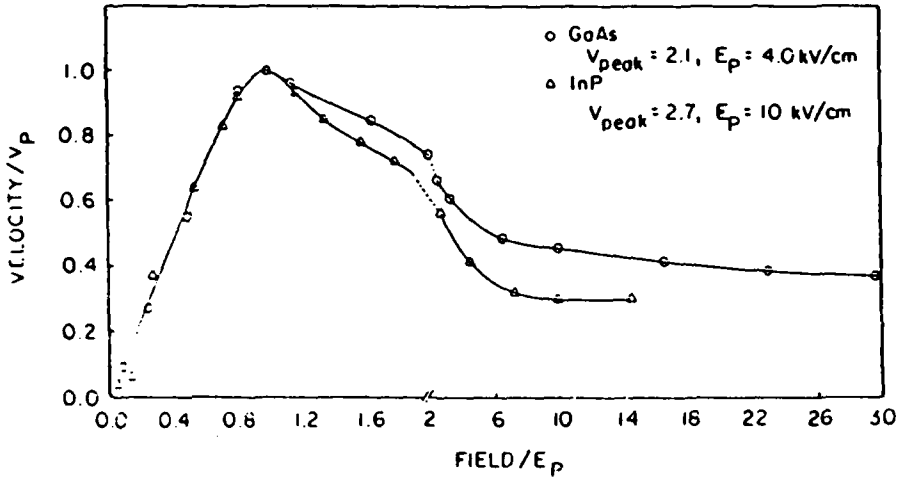


FIGURE 56. Normalized steady state field dependent carrier velocity. Simple linear scaling does not provide linear velocity scaling.

steady state velocity field relations. It is seen that linear scattering rate scaling does not yield linear velocity scaling.

The next item of interest concerns the transient response. Here with reference to equations (99) through (101) it is clear that constant λ scaling with

$$t' = t/t_{ref}$$

implies

$$v_\Gamma(t,F) = v_o(\lambda t, F/\lambda) \tag{107}$$

or in another form

$$v_o(t,F) = v_\Gamma(t/\lambda, \lambda F) \tag{108}$$

Consider the situation where the material under consideration is undergoing self-excited oscillations. Then the scaling law states that if GaAs has an upper frequency limit of 150GHz, and it were possible to design a scaled semiconductor material with λ = 2, then the latter would have an upper frequency limit of 300GHz. Similar remarks can be made with respect to small signal and driven oscillators. Simulations with InP self-excited oscillations have been performed and frequencies in excess of those found in gallium arsenide were obtained.

With regard to the more familiar transient transport calculations, as extrapolation from equations 107 and 108 yields the results shown in figure 57. Thus, transport in materials with enhanced scattering rates appears to approach equilibrium sooner.

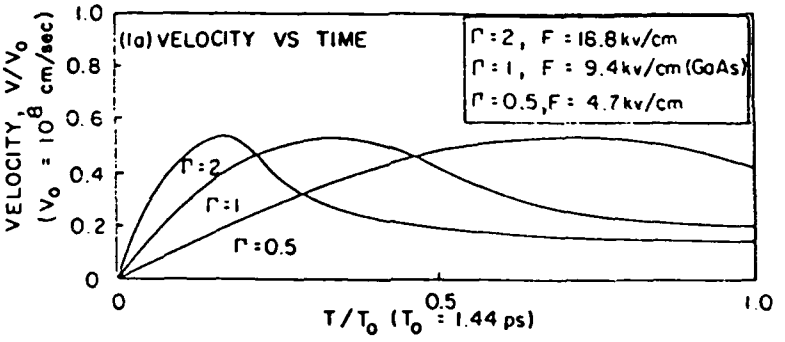


FIGURE 57. Scattering rate scaling as applied to velocity transients.

These conclusions, of course, must be folded into the fact that at higher values of field the carriers are driven to higher values of velocity and the scattering rates which are field dependent also increase. In terms of a comparison to real materials, we are not interested in scaled values, but rather in the response of two dissimilar materials to the same value of field. A simple sketch based upon previous comments is displayed in figure 58, which displays an estimate of the velocity for the Γ=1 element at a field of 18.8 kv/cm (dashed line). Note, the higher peak carrier velocity. Also shown in figure 58, is the velocity versus distance relation obtained through application of equation (98). The

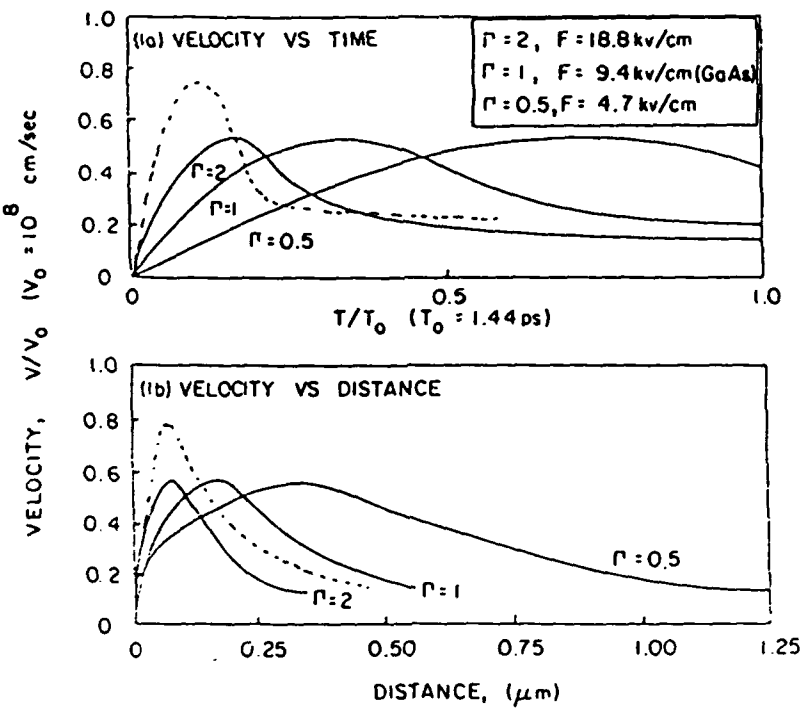


FIGURE 58. Scattering rate scaling as applied to velocity transients.

question is, how well do these very general comments, which can be obtained from some very general scaling of the Boltzmann transport equation, stand up against detailed numerical calculation. The answer is displayed in figures 59 and 60, where the general scaling principles appear to hold up fairly well.

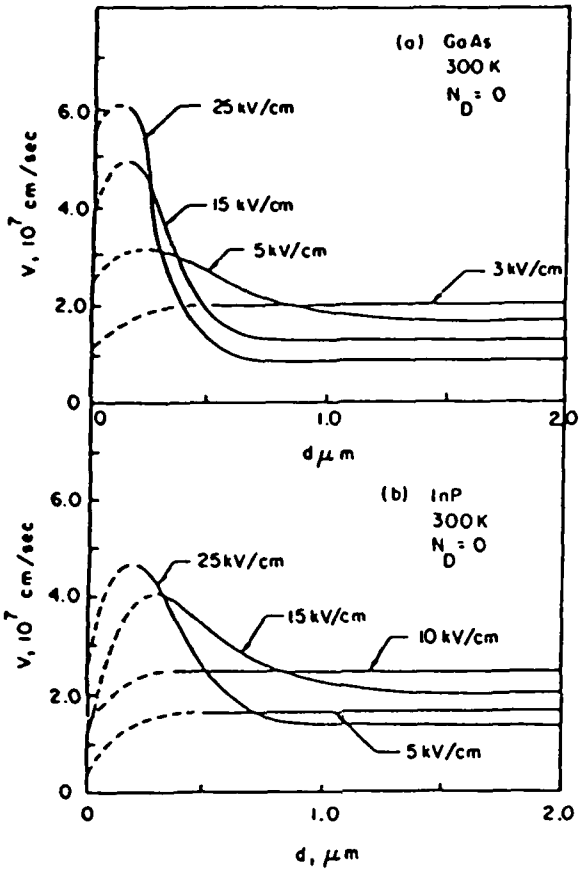


FIGURE 59. Drift velocity vs position in GaAs and InP for electrons release at $\chi = 0$ into various uniform fields: (a) GaAs, from Ruch [15]; (b) GaAs and InP, from Maloney and Frey[31].

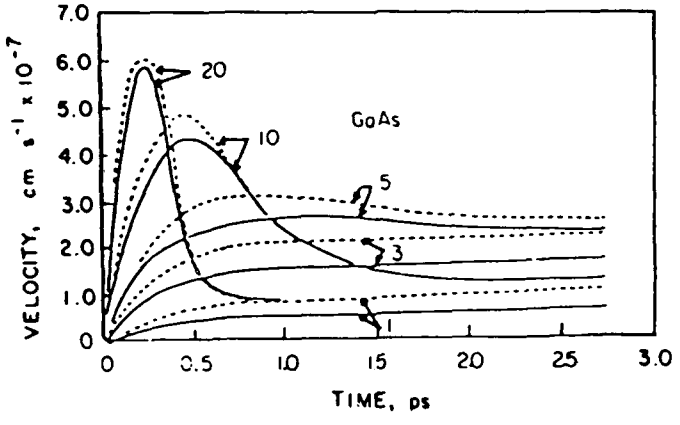


FIGURE 60a. GaAs transients. Solid curves denote impurity scattering; $N_p = 10^{17} cm^{-3}$.[32]

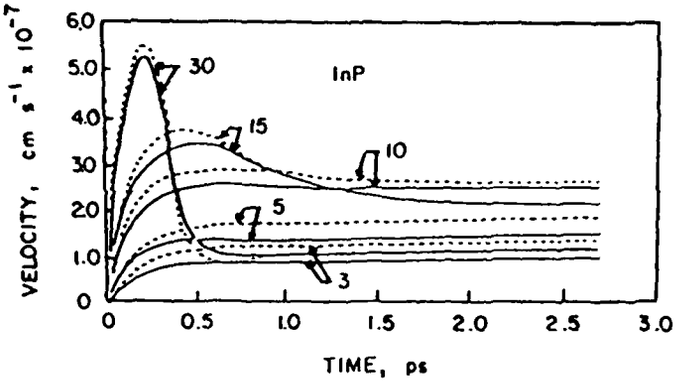


FIGURE 60b. As in (a), but for InP.[32]

The results of the above figure appear to suggest that for a fixed submicron length the transit time versus distance should show advantages for GaAs against InP. This is borne out in figure 61. Indeed, these results should not be taken as noncontroversial. There appears to be some differences between these results and those of Maloney and Frey who conclude that InP will always show some speed advantage. Indeed, based upon the Maloney and Frey results, we concluded recently, the results shown in figure 62. Based on figure 62, a calculation of

$$f_T = \frac{1}{(2\pi \text{ transit time})} \tag{109}$$

for channel length of 0.4μm f_T yields 114GHz for GaAs. It is

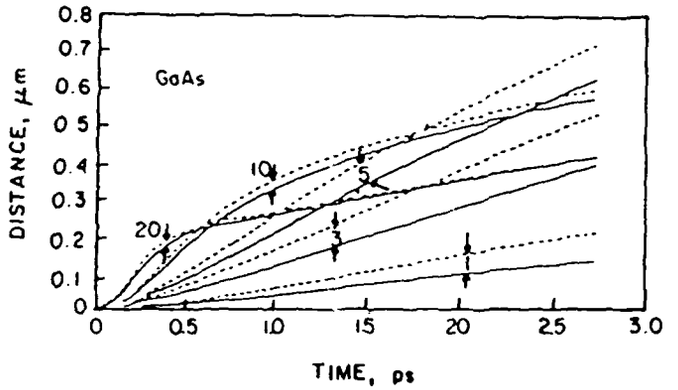


FIGURE 61a. Variation of drift distance with time (GaAs). N_p=10^{17}cm^{-3} ---N_p 0. Fields in kV/cm.[32]

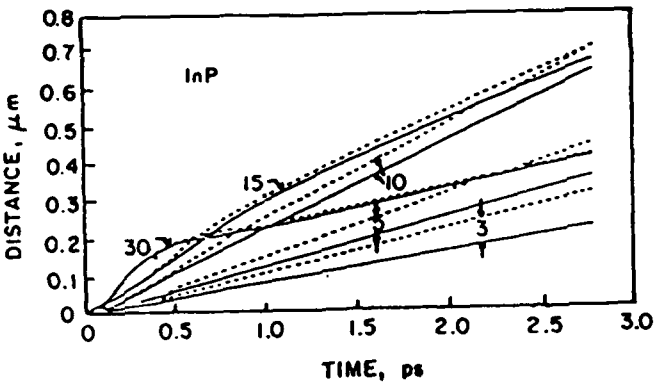


FIGURE 61b. As in (a), but for InP.[32]

slightly higher for InP. At 2500 A, it is approximately 265GHz for GaAs. For InP it is approximately 227GHz. But the numbers derived form figure 61 would indicate that the initial transient must be included in the material assessment. Thus, clear predictions based upon uniform field calculations cannot be given.

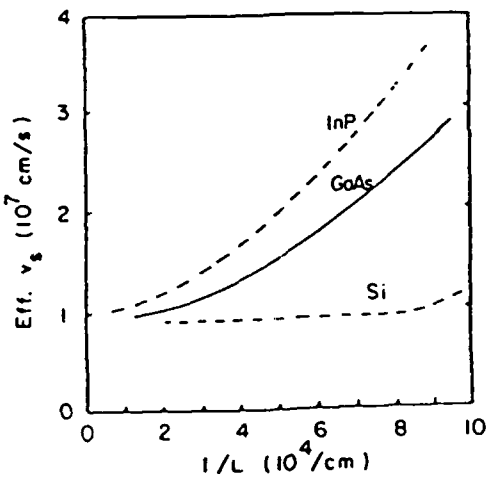


FIGURE 62. The effective, or time-of-flight, saturated velocity as a function of the inverse channel length. The InP curve is estimated from the data of ref. 31.[33]

The next situation of interest concerns itself with extrinsic scaling, where we are concerned with the effects of device length for nonuniform field situations. To deal with this situation we return to equation 52 and concentrate on the coefficients a and the normalized scattering rate. (A similar analysis holds for the energy balance equation). For extrinsic scaling.

$$t_{ref} = \frac{\text{Device Length}}{v_{ref}} \tag{110}$$

Thus

$$a = \frac{eF_{ref}}{m_{ref}\, \Pi_r\, v_{ref}} \tag{111}$$

and

$$\Pi_r = \frac{v_{ref}}{x_{ref}}\, \Pi_{r'} \tag{112}$$

With regard to the coefficient a, if the scattering rates are constant, and F_{ref} is constant, then a is independent of device length. Insofar as b is independent of device length, as the device length begins to decrease, the "non-drift and diffusion" terms

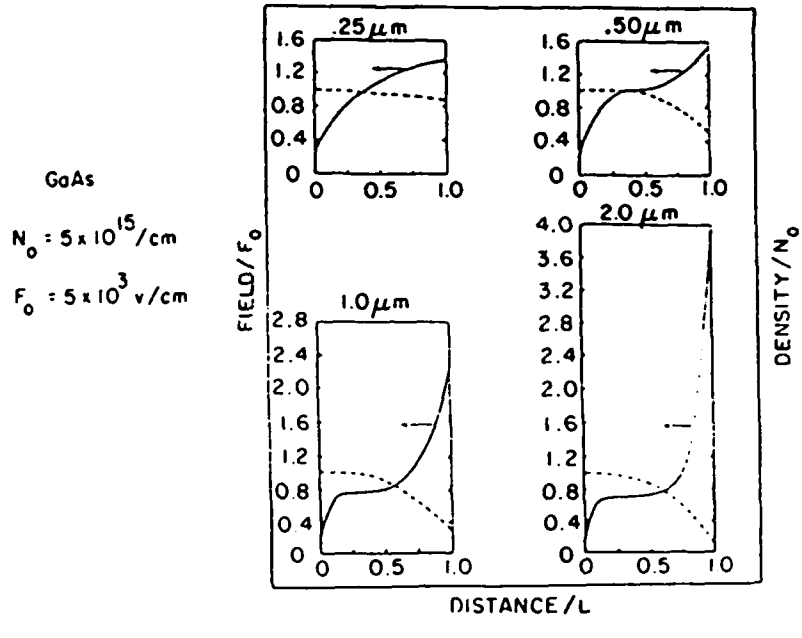


FIGURE 63. Effect of device length on electron transfer, for a constant average field.

begin to dominate. Additionally, the coefficient f in Poisson's equation is length dependent and as x_{ref} increases, the gradients in F increase. These contributions, when coupled to the fact that electrons need a specific device length to traverse before they undergo electron transfer, leads to the following key results. Figure 63 shows the calculated electric field distribution for GaAs subject to an average field of 5kv/cm. It is seen that the highest nonuniform field distribution occurs as the device length increases.

We note that electron transfer tends to be synonymous with low current levels. This is displayed in figure 64. Additionally, two other important features arise. First, there is the increased drive current for the submicron dimensions. Second, there is the absence of negative differential conductivity. A related question of importance at this time concerns the choice of material. If a material with an increased scattering rate were chosen, then simple scaling theory suggests that to achieve similar current levels, shorter device lengths are needed, and higher fields. If this

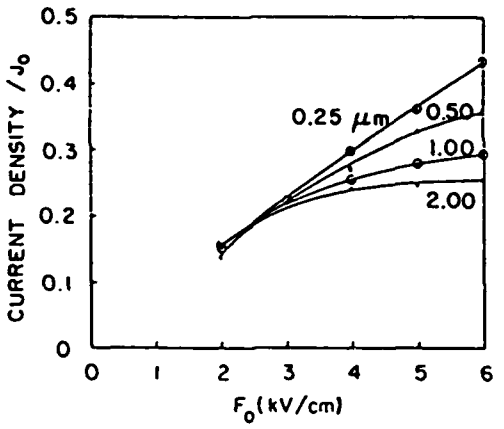


FIGURE 64. Current density versus average field versus device length.

conclusion holds true then the achievement of the high InP velocities may place extreme demands on available technology, particularly, if very small gate lengths are required, as appears to be the case.

The next type of scaling we consider is that of extrinsic carrier density scaling. For this case, the important equation to examine is Poisson's equation, and the relevant scaled quantity for this case is the coefficient f of equation (64). This coefficient tends to indicate that for a given value of average electric field all results are unchanged if the product of carrier density and device length are kept constant. 'NL' products are common in

examining transport in long GaAs diodes. They are, however, irrelevant for submicron length devices in that they ignore the fact the intervalley transfer is length dependent, as the results in figure 65 indicate.

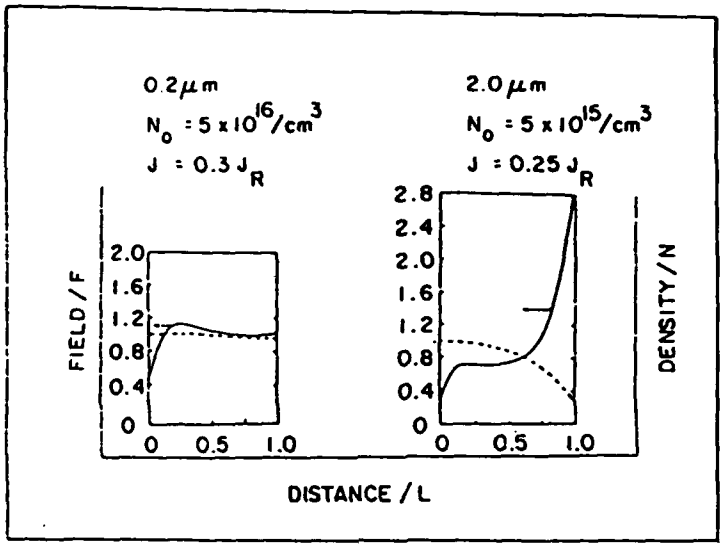


FIGURE 65. Field and Γ-valley carrier density for an average field of 5kv/cm. For this calculation $J_R = N_0 e v_R$, where $v_R = 1.0 \times 10^8$ cm/sec.

One very general conclusion can be drawn from the above scaling arguments, it is that the achievement of high speed submicron length devices does not necessarily require materials with high steady state velocities. High steady state velocities are irrelevant for submicron devices. Rather what is needed, are high mobility materials. For this case, it is clear that of the three materials of interest, GaAs, InP and Si, GaAs has the clear advantage. But there has been considerable interest in a variety of other materials.

MATERIALS CHOICE

The situation with material choice is best illustrated by several examples, but again the material choice is based upon device application. If the device conceived is such that submicron effects do not enter, then the steady state field dependent velocity curves are the ones of interest. as illustrated in figures 66 through 69. Note, particularly high values of field and velocity associated with the InGaAs alloys (figure 66), in addition to very high values of mobility. Also note the reduced mobility for the AlGaAs ternary, as Al is introduced; for InPAs as arsenic is introduced; and GaAsP, as phosphorous is introduced. The method

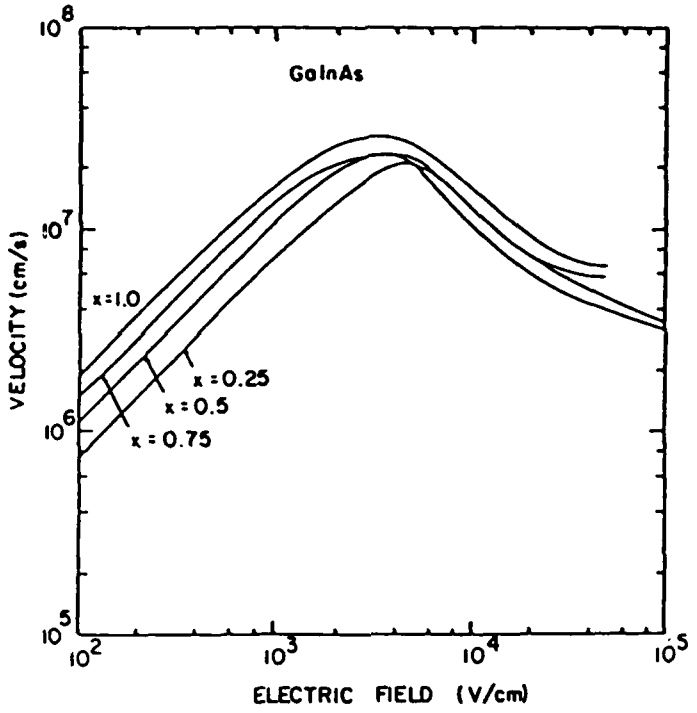


FIGURE 66. Velocity-Field Curves for $Ga_{1-x}In_xAs$ at 300°K with a doping level of 10^{16} cm^{-3} (From Ref. 31).

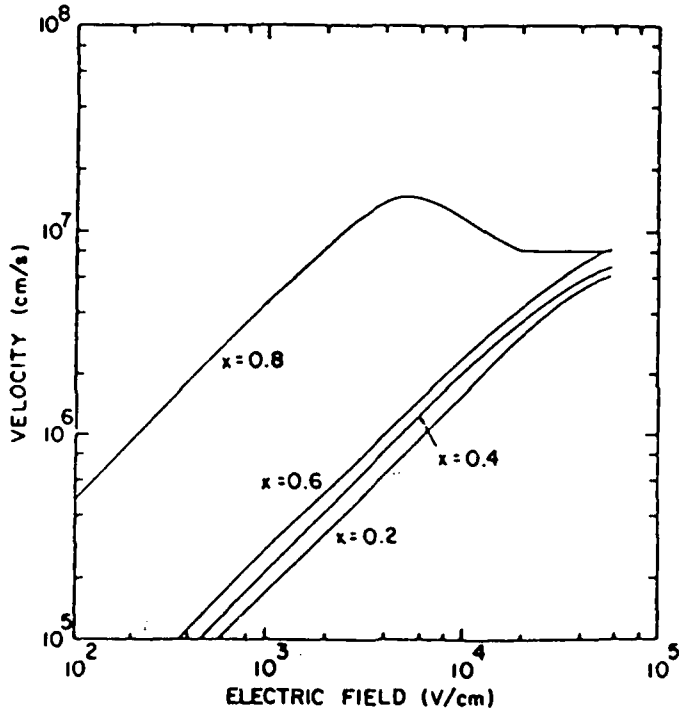


FIGURE 67. Velocity-Field Curves for $Al_{1-x}Ga_xAs$ at 300°K with a doping level of 10^{16} cm^{-3} (Ref. 31).

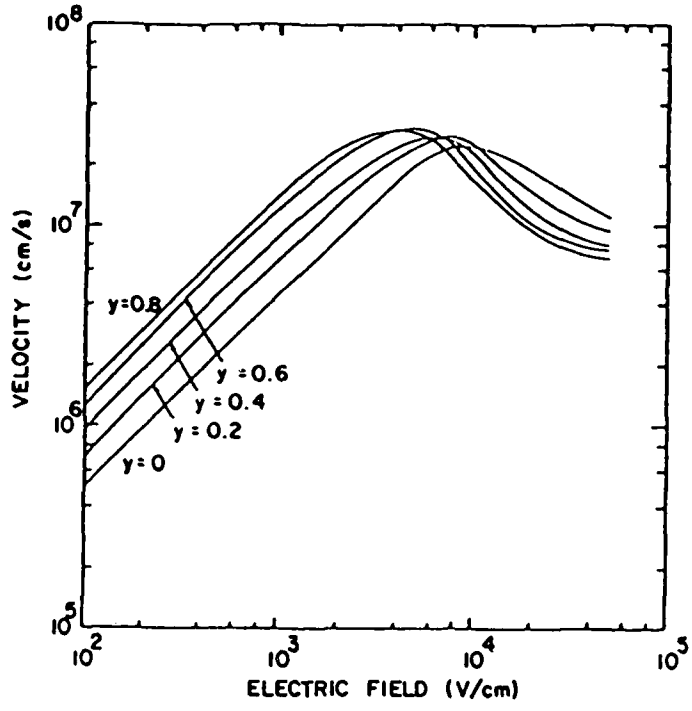


FIGURE 68. Velocity-Field Curves for $InP_{1-y}As_y$ at 300°K with a doping level of 10^{16} cm^{-3}. (Ref. 31)

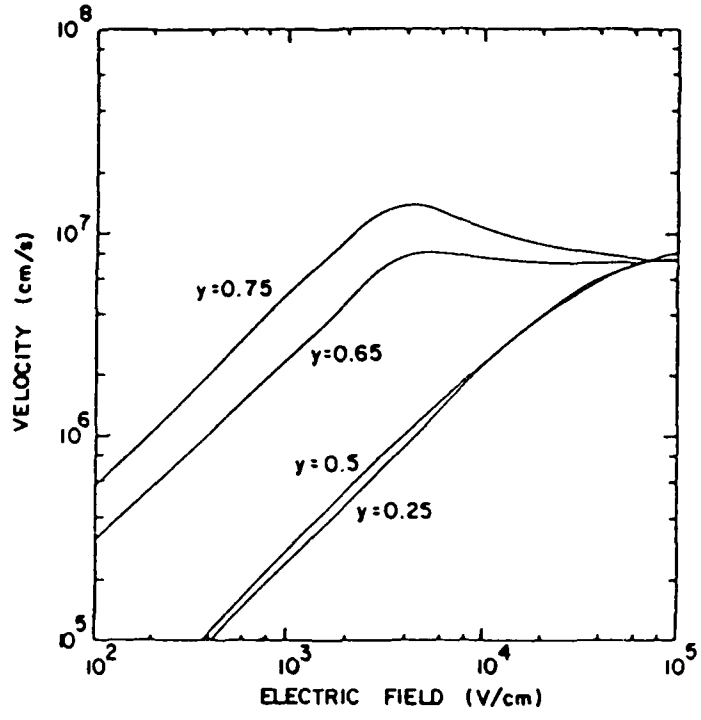


FIGURE 69. Velocity-Field Curves for $GaP_{1-y}As_y$ at 300°K with a doping level of 10^{16} cm^{-3} (Ref. 31).

for calculating these field dependent curves is essentially similar to the technique used for calculating the field dependent curves for GaAs, InP, etc.

Because of the narrow band gap of some of these materials, nonparabolic effects enter and additional care must be exercised in performing the calculations, but basically a set of material constants is needed for each binary component in the ternary element. We illustrate this in figure 70 for the ternary $Ga_xIn_{1-x}A_s$, where x represents the mole fraction of the GaAs element.

Figure 70 displays the band gap variation of the three principle portions of the conduction band for GaInAs. We note that InAs has the same Γ-L-X ordering as GaAs. Such is not the case with AlAs. The band gap variation of these curves is represented by (photoluminesence studier)[39]

$$E_\Gamma = 0.422 + .7x + 0.4x^2 \quad \text{at} \quad 2°K \tag{113a}$$

$$E_\Gamma = 0.324 + 0.7x + 0.4x^2 \quad \text{at} \quad 300°K \tag{113b}$$

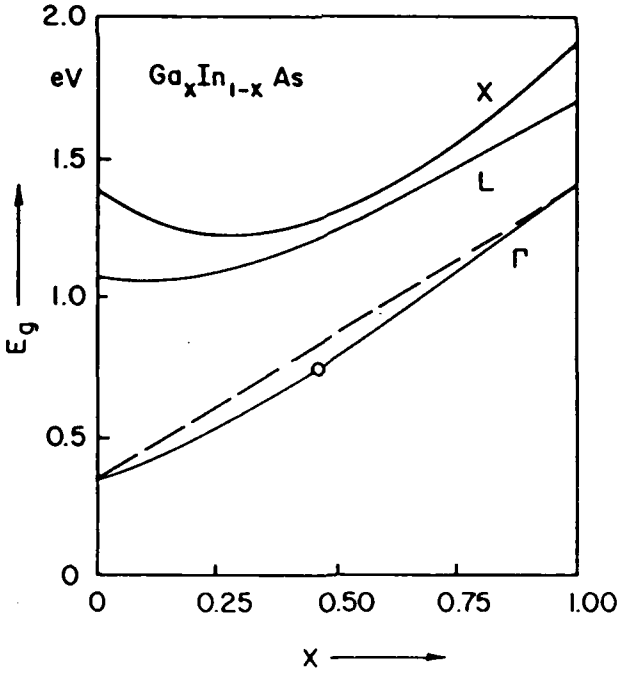


FIGURE 70. Energy gap vs composition for GaInAs.

where at room temperature for GaAs (x=1), E_Γ = 1.43ev and for InAs

(x=0), E_Γ = 0.35ev. There is a clear presence of bowing in the band gap variation with composition, but linear approximations for the L and X band for X>.47 are useful (note: $Ga_{.47}In_{.53}As$ has the same lattice parameter as InP)

$$E_L = 0.38 + 1.73x \tag{114}$$

$$E_X = 0.3929 + 1.93x \tag{115}$$

The above band gap variations is specific to GaInAs; there are, however, fairly general rules[40] often used for calculating other terms. The lattice constant a is computed from

$$a = xa_{GaAs} + (1-x)a_{InAs} \tag{116}$$

The effective mass is:

$$\frac{1}{m^*} = \frac{x}{m^*_{GaAs}} + \frac{(1-x)}{m^*_{InAs}} \tag{117}$$

for each section of the conduction band. For the dielectric constants[40]

$$\frac{\epsilon_{o,\infty}-1}{\epsilon_{o,\infty}+2} = x\cdot\left(\frac{\epsilon_{o,\infty}-1}{\epsilon_{o,\infty}+2}\right)_{GaAs} + (1-x)\left(\frac{\epsilon_{o,\infty}-1}{\epsilon_{o,\infty}+2}\right) \tag{118}$$

For the transverse and longitudinal optical frequencies

$$\omega_T = (x\omega_T{}^2 GaAs + [1-x]\omega_T{}^2 InAs)^{\frac{1}{2}} \tag{119a}$$

$$\omega_L = \omega_T\sqrt{\frac{\epsilon_o}{\epsilon_\infty}} \tag{119b}$$

To calculate the acoustic velocity U[40]

$$\left(\frac{U}{a}\right)^2 \omega_T^2 = x^2\omega_{AGaAs}^2\omega_{TGaAs}^2 + (1-x)^2\omega_{AInAs}^2\omega_{TInAs}^2$$

$$+ \frac{x(1-x)\ M_C(M_A+M_B)}{(M_A+M_C)(M_B+M_C)}\omega_{TGaAs}^2\omega_{TInAs}^2 \qquad (120)$$

where

$$\omega_A = \frac{U_{GaAs}}{a_{GaAs}}, \qquad \omega_A = \frac{U_{InAs}}{aInAs} \qquad (121)$$

and M_A, M_P, M_C are the masses of the constituent atoms. For example, in GaInAs::

$M_A = M_{GA} = 69.72$ gm/mole
$M_B = M_{IN} = 114.82$ gm/mole
$M_C = M_{AS} = 74.92$ gm/mole

Intervalley phonons are obtained from a linear extrapolation of the intervalley phonons of each of the constituent elements which in turn are obtained by application of a set of selection rules[41]. The selections rules were not obtained under hot carrier conditions, and so it is unrealistic to suggest that these be rigorously applied. Nevertheless, consider the intervalley phonon from Γ(00) → X(100). The selection rule is that an LO phonon at X is involved when $M_{III}<M_V$ (e.g., GaAs). An LA phonon is involved when $M_{III}>M_V$ (e.g., InAs). The phonon frequency for the ternary is taken as:

Intervalley Γ-X phonon frequency

$$\hbar\omega_{\Gamma X} = x\hbar\omega_{\Gamma X}^{GaAs} + (1-x)\hbar\omega_{\Gamma X}^{InAs} \qquad (122)$$

The intervalley phonon from Γ(000) to L(111) is an average of LO and LA at L. Again,

$$\hbar\omega_{\Gamma L} = x\hbar\omega_{\Gamma L}^{GaAs} + (1-x)\hbar\omega_{\Gamma L}^{InAs} \qquad (123)$$

For the intervalley phonon from L(111) to X(100), an average of the LA, TO and LO at L is used. Again,

$$\hbar\omega_{LX} = x\hbar\omega_{\substack{GaAs \\ LX}} + (1-x)\hbar\omega_{\substack{InAs \\ LX}} \tag{124}$$

From X(100) to X(100), from L(111) to L(111) for equivalent intervalley scattering, a linear extrapolation is used. Similarly for the nonpolar intervalley phonon (LO+TO) a linear extrapolation is used. The other remaining quantities of interest are the deformation coupling coefficients all of which undergo extrapolation.

Another feature of importance here concerns material choice associated with nonparabolic bands. In all of the calculations we have assumed a parabolic dispersion relation, namely

$$E = \frac{\hbar^2 k^2}{2m^*} \tag{125}$$

A simple extension to nonparabolic is generally given as

$$E + \frac{E^2}{Eg} = \frac{\hbar^2 k^2}{2m^*} \tag{126}$$

where Eg is an effective energy gap. The significance of this is

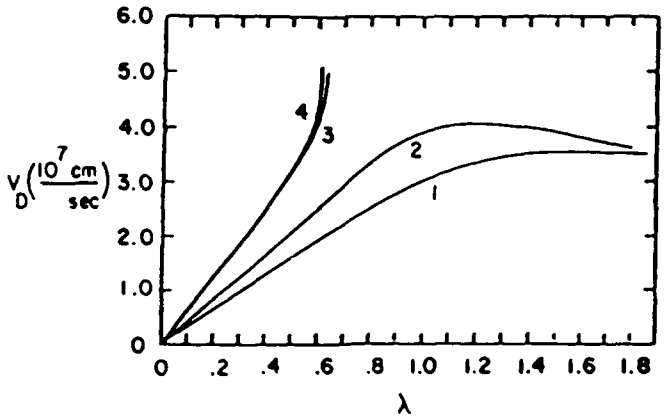


FIGURE 71. Drift velocity as a function of normalized field, λ: (1) InSb at 290°K, nonparabolic calculation; (2) InAs at 300°K, nonparabolic calculation; (3) InSb at 290°K, electron, parabolic band; (4) InAs at 300°K, parabolic band[42].

that the effective mass increases with increasing energy. The consequences of this is that negative differential conductivity in the absence of electron transfer can occur. On the basis of the above expression it is clear that nonparabolic effects, while important for GaAs, will be even more pronounced for InAs. The effects of nonparabolicity on the polar phonon scattering were discussed at a very early stage by Matz[42], and his results are shown in figure 71, where we see a clear contribution to NDR.

A general discussion of NDR due to nonparabolic energy bands was presented by Harris and Ridley[43] who applied their results to PbTe at 77 K. The general applicability of their conclusions are discussed below. Harris, et al[43] used the displaced Maxwellian approximation, and examined scattering due to acoustic and polar optical scattering. The results of their study are summarized in Figure 72. The symbols in their paper have the following significance. W is the ratio of the low field acoustic to polar optical mobilities:

$$W = \frac{\mu_{a^\circ}}{\mu_{p^\circ}} \tag{127}$$

when $W = \infty$ scattering is purely polar optical, when $W = 0$ scattering is purely acoustic. When $W = 1$ there is a mixture of acoustic and polar scattering. The bold lines signify calculations with nonparabolic contributions. The dashed lines are for purely parabolic bands. The nonparabolic contributions indicate an avoidance of runaway.

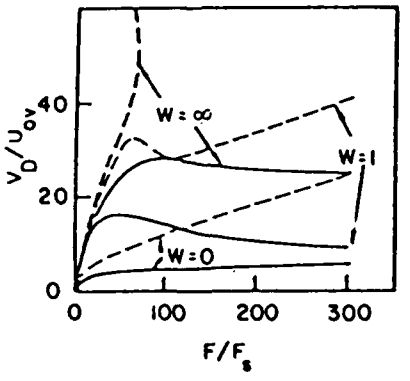


FIGURE 72. Field dependent variation of velocity for different combinations of acoustic and polar optical scattering. Dashed line denotes parabolic model, solid line denotes nonparabolic model[43].

Another important contribution to transport is the alloy scattering. Hauser, et al[41] treated a completely random array of allow scatters whose scattering rate increased with increasing electron energy. The results indicate a decrease in peak velocity and in the magnitude of the negative differential mobility.

The velocity - field curves of a variety of ternary compounds were shown in figures 66 through 69. We briefly reconsider AlGaAs. The interesting feature of this material is that for pure AlAs, the lowest portions of the conduction band is at X (see figure 73) Of particular interest, is the crossover, as reflected in figure 73. Figure 74 displays a normalized conductivity versus aluminum concentration where it is seen that (1) the lowest conductivities occur for highest aluminum concentration, and (2) the precipitous

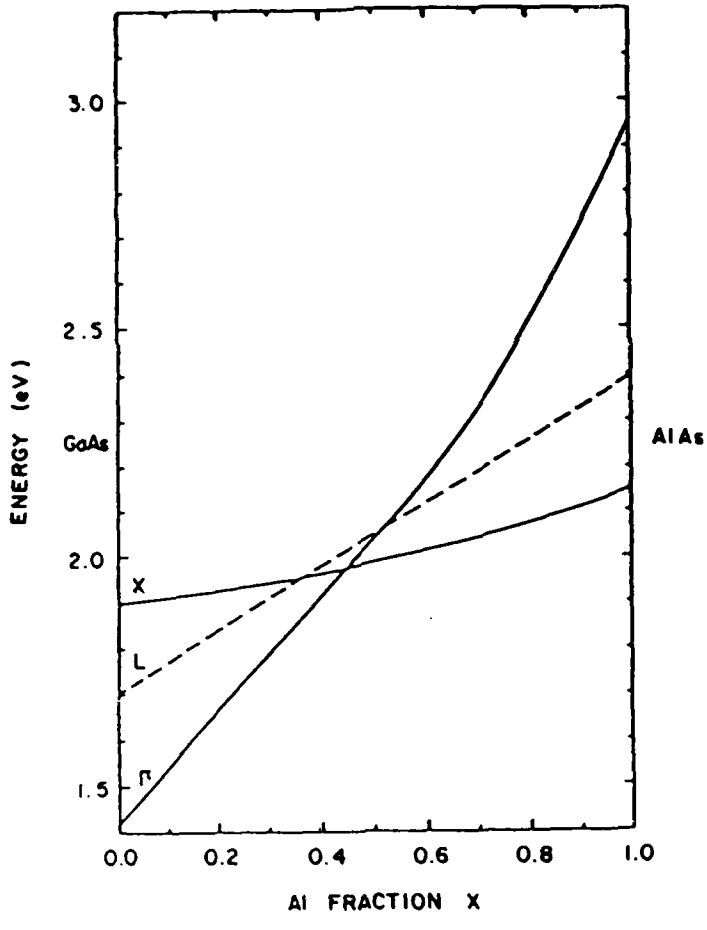


FIGURE 73. Energy band gaps (Γ,X, and L) of $Ga^{1-}{}_xAl_xAs$ (at 300° K) calculated from Eqs. (6)-(8), as a function of Al concentration x. Dashed line represents the L band[45].

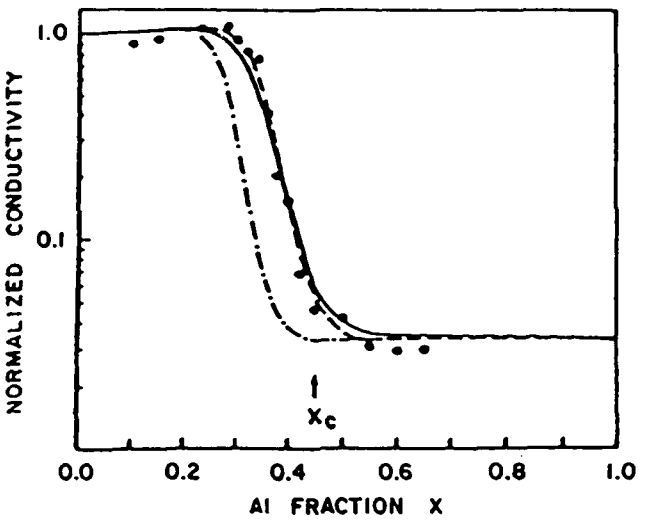


FIGURE 74. The data points represents normalized conductivity of $Ga_{1-x}Al_xAs$ vs Al concentration. A direct-to-indirect bandgap crossover occurs at $x_c = 0.45$. The solid line corresponds to a three-band (Γ, X, and L) calculating, only Γ and X are considered in the calculation resulting in a dashed line. The dot-dash line corresponds to $x_c = 0.37$, as explained in Ref. 45.

drop in conductivity occurs near the crossover. These results are also reflected in the field dependent velocity relation (figure 67).

At this junction we reiterate a point made earlier for parabolic bands, namely that central valley transport appeared to dominate device behavior. In figure 75a the calculated field

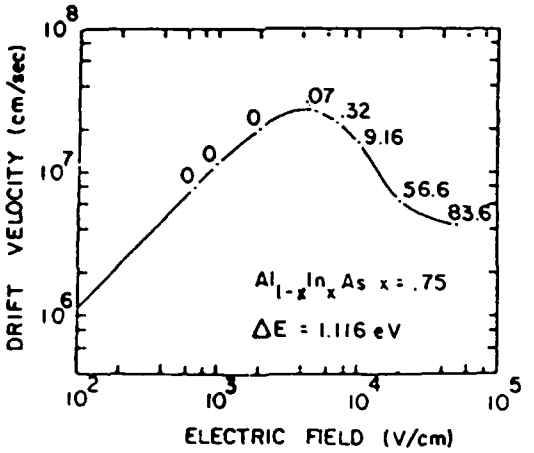


FIGURE 75a. Calculated velocity-field curve for $Al_{0.25}In_{0.75}As$ values along the curve show the percentages of electrons in the upper (111 and 100) valleys[47].

dependent velocity for $Al_{0.25}In_{0.75}As$ is shown along with the percentage of carriers in the subsidiary valleys. In Figure 75b the contribution from the Γ-valley velocity relative to the total velocity is shown. It is seen that transport is dominated by the central valley. Note: the calculated energy gaps for AlInAs are displayed in figure 75c. There is a discrepancy between the L-valley AlAs results of this calculation and that of figure 73. The calculations for gallium indium arsenide show a similar behavior, as displayed in Fig. 76.

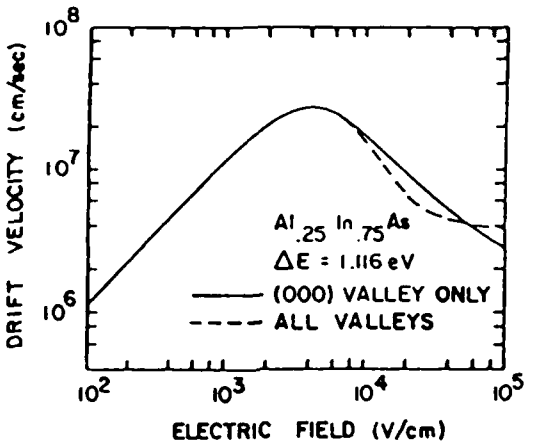


FIGURE 75b. Comparison of velocity-field curves for $Al_0l_{.25}In_{0.75}As$ use g central valley only and using all valleys[47].

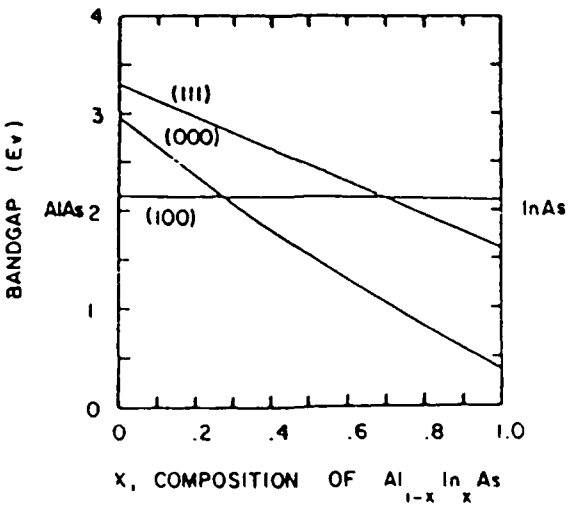


FIGURE 75c. Calculated energy gaps vs composition for $Al_{1-X}In_XAs$[47].

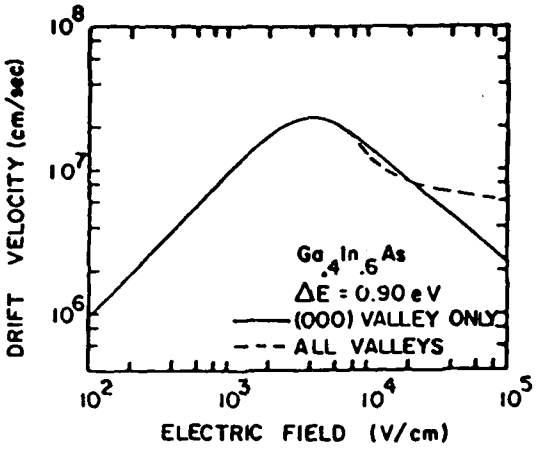


FIGURE 76. As in figure 75b, but for $Ga_{0.4}In_{0.6}As$[47].

To tie the varieties of field dependent transport coefficients together, several figures of merit have been proposed. One, put forth by Hauser, classifies materials in terms of the peak drift velocity obtained from polar phonon scattering including nonparabolic effects. This was estimated by Hauser, et al[42] as

$$V_P \leq V_{MAX} = \left[\frac{\hbar\omega_o}{m^2} + \tan\left(\frac{\hbar\omega_o}{2kT}\right)\right]^{\frac{1}{2}} \qquad (128)$$

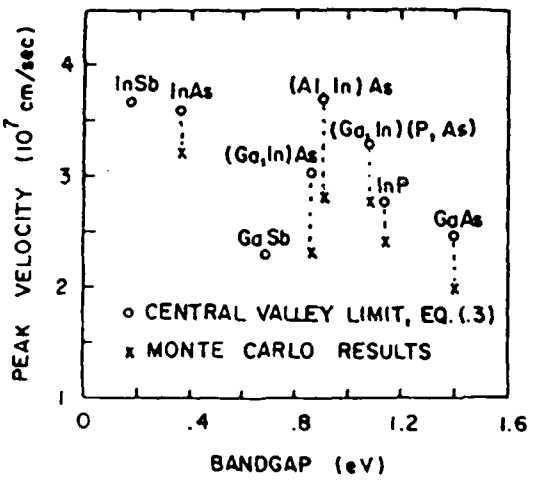


FIGURE 77. Calculated upper limits to peak velocity from equation (128)[42].

The results of the above expression for a variety of different materials are summarized in Figure 77, where we note that the highest peak drift velocity is for the material InSb.

In another study (Ferry[43]) summarized in Figure 78 the high field saturated drift velocity was plotted as a function of $8\hbar\omega/3\pi m^*$. The curve in figure 78 follows the relation:

$$V = (8\pi\omega_o/3\pi m^*)^{\frac{1}{2}} \tag{129}$$

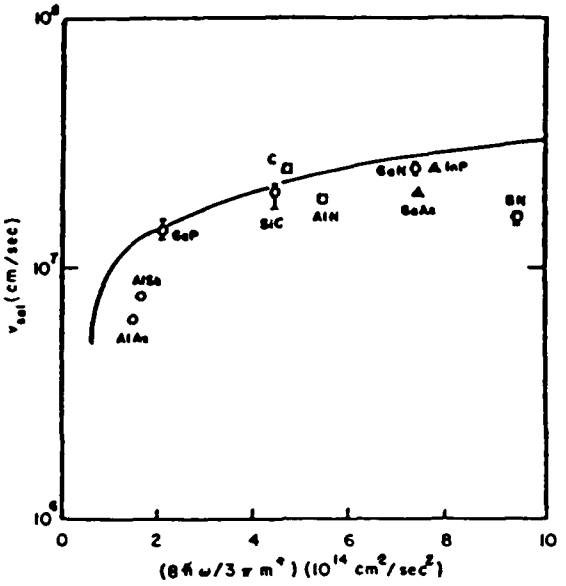


FIGURE 78. Saturated velocities calculated here are shown as functions of the parameter relating to energy relaxation.

BOUNDARY CONDITIONS TO DEVICES

We close this lecture series with a more detailed discussion of field nonuniformities.

While most of the previous discussion has tended to focus on the importance of field nonuniformities for realistically assessing device performance only a small fraction of the discussion was devoted to the single most important contribution to these field nonuniformities - boundary conditions. Here we provide a brief review of its influence on the behavior of two-terminal NDC devices. Figure 79 displays typical boundary-dependent data from three different gallium arsenide two-terminal devices. The lower portion of each diagram displays current versus voltage

characteristics, while the upper portion shows voltage versus distance at one bias point. Figure 79a shows measurements for a device in which the metal contacts are far removed from the active region of the device. The current-voltage relation is relatively linear until a point where current oscillations occur. The field profile just prior to the oscillation is relatively uniform within the active region of the device, and is near zero at the ends of the active region. Figure 79b represents a set of measurements in which the metal contact abuts the active region of the device. The current-voltage characteristics remain linear to threshold which again is manifested by a current oscillation. Notably different here is the lower average field prior to the instability and the enhanced voltage drop at the cathode. Figure 79c displays results for another device with a metal contact abutting the active region. For this case there is a sublinear current voltage characteristic and no instability. The probed voltage versus distance shows a large voltage drop at the vicinity of the cathode.

The electrical characteristics associated with Figs. 79a

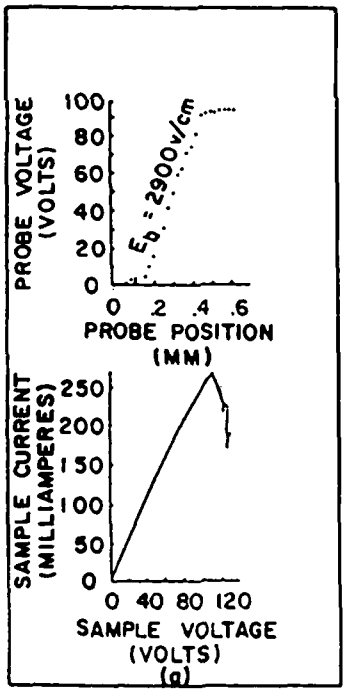


FIGURE 79a. Probed voltage versus distance, and current versus voltage for a two-terminal GaAs device with low boundary fields. From Ref. 48.

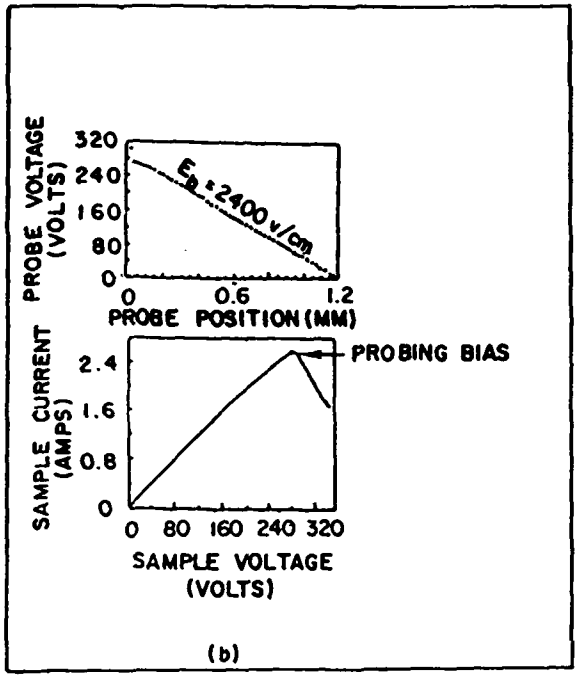


FIGURE 79b. As in Fig. 79a but for a metal contact abutting the cathode.

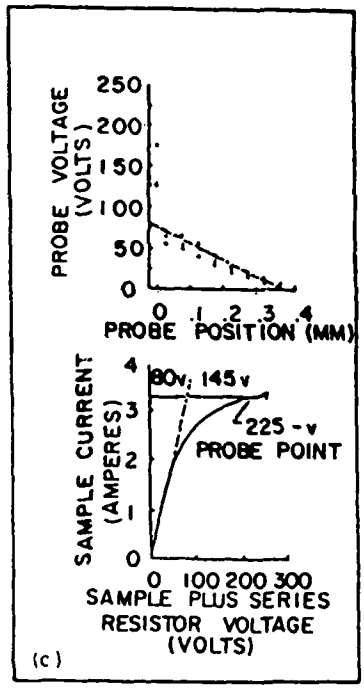


FIGURE 79c. As in Fig. 79a but for a metal contact abutting the cathode.

through c have been described as representative of "ohmic contacts", "slightly blocking contacts" and "strongly blocking" contacts, respectively. One of the earliest models employed for explaining these results assumed a "pinned" value of cathode electric field. Other models in which the cathode conductivity or doping profile was varied have also been suggested with varying degrees of success. The "pinned" cathode field model, discussed below, developed partially as a consequence of the way the equation's current instabilities was written. Here the continuity equation is rewritten in one dimension,

$$J(t) = q\left(nv - \frac{D\partial n}{\partial x}\right) + \frac{\epsilon \partial F}{\partial t} \tag{130}$$

where J(t) is the total current, and rewritten, again using Poisson's equation, as

$$J(t) = qN_o v + \epsilon\left(nv - \frac{v\partial F}{\partial x} + \frac{D\partial^2 F}{\partial x^2} + \frac{\partial F}{\partial t}\right) \tag{131}$$

This is a second order nonlinear partial differential equation on E requiring two boundary conditions and one initial condition.

Solutions to equation (131) have been used successfully to simulate device results similar to those of figure 79 for gallium arsenide. Qualitatively similar results occur for solutions with cathode fields F(=0,t) falling within any of the following three groups:

$$< 0 \leq F(x=0,t) \leq F_{TH}, \quad F_{TH} \leq F(x=0,t) < 4F_{TH}, \quad F\ (x=0,t) > 4F_{TH}.$$

Where F_{TH} is the threshold field for negative differential mobility. The simulations with pinned fields falling in either group 1, 2 or 3 yield electrical characteristics similar to those of figures 79a, b and c, respectively. The crucial feature of this model is that the cathode field is pinned, necessitating that any instabilities in current occur at a critical value of current density. The field profiles associated with cathode fields in the range

$$< 0 \leq F(x=0,t) \leq F_{TH}, \text{ and } F_{TH} \leq 4F(x=0,t) < 4F_{TH}$$

are sketched in figure 80. For reference, a velocity field curve with velocity scaled to current as $qN_o v(F)=J_n(F)$, and with a region of negative differential mobility is also included. Figure 80 is understood as follows: The second column of each section shows the electric field versus distance profiles. F(x) begins with a value F_c at the cathode and extends downstream to a value F_b. By current continuity, the current everywhere within the

device is given by $J=q\,N_o v(F_b)$. For $J < J_n(F_c)$, first row, a region of charge depletion forms near the cathode for F_c below and within the region of negative differential mobility (NDM). Increasing the current until $J \simeq J_n(F_c)$, second row, introduces charge neutrality everywhere for $F_c < F_{TH}$. However, because F is a double valued function of V, for F_c within the NDM region; approximate charge neutrality exists near the cathode for $J \simeq J_n(F_c)$, and for regions sufficiently far downstream from the cathode. Charge neutrality breaks down between these two regions. Finally, for $J > J_n(F_c)$, third row, an accumulation layer forms near the cathode. For figure 80a, the accumulation layer is stable until the bulk field exceeds F_{TH}. For Fig. 80b,

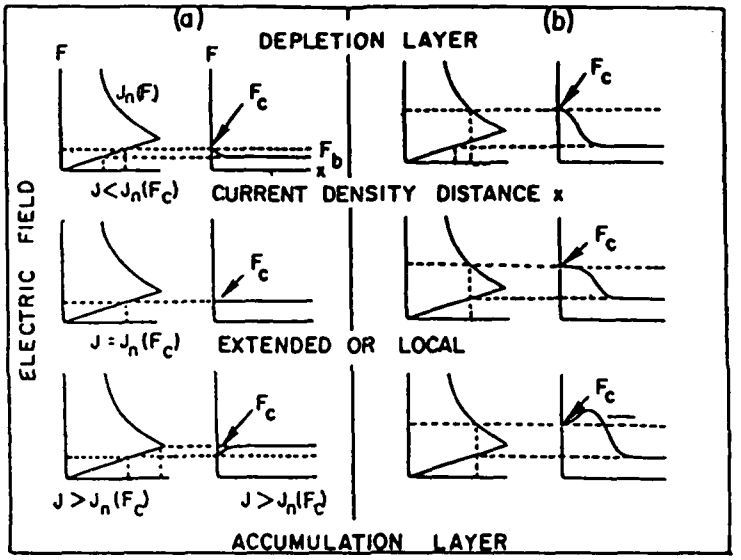


FIGURE 80. Boundary and bias dependent field profiles for materials with a region of negative differential mobility[48].

the accumulation layer, followed downstream by the depletion layer is often unstable and leads to cathode originated instabilities.

The situation corresponding to Fig. 79c is often represented by very high cathode fields. The field profiles are those appropriate to a wide region of charge depletion near the cathode. The field profiles are stable.

The pinning of the cathode field is not necessarily common, however, to all semiconductor devices. For example, while it was also applied to InP devices, where it worked for a significant number of cases, a broad class of InP device behavior could not be accounted for through it's use[13,41]. The latter showed anonymously high efficiency and low dc current levels. Spontaneous Gunn type

oscillations did not occur. Rather, device operation required a tuned circuit. InP device operation was also thought to depend critically on the cathode boundary condition, and the experiments were explained through a fixed cathode conduction condition. The distinction between "pinned" cathode field and "pinned" cathode conduction current is placed in perspective in figure 81 and in the following equation

$$J(t) = J_c(F_c) + \frac{\epsilon dF_c}{dt} \tag{132}$$

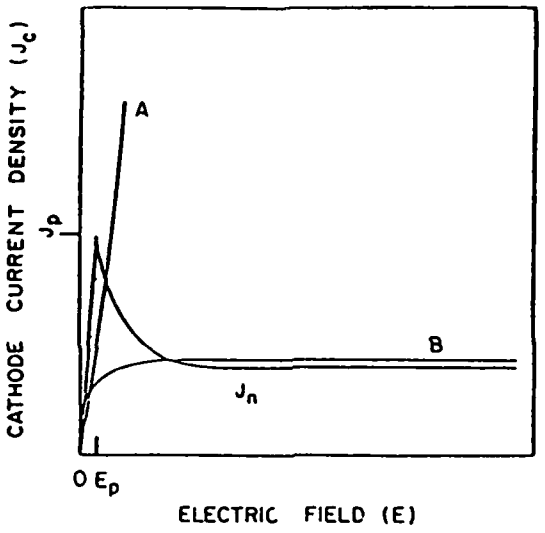


FIGURE 81. Cathode current density, from Equation (133).

Equation 132 is the equation for total current through the boundary to the device. $J_c(F_c)$ represents the current-field relation at the cathode, also referred to as a control characteristic which may be expected to differ from that of the semiconductor device. Two such types of curves are represented by curve A and curve B of figure 88. Curve A is closely related to the pinned cathode field model while curve B is associated with the pinned cathode current model. The similarity in "form" of curves A and B to, respectively, moderate barrier height tunneling and thermionic emission dominated contacts is deliberate, and the equation used to arrive at these curves is shown below.

$$J_c(F_c) = -J_R \left(\exp[-qF_cL_c/nkT] - \exp[-(1/n - 1)(qF_cL_c/kT]\right) \tag{133}$$

which was adapted from studies on the unalloyed metal/semiconductor contact[50]. It's use here presumes a similar description. For the unalloyed contact "n" is the ideality factor and describes the contact as dominated by thermionic emission ($n \geq 1$) or by tunneling ($n \gg 1$). J_R is the reverse current flux and may be related to the barrier height phenomenologically through the Richardson equation.

Detecting a particular contact effect on a device is a difficult procedure. For long devices current voltage characteristics as represented by figure 79 are often signatures of a contact classification. For short devices proximity effects introduce an additional complication and current-voltage measurements are less valuable. One type of measurement which may serve to provide information about the boundary is a noise measurement.

Here the situation to envision is that if the field is pinned within the negative differential mobility region, increasing the bias will result in an amplification and the noise will increase. If an increasing the bias results in carrier injection into the device, the field at the cathode is likely to decrease and the noise is expected to decrease. While these results should be folded in with the field dependence of velocity and diffusion, a simple analytical noise calculation assuming a three piece approximation to represent GaAs has been performed.

In this calculation, the "impedence field method"[51] is applied to calculating noise due to thermal velocity fluctuations amplified within the device. The mean squared noise voltage per unit band width is computed,

$$\left\langle \frac{\delta V_N^2}{\Delta f} \right\rangle = 4q^2 \int_{vol} |\nabla Z|^2 \, ND \, d(vol), \tag{134}$$

where ∇Z is the impedence field vector. The calculation was performed for a ten-micron long element with a doping of $10^{15}/cm^3$. The element sustains the field profile with a cathode depletion where it is seen that the NDM region increases with increasing bias. The calculations, which are discussed in detail elsewhere, [Ref. 29] are expressed in terms of the noise

$$NF = 1 + \left\langle \frac{\delta V_N^2}{\Delta f} \right\rangle \cdot \frac{1}{4k_o T |R|} \tag{135}$$

figure 51 where R is the real part of the device impedence. The results of the calculation are displayed in figure 82 where the noise figure is sketched as a function of bias current and transit angle. The

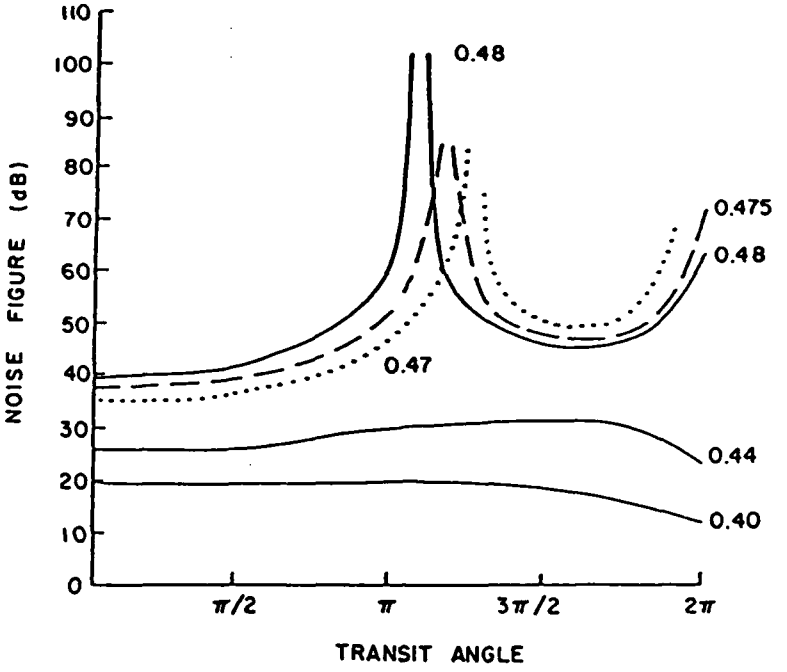


FIGURE 82. Noise figure versus transit angle and current.

results appear as a signature of the effects of the cathode boundary. First, at low values of transit angle $\theta = \omega T(\Delta)$, where $T(\Delta)$ is the transit time across the negative differential mobility region, the noise figure increases with increasing bias. This corresponds to an increase in the length of the negative differential mobility region and enhances amplification of any fluctuation originating there. More interesting structure is present at higher frequencies and higher bias where the noise figure increases and then shows a singularity. On the other side of the singularity there is a "U"-shaped region ending again at a singularity. The strong increase in noise figure represents the approach of $|R| \rightarrow 0$. Here at low frequencies the real part of the impedence is positive, and becomes negative at frequencies somewhere between $\pi \leq \theta < \omega T$. In going from positive to negative values it passes through zero, hence the singularity. The frequency range for small signal negative resistance increases with increasing bias, reflecting the broadening of the negative differential mobility regions - a broad "U"-shaped region appears. Both the increasing noise figure at low frequencies, and the "U"-shaped region at high frequencies are characteristics of an increasing depletion layer width. Note that increasing the bias still further will result in an electrical instability.

The discussion of the above sections dealt with devices whose lengths were typically 10μm-long or longer and the analysis was through the drift and diffusion equation. For near and submicron length devices the Boltzmann transport equation is required. A set of one dimensional steady state calculations are displayed below, using the moment equations discussed earlier. The boundary conditions for this problem are stated below[14]

$$\frac{\partial}{\partial X} \log N_1 = -A \tag{136}$$

$$V_1 = -\mu_c F \tag{137}$$

$$T_1 = B \tag{138}$$

$$\frac{\partial N_2}{\partial X} = \frac{\partial V_2}{\partial X} = \frac{\partial T_2}{\partial X} = 0 \tag{139}$$

The boundary condition represented by Eq. (136) dictates the slope of the Γ valley carrier density downstream from the cathode. Thus for "A" positive (negative) local charge accumulation (depletion) occurs at the cathode boundary. The carrier velocity within the interior of the device is governed by the assigned scattering rates. The velocity of the entering carriers will, in general, differ from that within the bulk. The assigned entrance velocity is governed by Eq. (137). The temperature of the entering carriers is assigned a constant value, represented by Eq. (138). "B" is generally greater than, or equal to 300 K and is a measure of the mean thermal energy of the central valley carriers. As in the discussion of Fig. 80, the results are placed into three categories, "ohmic", "slightly" and "strongly" blocking contacts. The "ohmic" results are shown in Fig. 83, for a 2.0μm-long structure with a doping of $5 \times 10^{15}/cm^3$.

For "ohmic" boundaries an appropriate set of constants are: A=0.2, μ_c=12,000cm^2/v_{sec}, and T_1 = 300°K. For this case carriers enter the device with speeds greater than that associated with the central valley mobility. Additionally there is an accumulation of carriers at the boundaries, resulting in low values of cathode field. The field starts off at nearly 1kv/cm and approximately 2500 A must be traversed before significant transfer occurs. Increased transfer results in a lowering of the mean carrier velocity, necessitating an increase in mobile charge as the

mode is approached. The average velocity across this device obtained from the relation

$$\langle v \rangle = \frac{J}{qN_o} \tag{140}$$

is $\langle v \rangle = 1.78 \times 10^7$ cm/sec, with $N_o = 5 \times 10^{15}/cm^3$.

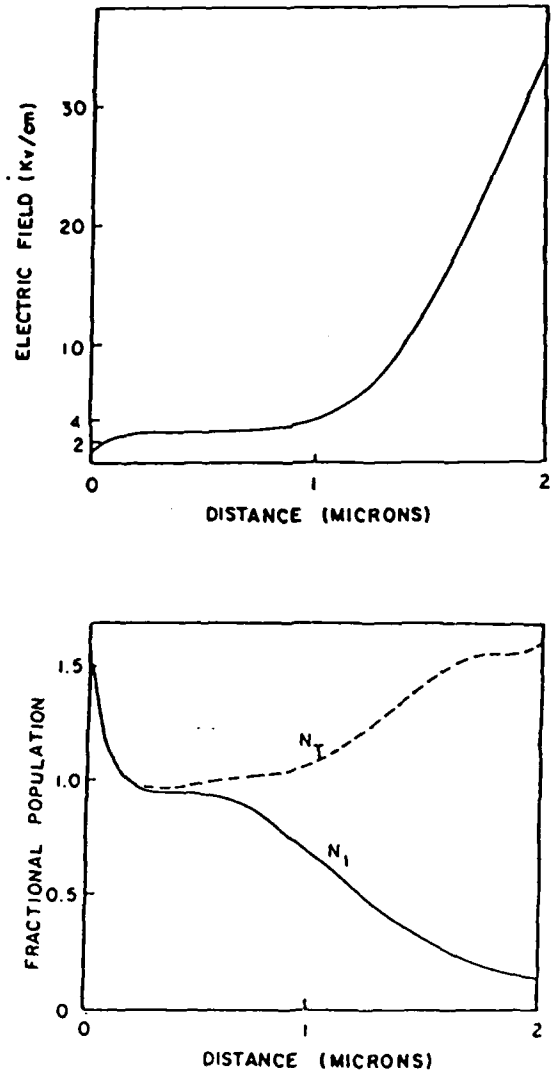


FIGURE 83. Distribution of field and carriers for ohmic boundaries, with an applied bias of 2 volts. From [14]

Lowering the cathode mobility to a value below that associated with the Γ-valley results in a more rapid dispersal of carriers, and cathode adjacent charge depletion, associated with slightly blocking contacts, occurs. This is seen in figure 84 for A=-0.11, μ_c=6000 cm^2/V_{sec}, and T=300°K. It is noticed that the cathode field for this case is approximately 4kv/cm, which is higher than that associated with the "ohmic" contact condition of Fig. 83. There are, however, important similarities between figures 83 and 84. In both cases the carriers adjacent to the cathode are, for all practical cases, Γ-valley electrons. Very little transfer, which is determined by carrier energy

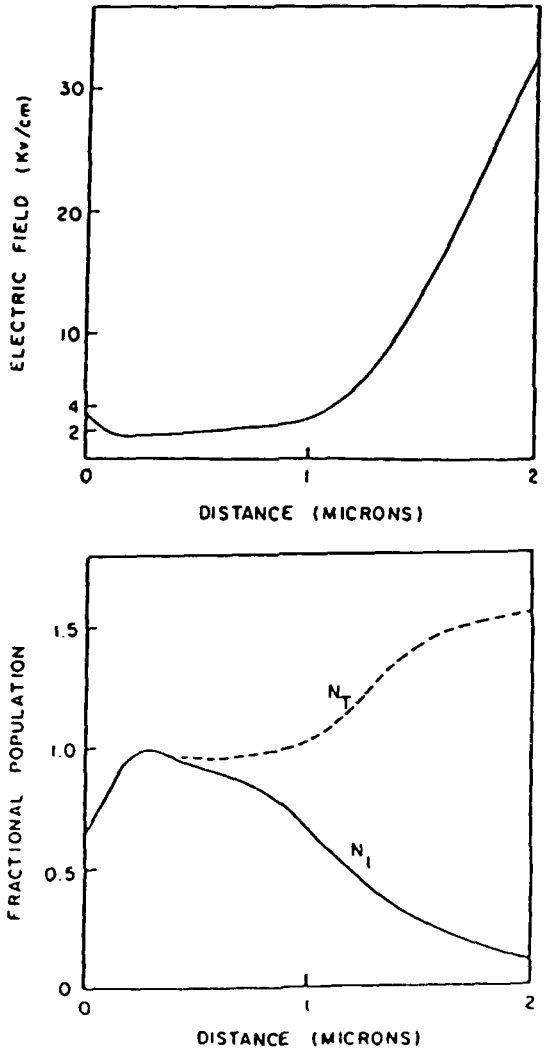


FIGURE 84. As in Fig. 83, but for a partially depleted cathode boundary. [From 14]

(temperature) has occurred. In addition, sufficiently downstream from the cathode the carriers appear to be ignoring the cathode condition and are dominated by the downstream voltage drop, whose spatial distribution is about the same for both. The average velocity for this case is $\langle V \rangle = 1.75 \times 10^7$ cm/sec, slightly below that of figure 83.

A significant change occurs when the mean energy of the Γ-valley entering carriers is elevated. For the parameters of figure 84, but with T_1=1200 k, a substantial amount of transfer occurs at the cathode, resulting in a lowering of the current through the device. The cathode is approximately 7kv/cm, higher than that associated with figure 84, and the downstream field is lower (see figure 85). The average velocity of the carriers in this case is lowered to 1.28×10^7cm/sec, even though the central valley carriers are traveling at higher speeds.

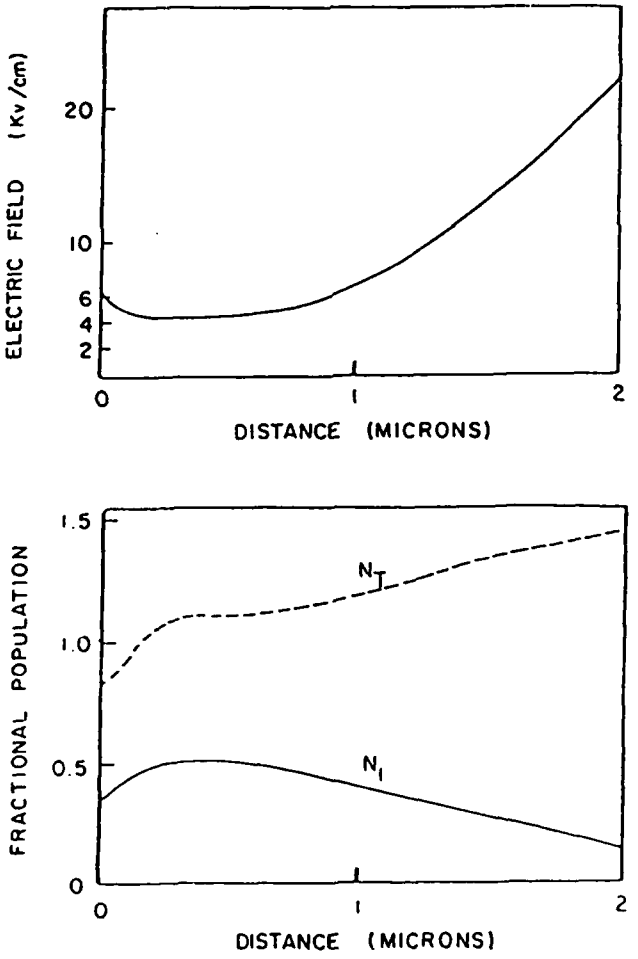


FIGURE 85. As in Fig. 84 but with an elevated cathode electron temperature. From [14]

The presence of moderately high cathode fields is attractive if a sufficient number of carriers can be retained in the central valley where they can sustain high transit velocities. While this case is discussed in more detail below, the simple ruse of injecting excess carriers into a device with the contact conditions of figure 85 does not always yield the sought after current levels. This is illustrated in figure 86, where now A=0.2. Here, the excess charge serves to lower the cathode field, which does not change the downstream characteristics in any significant way. The average velocity for this case is virtually unchanged when compared to figure 85. In this case v_{AV}=1.30x10^7/sec.

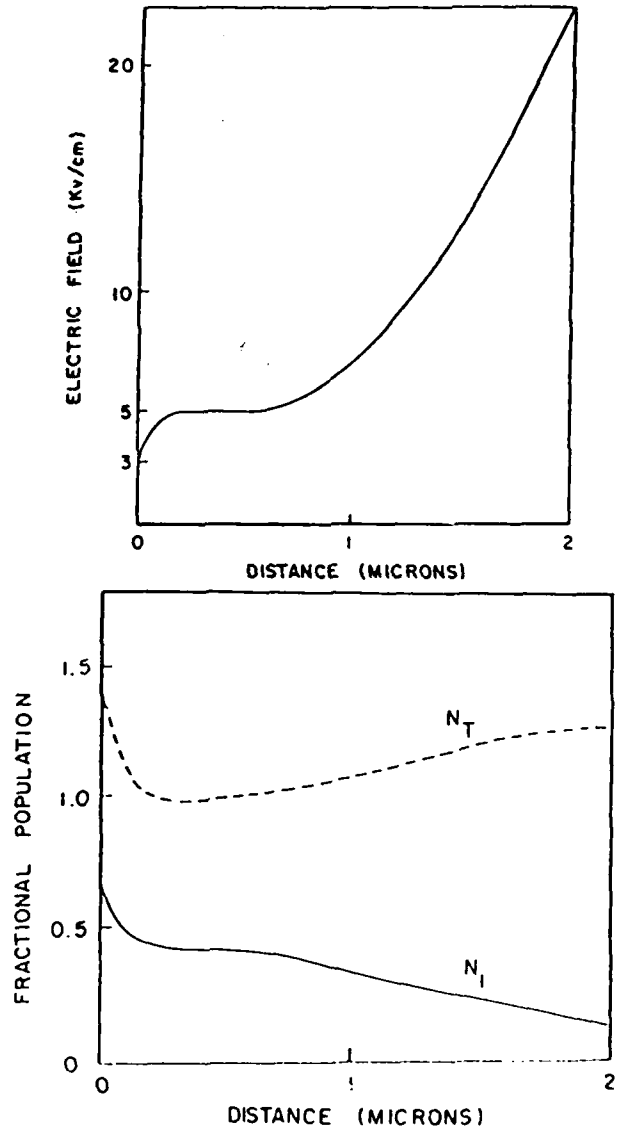
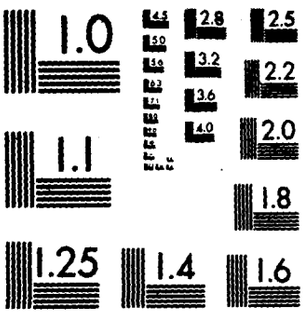


FIGURE 86. As in Fig. 85, but for injecting boundaries [from [illegible]

AD-A193 380 STUDYING THE PHYSICS AND OPERATION OF MULTI-TERMINAL NEAR-MICRON AND SUB-.. (U) SCIENTIFIC RESEARCH ASSOCIATES INC GLASTONBURY CT H L GRUBIN ET AL. 6/6

UNCLASSIFIED 15 FEB 88 SRA-R88-920010-F N00014-81-C-0452 F/G 9/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The situation of a strongly blocking contact is illustrated in figure 87 for A=9.0, μ_c = 1000 cm/V_{sec}, and T=3000 K. For this case virtually all carriers are swept away from the cathode, with the satellite valley carriers accounting for most of the transport in the transition region. The cathode field is extremely high, approaching 60kv/cm. The average velocity is approximately zero.

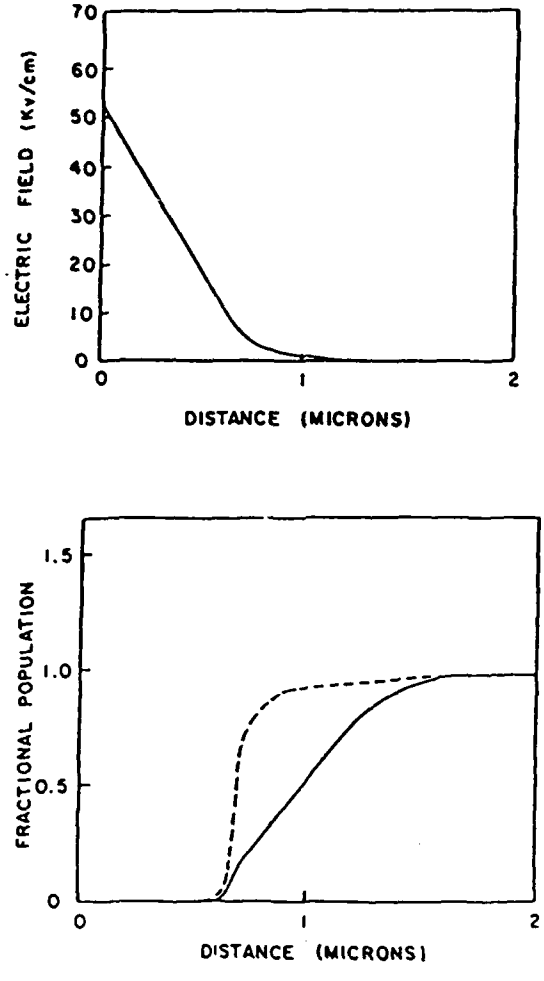


FIGURE 87. As in Fig. 86, but for a strongly blocking contact. [From 14]

In addition to determining the distribution of carriers, etc, boundaries through their influence on the field distribution, also affect current transients. This was shown for an accumulation boundary in Fig. 20. It is shown for a depleted boundary below, where a calculation was performed for a five-micron element with a

high cathode temperature, T_1=1200. The transient results which are displayed in figure 88, and are expressed in terms of a mean velocity. The detailed field profiles are shown in figures 88 and 89. We note, approximate saturation in current, as displayed in Fig. 88. A significant voltage drop in the vicinity of the cathode, as displayed in Figs. 89a and 90a. Note that as the

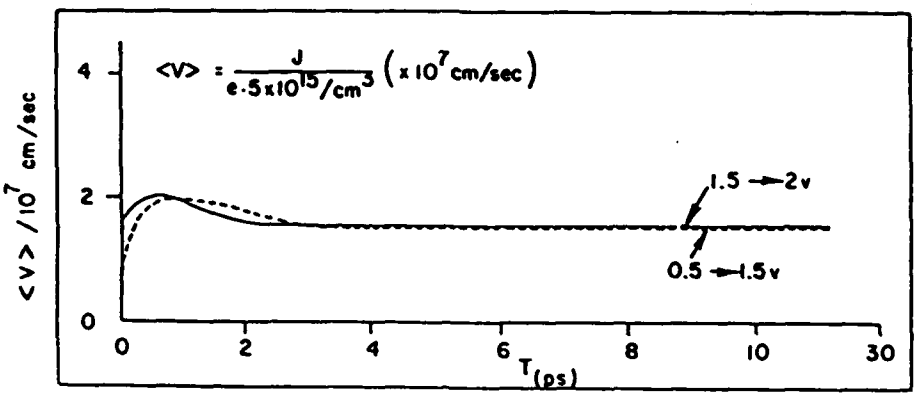


FIGURE 88. Transient current versus time for the indicated voltage change.

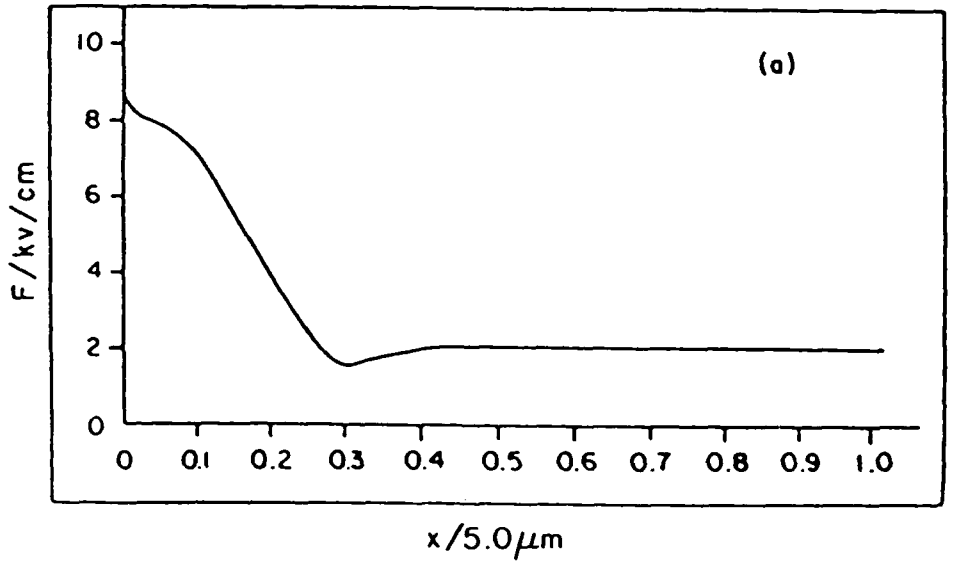


FIGURE 89a. For a bias of 1.5v, steady state distribution of: field.

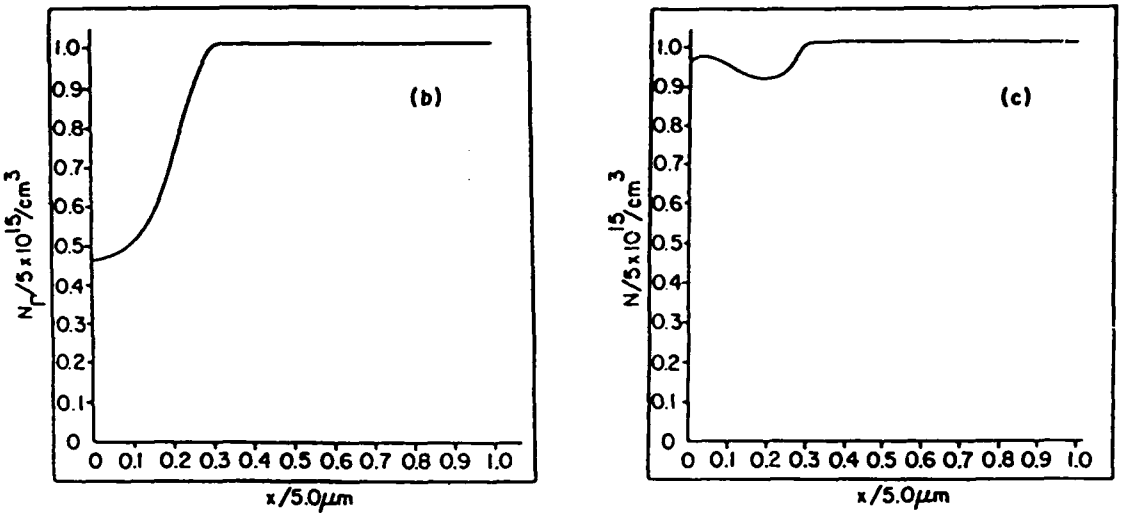


FIGURE 89b, c. For a bias of 1.5v, steady state distribution of: (b) Γ valley carriers; (c) total carriers.

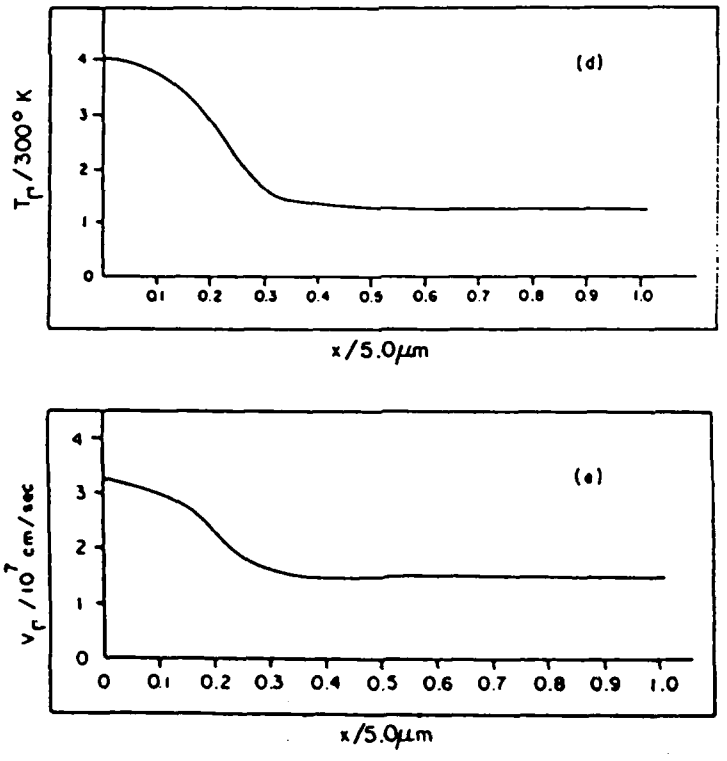


FIGURE 89d, e. For a bias of 1.5v, steady state distribution of: (d) Γ valley temperature, (e) Γ valley velocity.

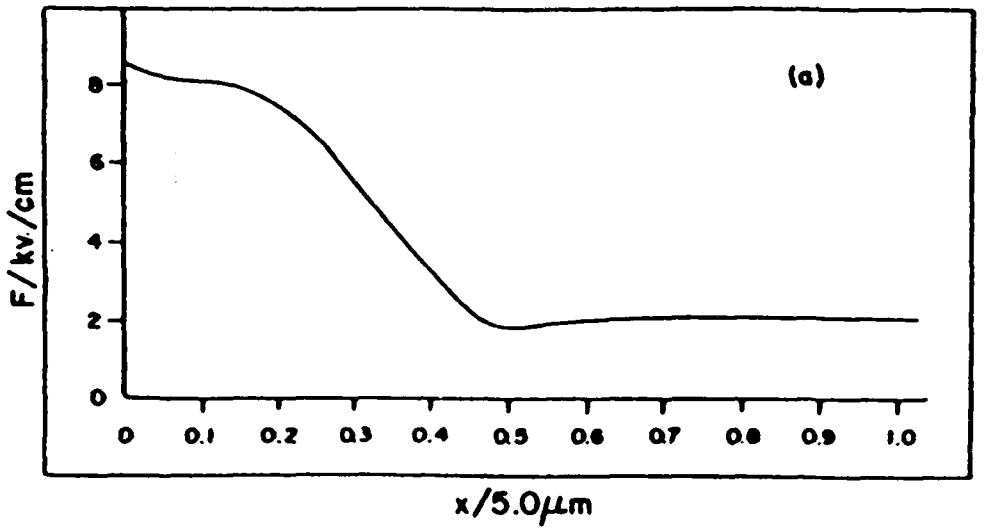


FIGURE 90a. As in Fig. 89, but for a bias of 2 volts.

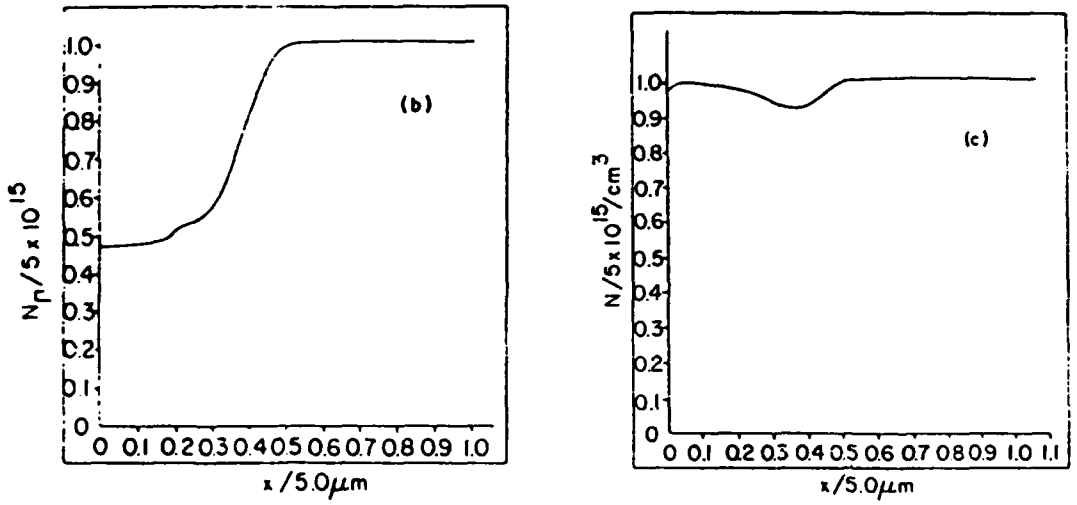


FIGURE 90b, c. As in Fig. 89 but for a bias of 2 volts.

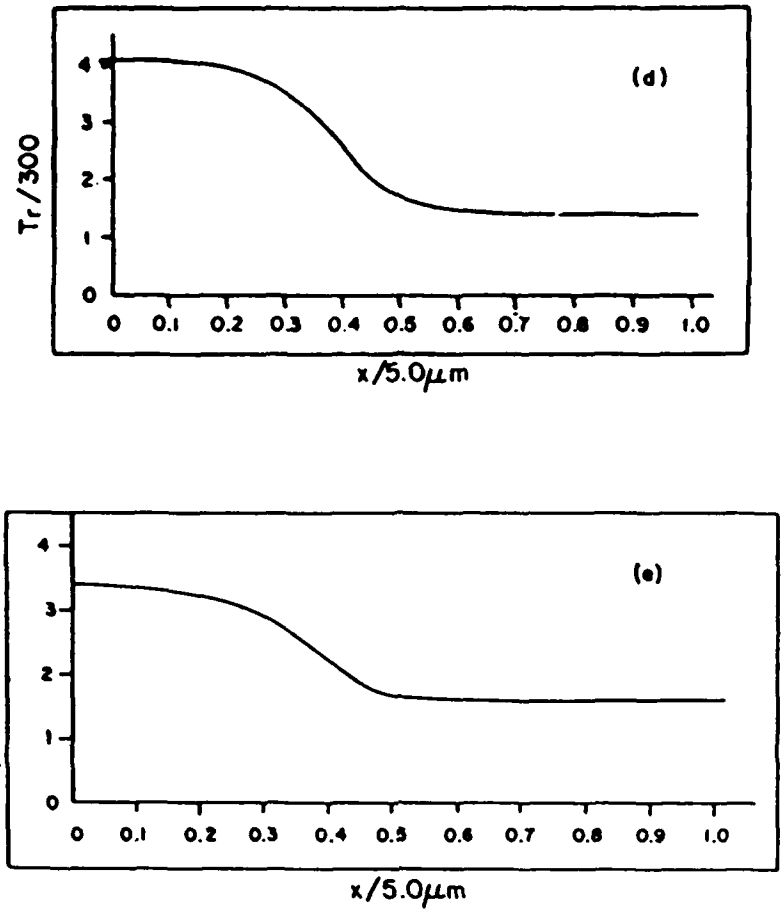


FIGURE 90d, e. As in Fig. 89 but for a bias of 2 volts.

voltage increases a region of local charge neutrality is forming at the cathode. Insofar as the Γ valley carriers are significantly depleted, the cathode region is rich in satellite valley carriers. But the distribution of velocity, as displayed in Figs. 89e and 90e, indicate again that most of the current is carried by the Γ valley carriers. What is clear from these figures is the apparent migration of the depletion layer toward the anode boundary, a situation similar to that obtained from the drift and diffusion equations.

ALTERNATIVE SOURCES OF FIELD NONUNIFORMITIES

All of the above calculations have tended to focus on the role of the cathode on device performance. Several calculations shown below illustrate the influence of selective reductions of "notches"

in doping. One calculation is for a 2μm-long device with an 8000 A notch at the anode boundary (figure 91). The second was for a device with a narrow notch (figure 92). In both cases the electric

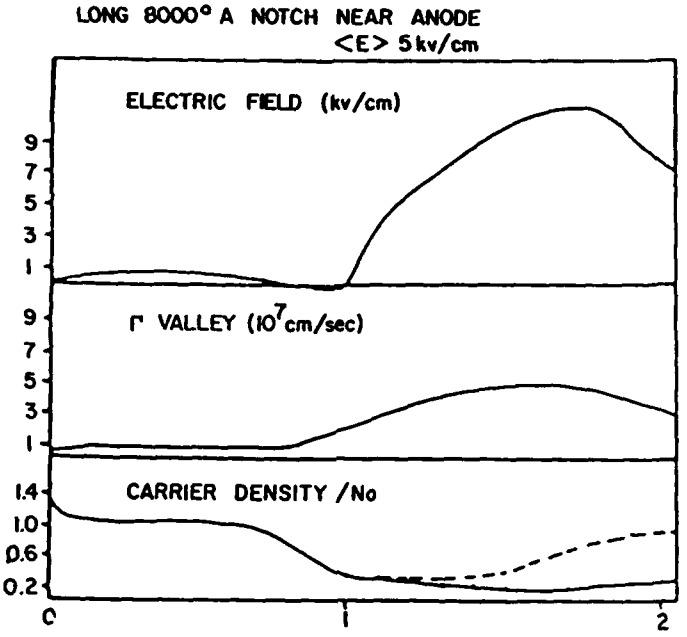


FIGURE 91. Long 8000°A notch anode. Ψ = 1v.

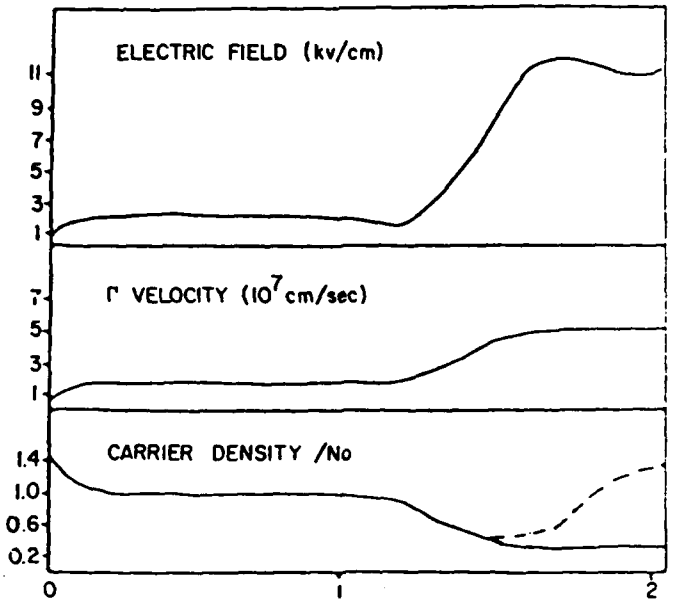


FIGURE 92. Narrow 4000°A notch is near anode Ψ = 1 v.

field showed an excess voltage over the notch along with significant transfer. The current level was higher for the narrow notch device and again most of the current level was carried by the Γ-valley carriers.

The situation when a section of periodic notches is introduced is shown in figures 93 through 95. In figure 93, the device is 2μm-long with an applied bias of 1.0v. Two notches were used

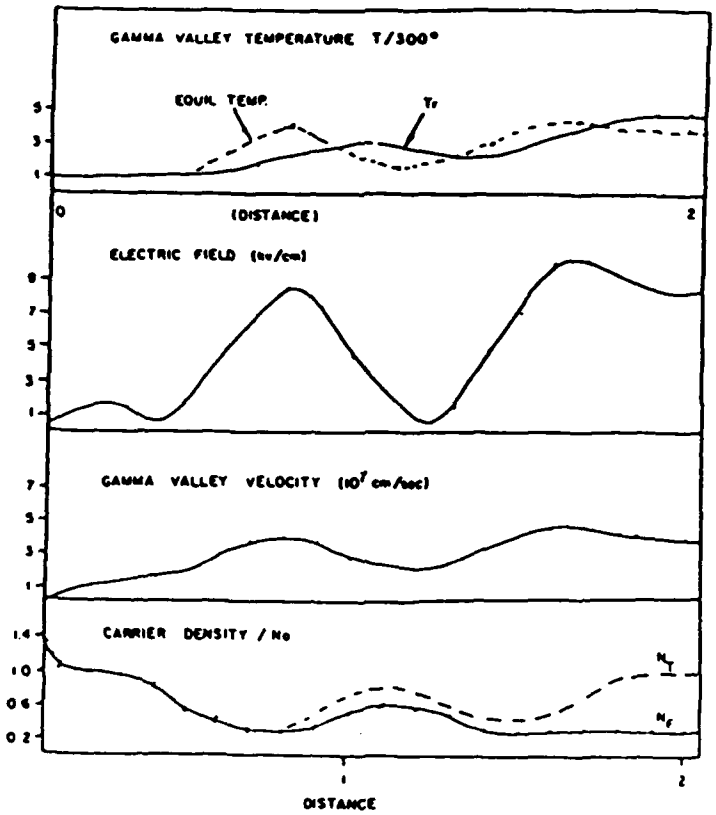


FIGURE 93. Double notch 2μm-long device. $\Psi = 1$ v.

here. Comparing the results of Fig. 93 with that of 94, it is seen that the Γ valley velocity is higher in the former case, as is

the current level. Figures 94 and 95 are for 3 and 4 notches within a one-micron long device. Each with an applied bias of .8v. In each case as the number of notches increases, the modulation of carrier density decreases, although not the field, as indicated in Fig. 95b.

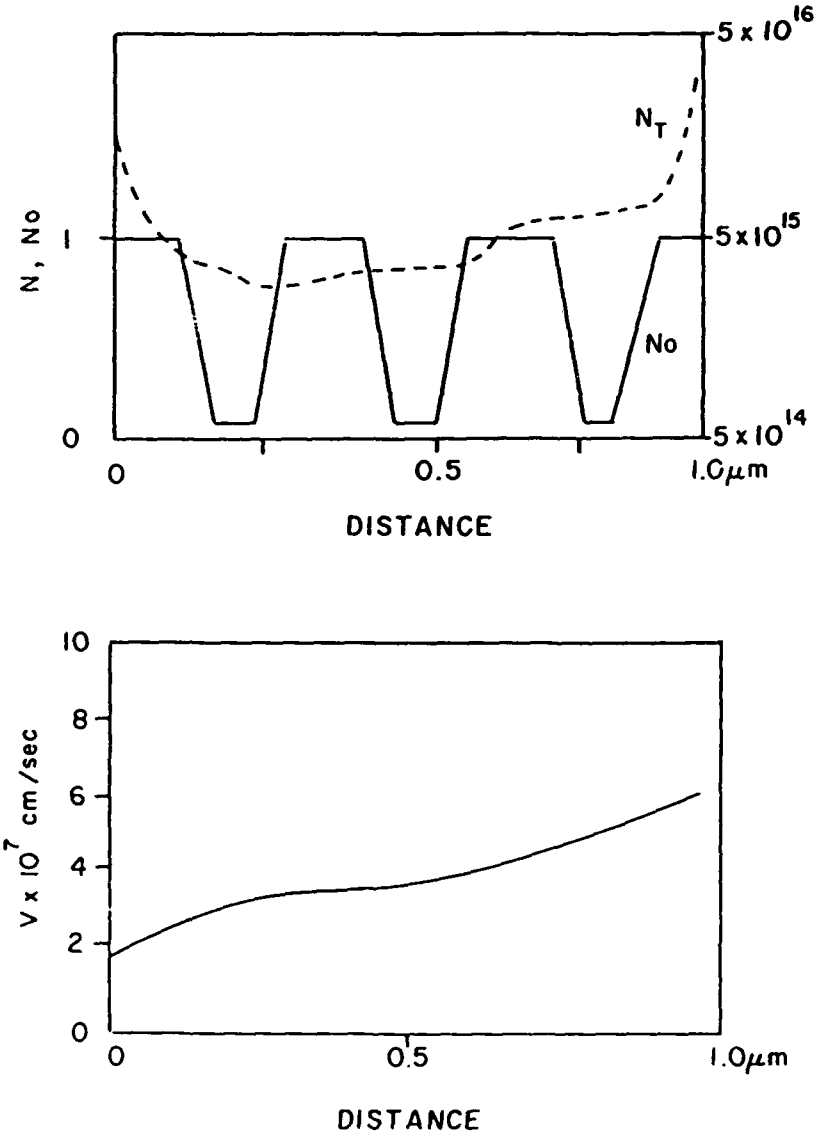


FIGURE 94. Repeated overshoot , (a) carrier density and doping profile, (b) gamma valley carrier velocity. <F> = 8Kv/cm.

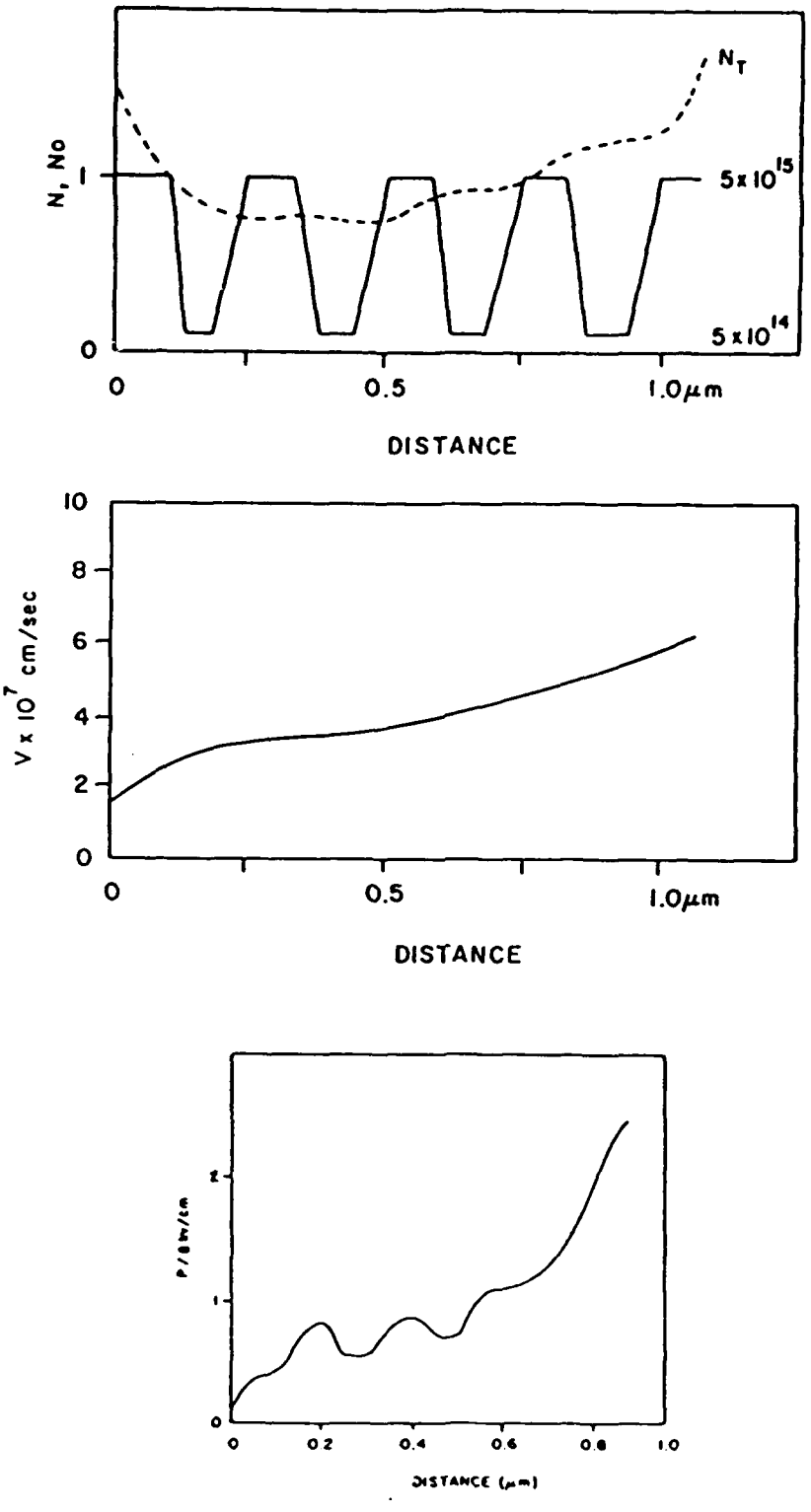


FIGURE 95. As in figure 94, but for four notches.

SUMMARY

The experimental situation is such that, with the exception of long compound semiconductor devices, there is very little data on the role of boundaries and contacts to submicron devices. The reason for the paucity of data lies in the fact that most submicron devices are three terminal devices designs and the third terminal tends to mask the role of the contact boundaries. This is unfortunate in that it is likely that two terminal device measurements will indicate what can be achieved in controlling the entrance dynamics of the carriers. To date, most two terminal device measurements on simple device structures have concentrated on the role of transport within the device, and raise the question of whether "ballistic" motion is possible. Based on the history of vacuum tube dynamics it should be recalled that, if transport is ballistic, the electrical characteristics will be controlled by the contacts.

The situation in submicron devices is further complicated by communication between the up- and downstream contacts. Thus is may be expected that the influence of a blocking contact on the electrical characteristics of long and submicron devices will be different. For submicron devices simple current voltage measurements may be rendered useless as a diagnostic tool. This is certainly not the case in long devices.

The role of numerical simulations in these boundary and device studies has been to act as surrogates for measurements that are not feasible. In one case, obtaining cathode boundary fields from measurements was not possible. Thus for long devices the sensitivity of the numerical results to numerical changes in the boundary conditions, when coupled to experiments, provided the key to the role of contacts on device behavior. For submicron devices, the difficulties of direct correlation of experiment with specific transport phenomena are apparent and simulation through parametric studies will provide a key to the role of boundaries. But the description of transport on a submicron scale is still inadequate and the descriptive role of boundaries is correspondingly weak. For example, most space charge dependent problems still treat the background as a "jellium" distribution. The discrete nature of impurities is ignored, as are structural variations in the contacts. The extent to which this affects such measurements as current-versus-voltage is yet to be determined. Notwithstanding these uncertainties, a considerable amount of information can be obtained by extrapolation from the ideal cases which can provide bounds on the limits of transport through both the boundary and active region of the device.

ACKNOWLEDGEMENTS

The author is grateful for the support of the Office of Naval Research and the Army Research Office.

REFERENCES

1. See, e.g. D.H. Austin, A.M. Johnson, P.R. Smith, and J.C. Bean, 1980, Appl. Phys. Lett. 37:371; and C.V. Shank, R.L. Fork, B.I. Greene, F.K. Reinhart and R.A. Logan, 1981, Appl. Phys. Lett. 38:104.

2. R. Hammond, This Lecture Series.

3. S. Laval, C. Bru, C. Arnodo and R. Castagne, Proc. 1980 IEDM, 626.

4. R. G. Chambers, 1952, Proc. Phys. Soc. (London) A65:458.

5. See, e.g., A.B. Pippard, 1962, "The Dynamics of Conduction Electrons" in "Low Temperature Physics", Eds: C. DeWitt, B. Dreyfus and P.G. DeGennes, Gordon and Breach, NY.

6. See, e.g., G.E.H. Reuter and E.H. Sondheimer, 1949, Proc. Roy. Soc., A195:336.

7. See, e.g. T. Kjeldaas, Jr. and T. Holstein, 1959, Phys. Rev. Lett., 2:340.

8. See, e.g, J.R. Barker, 1980, in " Physics of Nonlinear Transport in Semiconductors" (D.K.Ferry, R. Barker and C. Jacoboni, eds), 126, Plenum Press and D.K. Ferry, ibid.

9. A. Sommerfeld, 1956, "Thermodynamics and Statistical Mechanics", Academic Press, NY.

10. K. Thornber, 1978, Solid State Electronics, 21:259.

11. R. W. Keyes, 1958, J. Phys. Chem. Solids, 6:1.

12. See, e.g., H. Frohlich and B.V. Paranjape, 1956, Proc. Phys. Soc. B, 69:21; and C.J. Hearn, 1979, in "Physics of Nonlinear Transport in Semiconductors", (D.K. Ferry, J.R. Barker, and C. Jacoboni, eds.) Plenum Press.

13. H. D. Rees, see, e.g., Institute of Physics Conf. Series # 22:105; and K.W. Gray, J.E. Pattison, H.D. Rees, B.A. Prew, R.C. Clarke and L.D. Irving, 1975, Proc. Fifth Biennial Cornell Electrical Engineering Conf., 215, Cornell University Press, Ithaca, NY.

14. H.L. Grubin and J.P. Kreskovsky, 1983, Surface Science 132:594.

15. J.G. Ruch, 1972, IEEE Trans. Electron Devices, ED-19:652.

16. P.N. Butcher and C.J. Hearn, 1968, Electron Lett. 4:459.

17. G.J. Iafrate, R. Malik, K. Hess and J. Tang (to be published).

18. W.F. Shockley, 1954, The Bell System Telephone Journal, 33:799.

19. M.P. Shaw, P.R. Solomon and H.L. Grubin, 1969, IBM J. Res. Devel. 13:587.

20. R. Fauquembergue, M. Pernisek and E. Constant, 1984, "The Physics of Submicron Structures" (H.L. Grubin, K. Hess, G.J. Iafrate and D.K. Ferry, eds.), 171, Plenum Press NY.

21. R.K. Cook and J. Frey, 1971, IEEE Trans. Electron Devices, ED-28:951.

22. H.D. Rees, 1969, IBM J. Res. Devel. 12:537.

23. H. Kroemer, R.E. Hayes, and R.M. Raymond, 1978, IEEE Trans. Electron Devices, ED-35.

24. R.F. Cooke and J. Frey, 1982, COMPEL, 1,2.

25. P.C. Chao, P.M. Smith, S. Wanuga, W.H. Perkins and E.D. Wolf, 1983, IEEE Electron Dev. Lett., EDL-4:326.

26. R.W.H. Englemann and C. Liechti, 1977, IEEE Trans. Electron Devices, ED-24.

27. H.L. Grubin, D.K. Ferry and K.R. Gleason, 1980, State Electronics, 23:157.

28. R.S.C. Cobbold, 1970, "Theory and Application of Field Effect Transistors", Wiley-Interscience, NY.

29. H.L. Grubin and J.P. Kreskovsky, 1983, SRA Report R930007-F.

30. K.K. Thornber, 1980, J. Appl. Phys. 52:279 and 52:2127.

31. T.J. Maloney and J. Frey, 1977, J. Appl. Phys. 48:781.

32. G. Hill, P. N. Robson, A. Majerfeld and W. Fawcett, 1977, Electronics Lett.

33. D.K. Ferry and H.L. Grubin, 1981, Microelectronics Journal, 12:5.

34. J.W. Harrison, J.R. Hauser, T.H. Glisson and M.A. Littlejohn, 1977, Report No. AFAL-TR-77-129, figure 22.

35. Ibid, figure 26.

36. Ibid, figure 23.

37. Ibid, figure 25.

38. W. Porod and D.K. Ferry, 1983, Phys. Rev. B27:2587.

39. Y.F. Biryulin, S.P. Vul, V.V. Chaldychev, Y.V. Shmartsev, 1983, Phys. Semicond., 17:65.

40. J.W. Harrison, 1972, "Electron Mobility in Alloys of Direct Band Gap III-V Compounds", Ph.D Thesis, North Carolina State University.

41. See, e.g., J.L. Birman, 1962, Phys. Rev 127:1093; and 1963, Phys. Rev. 131:1489.

42. D. Matz, 1965, Phys. Rev. 168:843.

43. J.J. Harris and B.K. Ridley, 1973, J. Phys. and Chem. Solids, 34:197.

44. J.R. Hauser, M.A. Littlejohn and T. Glisson, 1976, Appl. Phys. Lett., 28:458.

45. H. Temkin and V.G. Keramidas, 1980, J. Appl. Phys. 51:3268.

46. J.R. Hauser, T.H. Glisson and M.A. Littlejohn, 1979, Solid State Electronics, 22:487.

47. D.K.Ferry, 1972, Phys.Rev.12, 2360.

48. M.P. Shaw, H.L. Grubin and P.R. Solomon, 1979, "The Gunn-Hilsum Effect", Academic Press, NY.

49. H.L. Grubin, 1976, IEEE Trans. Electron Devices, ED-23:1012.

50. See, e.g., V.L. Rideout, 1975, Solid State Electronics 18:541.

51. W. Shockley, J.A. Copeland and R. P. James, 1966, "Quantum Theory of Atoms, Molecules and the Solid State", (P.O. Lowdin, ed.) 537, Academic Press, NY.

END

DATE

FILMED

6-1988

DTIC